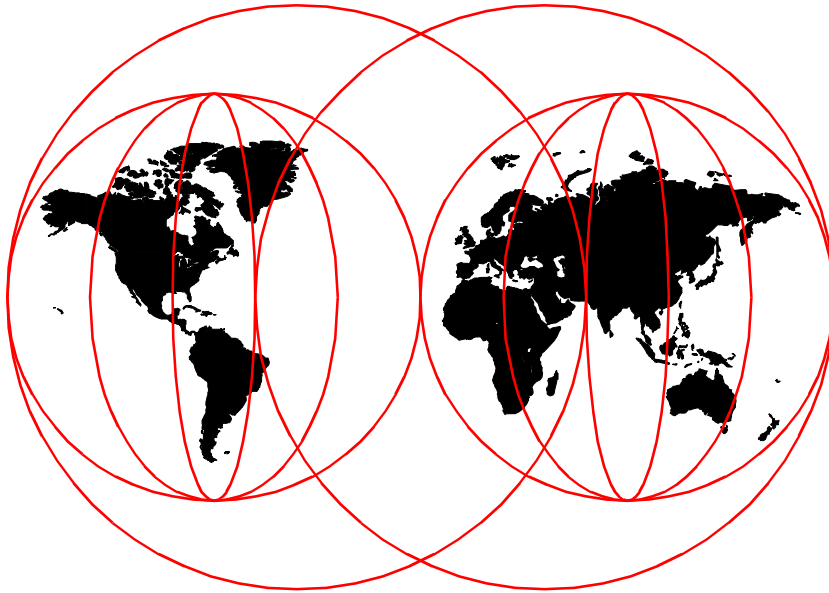


Server Consolidation on RS/6000

*KyeongWon Jeong, Bradley Wilkinson, Eric Sun, Murray White, Peter Reisner,
SangHo Lee, Trina Bunting*



International Technical Support Organization

www.redbooks.ibm.com

SG24-5507-00



International Technical Support Organization

Server Consolidation on RS/6000

November 1999

Take Note!

Before using this information and the product it supports, be sure to read the general information in Appendix B, "Special notices" on page 333.

First Edition (November 1999)

This edition applies to IBM RS/6000 for use with the AIX Operating System Version 4 and is based on information available in August, 1999.

Comments may be addressed to:
IBM Corporation, International Technical Support Organization
Dept. JN9B Building 003 Internal Zip 2834
11400 Burnet Road
Austin, Texas 78758-3493

When you send information to IBM, you grant IBM a non-exclusive right to use or distribute the information in any way it believes appropriate without incurring any obligation to you.

© Copyright International Business Machines Corporation 1999. All rights reserved.
Note to U.S Government Users – Documentation related to restricted rights – Use, duplication or disclosure is subject to restrictions set forth in GSA ADP Schedule Contract with IBM Corp.

Contents

| | |
|---|------|
| Figures | ix |
| Tables | xiii |
| Preface | xv |
| The team that wrote this redbook | xv |
| Comments welcome | xvii |
| <hr/> | |
| Part 1. Why consolidated servers? | 1 |
| Chapter 1. Introduction | 3 |
| 1.1 IT trends and directions | 3 |
| 1.1.1 Direction of business | 3 |
| 1.1.2 Expanded server consolidation for new IT trends | 5 |
| 1.2 RS/6000 and AIX | 7 |
| Chapter 2. Benefits and types of server consolidation | 17 |
| 2.1 Benefits from consolidation of servers | 17 |
| 2.1.1 Single point of control | 20 |
| 2.1.2 Giving users better services | 21 |
| 2.1.3 Regaining flexibility | 21 |
| 2.1.4 Minimize learning and optimize use of skilled resources | 21 |
| 2.1.5 Avoid floor space constraints | 22 |
| 2.1.6 Reduction of TCO | 22 |
| 2.1.7 To provide enhanced functionality | 28 |
| 2.1.8 To build strategic infrastructure | 29 |
| 2.2 Types of consolidation | 29 |
| 2.2.1 Centralization | 32 |
| 2.2.2 Physical consolidation | 33 |
| 2.2.3 Data integration | 35 |
| 2.2.4 Application integration | 36 |
| 2.3 Good candidates for server consolidation | 38 |
| 2.4 Decision criteria | 40 |
| 2.4.1 Risks | 40 |
| 2.4.2 Package availability | 42 |
| 2.4.3 Time to implement | 42 |
| 2.4.4 Success factors | 43 |
| Chapter 3. Why consolidate to the RS/6000? | 45 |
| 3.1 RS/6000 strengths as a central server | 45 |
| 3.1.1 Architecture | 45 |

| | | |
|---|---|------------|
| 3.1.2 | Components | 47 |
| 3.1.3 | System management | 48 |
| 3.1.4 | Availability | 58 |
| 3.1.5 | Scalability | 58 |
| 3.1.6 | Security | 61 |
| 3.1.7 | Network | 64 |
| 3.1.8 | Flexibility | 68 |
| 3.1.9 | Applications | 68 |
| 3.2 | Reasons for consolidating to RS/6000 | 69 |
| 3.2.1 | Why do customers consolidate onto an SP system? | 69 |
| 3.2.2 | What is the SP System being used for? | 71 |
| 3.2.3 | RS/6000 SP financial benefits | 72 |
| Chapter 4. Consolidation methodologies | | 75 |
| 4.1 | The IBM ALIGN methodology | 78 |
| 4.1.1 | Step 1: Qualification | 80 |
| 4.1.2 | Step 2: Customer environment profile | 80 |
| 4.1.3 | Step 3: "Islands" analysis | 80 |
| 4.1.4 | Step 4: Detailed analysis/solution design | 81 |
| 4.1.5 | Step 5: Implementation | 81 |
| 4.1.6 | Step 6: Validation | 81 |
| 4.1.7 | ALIGN examples | 81 |
| 4.1.8 | Advantages of IBM methodology | 83 |
| 4.2 | Business solution assessment | 86 |
| 4.2.1 | Overview | 87 |
| 4.2.2 | BSA case study | 89 |
| Chapter 5. Customer consolidation scenarios | | 93 |
| 5.1 | IBM Yasu lab | 93 |
| 5.2 | An investment services company | 95 |
| 5.3 | An automobile manufacturer | 97 |
| 5.4 | A medical facility | 99 |
| 5.5 | A life insurance company | 100 |
| 5.6 | An insurance company | 101 |
| 5.7 | A university | 104 |
| 5.8 | An aircraft manufacturer | 105 |
| 5.9 | A financial services company | 108 |
| 5.10 | A casino's success: Betting on RS/6000 technology | 110 |
| Part 2. Server consolidation solutions for RS/6000 | | 113 |
| Chapter 6. System management | | 115 |
| 6.1 | Software installation management | 115 |

| | | |
|--------|--|------------|
| 6.1.1 | Software installation management solutions | 115 |
| 6.1.2 | Operating system installation and updates | 116 |
| 6.1.3 | Other software installation products | 117 |
| 6.2 | User management | 118 |
| 6.2.1 | User management solutions | 119 |
| 6.2.2 | Other user management solutions | 124 |
| 6.3 | Performance management | 127 |
| 6.3.1 | Performance measurements | 127 |
| 6.3.2 | Performance management solutions | 128 |
| 6.3.3 | Other performance management solutions | 133 |
| 6.4 | Disk space management | 139 |
| 6.4.1 | Disk space management solutions | 139 |
| 6.5 | Automation | 145 |
| 6.5.1 | Automation solutions | 146 |
| 6.5.2 | Other automation solutions | 148 |
| 6.5.3 | Batch processing | 151 |
| 6.6 | Event management | 153 |
| 6.6.1 | Event management solutions | 153 |
| 6.6.2 | Other event management solutions | 158 |
| 6.7 | Print services management | 159 |
| 6.7.1 | Print services management solutions | 159 |
| 6.8 | Configuration management | 162 |
| 6.8.1 | Configuration management solutions | 163 |
| 6.9 | Change management | 164 |
| 6.9.1 | Change management solutions | 164 |
| 6.10 | System recovery | 166 |
| 6.10.1 | Centralized back up and recovery | 166 |
| 6.10.2 | Centralized backup and recovery solutions | 168 |
| 6.10.3 | Software recovery | 179 |
| 6.10.4 | Hardware recovery | 181 |
| 6.10.5 | Disaster recovery | 183 |
| | Chapter 7. Workload management | 185 |
| 7.1 | Introduction | 185 |
| 7.2 | AIX Workload Manager | 185 |
| 7.2.1 | Overview | 185 |
| 7.2.2 | Concepts and configuration | 186 |
| 7.2.3 | Administration | 194 |
| 7.2.4 | Benefits | 207 |
| 7.2.5 | Hints and Tips | 207 |
| 7.2.6 | Examples | 209 |
| 7.3 | LoadLeveler | 215 |
| 7.3.1 | Overview | 215 |

| | | |
|-------------------------------------|---|------------|
| 7.3.2 | Concepts and configuration | 215 |
| 7.3.3 | Why use LoadLeveler? | 218 |
| 7.3.4 | Benefits | 218 |
| 7.4 | SecureWay Network Dispatcher | 218 |
| 7.4.1 | Overview | 219 |
| 7.4.2 | Interactive Session Support | 219 |
| 7.4.3 | Dispatcher | 220 |
| 7.4.4 | Content-based routing | 231 |
| 7.4.5 | Remote administration | 231 |
| 7.4.6 | Why use SecureWay Network Dispatcher? | 231 |
| 7.4.7 | Benefits | 232 |
| Chapter 8. High availability | | 233 |
| 8.1 | Overview | 234 |
| 8.2 | High availability solutions | 234 |
| 8.2.1 | High Availability Cluster Multi Processing (HACMP) | 234 |
| 8.2.2 | High Availability Geographic Cluster (HAGEO) | 240 |
| 8.2.3 | High Availability Control WorkStation (HACWS) | 241 |
| 8.2.4 | HACMP/ES | 243 |
| 8.3 | High availability options at the operating system level | 247 |
| 8.3.1 | Disk mirroring/logical volume mirroring | 247 |
| 8.3.2 | Redundant Array of Independent Disks (RAID) | 247 |
| 8.4 | Integrating an existing cluster into a consolidated environment | 248 |
| 8.4.1 | Physical hardware and software relocation | 248 |
| 8.4.2 | Migrating a cluster to new hardware | 249 |
| 8.5 | Planning and integrating a cluster from conception | 249 |
| 8.5.1 | Cluster nodes | 249 |
| 8.5.2 | CPU options | 250 |
| 8.5.3 | Node considerations | 250 |
| 8.5.4 | Cluster networks | 251 |
| 8.5.5 | Cluster disks | 252 |
| 8.5.6 | Resource planning | 252 |
| 8.5.7 | Application planning | 253 |
| 8.5.8 | Other cluster consolidation issues | 254 |
| 8.6 | ClusterProven and Advanced ClusterProven | 254 |
| 8.6.1 | ClusterProven verification process | 256 |
| 8.6.2 | ClusterProven verification criteria | 256 |
| 8.6.3 | Cluster design guidelines | 256 |
| 8.6.4 | Advanced ClusterProven overview | 257 |
| 8.7 | Enterprise Storage Server | 258 |
| 8.7.1 | Positioning | 259 |
| 8.7.2 | Benefits | 260 |
| 8.8 | Customer example | 264 |

| | |
|--|-----|
| Chapter 9. Server consolidation for key business applications | 269 |
| 9.1 Overview | 269 |
| 9.1.1 Server consolidation for data/application integration. | 269 |
| 9.1.2 Expanded server consolidation for business integration | 270 |
| 9.2 Cross-industry solutions for business applications | 271 |
| 9.2.1 DBMS. | 272 |
| 9.2.2 OLTP mission-critical applications | 275 |
| 9.2.3 Decision Support Systems (DSS) | 281 |
| 9.2.4 ERP | 283 |
| 9.2.5 Groupware: Lotus Notes. | 286 |
| 9.2.6 e-business solutions. | 287 |
| 9.3 Server consolidation for data integration | 290 |
| 9.3.1 VLDB using parallel DBMS for DB integration. | 290 |
| 9.3.2 DB connectivity tool for data integration | 300 |
| 9.3.3 Sizing and capacity planning tools | 303 |
| 9.3.4 Lotus Domino Enterprise data integration | 308 |
| 9.3.5 Storage solution for data integration | 309 |
| 9.4 Server consolidation for business application integration | 311 |
| 9.4.1 Extended ERP | 312 |
| 9.4.2 Enterprise Application Integration (EAI) | 314 |
| 9.4.3 EAI solutions - MQ integrator and BEA eLink | 315 |
| 9.4.4 Server consolidation for Web application integration. | 318 |
| 9.4.5 Application development and automated testing tools. | 324 |
| 9.4.6 Outsourcing vs. server consolidation | 325 |
| | |
| Appendix A. Information sources | 327 |
| A.1 Independent software vendors | 327 |
| A.2 IBM Softwares. | 330 |
| | |
| Appendix B. Special notices | 333 |
| | |
| Appendix C. Related publications | 337 |
| C.1 IBM Redbooks publications | 337 |
| C.2 IBM Redbooks collections | 338 |
| C.3 Other resources | 338 |
| C.4 Referenced Web sites | 338 |
| | |
| How to get IBM Redbooks | 341 |
| IBM Redbooks fax order form | 342 |

| | |
|--------------------------------------|-----|
| Glossary | 343 |
| Index | 345 |
| IBM Redbooks evaluation | 353 |

Figures

| | |
|---|-----|
| 1. IT trends and directions | 4 |
| 2. Server consolidation opportunity for new business solutions | 5 |
| 3. New IT trend and business application evolution. | 7 |
| 4. Reasons for consolidation | 18 |
| 5. TCO in a distributed world | 23 |
| 6. True total cost of computing. | 24 |
| 7. How to reduce TCO | 27 |
| 8. Four types of server consolidation. | 32 |
| 9. Centralization | 33 |
| 10. Physical consolidation | 34 |
| 11. Data integration | 36 |
| 12. Application integration | 37 |
| 13. System integration | 38 |
| 14. Why do customers consolidate onto an SP? | 70 |
| 15. What is the RS/6000 SP being used for? | 72 |
| 16. Customers' consolidation concerns | 75 |
| 17. Server consolidation methodologies comparisons. | 78 |
| 18. ALIGN server consolidation methodology | 78 |
| 19. Server consolidation business model | 79 |
| 20. ALIGN methodology process steps | 80 |
| 21. BSA overview. | 87 |
| 22. IBM Yasu lab | 95 |
| 23. Investment services company | 97 |
| 24. User management on a multi-node RS/6000 SP | 119 |
| 25. Managing a central user database on an RS/6000 SP | 121 |
| 26. Workload response time of BEST/1. | 135 |
| 27. Sample EcoTOOLS service level management report | 138 |
| 28. Example: Virtual Shared Disk configuration. | 140 |
| 29. GPFS overview | 142 |
| 30. PIOFS supports the simultaneous access. | 144 |
| 31. Tivoli Enterprise software components | 147 |
| 32. Candle Command Center overview. | 149 |
| 33. Infoprint Manager. | 160 |
| 34. NetBackup Java interface | 173 |
| 35. NetBackup GUI Scheduler interface | 174 |
| 36. Basic WLM elements | 186 |
| 37. Example of classes and class assignment rules | 189 |
| 38. Example of share distribution automatically adjusting resources | 192 |
| 39. Web-based System Manager: Main menu for WLM | 198 |
| 40. Web-based System Manager: Create Class dialog | 199 |

| | |
|---|-----|
| 41. Web-based System Manager: Change Shares dialog | 200 |
| 42. Web-based System Manager: Class Assignment Rules dialog | 200 |
| 43. Web-based System Manager: New Class Assignment Rule dialog | 201 |
| 44. Web-based System Manager: Manage Configuration dialog | 202 |
| 45. Web-based System Manager: Change Configuration dialog | 203 |
| 46. SMIT: Main menu for WLM | 204 |
| 47. SMIT: Adding a class | 204 |
| 48. SMIT: Assign process attribute values | 205 |
| 49. Output of the ps command demonstrating a class column printed | 206 |
| 50. CPU usage results of example 2 | 211 |
| 51. CPU usage results of example 3 | 212 |
| 52. Memory usage results of example 4 | 214 |
| 53. Loadleveler job status flow | 216 |
| 54. Interactive Session Support (ISS) | 220 |
| 55. Dispatcher | 222 |
| 56. Dispatcher component | 224 |
| 57. ISS and Dispatcher together | 227 |
| 58. High availability overview | 233 |
| 59. A two node Hot-Standby cluster configuration | 237 |
| 60. Mutual takeover configuration | 238 |
| 61. Typical HAGEO configuration | 241 |
| 62. Typical HACWS configuration | 242 |
| 63. HACMP/ES Version 4.3 | 244 |
| 64. RSCT overview | 245 |
| 65. ClusterProven logo | 254 |
| 66. Server consolidation classification | 270 |
| 67. Consolidating business applications | 271 |
| 68. RS/6000 SP as a flexible consolidation server | 272 |
| 69. 3-tier architecture | 276 |
| 70. 2-tier and 3-tier architecture | 277 |
| 71. RS/6000 SP system for the typical OLTP 3-tier environment | 281 |
| 72. Business Intelligence framework | 283 |
| 73. Conceptual model of SAP R/3 structure | 284 |
| 74. Typical ERP client/server configuration | 285 |
| 75. Typical ERP SP consolidated configuration | 286 |
| 76. Logical model of Lotus Notes on the RS/6000 SP | 287 |
| 77. Enterprise-Wide Domino configuration | 288 |
| 78. WebSphere Application Server architecture | 290 |
| 79. Four physical nodes using DB2 and the SP Switch | 291 |
| 80. Partitioning map | 293 |
| 81. RS/6000 SP configuration with HACMP | 294 |
| 82. Oracle Parallel Server | 295 |
| 83. Scalability with DSS using DB2 PE | 299 |

| | |
|---|-----|
| 84. Web Server Logic programming environment evolution | 302 |
| 85. Average CPU utilization by workloads of BEZ | 304 |
| 86. Workload CPU utilization | 307 |
| 87. Domino NotesPump server | 309 |
| 88. Storage Area Network (SAN) for data integration | 311 |
| 89. Extended ERP solutions | 313 |
| 90. Efficient Message Routing of MQSeries | 316 |
| 91. BEA EAI solution framework | 317 |
| 92. IBM - BEA product comparison | 318 |
| 93. ISP Internet access and Web content hosting services | 319 |
| 94. Network balancing of ISS and SecureWay Network Dispatcher | 321 |
| 95. WebSphere application server architecture | 322 |
| 96. Web application service architecture | 324 |

Tables

| | |
|---|-----|
| 1. Top reasons for consolidation | 71 |
| 2. Features and benefits of Tivoli Software Distribution. | 118 |
| 3. Benefits of highly-available consolidated server systems | 236 |
| 4. Supported RS/6000 models as cluster nodes | 250 |

Preface

Server consolidation can have a positive impact on business results. It is broader than just physically combining assets. It involves the reduction in the number of servers in use within an organization by centralizing many applications, or data, onto fewer servers to reduce costs, increase the efficiency of system management, security, and resource utilization.

This redbook is intended to help consultants and IT managers, their technical team, and RS/6000 sales teams in IBM who need to identify requirements and opportunities and plan for server consolidation on RS/6000 platforms.

This redbook gives a broad understanding of server consolidation, and it will help in designing a solution on RS/6000 platforms for server consolidation.

The team that wrote this redbook

This redbook was produced by a team of specialists from around the world working at the International Technical Support Organization, Austin Center.

KyeongWon Jeong is a Senior Software Engineer at the International Technical Support Organization, Austin Center. He writes extensively on AIX and BI areas and education materials. Before joining the ITSO, he worked in IBM Global Learning Services of IBM Korea as a Senior Education Specialist and was a class manager of all AIX classes for the customers and interns. He has many years of teaching and development experience.

Bradley Wilkinson is a managing consultant with the IBM business partner OSIX in Australia. He has several years experience in the RS/6000 support and service field and holds a degree in Computer Science from the University of Newcastle, Australia. His areas of expertise include AIX, RS/6000, SP, and related products.

Eric Sun is a Technical Sales Specialist at IBM Philippines. He has many years of pre-sales experience in the Information Technology field. He holds a degree in Industrial Management Engineering from De La Salle University. His areas of expertise include Unix and Server Consolidation.

Murray White is a UNIX Technical Team Leader for IBM in Auckland, New Zealand. He has many years of experience in the fields of computer programming, systems administration, and UNIX consultation. Since joining IBM Global Services, he has gained expertise in UNIX system administration on different platforms including AIX, HP-UX, and SGI.

Peter Reisner is an AIX systems specialist working for OSIX, an IBM business partner in Australia. His fields of expertise are distributed systems administration and integration, storage management, high availability, and percussive maintenance. His experience includes Australia/New Zealand distributed systems support, TSM, and HACMP support on RS/6000 SP systems.

SangHo Lee is a Senior IT Specialist at IBM Korea. He is a team leader of Benchmarking Test Support Team in RS/6000 ATSC at IBM Korea. He has been working on RS/6000 and SP systems after joining IBM. He was responsible for RS/6000 technical support and pre-sales support in high performance computing for scientific and commercial environments.

Trina Bunting is a member of the Advanced Technical Support Group in Roanoke, Texas. She is the Server Consolidation Skills leader for North America. She joined IBM in 1996 and was part of Technical Services and AIX Support Line before joining the Advanced Technical Support group.

Thanks to the following people for their invaluable contributions to this project:

Tetsuya Shirai, IBM Austin
Catherine Cook, IBM Austin
Ole Conradsen, IBM Austin
Heinz Johner, IBM Austin
Scott Vetter, IBM Austin
Jim Beesley, IBM Austin
Kenneth Rozendal, IBM Austin
Andre L. Albot, IBM Austin
Jack Alford, IBM Austin
Bob Arbeitman, IBM Austin
Lalita Malik, IBM Poughkeepsie
Marcelo Barrios, IBM Poughkeepsie
Michael MacIsaac, IBM Poughkeepsie
Kathleen Smith, IBM Poughkeepsie
Curt Christopher, IBM Poughkeepsie
Kjell E Nystrom, IBM San Jose
Chris Gage, IBM Raleigh
Steve Weeks, IBM UK
Patrick Mulot, IBM France
Yukio Ohya, IBM Japan
Helena Restrepo, BMC
Tom Murphy, BMC
Joe Burns, Candle
Larry Phillips, Candle

Boris Zibitsker, BEZ
Deane Tierney, Compuware
Al Burstiner, Computer Associates
Shawn Klein, ADIC
Cathy Won, Legato
Nur Premo, Veritas

Special thanks go to the editors for their help in finalizing the text and publishing the book:

John Owczarzak
Rose Harlan

Comments welcome

Your comments are important to us!

We want our Redbooks to be as helpful as possible. Please send us your comments about this or other Redbooks in one of the following ways:

- Fax the evaluation form found in “How to get IBM Redbooks” on page 341 to the fax number shown on the form.
- Use the online evaluation form found at <http://www.redbooks.ibm.com/>
- Send your comments in an Internet note to redbook@us.ibm.com

Part 1. Why consolidated servers?

2 Server Consolidation on RS/6000

Chapter 1. Introduction

Server consolidation is a dominant, worldwide trend in IT deployment. Because new patterns of competition, integration of markets, and escalating pressures for efficiency are shifting the focus of corporate strategies away from decentralization, this is a way to adapt quickly to business changes, cut operating and maintenance costs, and offer enhanced service levels to end users.

At its core, server consolidation is an enabling solution for:

- Optimizing existing IT infrastructures
- Integrating existing architectures across applications/data
- Providing a foundation for new solution investment and implementation

1.1 IT trends and directions

Server consolidation is an industry trend fueled by the challenge many IT managers face in delivering higher IT service levels while increasing the cost effectiveness and efficiency of their operations and infrastructure. This challenge exists, in part, as a result of the acceptance of the distributed client/server computing model.

1.1.1 Direction of business

Distributed computing has often created environments within enterprises that are difficult to manage. Often there are many servers from different vendors and of different architectures that require skills in multiple operating systems. End user departments are installing servers and upgrades, managing those servers, and if they are backing up data, the backups are usually on slow devices and may be rather unreliable. End users are also doing problem determination and fixing software problems. There is little sharing of computer resources among these different environments.

The reality of distributed computing has been complex and costly. The perception of lower hardware costs rapidly gave way to the reality that distributed servers multiply at an alarming rate, software management becomes nearly impossible, and the productivity of end users is often sacrificed to problem resolution. This has led to total costs of distributed client/server computing that are much higher than expected and that are much higher than in traditional, centralized environments.

Over the past few years, the vast majority of large corporations have recognized the hidden cost of distributed servers, and many ambitious client/server implementation plans have been greatly modified or abandoned. To restore order, IT managers have turned to server consolidation.

Server consolidation is a way to centralize business computing workloads to reduce cost, complexity, and management overhead and, in general, to free up business professionals to do business instead of IT management tasks.

Recent research indicates a strong trend toward high-end server consolidation of enterprise applications driven by the travails of managing numerous application, database, and file/print servers throughout the enterprise.

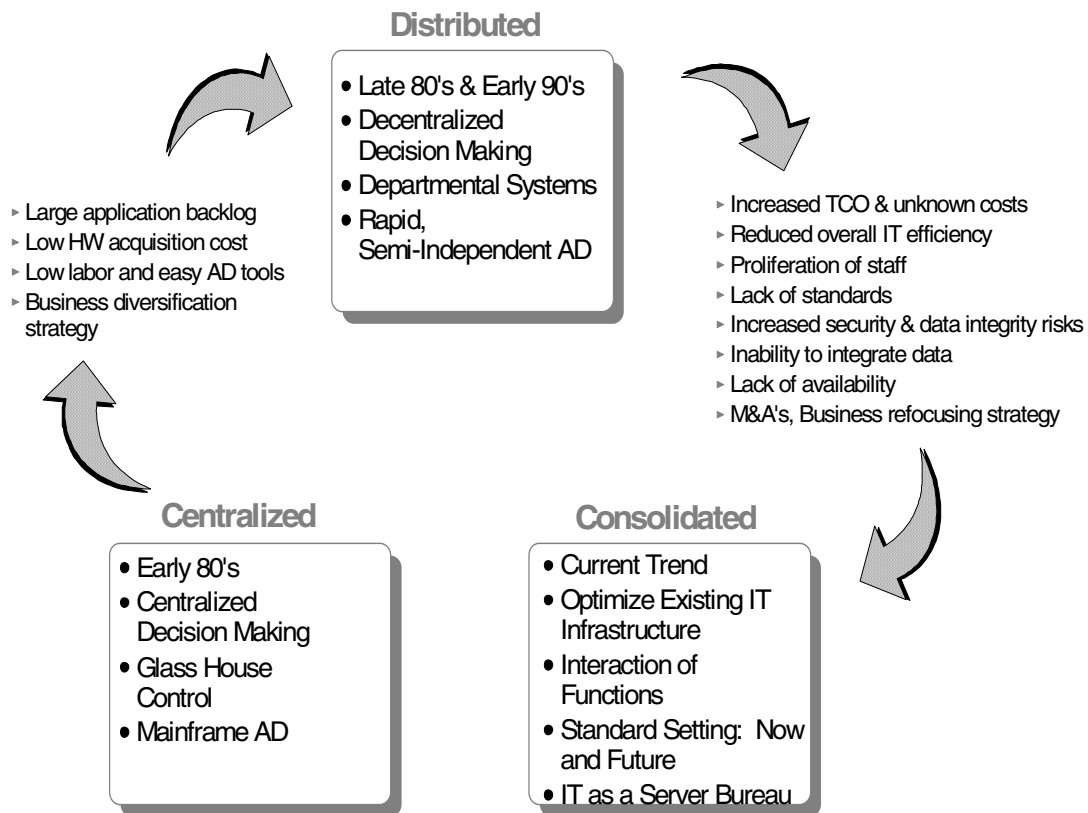


Figure 1. IT trends and directions

Some trends in business today are:

- 92 percent of North American corporations have hybrid or centralized information system models (*Computerworld*).
- 80 percent of multi-national corporations are recentralizing information systems (*The Research Board*).
- 68 percent of multinational corporations are recentralizing finance systems (*The Conference Board*).
- 72 percent of North American corporations are consolidating or recentralizing PC-LAN infrastructures (*Computer Economics*).

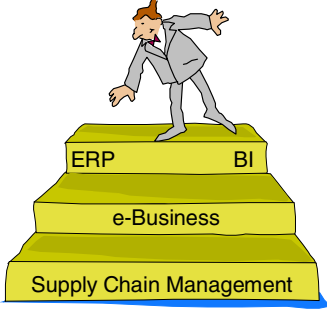
The direction of business today is focused on a strategic use of the information and cost-effectiveness of IT expenditure. Therefore, in order to meet future challenges, the assumptions and solutions built in the 1990s have to be reconsidered.

1.1.2 Expanded server consolidation for new IT trends

As customers have begun to understand the potential advantages and implications of network computing and e-business, the concept of server consolidation has expanded to include a broader range of activities and strategies. This expanded scope plays to IBM's traditional strengths in providing the full range of products, services, and financing that customers require in order to successfully implement enterprise-wide consolidation projects

Support strategic enterprise growth

- New business functions (ERP, e-Business, BI, Supply-chain management)
- Information as a strategic business tool
- Application scalability



How to Handle Fifty Million Unexpected Guests

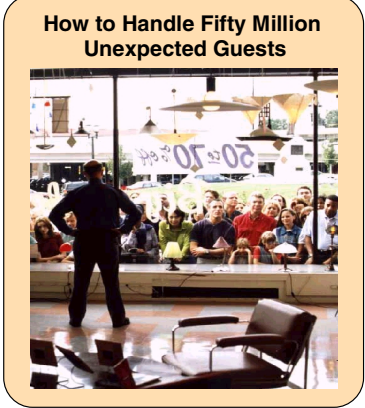


Figure 2. Server consolidation opportunity for new business solutions

Business integration

What is business integration? Forces, such as global competition, e-business, increasing customer sophistication, and cost pressures all present challenges to existing business models. The solution is a constant and ever-adapting business strategy. One of the keys to achieving this is Information Technology. New business opportunities and new strategies allow the integration of business applications. The challenge is to achieve integration in the diverse computing environment that pervades business today.

Business integration can provide significant value to your company by:

- Uniting diverse businesses so that you can deliver products to market faster.
- Deriving a single customer view so that you can service your customers' needs better.
- Cross-selling products and services through a better understanding of your customers' buying patterns.
- Linking the Web to your business strategy so as to reach new customers and provide new services to existing customers.
- Revitalizing your supply chain to reduce costs and get your production to market faster.
- Responding faster to business change.

The benefits of business integration are most obvious in the context of the following major business opportunities. However, the benefits of such solutions reach well beyond these areas.

- Merger and acquisition integration
- Automating Business Processes
- Supply chain integration
- Customer relationship management integration
- Enterprise Resource Planning (ERP) application integration
- e-business application integration

| Core Business | | e-business | | | Deep Computing | |
|-----------------------|------------------------------------|----------------------|-------------|------------|-----------------------|------------------------------------|
| Office Infrastructure | Transaction Processing (OLTP, ERP) | Collaboration | Web Serving | e-commerce | Business Intelligence | Scientific/ Technical/ Engineering |
| \$42.5B Market | | \$7.3B Market | | | \$15.9B Market | |
| 3% Growth | | 17% Growth | | | 9% Growth | |
| 26% Share | | 19% Share | | | 18% Share | |
| Mature | | Emerging | | | Evolving | |

Figure 3. New IT trend and business application evolution

1.2 RS/6000 and AIX

RS/6000 is an industry leader

RS/6000 systems serve customers in virtually every industry by connecting to millions of users and by helping run businesses, governments, and other institutions on which we rely on every day. Enterprises depend on RS/6000 to propel their e-business initiatives, to gain and maintain competitive advantage, and to manage their organizations effectively.

With its scalable parallel system, RS/6000 is a leader in parallel processing. The RS/6000 SP is a proven superstar. In the seven years since its debut, installations have grown more than tenfold with organizations worldwide benefiting from the SPs combination of UNIX, parallel computing, and IBMs unmatched record for mission-critical reliability, support, and security. Today, SPs are being used by customers to meet their daily challenges: To mine corporate data in support of enterprise strategies, to take advantage of the Internet to reach millions of customers, to support complex research and development projects, and to manage vital business operations.

The new RS/6000 Enterprise server model S80

The latest model for enterprise computing is the RS/6000 Model S80. The RS/6000 S80 server improves on the previous S7A line at the high end. The S80 doubled to tripled current S7A performance, doubled the number of Power PC processors from 12 to 24, and benefited from advanced copper chip-interconnect technology.

On July 1, 1999, IBM announced a new world record for Internet performance. The Web-serving record was set on an RS/6000 S80 Enterprise Server, a six- to 24-way, 64-bit system expected to become available in October, 1999. Fast servers that are reliable and can handle large numbers of users are considered critical to companies transitioning their businesses and their customer relationships to the Internet. A 12-way S80, running AIX pre-release Version 4.3.3, the IBM UNIX operating system delivered the best SPECweb96 benchmark result. SPECweb96 measures the maximum number of Hypertext Transfer Protocol (HTTP) operations per second.

On October 5, 1999, IBM announced that the RS/6000 S80 Enterprise Server set a new world record for transaction processing performance. Based on the TPC-C benchmark, the S80 is the world's most powerful single-server system for transaction processing. The S80 is also the best value among large enterprise servers with the lowest cost per transaction among the 10 most powerful systems. The test was performed using the IBM UNIX operating system, AIX 4.3.3, and the Oracle 8i Version 8.1.6 database.

Extensive reliability testing was performed to see if the application would crash. IBM was pleased that it didn't. Both performance and dependability of the RS/6000 system running Oracle have since been outstanding, and IBM is looking forward to the increased performance offered by the RS/6000 S80 system.

These results demonstrate clearly that IBM's long-standing technology and marketing alliance with Oracle delivers customer value. For detailed product specifications and performance information please visit:

http://www.rs6000.ibm.com/hardware/enterprise/s80_specs.html

RS/6000 continues to win awards

In July 1997, the RS/6000 model F50 was declared the most complete pre-configured (Web) solution in an InfoWorld comparison of leading Web servers, taking top honors with a score substantially higher than its nearest competitor.

Based on the success mentioned above, the RS/6000 model F50 was named InfoWorld 1997 Product of the year in both the server and solution categories. The InfoWorld Test Center used F50 as a test platform for the center.

RS/6000 was awarded the Computerworld Reseller Choice Award for 1997 in the mid-range, multi-user workstation category.

RS/6000 received awards and nominations in the categories of workstations, network servers, and middleware from *Network Computing* and *CIO* magazine.

AIX - The best rated UNIX with superior e-business support

With its comprehensive 64-bit support and superior Internet features, AIX is able to take advantage of advanced e-business applications that require large scale database handling, advanced security features, and extensive Java support. At the same time, AIX supports existing 32-bit applications with no change or operational disruption.

AIX's industrial-strength functionality includes easy-to-use installation and administrative tools, high-function file and data management capabilities, and outstanding PC and mainframe connections. Its Web-based systems manager provides access and management from any point on an intranet or the Internet. Access to documentation in the HTML format gives the user an intuitive interface and eases search and navigation.

Although e-businesses needs protection and secure communication, the IP Security (IP Sec) supplied with AIX 4.3 is even better. It supports unlimited filter rules that control network traffic by characteristics, such as: Source and destination address, interface, type of TCP/IP, subnet mask, specific protocol, and port, such as ftp, mail, and so on.

eNetwork Lightweight Directory Access Protocol (LDAP) V1.1.1 with IBM DB2 Director is included in the base operating system. This duo supports up to four million directory entries with sub-second search response time, and the directory can be local. Unlimited directory replication is included for no additional charge.

The pervasiveness of Java in the industry now requires superior Java support on the server. The new Java Development Toolkit (JDK) V1.1.4, with IBM Just in Time (JIT) Compiler V2.01, is now part of the base operating system so that Java loads automatically. Plus, Java performance has been improved up to 20 percent with the new JDK and JIT Compiler.

AIX is packaged with commonly used facilities such as the Adobe Acrobat Reader, the Netscape Navigator, Netscape FastTrack, Lotus Domino Go Webserver, IBMs Ultimedia Services, Novell Network Services with Novell Directory Services, and many more. The AIX environment supports an extensive array of middleware packages, such as Distributed Computing Environment (DCE), high availability options, object technology for streamlined development, and systems management products as well as the

industry-leading database products, such as IBM DB2, Oracle, Informix, Progress, CA-Ingress, Sybase, the Pick databases, and more.

The AIX development environment has been certified by the Information Technology Association of America as having the core capabilities needed to address the Year 2000 challenge. AIX development participated in a rigorous evaluation of its approach to date conversion with extensive analysis in eleven areas deemed necessary to a successful Year 2000 conversion. Products developed using these processes are the AIX operating system.

In a VAR Business Magazine poll (October 1996), AIX was co-winner sharing top honors in reliability/availability, performance, and security.

AIX was named the best UNIX system and top-rated commercial UNIX operating system by D.H.Brown (*AIX Sustains IBMs Leadership in Open Systems*, May 1997).

Competitive edge in the Web-connected world

For companies using the Internet or an extranet to give their enterprise applications e-business reach, the RS/6000 with AIX Version 4.3 is the superior platform.

RS/6000 is a Web-serving powerhouse. At the Nagano Olympic Games, it did set a world record - over 103,000 hits per minute, 103,429, to be precise, at 9:00 p.m. Japan Standard Time on February 20, 1998. In all, the RS/6000 SP flawlessly handled nearly 650 million Web site hits and was the most heavily used Internet-based technology application in the world to-date. The RS/6000 was also the power behind the Results System that processed timing and scoring information from each event - in close to real-time - and delivered it to scoreboards, sportscasters, broadcast media, World News Press Agency, and the Info 98 site (an intranet site also run on an RS/6000). And it wasn't all SPs in Nagano. The World News Press Agency application, responsible for delivering comprehensive reports to 123 global news agencies, was hosted by an RS/6000 43P functioning as a transmission server.

IBMs Web site for the Deep Blue vs. Garry Kasparov rematch won the 1997 Cool Site of the Year award in the Live Internet Event category. Winners in the Web's best-known awards competition are chosen by the votes of Internet users.

The RS/6000 model F50, using the Internet Connection Secure Server solution, took top honors in the InfoWorld Comparison of Web servers published July, 1997.

For a fast, easy start-up on the World Wide Web, RS/6000 Internet POWERSolutions combine the RS/6000 hardware, AIX, and a choice of industry-leading application software from Netscape, Lotus, and IBM. The functions necessary to get started doing business on the Internet or to set up an intranet for internal communications and collaboration are provided, thus, helping to make implementation faster and less costly. RS/6000 and AIX offer a choice of solutions based on requirements for security, firewall protection, and other functions, such as local replication of Web page content and access to existing business systems.

It has managed Web sites supporting the U.S. Open and Wimbledon in the tennis world and the PGA and Masters golf tournaments.

The IBM Global network is one of the world's largest Web services provider. Its platform? Of course, the RS/6000.

Applications, applications, applications

The RS/6000 runs thousands of applications provided by vendors committed to competitive business solutions on open systems. The RS/6000 and AIX support the key UNIX standards, and IBM has taken a leading industry role in defining and implementing standards, such as the Common Desktop Environment (CDE), the Distributing Computing Environment (DCE), and many others.

The RS/6000 is ready for Java. The RS/6000 and AIX are fully enabled to take advantage of this explosive growth. Java server-side applications can take advantage of the security, reliability, and high performance of the RS/6000 as they deliver their "write once, run anywhere" functions to users worldwide.

As of February 19, 1998, the RS/6000 SP has moved itself to the top of transaction processing. The RS/6000 SP, with Oracle8, IBMs TXSeries, and IBM Serial Storage Architecture (SSA) disk, set new world records for the number of transactions-per-minute-C and the number of concurrent users performing transactions on the system. The RS/6000 SP-TXSeries-Oracle transaction processing solution set these records by simulating a real-world, high transaction customer scenario as measured by the Transaction Processing Council (TPC) transaction throughput benchmark.

For organizations looking to deploy business intelligence applications to gain a competitive edge, the RS/6000 supports a broad range of the industry-leading software, such as IBMs DB2 OLAP Server, Arbor Essbase Server, Oracle Express Server, Sybase IQ, SAS, Microstrategy, and Informix Metacube. And the RS/6000 is an industry leader in price/performance. The

RS/6000 models F50 and H50 received the top two price/performance ratings in the 100 GB class, and the RS/6000 SP holds the top position in the 1T class for price/performance (Ideas International as of 1998).

IBM works with the leading providers of software worldwide to help them optimize their offerings for the RS/6000 family. IBM and some key ISVs operate joint competency centers to help ensure customers get the expertise and support needed for successful application implementation. In addition, IBM operates application porting centers worldwide to assist vendors who are making their products available on the RS/6000.

RS/6000 specialists worldwide work with Business Partners (over 7,000 strong) to combine packaged products and unique development and services to insure the best possible solutions to customers' business problems.

The RS/6000 is a star of the growing Enterprise Resource Planning (ERP) software industry.

- The RS/6000 is one of the most popular UNIX platforms for companies implementing SAP's R3 solution for enterprise management.
- Customers implementing Symix's SyteLine ERP solution choose the RS/6000 twice as often as the next popular choice (HP). (Gartner Group, ERP Vendor Guide 1997: R-345-131, July 23, 1997.)
- Advanced Manufacturing Research reports in their advisory to clients (*Present Market Position*, July 7, 1997) that IBM's AIX is the most popular platform for QAD's Mfg/Pro sold in the U.S. in 1996.

Flexible, adaptable family of systems

The RS/6000 has the right system for the job at hand to help you develop and maintain a competitive advantage. The RS/6000 family has models that are economical for applications serving only a few users and models that have the capacity for the largest jobs with the right steps in between. It has the ability to grow with workload and opportunities.

The RS/6000 has the connections to access a wide variety of networked systems. Customers use RS/6000s to access mainframes running legacy applications, to connect their users and applications across the enterprise, and to benefit from the latest in Internet applications.

The RS/6000 supports industry attachment options allowing connection of low-cost PC-like devices as well as devices used for industry-specific applications. Devices, such as programmable logic controllers and digital control systems, allow companies to run entire manufacturing and process operations with an RS/6000 system.

The RS/6000 and the S/390 are a powerhouse combination. Using high-speed connections (ESCON, ATM networks, FDDI, and block multiplex), customers can take advantage of the unique strengths of both systems. In large e-business environments, where many Web-application servers are coordinating activity across multiple databases and services on behalf of thousands of simultaneously active clients, the SP can be used to house and manage Web-application servers, and the S/390 can be used for the transaction processing applications and database storage. In SAP R/3 environments, the S/390 can support transaction and database processing while the SP manages SAP application processing.

The RS/6000 supports storage options that provide state-of-the-art capacity, reliability, and management functions enabling your business to profit from leading-edge decision support and data mining applications.

High availability solutions

IBM understands the critical availability needs of commercial customers like no other vendor. In today's demanding e-business environment, 24-hour, 7-day availability is more than desirable; it is a competitive requirement.

HACMP, the RS/6000 high-availability software, is providing high-availability protection for everything from airlines reservations, credit verification, cellular phone service, fleet services, trading systems, and distribution to entertainment and gaming environments.

HACMP is top rated; D.H. Brown rates HACMP for AIX as the leader for multi-system high-availability. HACMP took home first place overall and was first in hardware/software failure recovery, HA configuration support, backup/recovery detection, and disaster tolerance. It tied for the highest honors for service processor features, concurrent access, and HA administration (D.H. Brown: *High Availability for Clusters: Functional Analysis*, May, 1997).

HACMP, configured through SMIT, can recover from a wide variety of failures. These include TCP/IP LAN network adapter failures, TCP/IP software subsystem failure, server failure, and disk, disk adapter, or disk cabling failure. By using application server scripts, HACMP has the capability to restart an application, transfer wide area communications resources, effect ADSM failover, and plan for additional sophisticated hardware/software failover actions.

Broadest range of scalable, compatible systems

The RS/6000 family offers an impressive RISC product line. The range begins with desk top systems and extends to the high-performance RS/6000 SP

systems with hundreds of parallel processors - all managed by AIX running industry-leading applications. Because AIX is binary compatible across systems, investments in people skills and applications continue to bring benefits as customers grow into larger and more powerful systems.

The 64-bit capability of the RS/6000 Enterprise Server S80 and AIX Version 4.3 extends the RS/6000 family's scalability through support of high-performance, 64-bit databases, therefore, meeting the need for the complex functions demanded in e-business and data mining environments.

No other product family satisfies so many challenging customer requirements across such a broad range of computing environments. This spells growth potential, extended system life, and, most important, helps protect investments in systems, applications, and the people skills needed to take advantage of them.

Award-winning Systems Management

AIX offers functions that protect system resources and make it easier to use. The Journaled File System provides protection against unexpected system outages. The Logical Volume Manager simplifies daily operational tasks, and AIX gives administrators the choice of using either traditional UNIX commands or easy-to-use Web and GUI interfaces. In fact, by using these interfaces, one can administer systems without knowing UNIX at all.

With the acquisition of Tivoli Systems by IBM, the RS/6000 is further enhanced as a UNIX systems management leader. The combination of Tivoli's standard-setting technology for the client/server environment with IBM's outstanding products for mission-critical applications management provides RS/6000 customers state-of-the-art control of their networked, multi-vendor environments. TME/10 provides functions, such as user administration, console management, software distribution and inventory, job scheduling, and help desk, in an integrated, distributed environment.

Outstanding reliability backed by worldwide support and service

RS/6000 uses an innovative design that includes state-of-the-art reliability features. Double-bit error detection and single-bit error correction memory and hot-swappable internal and external disk storage units combine to contribute to the excellent reliability of the RS/6000 product line. The service processor feature on the SMP models provides unattended recovery and automatic support calls, features that are unmatched in this price/performance category of UNIX servers.

A standard one-year, 24 hour/7 day hardware maintenance warranty is included with every RS/6000 purchase worldwide. A complete portfolio of

support services, ranging from telephone support and consulting to on-site support and specialized services, is available. These include:

Project support services

- Operating system porting/conversion
- Operating system migration assistance
- Systems integration
- IBM and non-IBM software customization
- IBM application development
- Site planning services

Continuing support services

- Customer Support Center services
- On-site software maintenance support
- Capacity planning
- Maintenance services including multi-vendor environments
- Technical/application specialists
- Network custom services
- Education

The May, 1996 VAR Business Magazine's annual mid-range server review rated the RS/6000 as the most reliable server with the best technical support.

Chapter 2. Benefits and types of server consolidation

As companies react to worldwide competition and economic trends, their agenda contains mergers, acquisitions, demands on IT for new applications to address global markets, greater customer intimacy, and ever faster access to more information.

At the same time, pressure is mounting to reduce costs, maintain or improve service levels, and maintain or improve the resilience of systems that become ever more critical to daily operations.

Users want new applications that are delayed or inadequate because of IT infrastructure. IT needs to provide a cost effective and reliable service, which is made difficult by constantly changing applications.

The proposition of Server Consolidation is this: Only by simplifying the IT Systems Architecture can one grow the scale and complexity of systems while containing the cost and energy consumed in their maintenance.

Many organizations are finding that, as the number of servers proliferates, the cost and operational complexity are also rapidly increasing. In many cases, there are concerns whether multiple distributed servers can provide the application availability, hours of service, responsiveness, and ability to grow with the requirements of the business. These characteristics are being increasingly demanded by business applications. To reduce these costs, many customers are attempting to consolidate their servers into a more manageable central location.

The main objectives of server consolidation are:

- Recentralizing servers
- Merging workloads onto a single large server
- Consolidate Architecture
- Optimize the IT infrastructure

2.1 Benefits from consolidation of servers

The benefits that fueled the movement to decentralization and distributed computing were often costly to achieve or did not materialize at all. Since 1990, many IT organizations have done a 180 degree turn in moving from a decentralized architectural structure to moving towards a more centralized

environment. The relative importance of this shift is that it tends to validate the needs and benefits of server integration.

Analysts and IT professionals say proliferation of systems is driving the movement to rein in servers. Other enterprise professionals have also launched their own consolidation strategies. What is driving consolidation are the changes in business practices, therefore, making it logical for users to collect their applications, LANs, and data-center functions.

Figure 4 summarizes the reasons for IT organizations to move toward server consolidation.

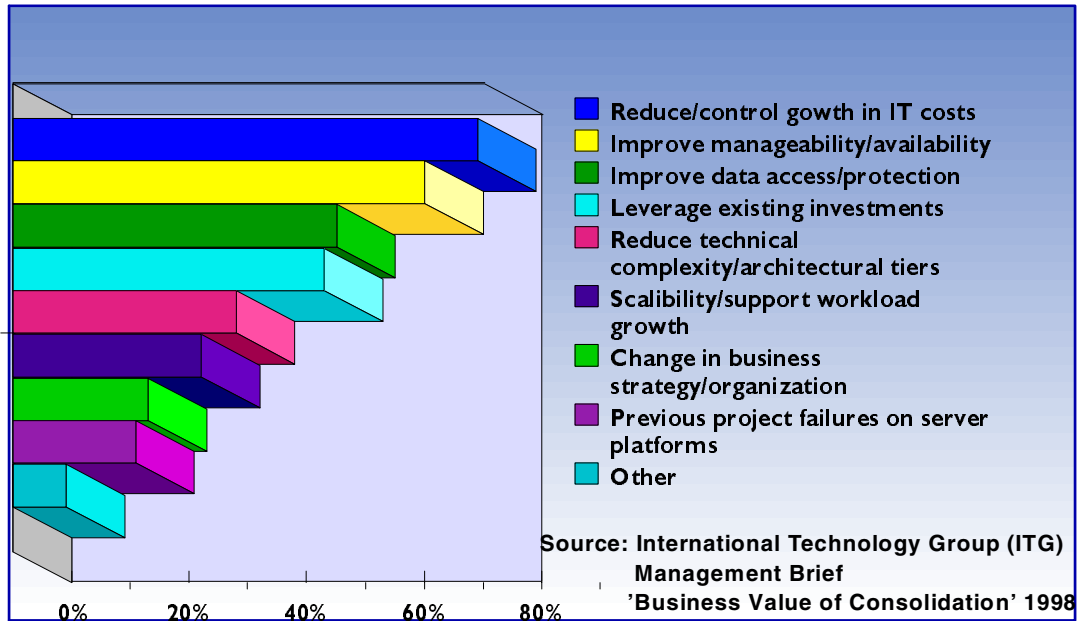


Figure 4. Reasons for consolidation

In a consolidated infrastructure, support and maintenance costs are dramatically reduced. Thanks to centralized and automated procedures, administration, back-up, and operating costs are also decreased. The costs incurred for the distribution and the management and the update of software are reduced, and there are savings in floor space and electrical consumption, given that fewer machines are necessary.

The main benefits of a server consolidation operation are:

- Single point of control
- Better services to users
- More flexibility
- Minimize learning through discovery and optimize use of deeply skilled resources
- Avoid floor space constraints

Some of the business benefits of server consolidation are:

- Operational cost reduction
- Improvements in service levels
- Automatic application of service level metrics with exception reporting
- Improved capacity planning (and, thereby, financial planning)
- Improved potential for future application rationalization
- Improved potential for future data consolidation

Generally speaking, the driver for server consolidation falls into two groups:

1. Improved service to customers or end users

- Improved availability of service (fewer or no outages, and shorter outages should they occur).
- Potential for improved service times to the end user due to a reduction of the time to communicate between clients and servers in single location.
- Opportunity for new services by enabling or improving interaction between different applications or between new applications and data in existing systems.
- Web serving/Electronic commerce based on secure access to data in existing production systems
- Improved security and integrity of data.
- Support for larger numbers of users accessing the same application by running on a larger server than may be cost justified in a distributed environment.

2. Lower operational costs

- Lower systems management costs
 - Network management
 - Configuration management

- Problem and change management
- Operational management for both automated and manual operations
- Security administration costs
- Single pool of skills in one location rather than many
- Reduction or elimination of user department operational costs
- Possibility of reducing some software licenses
- Possibility of reducing the number of systems, disk storage costs, elimination of maintenance charges, and so on
- Potential for greater scalability on a large centralized server, for example, IBM RS/6000 SP

Costs may be lowered further if multiple applications can be run on a single server and if the peak activity in any given workload occurs during periods of lower activity for the other workloads. In this case, the server size may be much less than the sum of the sizes of the servers that it replaces. This last benefit may only be achievable if the operating system used is capable of dynamically balancing the workloads according to both business priorities and the resource constraints of the server.

2.1.1 Single point of control

Enterprises with a strong mainframe heritage intuitively understand the value of central control. This is more than a statement of the organizational and power structure of a firm, but rather a recognition that benefits of disciplined standardization are best achieved through a central point of control. Indeed, rapidly growing firms, especially those growing through mergers and acquisitions, frequently felt that disparate distributed systems were so unwieldy to manage that they were losing control, which could constrain further corporate growth.

A single point of control allows enterprises to:

- Reduce or eliminate department operational costs
- Reduce some software licenses
- Reduce number of systems, disk storage costs
- Reduce maintenance charges
- Avoid multiple copies of the same application on distributed systems
- Reduce owner operational costs
- Offer better availability of service

- Improve systems management
- Have better version control management
- Have better software distribution
- Reduce risk and increase security

2.1.2 Giving users better services

With a consolidated infrastructure, end-users can count more easily on round-the-clock service, seven days a week. The response time is much better than with an overly distributed environment, and the data is more easily accessible while being highly protected. The control procedures are simpler, while security becomes even higher. And information sharing is improved, giving end-users increased data consistency. The availability of service is improved mainly due to a reduction in the time needed to communicate between clients and servers in a single location.

2.1.3 Regaining flexibility

Globalizing and optimizing resources means that evolution to meet new needs becomes easier: A higher volume of strategic corporate data and a growing number of end-users. The standardization of procedures, releases, and servers also makes it easier to install new application software, for example, Internet and intranet, electronic commerce, and so on. In today's fast moving environment, computing resource consolidation enables a trouble-free upgrade of the information system and less costly adaptation to organization or environment changes. Enterprises can react more quickly to market changes since storage is readily available and can easily be reallocated.

2.1.4 Minimize learning and optimize use of skilled resources

Locally controlled systems are somewhat unique in their combination of servers, storage systems, operating systems, and applications. Therefore, each site can have problems that have not been experienced at other sites within the firm. This problem becomes particularly acute when different hardware architectures and operating systems are involved, such as results from mergers and acquisitions, or even from excessive departmental freedom to select hardware and software without consideration of accepted corporate standardization.

Under the distributed alternative, systems management responsibilities are often only part-time extra duty assignments such that a critical mass skill level is rarely achieved. Furthermore, since other departments may employ

disparate architectures and applications, there is little opportunity to benefit from the experience of others. On the other hand, a central, consolidated server site with standardized hardware and software can readily justify dedicated experts with the in-depth training to quickly resolve problems. Therefore, re-hosting to a converged set of applications, operating systems, and platform architectures is a means to lead enterprise personnel to develop an expertise in optimizing operations as well as enabling quick problem resolution.

2.1.5 Avoid floor space constraints

As reported from discussions with server consolidation sites, surprisingly often limited floor space may seriously constrain growth. While a small server may be easily shoehorned into a closet, as compute demands increase, enterprises find that suitable floor space is hard to find for proliferating small servers. The solution is a central site outfitted with appropriate power, cooling, access to communications links, and so on, and populated with more powerful systems, each giving more performance in the same footprint.

2.1.6 Reduction of TCO

There are several costs associated with server consolidation: Hardware costs, software costs, disruption costs, and hidden costs. Added up, these compose the Total Cost of Ownership (TCO). All these costs are detailed in the following sections.

2.1.6.1 What is TCO?

TCO is composed of several components that contribute to the maintenance of systems.

Hardware costs

Numerous hardware costs can be generated by a server consolidation operation, for example, costs generated by:

- New servers
- New storage systems
- New infrastructure, for example, cabling, cooling, rooms, and so on
- Upgrade of existing hardware
- Upgrade of existing infrastructure

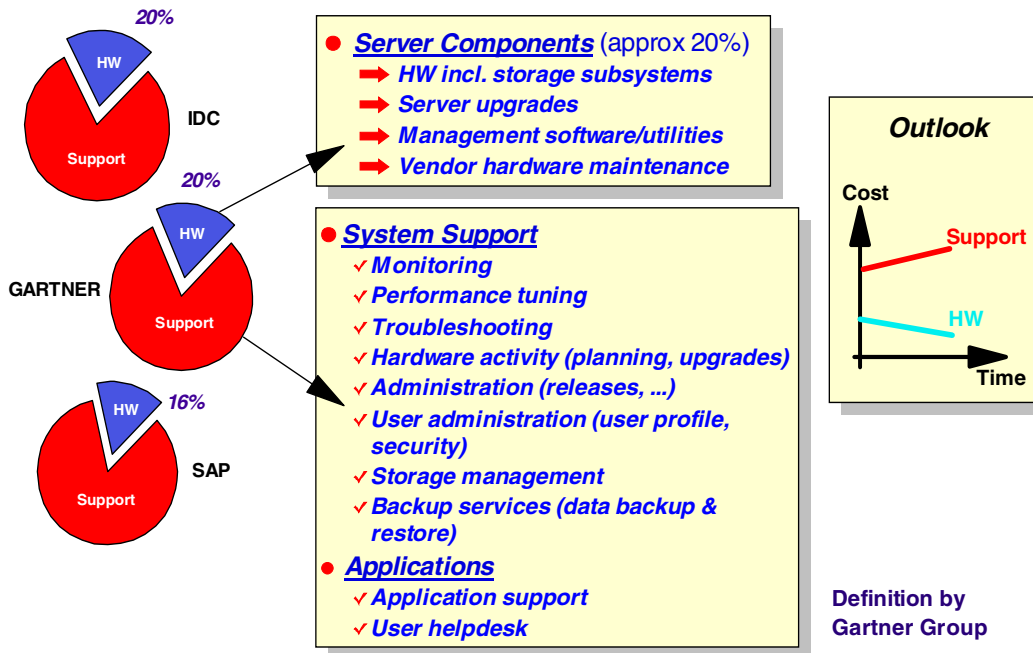


Figure 5. TCO in a distributed world

Software costs

Another area to examine is the cost of software licenses. This is usually a major expense for distributed servers. Typically, enterprises need to buy a license for each server. By consolidating workloads into larger servers, organizations can save money on reduced software licenses. Also, an area that has to be analyzed is the one where a business application is split among multiple distributed servers and where the different parts of the application run on different servers. Merging all parts into the same server decrease software costs by providing:

- Lower operational costs
- Better availability of service
- Simpler, more reliable operation
- Greater scalability
- Improved performance
- Improved systems management
- Lower overhead by removing communications between parts

True total cost of computing: Most distributed systems costs are not visible, whereas data center cost is fully visible. Consistently, studies have shown that in comparing total costs of a multiple server environment versus those of a more centralized solution, costs are less in the latter case.

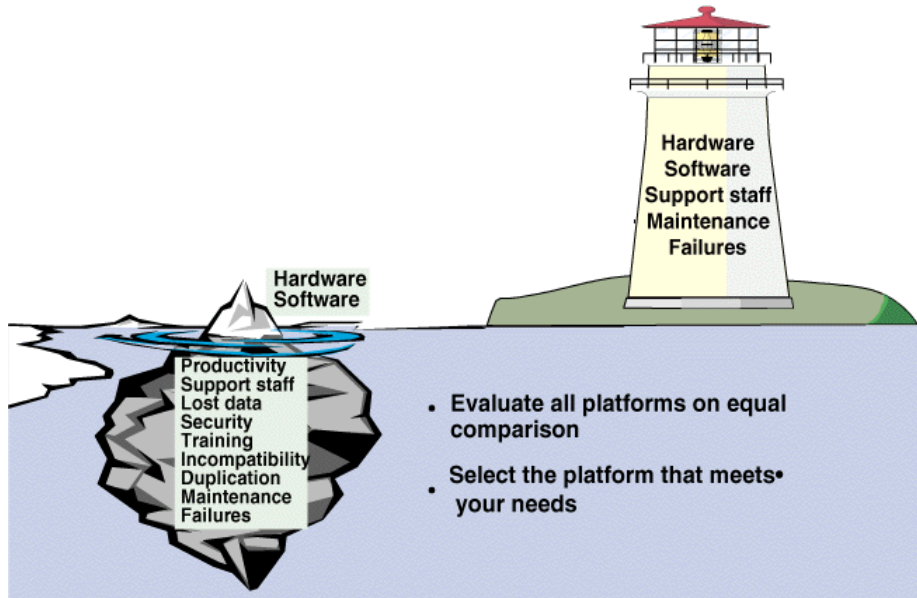


Figure 6. True total cost of computing

Economies of Scale and Scaling Economies: Integration of many servers onto one platform can provide clear economies.

Disruption Costs

Management is concerned about the dangers of business disruption. The manager is looking at a migration to new hardware (servers or storage systems), operating system, and networking platforms. Each of these changes has to be handled without (or with a minimum of) disruption in a flexible, vendor-independent manner. The same has to be applied for software before introducing new technology, integrating new applications, or applying changes to an existing environment.

There is real business value in a system or software package that allows enterprises to avoid or minimize the disruption costs. In general, when dealing with existing systems, less disruption is better. Today, systems are running the entire business of enterprises. They contain the information assets and business logic, and they are the foundation of the future.

The costs associated with the service interruption while the upgrade is in progress have to be sized. All these disruption costs can be minimized by high systems scalability. The enterprise should have:

- The capability to start with small servers and grow without disruption as its needs expand.
- Hot-swappable, scalable, and non-proprietary storage units.
- Solutions for duplication, with minimal disruption to application availability and use of system resources.
- The ability to incrementally add data and power to the system at a low cost with minimum disruption as workloads increase.

Hidden Costs

Many hidden costs are generated by server consolidation and have to be estimated before going to this operation.

Testing systems costs: While there are compelling benefits in concentrating independent servers into larger systems for workload consolidation and software standardization, there is a need to isolate development and testing systems from mission-critical production servers. This leads to separate operating environments in order to guarantee security and fault isolation. The need for independent development and test systems can be addressed either by partitionable servers or storage systems or small stand-alone systems from a broad, compatible product line.

Support costs: Grouping existing support staff or finding a service contractor can generate additional costs.

Backup costs: When central backups are considered, even if the equipment and support costs will be reduced while the reliability of the service will be increased, the cost of the storage consolidation has to be considered.

Upgrade costs: Finally, there are the costs associated with server consolidation manpower:

- Staff involved in the upgrade.
- Travel expenses to the remote site.
- Upgrades done by sub-contractors.
- Upgrades to distributed servers, which are typically serial in nature.
- Upgrades for centralized servers, which can be grouped for efficiency.

Reduced end user labor costs

The objective is to transfer some of the end user labor for systems management back into the data center. This means having people trained and specializing in file and database archiving. This also means having data center people managing the configuration centrally, thus, assuring that installation parameters are consistent and correct.

A special feature of the SP system allows you to manage security information centrally and have that data propagated throughout the system.

Netview, Trouble Ticket, and various RS/6000 tools operate on the SP system making it easier to manage.

Reduced software cost

A software license can be installed anywhere in the network. However, in today's environment, there are multiple copies of software spread through the network along with the licenses server. By consolidating software into the SP system, you can centralize use of the more expensive software subsystems. In addition, software distribution gets easier since the software is inside a single system. Configuration management is also simplified.

It is also often true that users have many software licenses because they have small servers. To scale up, the software license can go to one larger server and reduce the number of licenses because a single server can handle more users. Often, the "per user" license charge favors the customer when more users are added to one big node instead of many small nodes with the same per user count.

2.1.6.2 How to reduce TCO

Gartner Group has provided data that clearly shows that the majority of the Total Cost of Ownership is directly related to support costs.

Support costs are on the rise, while hardware costs are decreasing.

There are three ways to reduce TCO.

The total cost of ownership can be reduced:

- By reducing the complexity and costs associated with managing the server environment.
- By optimizing the capacity of the servers and utilizing the equipment more effectively.
- By optimizing the utilization of peripherals.

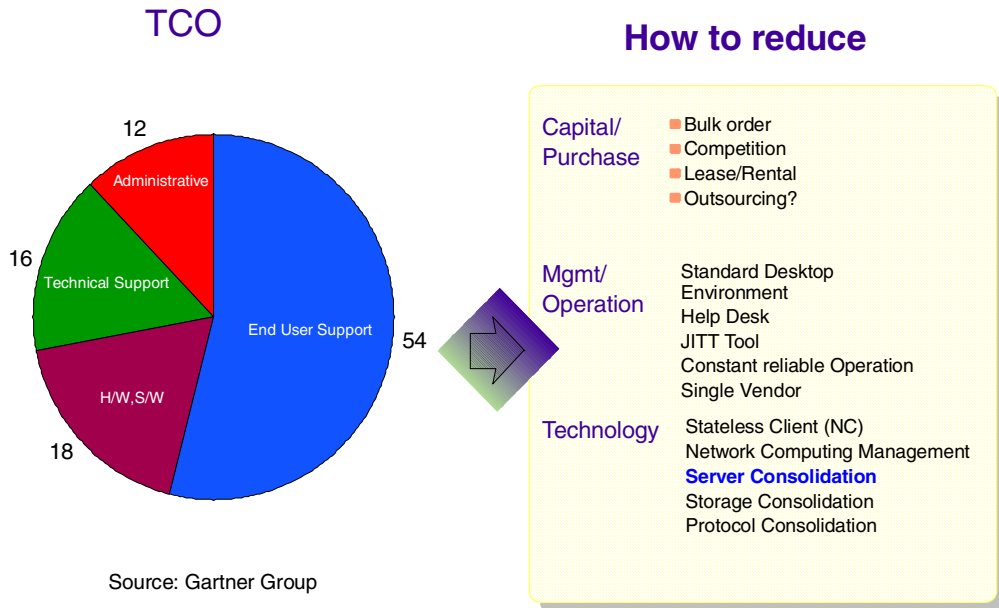


Figure 7. How to reduce TCO

The Gartner group has endorsed the view that strategies for reduction of the TCO can be grouped into three types: Capital/Purchase, Management/Operations, and Technology.

Capital / Purchase

For example:

- Bulk order
- Competition
- Lease/Rental
- Out-sourcing

Management / Operations

Enterprise Management - Integrated operations allows for consistent management of all facilities and IT services.

Continuous Operations - 24x7 operations have become the norm. Most systems and applications can provide business benefits from continuous operations.

Consistent performance - Providing consistent response time at peak load periods is very important.

Dependability - Commonly cited problems of distributed environments include frequency of outages and excessive requirements for manual intervention by IT staff.

For example:

- Standard Desktop Environment
- Help Desk
- JITT Tool
- Constant Reliable Operation
- Single Vendor

Technology

For example:

- Stateless Client (NC)
- Network Computing Management
- Server Consolidation
- Storage Consolidation
- Protocol Consolidation

D.H. Brown Associates conducted a study of RS/6000 SP customers who installed an SP system specifically for server consolidation. Refer to Chapter 3, "Why consolidate to the RS/6000?" on page 45 for the details about this study.

Twenty-four percent of these customers choose an SP system for the System Management features SP delivers. Reduction in overall system costs, floor space, and administrative cost were also strong reasons for their choice of SP. Scalability, the SP Switch architecture, and IBM support were also cited as reasons for the SP selection.

2.1.7 To provide enhanced functionality

To enhance their functionality, organizations are motivated by a need to provide better response times, increase access to data, and deploy new applications rapidly. This can be broken down into the following strategies:

- Extend system's scalability and flexibility for customers who are limited by their system

- Standardize applications and services
- Upgrade existing enterprise systems
- Increase their system's interoperability and data sharing

2.1.8 To build strategic infrastructure

Management of system flexibility encompasses needs, such as the ability to scale processing and storage capacity without adding physical devices or subsystems, as well as the flexibility to partition and allocate resources as needed. The system's flexibility can be hardware or software flexibility and can be considered in different ways:

- Flexibility in integrating existing hardware, such as servers or storage systems.
- Flexibility in deploying new applications.
- Flexibility in running legacy applications.
- Flexibility in integrating old and new applications. New applications can interface to current applications and, therefore, risks are minimized.

Standardize applications and services

Customers, user groups, and vendors have all recognized the need for standardization. The importance of standards relates to portability. If a program is implemented solely within the specification of a standard, it should be easily ported to another system claiming conformance to that standard. The smaller the scope of a standard, the more portable an application written to that standard is. The larger the scope of a standard, the easier it is to write an application written to that standard. Standardization can be on:

- Hardware vendors
- Hardware platforms
- Software products
- System interfaces
- Operating and system management procedures

2.2 Types of consolidation

Server consolidation is the trend to centralize business computing workloads to reduce cost, complexity, network traffic, management overhead and, in general, optimizing and simplifying existing IT infrastructures and providing a foundation for new solution investment and implementation.

There have been many studies of the total cost of distributed client/server computing, but all analysts agree on one thing: In a “pure” client/server implementation across a large enterprise, the hidden costs are staggering. The components are already familiar: High maintenance and support demands, reduced end-user productivity, increased training costs, multiple-site software licensing costs; poor security, increased downtime, and lost data.

Server consolidation can be placed into the following types:

- Centralization
- Physical Consolidation
- Data Integration
- Application Integration

Centralization and physical consolidation are relatively easier than data and application integration. The following are considerations when doing UNIX server consolidation:

Resources

- Is there enough CPU, memory, network, I/O resources to combine the workload?

Availability characteristics

- Is the application mission critical? Does it require 24X7 availability?
- Combining workloads of this nature introduces additional risk of downtime and is generally not advisable.
- Is the application part of a high-availability cluster?
- If the application requires 24X7 availability, and the customer is willing to dedicate a hot standby or HA spare, this application is probably not a good candidate for combining with other applications.
- What is the maintenance window for the application? If application A can be brought down only on Saturday night, and application B can be brought down only on Sunday, there won't be a common window for system reboots, and so on.

General

- Does the application use hard coded path names or hard coded TCP/IP ports?
- There may be conflicts in the names of binaries between applications, for example, /usr/bin/start could be the same for two programs. Hard coded path names may not allow test and production systems to run together.

- Does the application include modules or programs that must run in kernel mode, such as device drivers?
- Anything that runs in kernel mode can bring down the entire system. If the vendor does not do a good job of testing, introducing programs that must run in kernel mode can be risky. It is fairly common for applications of this nature to require a reboot to install new versions.
- Does the application make use of background processes or daemons? If the application uses these processes, does it provide routines to stop and start the daemons? Do these daemons run as root?
- Some applications, such as Netview/6000, have daemons that must run as root. When those daemons “misbehave,” the system sometimes requires a reboot to clear them. Applications that do not have memory resident portions are easier to combine.
- Does the application require a reboot for any maintenance operations?
- Does the application scale by providing redundant instances that can be maintained individually? For example, most DNS implementations include a primary and one or more secondary instances. These kind of applications can often be rebooted or changed, if required, during the day as another system can pick up the workload.
- Does the application require any operating system specific tuning, for example, network tuning options?

Security

- Does the application require root authority to start, run, or stop?
Applications of this nature require extensive use of the root userid, which can increase the risk of an outage.
- Is the application supported by an outside third party? Will a third party need to dial into the system?
- Is the application supported by a separate group than the underlying operating system? If the application requires root authority and is supported by a different group than the operating system, potential conflicts can occur.
- Does the application live outside a firewall, on a DMZ, or on any other restricted network?
- Does the application contain confidential data, such as payroll information, that would be maintained by a sub-set of systems administrators?
- Does the application require the use of modems? Modems on a system can increase the security threats.

The following four categories (shown in Figure 8) can be grouped into two consolidation phases, physical (centralization and physical consolidation) and logical (data integration and application integration).

Based on a review of our customers' consolidation architectures, we have defined four general classes (shown in Figure 8) of server consolidation implementation strategies.

What are the Strategies?

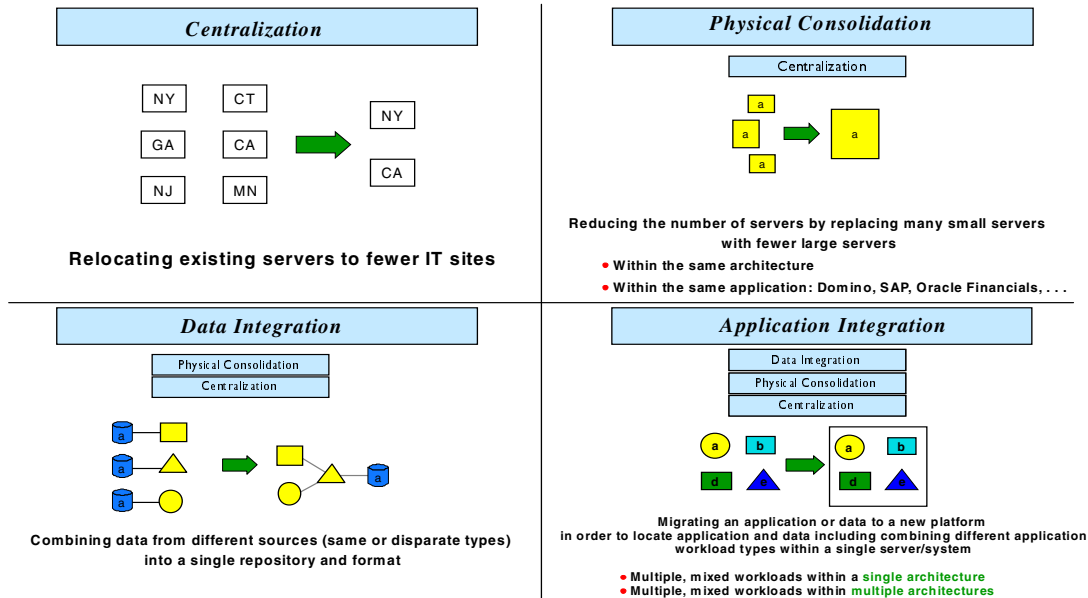


Figure 8. Four types of server consolidation

2.2.1 Centralization

Server consolidation means different things to different people. In its simplest form, servers are physically moved to a common location. Because this simplifies access for the IT staff, it helps reduce an operation's support costs, improve security, and ensure uniform systems management. This is an important predecessor to future consolidation activities.

Relocating existing servers to one or fewer IT sites

Although this is a type of server consolidation, it obviously offers minimal immediate opportunity. Centralization, data center consolidation, may be a

first step for an organization after a merger. After a merger, the resulting entity does not want to attempt merging applications; however, they will co-locate their systems as a first step.

For both servers and storage systems, two sub-categories of centralization are defined: Virtual centralization, which is mainly made through the network, and physical centralization, where hardware is physically moved to different locations.

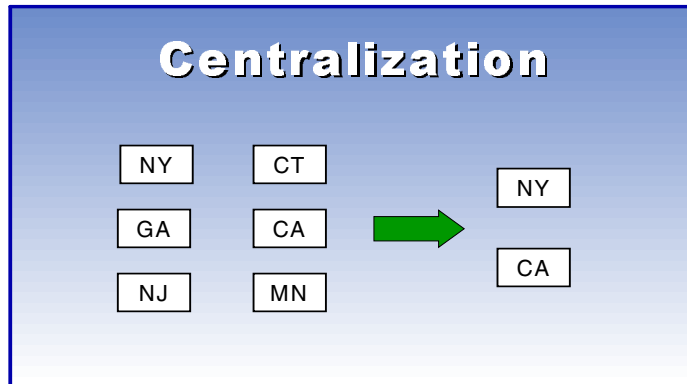


Figure 9. Centralization

Virtual centralization or remote management

Physically dispersed servers or storage systems are logically centralized and controlled through the network. Hardware remains physically distributed but is brought under a common umbrella of systems management and network management tools. An operation's costs can, therefore, be reduced, and system availability can be improved.

Physical centralization or server relocation

Existing servers or storage systems are physically relocated to one or fewer IT sites. Because this simplifies access for the IT staff, it helps reduce operations support costs, improves security, and ensures uniform systems management. This is a step in the right direction, but the payback is relatively low. However, it is an important predecessor to future consolidation activities.

2.2.2 Physical consolidation

Physical consolidation is the practice of simply replacing small servers with larger servers of the same breed. This consolidation does have advantages: It improves availability because there are fewer points of failure; it can reduce

the cost and complexity of system communications, and it simplifies operations.

It is fed by Moore's law: The fact that the number of transistors that can be put on a chip doubles approximately every 18 months means that more powerful "boxes" are always available.

Reducing the number of servers by replacing many small servers with fewer large servers

Physical consolidation may be implemented on a site, department, or enterprise basis. Two environments, Intel and UNIX, are proving to be good candidates for physical consolidation. In the Intel world, file/print servers are being consolidated onto newer, much faster, more reliable Netfinity boxes. The UNIX environment is also seeing older UNIX boxes with high hardware maintenance costs being replaced by newer, much faster, cheaper-to-maintain processors.

As described in the following sections, the physical consolidation category is divided into two major sub-categories: Server consolidation and storage consolidation.

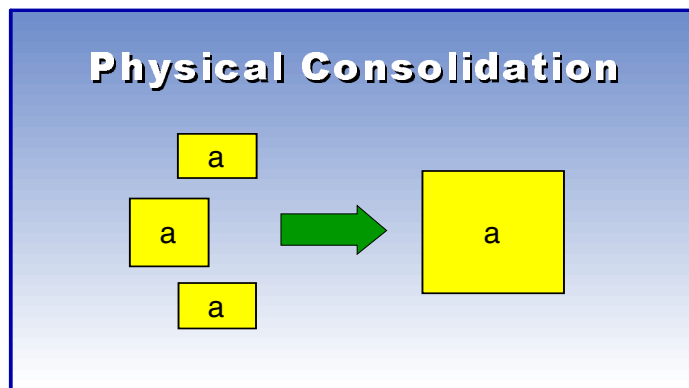


Figure 10. Physical consolidation

Physical server consolidation

The number of separate hardware platforms and operating system instances within a consolidation site may vary considerably by customer. Typically, some reduction in the number of distinct servers is accomplished when gathering distributed systems into a central installation or when a number of small servers are replaced with larger servers of the same platform. Based on

the enterprise's platform, four physical server consolidation cases can be considered.

- **Case 1:** Small servers from one platform to server(s) on the same platform
- **Case 2:** Small servers from different platforms to servers on different platforms (platform source and target are the same)
- **Case 3:** Small servers from one platform to server(s) on a different platform
- **Case 4:** Small servers from different platforms to server(s) on a different platform (platforms source and target are not the same)

Cases 1 and 2 are physical server consolidation, and there is no logical work to do. For cases 3 and 4, a platform migration has to be planned, and applications and data have to be ported from one platform to another. The objective of the physical server consolidation phase is not to share applications or data but to have an application that was running on one platform run on a new platform. Therefore, this operation has to be differentiated from application or data integration.

2.2.3 Data integration

Data consolidation is the process of taking information from several disparate sources and merging it into a single repository and a common format where each data element is made to follow the same business logic or by deploying the shared storage subsystem to manage disk requirements in a heterogeneous environment.

Enterprise servers are a logical safe haven for data that is now scattered in Local Area Networks throughout the enterprise. When all corporate data resides on the same robust system, the efficiencies can deliver immediate payback to end users. Data sharing throughout the enterprise is vastly simplified. Consolidation allows high levels of security and data integrity that are nearly impossible to achieve in a distributed environment.

In many client/server infrastructures, centralizing LAN data can bring dramatic improvements in data transfer speed. New enhancements in communications hardware will expand the high-speed connectivity options to server platforms of all types.

Combining data from different sources, such as the same or disparate types, into a single repository and format, for example, moving data from multiple, discrete sources, such as departmental servers, creates a data warehouse on an SP.

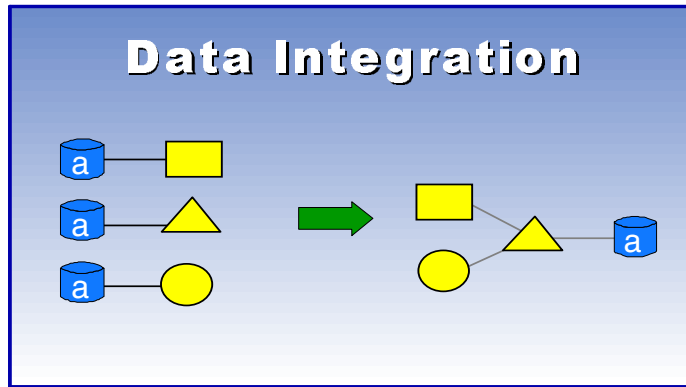


Figure 11. Data integration

There can be two kinds of data integration: Data integration from several servers (Case 1) and consolidated on a unique repository, and data integration from several repositories in one server and consolidated on a unique repository (Case 2).

Depending on the type of application integration selected, data integration can be performed separately or together with application integration.

2.2.4 Application integration

Application integration is the combining of multiple, similar applications, such as Web servers, onto one consolidated server.

Application integration is also the combining of different application workload types within a single server/system and migrating an application or data to a new platform in order to co-locate the application and data. For example, NT front-end applications accessing data from an SP are migrated to the SP in order to co-locate data and the application.

When a single server is able to run multiple workloads, multiple servers, which are dedicated to run individual workloads, can be consolidated. When the distributed servers and the consolidation server run the same applications and operating system, the migration is relatively straight forward.

Based on the consolidation platform, this migration can take different forms:

- The migration may just be the relocation of the application on the server.
- The migration may imply that application programs have to be recompiled in order to run on the new platform.

- It may also imply that application programs have to be redesigned and rewritten in order to run on the consolidation platform.

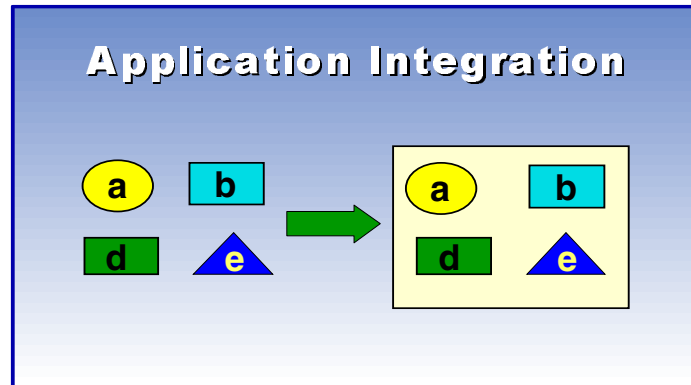


Figure 12. Application integration

The main objective of application integration is to migrate applications from one or several locations to a single location. Based on the consolidation platform, this migration can take different forms:

- The migration may not bring any additional costs beyond that of relocating the application on a new server.
- The migration may imply that application programs have to be recompiled in order to run on the new platform.
- The migration may imply that application programs have to be redesigned and rewritten in order to run on the consolidation platform. As for physical server consolidation, application integration has several cases.
- Application integration is combining different application workload types within a single server or system.
- Distributed systems do not run identical applications and system software and have to be integrated into a consolidation server running a different operating system.

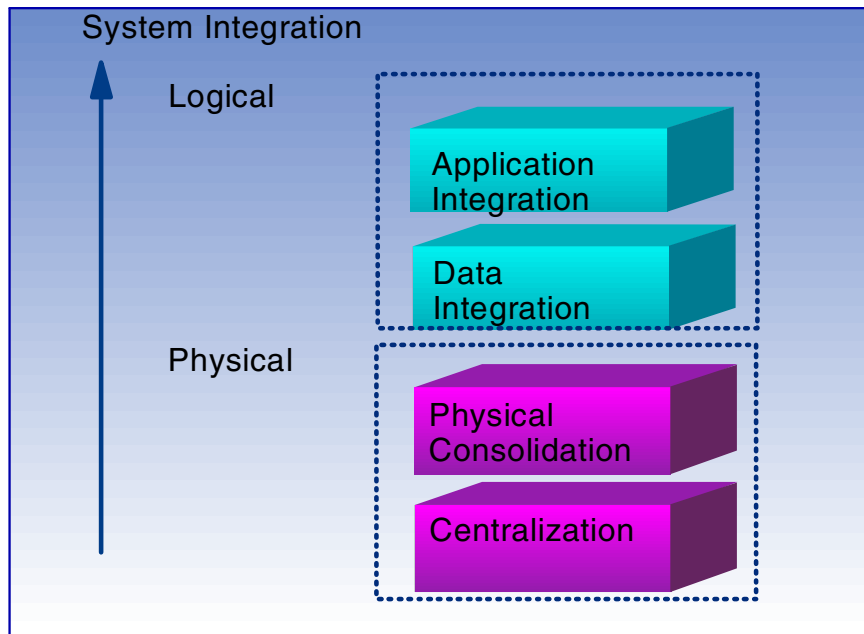


Figure 13. System integration

Degree of difficulty

From past projects, we have found that the amount of effort needed to do the four types of consolidation increases from centralization to application integration. The easiest form of consolidation would be centralization and physical consolidation. The more difficult type would be data integration and application integration. These would entail migrating databases and porting applications, when applicable, and would require more resources and time.

2.3 Good candidates for server consolidation

In general, the following factors make companies good candidates for server consolidation:

- Companies who have a directive to reduce IT costs.
- Companies who find that they need to integrate their distributed systems.
- Companies who find it hard to retain system administration personnel.
- Companies with server management issues.
- Companies who need to reduce staffing costs.

- Companies who have backup and recovery problems.
- Companies with network traffic problems.
- Corporate mergers and data center mergers.
- Data centers being given applications to manage.

Decisions made during the initial phase of any project can have a fundamental effect on quality, customer satisfaction, costs, and the overall success of the subsequent engagements. When an enterprise is considering an application(s) for server consolidation, there needs to be a very early assessment of the costs, effort, and benefits of performing the consolidation.

Also, it may not be immediately evident which type of server consolidation will be best for the enterprise:

- Physical movement of the distributed servers to a centralized location
- Movement of the individual servers to nodes of a single centralized machine, such as an IBM RS/6000 SP
- Complete integration of the server applications and data into a single RS/6000 Model S80

There are several methods that can be used to go forward. Many of these methods could be right, depending on the circumstances. The objective is to ensure that the correct method is selected, proper expectations are set, and that a viable high-level plan is created. These objectives will be used to define a path with the appropriate steps, projects, and proposals.

Here are some examples of server consolidation on RS/6000:

Case 1:

A manufacturing customer consolidated 18 Prime UNIX systems to an RS/6000 SP server out of concern for Year 2000 and concerns over the business viability of the vendor. After beginning the consolidation, the customer used the SP as the base for consolidating workloads from other UNIX servers and for allowing growth of existing workloads.

Case 2:

An information services company consolidated 10 UNIX servers with discrete databases from different vendors to one RS/6000 SP with IBM ADSM software.

Case 3:

A transportation company consolidated 40 UNIX servers to one RS/6000 SP server.

Business Benefits

- Reduced operating costs
- Flexibility in balancing workload across one single system vs. multiple servers
- Single systems management/administration for the new project
- Current applications on the already installed servers that would migrate/consolidate over time
- Scalable capacity for additional projects for the future (allowing capacity to grow while holding down systems administration costs)
- Can easily add capacity by adding nodes within single serial number, for example, typically does not require same justification as a new system purchase
- Can easily add pilot projects, such as a new Data Mart or Web/Intranet Serving

2.4 Decision criteria

When an enterprise decides to consolidate server systems, it usually has to analyze what the expected benefits of the operation are and what success factors are involved.

Many factors can influenced its choice:

- Risks
- Package availability
- Time to implement
- Costs
- Disruption and retaining costs
- Benefits and business value
- Evolution and direction of the business

Among these factors, we have already described the last four factors in the above sections in this chapter. In this section, we will focus on the first three factors.

2.4.1 Risks

Customers who want to consolidate servers will have to face problems and associated risks. Risks have to be identified, assessed, and managed in

order to guaranty business continuance. Risk management can be divided into four main processes or steps:

1. Risk identification
2. Risk analysis
3. Risk management
4. Risk assessment

These steps can be used as a risk assessment methodology. However, identifying the risks and gathering information on the risks are totally specific.

2.4.1.1 Risk Identification

The objective of this process is to clearly identify project risk areas. When risks areas are identified and accepted, each risk has to be analyzed. This is done during the next step.

2.4.1.2 Risk Analysis

The following elements need to be considered during the analysis of an identified risk:

- Description summary
- Impact summary
- Probability (likelihood: high, medium, low)
- Severity (consequences to the business: high, medium, low)
- Mitigation (countermeasures: prevention, reduction, transference, contingency, acceptance)
- Backup solution
- Status (open, closed)
- Owner (person responsible for ensuring that the risk is managed until it is eventually closed)

2.4.1.3 Risk Management

This section enumerates actions that constitute risk management:

- Planning (countermeasure actions)
- Resourcing (which will identify and assign the resources to be used for the work to carry out the risk avoidance or improve actions)
- Monitoring (including checking that execution of the planned actions is having the desired effect on the risks identified and watching for the early warning signs that a risk is developing)

- Controlling (taking action to ensure that the events of the plan really happen)

2.4.1.4 Risk Assessment

The results of the previous steps are documented in a risk report document.

2.4.2 Package availability

Another deciding factor is the capability to provide "ready to run" packages. This means integrated sets of packages that would not require significant tailoring or time to implement and could grow with the customer as their needs change in the future. These packages have to be able to be installed on different platforms.

2.4.3 Time to implement

The time to implement the solution has to be estimated, but it is not the main concern. As a matter of fact, the disruption time is much more important for the business. Also, a backout time always has to be planned.

How long will the business process be stopped? This is truly the question when talking about server consolidation. Therefore, the time to implement and the disruption time for a server consolidation solution depend on the kind of consolidation put in place.

The following are some time to implement versus disruption time scenarios:

- In case of physical hardware centralization (servers or storage systems), the hardware will not be available during the transfer. So, the implementation time should be equal to the disruption time.
- In case of physical server consolidation, business processes can continue to run on the source platform, and the workload will be transferred when the consolidated platform becomes available (including tests). The disruption time will be equal to the migration time from the source to the target platform.
- In case of data integration, the disruption time will be equal to the data migration time.
- In case of application integration, the best plan is to continue to work on the old platforms until the new solution is implemented and fully tested. Doing so, the disruption time will be equal to the applications' migration time.

2.4.4 Success factors

Almost universally, customers embarking upon a server consolidation have been quite pleased with the outcome. Nonetheless, some comments are worth noting regarding suggestions to make the transition smoother.

- Autonomous departments' fear of losing their independence and becoming subject to perceived burdensome central control has stalled some consolidation projects.
- While vendor analysis may well point out legitimate benefits of consolidation, centralization versus decentralization remains an emotionally charged argument as it reflects on organizational structure, company politics, and even the culture of the firm.
- Changes impacting organizational power structure are not well received when driven from outside the organization itself. Thus, it is very important to have an in-house customer champion or sponsor who is sensitive to internal organizational power structures and can overcome reluctance to a fundamental change in computing control.

The ability to identify a specific computing resource as belonging to a particular department can often ease the transition to a centralized server consolidation installation. That is, even though the departmental server may have been physically relocated to a central site, that departmental computing resource may still be physically identifiable as a specific server within a cluster, a particular partition within a larger system, or a distinct node within an RS/6000 SP.

Even when an organization fully supports the concept of consolidation, planning the migration may be difficult if there is not a clear understanding of the functions provided by distributed servers. A disciplined approach is needed to fully document current application inventory and plan the transition. A thorough application inventory will be necessary prior to beginning the consolidation. In addition, a well thought-out project management plan must be defined to insure continued delivery of all services during the transition.

Chapter 3. Why consolidate to the RS/6000?

This chapter presents the advantages of choosing an RS/6000 server, in particular the SP, for Server Consolidation.

3.1 RS/6000 strengths as a central server

The RS/6000 SP system delivers solutions to some of the most complex technical and commercial applications by simultaneously bringing dozens of RISC processing nodes to a computing problem. This parallel processing capability enhances computing performance and throughput over serial computing.

The basic SP building block is the processor node. Each node consists of a POWER3 microprocessor or a PowerPC symmetric multiprocessor (SMP), memory, disk, Micro Channel or PCI expansion slots for I/O, connectivity, and the SP Switch adapter. Node types with varying processor power and I/O attachment options may be mixed within a system.

3.1.1 Architecture

How does the RS/6000 SP fit in the broad spectrum of computer architectures? It is hard to classify the SP in a single category. Moreover, classification of computer architectures can be done in many ways. Computers can be classified based on the way streams of instructions interact with streams of data.

The above classification is widely accepted and practical implementations can be found for most computer types. As a matter of fact, the RS/6000 SP could be classified in more than one of categories. Modern computer architectures, in general, cross many of the lines that divide each category, thus, making computer classification an “art”.

Although Single Instruction Single Data (SISD) architecture is part of what the RS/6000 SP can offer, we will concentrate on the multiprocessor aspect of the system, adding notes along the way when an explicit discussion of uniprocessor systems is needed.

We understand that multiple streams of instructions are independent and executed by independent executing processor units. A MIMD machine is then a multiprocessor machine by definition.

Within multiprocessor systems, we can further divide the classification based on the way data is shared between processors. At this point, we have two

categories: Shared Nothing and Shared Data. We could have said *shared memory*, and that term would probably ring a bell, but sharing data can be achieved at memory level as well as at disk level; so, a more generic term is useful to avoid being too specific.

The term shared nothing, in this context, refers to data access. In this type of architecture, processors do not share a common repository for data (memory or disk); so, data sharing has to be carried out through messages. Systems using this architecture are also called message-based systems.

In contrast, shared data architectures have processing units sharing data in one way or another. One of the most popular shared data multiprocessor implementations is Symmetric Multiprocessing (SMP). In this architecture, all processors share a common memory bank and I/O units. All the processors, memory, and I/O units are connected through a system bus (there can be multiple system buses).

The advantage of an SMP architecture is having minimum impact to applications while taking advantage of parallel processing. Applications developed for uniprocessor systems can easily run on SMP machines while they are using a single processor. However, the system is able to handle multiple applications like this in parallel, which will not necessarily benefit the particular application, but the system in its entirety.

Inherent in SMP architectures is the concept of threads. Streams of instructions can be broken into several portions that are relatively independent and executed in different processors, thus, speeding up the application as a whole. This is in theory, because in reality the effectiveness of multi-threading an application will ultimately depend on the interaction between threads and data.

One limitation though, in SMP architectures, is scalability. While adding processors to an SMP system may improve its performance (by adding more executing units), the shared elements within this architecture will slow down the system to a point where adding more processors will prove to be counterproductive. However, this self-imposed limitation has been overcome year after year, thus, increasing the practical number of processors in an SMP machine to numbers that could not have been imagined a few years ago.

Shared nothing systems are also referred as message-based or message-passing systems. In such systems, communication is completed when data has been received by the target node.

If you analyze the RS/6000 SP from a system point of view, you see that processor nodes communicate each other through communication networks (Ethernet, SP Switch, or any other supported network) so, by definition, the SP architecture is a Shared Nothing architecture. However, each node could be classified as a SISD node (old uniprocessor nodes) or a MIMD node (such as SMP nodes). Furthermore, the RS/6000 SP flexibility allows to have clusters of nodes running within the RS/6000 SP umbrella.

Besides the fact that we can use the RS/6000 SP nodes as stand-alone machines for running serial and independent applications in what is called server consolidation, the RS/6000 SP can be also viewed as a cluster or even as a parallel machine (a massive parallel machine).

3.1.2 Components

An RS/6000 SP system consists of the following hardware components:

Components that can be monitored and controlled

- Frame
- Node
- Switch
- Control workstation

The control workstation, being the center of hardware control, is able to monitor and control the frames, nodes, and switches.

Frame

Hardware components in a frame that can be monitored include:

- Power LEDs
- Environment LEDs
- Switch
- Temperature
- Voltage
- State of the power supply

Hardware components in a frame that can be controlled include:

- Power on/off

Node

Hardware components in a node that can be monitored include:

- Three-digit or LCD display
- Power LEDs

- Environment LEDs
- Temperature
- Voltage
- Fan

Hardware components in a node that can be controlled include:

- Power on/off
- Key Mode Switch
- Reset

Switch

Hardware components in a switch that can be monitored include:

- Temperature
- Voltage
- Fan
- Power LEDs
- Environment LEDs
- MUX

Hardware components in a switch that can be controlled include:

- Power on/off
- Change multiplexor clock setting

3.1.3 System management

One of the main reasons that customers choose the RS/6000 SP is the ease of managing the system for the system administrator. In particular, system management is very important for a customer who is going to consolidate numerous servers into a single server.

The Parallel System Support Programs (PSSP) comes with the RS/6000 SP and enables the system administrator to manage the SP, but there are also several third party software tools to help manage the RS/6000 SP, such as Candle.

The purchase of a new computer system requires a significant investment of financial and human resources. As with any other investment, this requires care and attention to get the maximum benefit. For this reason, a clear system management strategy is required.

Below is a list of common system management tasks. Because systems differ, and system management requirements differ, a unique system

management strategy must be set up for each environment. Tasks that are important in one environment may not necessarily be important in another.

Software maintenance-If new software is required, it has to be installed and tested. After successful installation and testing, user access has to be enabled. For new machines, it must also be decided which software should be installed.

User management-New users have to be created with the appropriate access rights. For security reasons, unused user IDs must be disabled or removed. In addition, current users often require changes to their environment.

Backup/recovery-One of the most important tasks in any system is to make sure that all important data can be restored in case of a disaster. There should be recovery procedures in place for both the worst case and also for simple cases, such as restoring a file that has been accidentally deleted.

Storage management-In environments with a large amount of data, there will be a need for storage management products. In general, this involves migrating seldom used files onto less expensive media, such as tapes or optical media, while still keeping them accessible to users.

Network management-This includes all the tasks that are necessary to enable the communication between all machines within an enterprise. These are mainly routing, nameserving, and mailing.

Problem management-Any kind of problem must be handled by the system administration. Problems can be system halts, application problems, network breakdowns, simple user questions, and so on.

Security-Security is extremely important to the enterprise; unfortunately, it is one area that is often neglected. Security is especially important when there are connections to networks outside the enterprise and unauthorized access is to be prevented.

Accounting-With accounting, one can control the usage of system resources, such as CPU time, disk space, and so on.

Configuration-This is basically defining, configuring, and updating system resources, such as printers, network parameters, hardware, system parameters, and so on.

Change management-This involves planning, distributing, and installing changes to system software, applications, and data.

Other activities-Specific activities, such as providing documentation. Small programs for automation are often need to be written.

Planning-Since the system managers have the best knowledge about the environment and system usage, they must be involved to get a proper investment planning. AIX provides many standard tools to manage your environment, for instance, Visual System Management. Many additional applications also exist for this purpose.

System management tools and capabilities should be of prime concern in purchase decisions regarding the managing of a computer environment. IBM uses its considerable expertise in commercially robust systems to develop an extensive portfolio of parallel systems management software, which sets apart the RS/6000 SP from its competition. The software is based on the layered systems management architecture of the AIX operating system and incorporates standard protocols and products from the open systems arena and expertise gained from IBMs first SP1 systems.

PSSP is the suite of system management applications that enables system administrators to manage SP systems and their environments. PSSP uses a common graphical user interface to display node status information from a single, control workstation. It is this program, with its advanced, user-friendly graphical interface, that makes system management easier.

Along with AIX system management functions, PSSP enables installation, operation, and maintenance of all nodes in an SP system from the single workstation. Advanced error detection and recovery facilities help reduce the number of unplanned outages and helps minimize the impact of outages that do occur. Enhanced installation and migration tools can also help minimize scheduled outages.

AIX's menu-driven Systems Management Interface Tool (SMIT) is used for installation, configuration, device management, problem determination, and file system management, and presents a systems management interface that is consistent with all other RS/6000s.

There are many important topics in the system management area. To avoid duplication, we will focus on what we feel are the main topics of system management in this section. Please refer to Chapter 6, "System management" on page 115 for the other topics that are not covered.

3.1.3.1 Single point of control

A single, standard RS/6000 workstation is at the center of SP system management. It acts as a control workstation, that is, a focal point for systems

administration. The workstation is connected to each SP frame via an RS-232 line and to each node via an Ethernet LAN. Any X-Windows-capable system on a LAN can be used to log into the control workstation to perform system management, monitoring, and control tasks. The administrator does not need to physically sit at the control workstation.

One main task of the control workstation for the RS/6000 SP system is to provide a centralized hardware control mechanism. In a large SP environment, the control workstation makes it easier to monitor and control the system. The administrator can power a node on or off, update supervisor microcodes, open a console for a node, and so on, without having to be physically near the node.

3.1.3.2 Print management

The printer subsystem includes spoolers, real printers, virtual printers, backends, and queues. A printer can be attached directly to a local system, or a print job can be sent over a network to a remote system.

How it works

SMIT can be used to set up the printer. SMIT allows the printer device to be created, the virtual printer to be added and customized, and the queues to be manipulated.

Although it is possible to start SMIT with the `smit` command to access the printing menus, the simplest way is to use the `smit printer`.

The printing command places the file, plus any print flags, onto the queue. When the printer is ready, the `qdaemon` passes the file and flags to the printer backend, which formats the output for the specific printer. The data is then passed to the printer device driver that communicates with the actual device and prints the file.

It is also possible to bypass the queue and backend by redirecting the file to the device (using the `cat` command).

AIX uses the virtual printer concept, which allows more control over each printer and more customization. The following describes the new terminology.

Printer/Plotter Device

This is the special file in the `/dev` directory for the device. It can only be used by redirection (`cat filename > /dev/lp0`) and may be manipulated with `splp`. Printer commands will not be able to access a printer device unless a virtual printer has been created.

Virtual Printer

The virtual printer is actually a combination of a specific queue and queue device in the `/etc/qconfig` file along with the associated file in the `/usr/lpd/pio/ddi` directory (which contains formatting data). The virtual printer, when added through SMIT, will automatically create the queue and ddi file.

To create a queue and queue device for a printer that will use the standard piobe backend, add a virtual printer. To add a second queue device to an existing queue to allow load sharing, add a second virtual printer to an existing queue.

Queue

The queue is where the user directs a print job. It is a stanza in the `/etc/qconfig` file whose name is the name of the queue and points to the associated queue device.

Only add a queue when the backend will not be the standard piobe backend. Adding a virtual printer is the normal way to create a queue and queue device.

Queue Device

The queue device is the stanza in the `/etc/qconfig` file that normally follows the local queue stanza. It specifies the actual `/dev` file (printer device) that should be printed to and the backend that should be used.

There may be more than one queue device associated with a single queue. The virtual printer can handle this, as it associates itself with a queue:queue device pair. Only add a queue device to an existing queue if it is to use a backend other than the standard piobe backend. Adding a virtual printer will create a standard queue device entry to an existing queue.

3.1.3.3 Consolidated error log

Once servers are consolidated into a single server, there needs to be a repository of error logs to help the system administrator to easily monitor the system.

System logging

Since physical console devices are not attached to each of the nodes, many of the SP subsystems redirect console output to a log file. Rather than being consolidated on the Control Workstation, each of these logs remains on the individual nodes. This scales better, therefore, allowing users to examine these logs only when a problem is detected. The logs, collected in a set of directories in `/var/adm/ssp`, are managed through a nightly cron job that trims large logs and removes old ones. Software that provides a comprehensive

system for managing and viewing the nodes' error logs and other log files would be a useful addition to the existing collection of PSSP tools.

Error logging is the writing of information to persistent storage to be used for debugging purposes. This type of logging is for subsystems that perform a service or function on behalf of an end user. The subsystem does not communicate directly with the end user and, therefore, needs to log events to some storage location. The events that are logged are primarily error events.

Error logging for the SP uses Berkley Software Distribution (BSD) syslog and AIX Error Log facilities, plus SP log, to report events on a per node basis. The intent is to have the AIX Error Log be the starting point for diagnosing system problems.

To manage and monitor the error log, you can do the following:

- View error log information in parallel.
- View SP Switch error log reports.
- Use AIX error log notification.
- Effect of not having a battery on error logging.

In a typical RS/6000, a battery is installed to maintain NVRAM. On an SP system, there is no battery, and NVRAM may be lost when the node is powered off. AIX writes the last error log entry to NVRAM. During system startup, the last entry is read from NVRAM and placed in the error log when the errdemon is started. This last error log entry may be important in diagnosis of a system failure.

On SP wide nodes, the NVRAM does have power to it as long as the node is plugged into the frame and the frame is plugged into a working power source. On SP thin nodes, NVRAM is lost whenever the node is powered down. If the last error log entry is desired, the thin nodes should not be powered off. They should be re-IPLed in the normal key mode switch position if at all possible.

Managing and monitoring the error log

To manage and monitor the error log, you can do the following:

- View error log information in parallel.
- View SP Switch error log reports.
- Use AIX error log notification.

3.1.3.4 Consolidated accounting

The SP support for accounting builds upon standard AIX accounting function, thus, providing three main capabilities.

Accounting-record consolidation

SP accounting support builds upon standard UNIX System V accounting. Partial reduction of accounting data is done on each node before the data is consolidated on an accounting master. The accounting master will generally be the Control Workstation but can be any nodes. Nodes are configured into groups called classes. All of the data from a given class is consolidated. Classes can be used to impose different charges for nodes with differing capabilities.

The administrator uses the CMI or commands to specify the accounting configuration. The SP system-management code automatically configures accounting on the selected nodes. The command `nrunacct` (a modified version of the AIX command `runacct`) is scheduled to run nightly on each node to consolidate the accounting data for that node. Data from all of the nodes is then consolidated on the accounting cluster master, which performs additional processing. The output of this processing is data, in standard format, that can be used for charge-back purposes, usage monitoring, or capacity planning, and can be fed into the existing accounting applications that you may have.

Parallel job accounting

The LoadLeveler job-management program provides for job-based accounting by accumulating data on resource usage from all processes triggered by a specific job.

Node-exclusive-use accounting

Standard accounting generates charges for resources used, for example, processor, disk, and so on. If a user is given exclusive use of a set of nodes for the duration of a parallel job, standard accounting may not be an appropriate way to charge for processor usage. Instead, administrators may want to bill based on the wall-clock time that a processor is in use since it is unavailable to other users during that period regardless of what processor cycles are actually consumed by the running job.

The PSSP provides an optional mechanism with which to charge for nodes that have been assigned for exclusive use. If this support is enabled, an accounting record for a marker process is created before and after the job is run. All processor accounting records associated with that user and that fall within the window of the marker records are discarded. Instead, a special

charge is applied based on the actual number of seconds the job runs. The fee is specified by the administrator through the standard accounting charge-fee mechanism.

3.1.3.5 System Data Repository (SDR)

The SDR is a centralized, common-data repository that provides configuration data storage and retrieval from the Control Workstation and all nodes for persistent system data. System and node configuration data, job data and other system data, such as site environment data, is stored in the SDR. Command-line interfaces are provided for the SDR. Typically, data in the SDR is accessed through management applications.

The SDR is an SP subsystem that stores SP configuration and some operational information. The SDR stores:

- Information about the frames, nodes, and switches, and how they are configured.
- Job Manager operational data.
- VSD configuration information.
- The current values of `host_responds` and `switch_responds`, two indicators of the health of a node.

This information is stored on the Control Workstation but is made available through a client/server interface to other network-connected nodes.

SDR data model

The SDR logical model consists of classes, objects, and attributes. Classes contain objects. Objects are made up of attributes. Objects have no unique ID, but their combined attributes must be unique among other objects. Attributes can be one of three types: Strings, floating-point numbers, or integers.

Most SDR interaction is performed using the SDR command-line interface. Some clients, such as Job Manager, communicate directly with the SDR.

Authorization

SDR permits only two types of authorization, read-only and read-write.

Read-write authority is granted only if the following are both true:

- The user is running as root.
- The user is running from the Control Workstation or from a node whose adapter is in the SDR Adapter class.

If both of the above are not true, the user gets read-only permission.

3.1.3.6 Installation management

Another important aspect of system administration is the ability to easily install software on the different nodes on the system. The SP enables the system administrator to easily install software on the different nodes that have been consolidated and allows easy installation of new nodes.

The installation of SP nodes is based on the AIX net-install capability in which an image of a system is cloned on a target systems. This allows nodes to be easily restored to a known level of software. Node configuration data, such as hostname, default route, and each communication adapter's IP address and netmask, is entered into the SDR using SMIT or line commands. This data is used after a net install to automatically customize Object Data Manager (ODM) information. The Control Workstation acts as a net-install server for the nodes acting as boot/install servers (typically, one boot/install server for each frame). This two-stage structure allows the install of each frame to proceed in parallel. When a node is rebooted after installation, a user-supplied firstboot script may be executed, performing additional tailoring, such as setting name resolution, enabling AFS or Network Information System (NIS), setting paging space, installing additional LPPs, setting licenses or time zone, and so on.

Role of Control Workstation and Boot/Install servers

The Control Workstation plays the main role in an RS/6000 SP installation. It is from this system that the rest of the nodes are installed. The control workstation is an RS/6000 system running AIX. In order to perform its role as a Control Workstation, it has to be loaded with the IBM Parallel System Support Programs for AIX (PSSP) software package. With this software installed, it serves as a single point of control for installing, managing, monitoring, and maintaining the frames and nodes. In addition to these functions, the control workstation is normally set up as an authentication server.

After installing at least one node, it can configure one or more of the nodes to become boot/install servers. These nodes, serving as boot/install servers, can be used to boot up and install other nodes, offloading from the control workstation this time-consuming and process-hungry task.

NIM concept

The core of the SP node installation centers around Network Installation Management (NIM) in AIX.

With NIM, you can manage stand-alone, diskless, and dataless systems. In a broad sense, an SP node can be considered a set of stand-alone systems. Each node has the capability of booting up on its own because the RS/6000 SP system is based on a share-nothing architecture. Each node is basically a stand-alone system configured into the SP frame.

Working together with the SDR, NIM allows you to install a group of nodes with a common configuration or individually customize to each node's requirements. This helps to keep administrative jobs simpler by having a standard rootvg image and a standard procedure for installing a node. The rootvg volume groups are consistent across nodes, at least when they are newly installed. You also have the option to customize an installation to cater to the specific needs of a given node if it differs from the rest.

As NIM installations utilize the network, the number of machines you can install simultaneously depends on the throughput of your network (namely, Administrative Ethernet). Other factors that can restrict the number of installations at a time are the disk access throughput of the installation servers and the processor type of your servers.

The Control Workstation and boot/install servers are considered NIM masters. They provide resources, such as files, programs, and booting capability, to nodes. Nodes are considered NIM clients, as they are dependent on the masters for services. Boot/install servers are both masters and clients since they serve other nodes but are, in turn, dependent on the Control Workstation for resources. Not considering NIM masters outside of the SP complex, the Control Workstation is configured only as a NIM master. The master and clients make up a NIM environment. Each NIM environment can have only one NIM master. In the case of the Control Workstation and the boot/install server, only the Control Workstation is considered the master. In another NIM environment consisting of the boot/install server and its client nodes, only the boot/install server is the master.

A NIM master makes use of the Network File System (NFS) utility to share resources with clients. As such, all resources required by clients must be local file systems on the master.

The way NIM organizes itself is by object classes, object types and object attributes. Administrators might be tempted to NFS-mount file systems from another machine to the Control Workstation as /spdata/sys1/install/images to store node images and as /spdata/sys1/install/<name>/lppsource to store AIX filesets due to disk space constraints. This is strictly not allowed, as NIM will not be able to NFS export these file systems.

Not all of NIM's functionality is used in the SP complex. Only the minimum to install a stand-alone system is utilized. NIM allows two modes of installation: Pull or push. In the pull mode, the client initiates the installation by pulling resources from the master. In the push mode, the master initiates the installation by pushing resources to the client.

If you need further details about the way NIM works, refer to *AIX Version 4.3 Network Installation Management Guide and Reference*, SC23-2627.

3.1.3.7 Distributed Shell (dsh)

System command execution

A new command, `dsh`, is provided to execute remote commands on all nodes or groups of nodes in the SP system. A number of flexible ways to specify these nodes and groups of nodes is provided. The `dsh` command is based on `rsh`, hence, requiring `rsh` privileges. With `dsh`, the user can execute all commands in parallel, with output returning to a single destination. To control the use of system resources, the degree of parallel execution can be specified. As a default, `dsh` will only operate on those nodes that are *alive*, as indicated by the host responds indicator. Although this is a relatively simple tool, we have seen many benefits from its use. For example, the command `dsh -av ps -fu johndoe` could be used to report all processes belonging to user johndoe on all running nodes. If available, `dsh` will use the authenticated `rsh`.

The dsh command

The `dsh` is the parallel implementation of the `rsh` command (which executes the specified command at the remote host). The `dsh` command is build on top of the Kerberized `rsh` command; so, to use `dsh`, you should be an authenticated user to Kerberos, or you should work with the `.rhosts` file. The command specified with the `dsh` command is performed on all the nodes specified in the working collective with an `rsh` on those nodes.

3.1.4 Availability

There are many solutions for high availability including HACMP, HACMP/ES, HAGEO, HACWS, and so on. For the detail information, please refer to Chapter 8, "High availability" on page 233.

3.1.5 Scalability

3.1.5.1 Add nodes as needed

One distinct advantage of the RS/6000 SP is the ability to add new nodes easily when needed. This is very important when doing server consolidation since growth of applications and users is inevitable.

The various nodes may be mixed in a system and are housed in frames. Depending on the nodes used, an SP tall frame can contain up to 16 thin nodes. These frames can be interconnected to form a system with up to 512 nodes. A maximum of 64 SMP high-nodes can be installed per system. No other vendor is able to provide scalability as high as the RS/6000 SP.

3.1.5.2 SP Switch

At the heart of the RS/6000 SP is the Switch where all the nodes communicate.

The SP Switch network provides a high bandwidth for data transfer between nodes and node-to-node communications for parallel programs. It provides a low latency (time to transfer the first piece of data) interconnection between processors. Each node is connected to the switch by a Switch Adapter, which is architected to provide superior error detection and data handling. SP Switch bandwidth scales linearly as nodes are added. Building on the same architecture as the High Performance Switch, which is still supported, SP Switch provides improved reliability, availability, serviceability, and performance.

Effective parallel computing requires high-bandwidth, low-latency internode communications. The SP Switch, a state-of-the-art IBM innovation, can provide a high, bi-directional data-transfer rate between each node pair. Not only that, it maintains point-to-point communications time independently from the relative position of the nodes.

SP Switch Communication Network

During the initial development of the SP system, a high-speed interconnection network was required to enable communication between the nodes that make up the SP complex. The initial requirement was to support the demands of parallel applications that utilize the distributed memory MIMD programming model. More recently, the SP Switch network has been extended to a variety of purposes:

- Primary network access for users external to the SP complex (when used with SP Switch Router).
- Used by ADSM for node backup and recovery.
- Used for high-speed internal communications between various components of third-party application software, for example, SAP's R/3 suite of applications.

All of these applications are able to take advantage of the sustained and scalable performance provided by the SP Switch. The SP Switch provides the

message passing network that connects all of the processors together in a way that allows them to send and receive messages simultaneously.

Two networking topologies can be used to connect parallel machines: Direct and indirect.

In direct networks, each switching element connects directly to a processor node. Each communication hop carries information from the switch of one processor node to another.

Indirect networks, on the other hand, are constructed such that some intermediate switch elements connect only to other switch elements.

Messages sent between processor nodes traverse one or more of these intermediate switch elements to reach their destination. The advantages of the SP Switch network are:

- Bi-sectional bandwidth scales linearly with the number of processor nodes in the system. Bi-sectional bandwidth is the most common measure of total bandwidth for parallel machines. Considers all possible planes that divide a network into two sets with an equal number of nodes in each. Considers the peak bandwidth available for message traffic across each of these planes. The bi-sectional bandwidth of the network is defined as the minimum of these bandwidths.
- The network can support an arbitrarily large interconnection network while maintaining a fixed number of ports per switch.
- There are typically at least four shortest-path routes between any two processor nodes. Therefore, deadlock will not occur as long as the packet travels along any shortest-path route.
- The network allows packets that are associated with different messages to be spread across multiple paths, thus, reducing the occurrence of hot spots.

The hardware component that supports this communication network consists of two basic components: The SP Switch adapter and the SP Switch board. There is one SP Switch adapter per processor node and generally one SP Switch board per frame. This setup provides connections to other processor nodes. Also, the SP system allows switch boards-only frames that provide switch-to-switch connections and greatly increase scalability.

3.1.6 Security

3.1.6.1 Sysctl

Sysctl is an authenticated client/server system for running commands remotely and in parallel. It provides:

- Least-privilege capability -- Root authority can be dynamically delegated to non-root users based on their authenticated identities, the task they are trying to perform, access control lists, and any other relevant criteria. The root password need not be given out to as many people, thus, keeping it more secure.
- Distributed execution -- Sysctl applications can be executed on remote hosts with full authentication and authorization. Sysctl provides a secure, easy-to-program, remote-command execution mechanism for arbitrary AIX commands, scripts, and programs.
- Parallel execution -- Sysctl applications can be efficiently executed in parallel on many hosts.
- Programmability -- Customized Sysctl applications can be coded as shell, Tcl, or Perl scripts.

The SP system-administration software provides a set of commands that exploit the Sysctl facility. You may also want to design your own Sysctl applications. Candidates for such applications are tasks that require root authority, but those which administrators would like to delegate, for example, user management, backups, file system administration, and so on.

The Sysctl utility is a client/server system for running tasks and commands on remote systems. The command is authenticated through Kerberos; so, users of Sysctl require Kerberos authentication, although Sysctl allows non-authenticated users to have access to Sysctl services as well.

A Sysctl environment is another method with which to allow system administrators to execute root commands on local and remote hosts without having to distribute the root password. Essentially, Sysctl allows authenticated and authorized access throughout the SP System to servers running as user ID 0.

Authenticated users can run `sysctl` commands as root on those local and remote nodes where they are authorized. These commands are run in parallel on the target nodes. Sysctl initially provides four levels of authorization.

Sysctl especially provides added value in the following areas:

- **Security:** Sysctl is based on the Kerberos Authentication mechanism. Although not mandatory, generally a Sysctl user must be a *known* Kerberos user before getting access to Sysctl services. The default configuration of the SP does not allow normal users to access Sysctl.
- **Authorization:** Most of the authorization schemes are based on machine names and user IDs. TCP/IP authorization is usually *all or nothing*. You either give a user access to a particular machine or do not. Sysctl, however, provides a multi-level authorization mechanism for not only granting user access for Sysctl use, but for providing authorization levels for commands or groups of commands, therefore, identifying which are executable or usable and which are not.

3.1.6.2 Access Control Lists (ACLs)

PSSP includes a security infrastructure, which provides authentication services for users and servers. Authentication is the foundation for providing other security features, such as accountability, access controls, and least privilege. The authentication services provided in the SP system are based on MIT-Kerberos Version 4. Systems running AFS or an existing Kerberos Version 4 server can integrate the SP System into their existing Kerberos realm.

Kerberos functions as a third party to authenticate the identities of clients and servers. Utilizing Kerberos authentication prevents unauthorized access to system resources.

ACLs are used to control access to resources in the SP System. The servers for hardware monitoring and system control use authentication and ACLs to control access to their services. Secure versions of the remote commands, such as `rsh` and `rcp`, have been provided, thus, eliminating the need for insecure `.rhost` files.

Because the use of Kerberos requires changes in existing command libraries, it is expected that a generalized interface, such as the Generalized Security Services Application Programming Interface (GSS-API), will be exploited in the future. This will allow any authentication server that supports this API to be used including Kerberos Version 5, DCE Kerberos and AFS Kerberos, among others. A comprehensive system of ACL management should then be provided.

3.1.6.3 Kerberos

Kerberos is used for services authentication on the SP. Kerberos is basically the watchdog of the system. Its name is derived from Greek mythology, where Cerberus guarded the entrance to the underworld. Kerberos provides

authentication services that allow certain distributed services within the SP System and between it and other workstations to securely control access to their facilities. It is used to provide secure remote command capability used by the SP System installation programs and is available for general use in place of the standard AIX remote shell and remote copy commands, which are notoriously insecure. The root user must use Kerberos when installing the SP System, therefore, taking on the role of Kerberos administrator because the installation process includes the creation or modification of the Kerberos security database.

The System Monitor that administrators use to monitor and control the SP hardware resources uses Kerberos to ensure that only authenticated users can invoke its functions through either the graphical or command-line interfaces. Generally, a small number of administrative users would expect to use Kerberos for that purpose. More general use of Kerberos by other users is possible but not required. It is required for the exploitation of the authenticated distributed command execution facilities: Sysctl, dsh, the "p*" commands, and the enhanced rsh and rcp commands. Kerberos can best be compared with going to the motor races, say the Grand Prix of Monte Carlo. Going to Monte Carlo requires you to show your passport.

This is, in fact, equal to logging into the system. Now, in order to enter the race track, you must be a member of the Grand Prix Racing Club. In terms of Kerberos, you must be a known member of the Kerberos database to get access to the Kerberos services. Having a seat in the grandstand, getting access to restricted areas such as the pits, requires additional permits. You can get those permits by buying tickets. These tickets, however, are only valid for a certain amount of time. In terms of Kerberos, you must ask Kerberos to give you a ticket granting you permission to use certain services. As with the races, where you can only get a pit ticket when you are a member of the Grand Prix Racing Club, with Kerberos, you can only get a ticket when you are known to Kerberos.

Kerberos is delivered with the SP as a vehicle to enable the highest SP security. The main reason for implementing Kerberos on the SP, and on UNIX workstations in general, is to avoid the use of the .rhosts file of the root user. This does not mean that SP users are not allowed to use this file anymore; users who still want to have all network services available can do so. Kerberos is a method of authentication that is not related to any AIX authentication system nor is it used as an additional login user password verification. Guarding your root password is still critical to having a secure environment. As soon as someone other than the system administrator can log into the system as root, he or she can destroy the Kerberos database

anyway. Rather, Kerberos is used as a system for use of authenticated services within the SP.

The authentication services provided with the SP are based on MIT's Kerberos Version 4. These services use the Data Encryption Standard (DES) algorithm. Due to export regulations, DES is restricted for use to the United States only.

3.1.7 Network

Nodes communicate through the SP Switch and the network. Clients, workstations, and dumb terminals will access the main server through the network and run their applications, which have been consolidated.

Overview of the SP Switch

There are four basic physical components of an SP:

- Frame - A containment unit consisting of a rack to hold computers, together with supporting hardware, including power supplies, cooling equipment, and communication media, such as the system Ethernet.
- Nodes - AIX RS/6000 workstations packaged to fit in the SP frame; a node has no display head or keyboard; so, user human interaction must be done remotely.
- Switch - The medium that allows high-speed communication between nodes.
- CWS - The Control Workstation (CWS) is a stand-alone AIX workstation, with display and keyboard, possessing the hardware required to monitor and control the frames and nodes of the system.

Each frame may contain an SP Switch board. The nodes of the frames are connected to their respective switch board via a special SP Switch adapter and corresponding switch cables, and the switch boards of the system are connected via the same type of cables. Thus, a high-speed communication network is formed that allows the nodes to communicate with each other to share data and status. The primary purpose of this high-speed network is the support of solving problems in parallel.

Communication over the switch is supported by IBM software, which is shipped with the SP. The switch hardware, together with this software, is called the Communication Subsystem (CSS). On each node attached to the switch, the CSS software is continuously available to provide the following:

- A communication path through the switch to other nodes
- Monitoring of the switch hardware.

- Control of the switch hardware for startup, and in case of error, execution of any appropriate recovery action.

This software is responsible for sending and receiving packets to and from other nodes on behalf of applications. It also sends packets to switch hardware components as part of its monitoring and controlling functions. If a component of the switch network, for example, switch board, adapter, cable, node, or software, is not functioning correctly, the CSS software is responsible for recognizing this and reporting this *fault* via the AIX error log. The software will also take recovery action as deemed most appropriate for the health of the system, which may mean removing the offending component from the switch network.

Communication Network hardware

Reliability is of great importance. The communication subsystem detects, locates, and isolates all failures. Communication links are protected by error-detecting codes; switch chips are designed using error detection and fault isolation methods. Links and switch chips can be individually disabled to isolate faults, therefore, making some communication flow through alternative paths.

Packet data flow

Data is inserted in the SP Switch by a node through the node's switch adapter, which provides for switch network communication. The information that flows into the network is made up of single independent packets of data that contain the route information needed to traverse the SP Switch. The route information is used by switch chips in order to decide which of its ports to route the received packet.

Each packet is destined either to a node or to a single switch chip. In the first case it is typically a data packet that has been created by some node application, while in the second case, it is called a service packet and is used to configure the features of the switch chip. A service packet may also be created by a switch chip to notify a node of some event that is meaningful in network administration or by a node for node-to-node administrative communications.

Both data and service packets flow on the same network. This choice keeps architecture simple and gives to both packet types similar path redundancy. Service packets do not greatly affect the network performance since they are basically used whenever a switch network configuration is executing or in case of failure recovery. The protocol used is designed to keep service communication low.

Packets are of variable length (application dependent) and are identified by special beginning of packet (BOP) and end of packet (EOP) control characters. The routing part is also of variable length and contains, for each switch chip the packet reaches, the port number to exit through.

SP Security

With respect to security, the SP is a cluster of RS/6000 workstations connected on one or more LANs with the control workstation serving as a central point to monitor and control the system. Therefore, the SP is exposed to the same security threats as any other cluster of workstations connected to a LAN. The control workstation, as the SP's single point of control, is particularly sensitive: If security is compromised here, the whole system will be affected.

General security concepts

Before discussing technical details of computer security, it should be emphasized that security must be based on a clearly expressed and understood security policy. This is a management issue, and each organization should devote sufficient time to establish a reasonable and consistent security concept. This should include a discussion on which individuals or groups should have access to which resources and what kind of data has to be restricted in which way. In addition, the risks of potential security breaches have to be evaluated with respect to both the costs to implement reasonable security measures and the influence of these measures on the ease of use of the overall system. An important cost factor is the time required for continuous security administration and auditing, which is crucial to maintain security over time.

When computers were monolithic systems without any network connections to other computers, security was much more confined than today. Each computing system could be treated independently from all others. It had, at least conceptually, a single administrator or root user, a single, global user database and access control information, and terminals as well as data storage devices were directly attached to the system. If dial-up connections were used, they were normally much more restricted and reliable than today's Internet connections. To access the system from a terminal, users typically had to identify themselves by a user name and prove this identity by a password that was checked against the central user database. Access rights for all users were recorded and enforced in a central location.

In today's distributed environments, many autonomous computing systems are interacting. They are often connected through an unreliable (in terms of security) network, which is used for remote login or command execution, remote data access, and also for remote system management. Each of these

computers may have its own administrator and its own user database and file space. The services they offer are often based on the client/server model.

The client software component requests a service on behalf of the user, such as a login request or access to a remote file. The server, which probably runs on a different machine, processes that request after checking the client's identity and permissions. There are normally three steps involved in these checks:

- Identification - The identity of the user is presented, for example, by a user name or numerical user ID (UID).
- Authentication - Some kind of credentials are provided to prove this identity. In most cases, this is a personal password, but other means of authentication can be used.
- Authorization - The permissions of the client to perform the requested action are checked. These permissions and their management are different depending on the desired action. For example, file access is typically controlled through the UNIX mode bits, and login authorization is normally controlled by the user database, which also happens to store the password, in other words, the authentication information.

Apart from the fact that most security exposures are still caused by poorly chosen passwords, the insecure network introduces several new security threats. Both partners of a client/server connection may be impersonated, that is, another party might pretend to be either the client (user) or the server.

Connections over the network can be easily monitored, and unencrypted data can be stored and reused. This includes capturing and replaying of user passwords or other credentials, which are sent during the setup of such client/server connections.

A common way to prevent impersonation and ensure the integrity of the data that is transferred between client and server is to set up a trusted third party system, which provides authentication services and, optionally, encrypts all the communications to allow secure communication over the physically insecure network. A popular system that serves this task is Kerberos, designed and developed by the Massachusetts Institute of Technology (MIT) as part of the Athena project.

AIX security

Since AIX is running on the Control Workstation and all the SP nodes, it is also AIX that provides the basic security features for the SP. AIX 4.3.1 has been certified to conform to the C2 level of trust classification by the U.S. National Security Agency (NSA) and provides many facilities for discretionary

security control. A good description of the basic security features of AIX Version 4 can be found in the book *Elements of Security: AIX 4.1*, GG24-4433. Although written for AIX 4.1, most of its content are still valid for AIX 4.3. This publication discusses the following main topics:

- Managing user accounts and login control
- Access control for files and directories including device special files
- Network security and control of remote command execution
- Logging and auditing facilities

3.1.8 Flexibility

One of the most important reasons for choosing the RS/6000 SP for server consolidation is the flexibility of adding nodes to the system. Each frame can accommodate 16 thin nodes, or eight wide nodes, or four high nodes, or any combination. Up to 512 nodes can be added on the SP. All the nodes can be centrally managed on the Control Workstation, which makes the system administrator's job a lot easier.

SP nodes are available anywhere three form factors: Thin, wide, and high. The SP is generally available with from two to 128 nodes (512 nodes through special order), which are packaged anywhere from one to nine logical frames. The number of physical frames can be more, depending on the types of the nodes.

The SP is designed to enable more than a thousand nodes.

Each node is a separate system having its own CPU, memory, and internal disk. This is called shared-nothing architecture. Each node can run a separate application. For example, an SP frame can contain nodes that run SAP applications, and other nodes can function as the SAP database server. Other nodes can be SAP training nodes, development nodes, QA Test nodes, and other nodes can run the Lotus Domino Server, and so on.

3.1.9 Applications

There are numerous applications available on the RS/6000 SP. Among the most popular are SAP R/3 and Oracle Financials for ERP and Lotus Domino for Groupware and enterprise messaging.

Lotus Notes Domino is one of the top choices for application integration for server consolidation. We frequently find customers consolidating their NT

Domino servers to the RS/6000 SP once they have realized the benefits from doing server consolidation.

All the UNIX databases, such as Oracle, DB2, Sybase, Informix, and Adabas run on the SP.

In addition, there are many application development tools available for developing your own applications to run on the RS/6000 SP.

There are also many applications available for Business Intelligence, Decision Support Systems, datawarehouse systems, and Supply Chain Management, such as i2's Rhythm.

Compuware has tools for migrating databases from different systems called File Aid. They also have development tools (UNIFACE), and they also have testing and monitoring tools (QACenter, QALoad, and EcoTOOLS).

Another third party vendor, BMC, has tools for scheduling (Control-M). They also have tools for monitoring (Patrol), tools for capacity planning (BEST/1), and system backup (SQL BackTrack).

BEZ is also a third party vendor that provides tools called BEZPlus, Investigator, SerView DBA, and CorpView DBA, for capacity planning and monitoring of databases.

Aside from these commercial applications, there are many scientific and engineering applications available for the RS/6000 SP.

Best of all, RS/6000 has all the applications required to run your company on the net.

From Lotus Domino to Netscape and Checkpoint Firewall, Comercepoint and MQSeries, the RS/6000 is a platform that is ready for e-business.

3.2 Reasons for consolidating to RS/6000

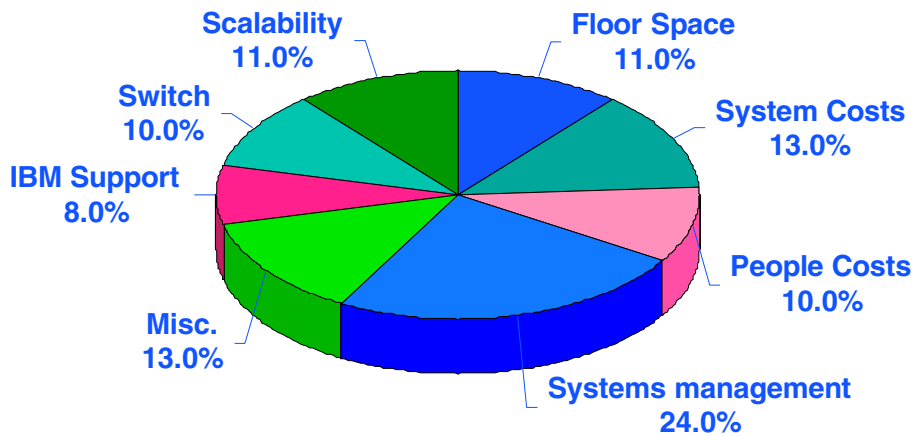
In this section, we will focus on the reasons why customers choose RS/6000 for server consolidation.

3.2.1 Why do customers consolidate onto an SP system?

D.H. Brown Associates' 1997 IBM SP Server Consolidation Study

D.H. Brown Associates conducted a study in 1997 on behalf of IBM's RS/6000 Division. The study focused on sites that had selected IBM RS/6000 Scalable Parallel Processing System, the SP, as their server consolidation platform.

Per IBM, many SP systems have been installed specifically for server consolidation. The D.H. Brown Associates study interviewed a sample of SP users, selected from a customer list provided by IBM, to understand why they made the decision to install an SP for server consolidation and what their initial experiences were. While some of the responses given to D.H. Brown Associates naturally dealt with IBM SP specific issues, most customers' experiences were of a general enough nature to apply to other target platforms. Detailed responses were provided to IBM in consideration of their sponsorship of the study.



D. H. Brown Associates, Inc.
RS/6000 SP System Server Consolidation
Study
October 1997

Figure 14. Why do customers consolidate onto an SP?

While D.H. Brown Associates initially sought to gather quantitative data confirming total cost of ownership savings, it soon became clear that quantitative data would be difficult to collect.

Many customers were growing quickly and did not wish to freeze growth just to remain constant for an *Apples to Apples* cost comparison. Furthermore, other customers were not even sure of the true costs of their distributed servers. Indeed, the very fact that they couldn't properly determine the costs of distributed servers was a powerful argument justifying centralization.

Even among those customers who felt they had some evidence of cost savings, there was not a clear consensus; each individual installation had its own story. Some found consolidated hardware platforms offered significant savings over distributed servers; others felt savings accrued primarily from software licensing terms and conditions.

From a list of 89 companies considered, 45 provided information used in D.H. Brown Associates' survey results. The customer list represented a broad mix, covering SP systems with only a few processors to those with well over 50 SP nodes. Some of the customers had recently installed an SP; others had over two years in production. The companies interviewed, predominately in the U.S., but also a few in Canada, were asked to identify the top reasons why they chose an SP for server consolidation as well as identify any problems encountered. The categories were defined to summarize responses, not to limit choices, therefore, customers stated their reasons in their own words.

Overall, the customers interviewed expressed strong satisfaction with their decision to implement server consolidation. Server consolidation appeared well-suited for customers looking to reduce systems management complexity as well as those planning for significant growth. Table 1 provides the percentages of the reasons involved.

Table 1. Top reasons for consolidation

| Reason | Reasons% |
|---|-----------------|
| Systems Management | 23% |
| Administrative Staff Efficiency (People Cost Savings) | 13% |
| Systems Cost Savings | 12% |
| Floor Space | 11% |
| Scalability | 11% |
| Interconnect | 10% |

3.2.2 What is the SP System being used for?

It is very interesting to note that most customers running the RS/6000 SP are using it for server consolidation. Twenty-three percent of the customers using SP are into Server Consolidation; 20 percent use the RS/6000 SP for commercial OLTP applications, and 16 percent of the customers use the RS/6000 SP for Decision Support Systems, Data Warehousing, and Datamarts.

What is the SP being used for ?

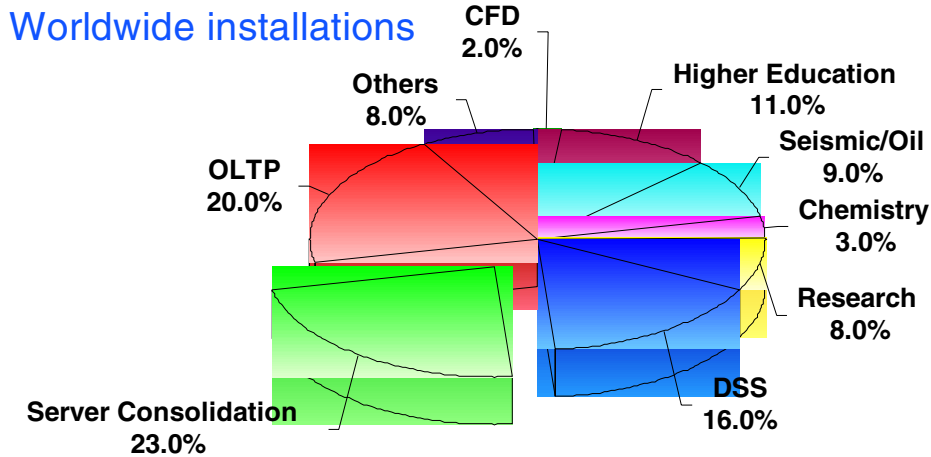


Figure 15. What is the RS/6000 SP being used for?

The rest of the customers use SP for higher education, seismic applications, computational fluid dynamics, research, and chemistry fields.

3.2.3 RS/6000 SP financial benefits

The key deliverable for the SP financial justification is the cost benefit analysis of a Server Consolidation project. These projects are driven mostly by a customer's desire to reduce the overall cost of computing for their application environment.

Many independent studies have concluded the largest component to the total cost of deploying distributed servers are system administration and management expenses.

The majority of these expenses are hidden because they are not easily accountable and are usually born directly by the end user. These expenses can manifest themselves in ways that reflect negatively on the systems and support staffs: Hard to manage, delays for application roll-outs, unreliable systems, and so on.

Therefore, interesting and lucrative benefits can be derived from a detailed comparison of the SP's system administration and management productivity gains versus other UNIX and non-UNIX alternative platforms. Clearly, system expenses are important, but they sometimes can be less than 50 percent of a project's true lifetime costs.

The key objective of the SP financial justification is to convert intangible benefits of the SP to tangible dollar savings. When these savings are combined with the system costs and calculated over a project's lifetime, the result is a number that reflects the overall cost of the project.

Repeating the process for alternative SP solutions, you can determine how significant the SP savings will be. This analysis is a highly effective way to educate, enlighten, and ultimately convince decision makers of the SP's significant financial benefits.

In general, the SP has three key platform benefits:

- Unlimited scalability
- Single system image
- Supports countless applications and middleware software

The SP excels at running parallel applications. Customers are able to start and then grow the system to the desired capacity in support of their business requirements. Most importantly, the incremental costs associated with additional growth are flat and predictable.

Traditional platforms, that is, non-parallel systems, tend to have compounding costs as they scale up to the business requirements and eventually cannot even achieve the necessary capacity and performance.

Parallel applications can be grouped broadly into two types: Numerically intensive (for example, forecasting, simulation, and optimization) and data intensive (for example, databases and transaction processing). Many business problems will combine both types of applications. Some real life examples include financial risk analysis for a bank lending department, asset portfolio analysis for a mutual fund investment firm, and understanding customer purchasing behaviors for a retail consumer goods company. In all these cases, the SP financial justification is usually based upon the projected return on investment for the application, system, support costs, and all other charges required to solve the business problem. While the customer may find these benefits to their liking, they usually go without quantification.

Since SP has been available, many customers have discovered SP is not just for parallel applications. Instead, they run many different applications on the SP as if it was a collection of independent servers administered and managed as a single machine. These customers are able to successfully manage large and complex environments with a minimal staff.

There are many benefits of running applications in this manner on the SP. The most significant benefit is reducing the cost of system administration, management, and infrastructure expenses which, in turn, increases all the applications' return on investment. These expense reductions can be projected by understanding the customer's current environment, their business requirements for one or more applications, and how the SP would be administered and managed for each application.

In fact, the financial benefits of consolidating applications to the SP are analogous to the benefits of running parallel applications on the SP. Many economies of scale are realized by the SP. The point at which the SP financially out performs alternative systems varies by application, scale, and customer skill set. Generally, systems with high, ongoing support requirements, for example, database servers, Lotus Notes servers, and SAP, are justified on small SP systems. Other applications, for example, file servers and communications gateways, require larger and more complex machines to be justified.

Clearly, no two customer environments are identical, and the financial justification will enable a precise understanding of the actual cost benefits of the SP, taking into account all possible considerations.

All these calculations can help in justifying the cost of the SP for server consolidation and can be done with the help of your IBM representative.

Chapter 4. Consolidation methodologies

IBM is a leader in server consolidation with its proven methodologies, patented tools and breadth of offerings, skills, and services. IBM's leadership in this market helps customers optimize and simplify their existing IT infrastructure and provide a foundation for new solution investment and implementation that supports the customer's business goals and objectives.

IBM's solution offerings distinguish themselves by focusing on "how to" skills and experiences to implement these projects successfully. IBM offers an end-to-end solution that can incorporate server, storage, and network components, utilize the appropriate offerings from IBM's Service Offerings, use IBM Global Financing to ensure the whole package is affordable, and leverage IBM Business partner skills.

Customers would like to optimize their IT infrastructure ... But what can they do about that huge portfolio of hardware, software, business, and decision support applications???

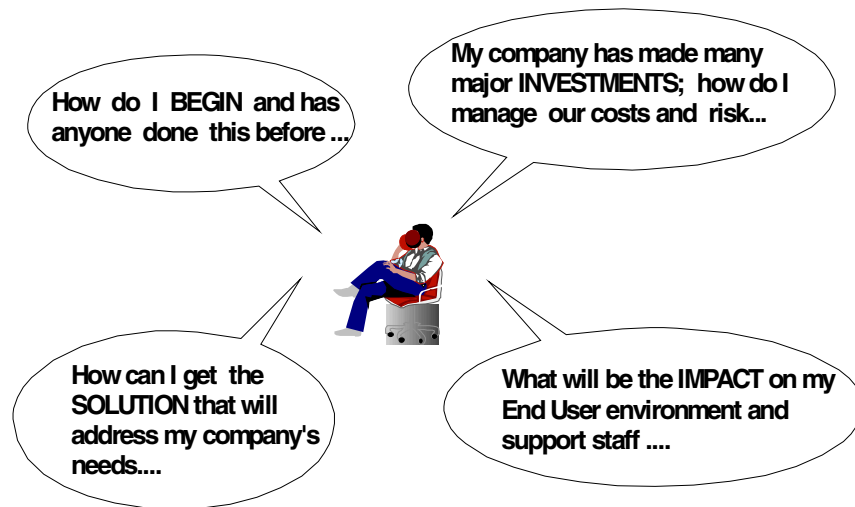


Figure 16. Customers' consolidation concerns

IBM has developed methodologies, Intellectual Capital, and toolsets for evaluating inventory, defining "from point A to point B", implementing

consolidation, and predicting and measuring cost benefits. This allows IBM to add value effectively and consistently across IBM's World-Wide Server Consolidation Solutions team.

The methodologies and modeling/predictor tools, such as the IBM ALIGN methodology, Business Solution Assessment methodology (BSA), Scorpion methodology, the IBM Cost of Ownership Management tool (ICOM), the Server Consolidation Savings estimator, and the Gartner Cost of Ownership tool, successfully assist customers in accessing their current business needs and in determining which IT process makes sense to consolidate and which ones do not.

The IBM ALIGN methodology

IBM developed a unique industry leading server consolidation methodology called ALIGN. This "best-candidate" identification methodology was designed to address the full spectrum of IT optimization and not just the reduction in servers. It is, however, extremely flexible and can be used for any size project or sub-project.

The ALIGN methodology analyzes your existing environment at a high level with a view to locating islands with a propensity for consolidation. These groups must then be further analyzed to produce applicable alternative consolidation scenarios. Assuming these alternatives are accepted in principle, IBM will then provide a formal proposal to address these issues.

BSA methodology

The Business Solutions Assessment methodology is being introduced specifically to address the requirements of customers looking at rehosting applications from one architecture to another for enhanced function with reduced complexity and cost. BSA includes modeling, predictor, sizing, and porting tools, a Web-based application questionnaire, business requirements analysis, application analysis, and portability assessment. BSA can be especially effective when users are considering porting an existing application to a different architecture or platform.

The Scorpion methodology

Scorpion is an IBM approach to simplifying the IT infrastructure. It is aimed at CIOs, IT architects, and board-level executives wanting to address the end-to-end cost and complexity of the IT infrastructure. It aims to identify the end-to-end costs of major services, front and back-office OLTP and ERP services, data warehousing, groupware and messaging, and intranet and Internet Web services. It includes an assessment of a wide variety of system architectures including NT, Solaris, AIX, and OS/390 as well as Oracle, DB2, CICS, and Java as major building blocks of the IT infrastructure. It

recommends migration to a large-scale hub service environment to achieve significant economies of scale through simplification.

ICOM

The IBM Cost of Ownership Management tool helps you guess the annual total cost of ownership per server and to calculate the cost savings of owning a smaller number of servers.

Server Consolidation Savings Estimator

The Server Consolidation Savings Estimator is a tool that generates estimated potential savings from server consolidation. This tool will enable customers to input data and obtain a cost savings estimate. The analysis from this tool is very preliminary, given the incomplete nature of the data used to calculate the results. Further potential benefits being reported are based upon certain assumptions that may not apply to your organization.

This tool is available on the Internet at: <http://www.rs6000.ibm.com/estimator/>

Gartner Cost of Ownership

The Gartner Cost of Ownership tool developed by the Gartner Group allows you to analyze and manage cost and other issues within your computing and business environment. The tool provides an automated method to measure, manage, and reduce the total cost of ownership.

The server consolidation tools focus on at least two key elements of the customer's server consolidation decision process:

- The identification of the appropriate IBM platforms that best meets server consolidation needs.
- An initial, high-level estimate of the potential financial benefits of consolidating on these platforms.

These tools could be used either as stand-alone implementations in smaller opportunity situations that do not justify the investment of time and resources required to complete the full ALIGN methodology, or as preliminary "interest generation" tools in larger opportunities where the full ALIGN methodology would ultimately be employed.

| | BSA | ALIGN | Scorpion |
|--------------|------------------------------------|---|-----------------------------------|
| Objective | Assess application rehosting | Optimize IT server infrastructure | Optimize entire IT infrastructure |
| How it works | Structured interviews (P&P) | Structured data gathering & Approach DB | Structured Interviews (P&P) |
| Duration | 3-5 weeks | 3-5 weeks | 3-5 days |
| Deliverable | Customer present. and report | SCON Projects proposals | Customer present. and report |
| Major skill | Porting& IT & Business Consultants | SCON Sales & Tech. | IT & Business Consultants |

Figure 17. Server consolidation methodologies comparisons

4.1 The IBM ALIGN methodology

ALIGN is a server consolidation methodology that is used to evaluate a company's existing IT infrastructure and weight potential cost savings to the point where the task is broadly repeatable in any customer engagement.

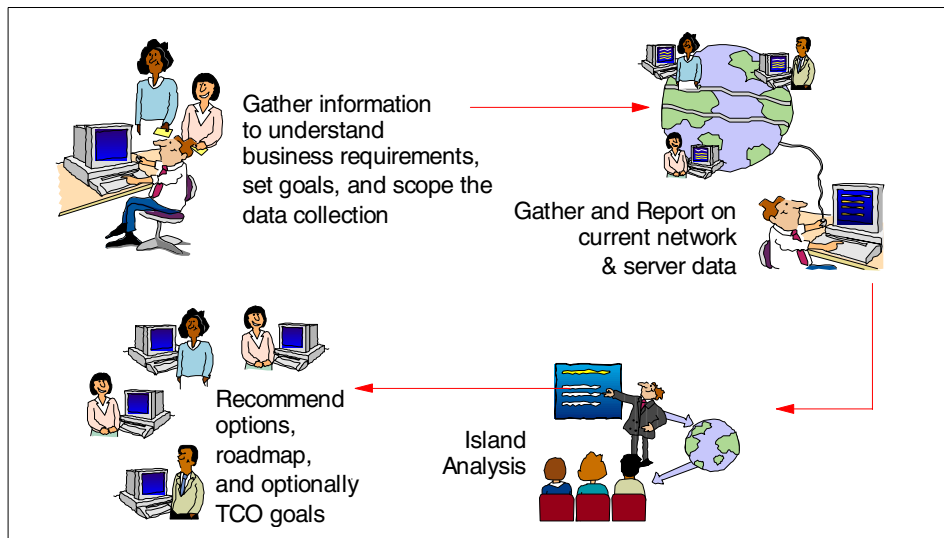


Figure 18. ALIGN server consolidation methodology

The ALIGN methodology allows users to develop a database capturing links between existing applications and specific servers and architectures as well as other application attributes. For example, the ALIGN generated database can contain mission-critical application information, such as existing service levels, source-code dependencies, structure of existing databases, and TP-monitor characteristics. One can then apply sophisticated analytical tools to the resulting data, yielding key insights into potential areas for cost savings and service-level improvements.

The ALIGN process focuses on analyzing servers and application inventory or *Islands of Consolidations*. This data can be used to identify the possible groups of applications that could be targeted for consolidation. This approach is different from consolidating small servers onto large servers. It allows a variety of valid business drivers to be taken into account and considered along with the constant changes in technology that always provide the IT architect with necessary input to your IT strategy.

Simplified Server Consolidation Business Model

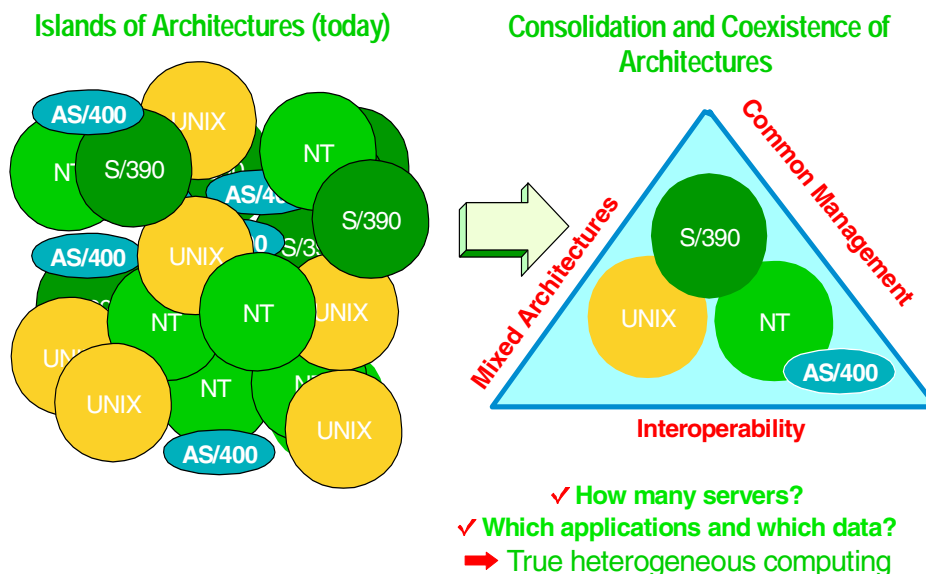


Figure 19. Server consolidation business model

ALIGN is designed to be highly customizable for a wide variety of customer situations. One key output is an asset database, another is user training in the ALIGN methodology. ALIGN draws on IBM's unmatched breadth of service resources based on IBM's and its consultants' decades of architectural-consulting experience.

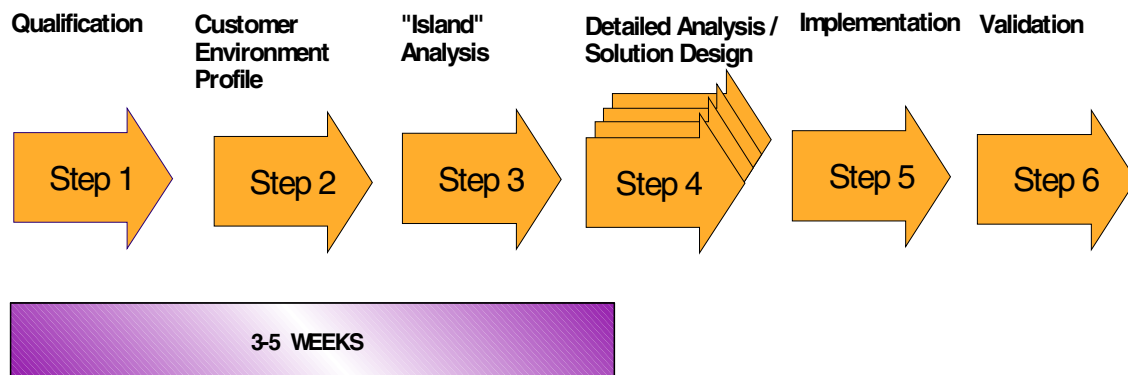


Figure 20. ALIGN methodology process steps

4.1.1 Step 1: Qualification

Assessments are structured as listening, studying and learning activities that can provide value to customers, IBM, and Business Partners for Server Consolidation solutions. It is important to understand and document business issues and IT strategies.

4.1.2 Step 2: Customer environment profile

The information gathered during the assessment is essential to ensure that the recommended solution will fulfill the business needs of the customer. It is important to ensure that the alternatives selected and the recommended next steps are directly linked to what the customer considers a business benefit.

4.1.3 Step 3: "Islands" analysis

In this step, we analyze the existing environment at a high level with a view to locating groups with a propensity for consolidation. Once these groups have been identified, they are mutually agreed upon by the customers, IBM, and Business Partner sales and technical staff.

4.1.4 Step 4: Detailed analysis/solution design

IBM, Business Partner sales and technical staff design the systems to meet the requirements, though these are strongly restrained to meet the strategic architecture that matches the customer's consolidation requirements.

4.1.5 Step 5: Implementation

"Business as usual". IBM, Business Partner sales and technical staff deliver the systems as specified and hand them over to ongoing support and maintenance.

4.1.6 Step 6: Validation

IBM, Business Partner sales and technical staff and customers assess the overall server consolidation effort. Validation is evaluated in terms of the transition costs, changed support costs, and with actual savings documented to the customer.

4.1.7 ALIGN examples

This section provides four scenarios in which ALIGN was used to evaluate existing IT structures.

An Insurance Company

Situation:

A very large number of Midrange/Intel Servers deployed in over 30 locations

- 805 UNIX Servers
- 2,500+ INTEL NT and Novell Servers

Concerns:

- Number of servers growing rapidly

Solution:

- IBM's ALIGN methodology used to evaluate the UNIX environment to identify islands of Consolidation opportunity
- Informal analysis of Intel environment to make recommendations
- Several recommendations made and now pursuing detailed analysis/proposals for consolidation including:
 - S/390 for notes consolidation

- Leverage of SP Switch
- Re-evaluate NT strategy for I-net solutions

A Life Insurance Company

Situation:

Consolidate 10 business-critical applications from SUN, RS/6000, and Intel servers to IBM SP and ESCON attached mainframe

Benefits:

A single point of management has made all the difference to the business by providing improvement in information sharing, customer service and quality, and a reduction in overall costs.

- Improved customer service
- Ability to handle rapid growth
- Simplified, cost-effective systems management

A Hospital

Situation:

Chose IBM SP over HP and chose Sun to consolidate 13 IBM and DG servers. This was done to provide a platform for future consolidation and growth

Benefits:

- Saved 30 percent overall costs in I/S budget
- Grew I/S infrastructure by 300 percent without adding additional support staff
- Implemented critical new applications without increasing floor space or support requirements

An Automobile Manufacturer

Situation:

Chose IBM SP over Sun and chose HP or NT as a consolidation platform for SAP, PeopleSoft, and data warehouse

IGS services used for SP, HACMP, ADSM, and Tivoli planning for implementation and customization.

- SP with 35 nodes, switch, HACMP, and ADSM
- ADSM and SP Switch to provide faster, more reliable backup
- Benefits:

This company saved \$200,000 per year by reducing support staff requirements by 50 percent.

4.1.8 Advantages of IBM methodology

IBM is the only vendor in the market today that can truly deliver server consolidation solutions that include all the elements customers require to implement these solutions quickly and successfully. This packaging includes not just the hardware and software elements, but also the services skills and expertise to make it work for the customer coupled with the financial packaging of IBM Global Financing to help customers meet their overall financial requirements and dispose of existing assets in a minimally disruptive fashion. IBM Global Financing is a powerful partner in building customer solutions. For example, at some point during the server consolidation project, you will have duplicate equipment, for example, old equipment and new equipment, which can affect your budget. By involving IBM Global Financing, we can design a financing stream that smooths over the financial life cycle of the project.

IBMs Server Consolidation methodology is designed to maximize return and minimize risk. It allows the customer to understand how and where to begin their server consolidation activity, accelerate return on investments, understand associated costs and risk, and clearly delineate the roles and responsibilities of all parties.

At every stage of server consolidation, IBM can work with your IT team to provide "peace of mind" guarantees and ensure that the right balance of cost and quality of service is addressed in every aspect of design, planning, implementation, support, and continuing systems maintenance and management.

4.1.8.1 Proven experience and expertise

IBM provides a dedicated team of server consolidation specialists and certified IBM Business Partners populated with Intel, Unix, AS/400, and S/390 skills that deliver the right technology to address the business requirement rather than force fit a particular architecture server platform that may not meet the needs of the customer's mission.

IBM has an excellent world-wide portfolio of references including server consolidation of different platforms to achieve cost-saving, service delivery improvement, data-access, or preparation for cost-effective out-sourcing.

4.1.8.2 Cross-Platform

IBM is the only IT vendor in the market today that is not predisposed to any one architecture or "box" solution. The methodologies and technology available allow IBM to integrate and address all of the architectures their customers have installed today and propose server consolidation solutions that address all of these environments in a way that best meets their customers' unique IT requirements and environments.

4.1.8.3 Application level when needed

Customer's are motivated by a business need to provide a higher service level to existing applications including faster response times, continuous availability, and/or increased access to data.

It is important to try to reduce the size of the infrastructure required for any given application and to reuse the architecture base wherever possible to ensure economies of scale.

4.1.8.4 Both data and server environment are considered

The key to an optimized environment is typically the use of the least number of architectures and servers required for a given application. This not only reduces support requirements but also minimizes complexity and reduces points of failure due to incompatibility and complex bridges. Also by simplifying the server environment, data access and turn-around time is improved.

4.1.8.5 IBMs services and support

In general the IBM support for the RS/6000 is comprehensive and flexible, allowing you to tailor a contract that provides the coverage and the degree of support you need. This can be temporarily modified, for example, to provide 24-hour support over the week of a complex upgrade, therefore, ensuring availability of both defect and "how-to" support at absolutely any time.

All products and services are not packaged and available in every country. Contact your local IBM sales representative for more information.

4.1.8.6 Key service offerings

Align analysis

- Steps 1-3 of the ALIGN process

- The entire ALIGN project

Project Definition Workshop

- Total project planning with customer

SmoothStart

- Initial installation and configuration service including basic AIX training for system administrators

ProjectWatch

- Regular monitoring and recommendation for best use of the system including its applications within the customer environment

SystemWatch

- SP system and system software

Migration Service

- Installation of additional nodes and upgrades to system software for the SP to minimize disruption of the customer environment due to functionality

Systems Management Transformation Projects

This service is only for RS/6000 SP customers in Europe, the Middle East, and Africa (EMEA). This will address all areas of Systems Management, assisting you in setting goals, designing solutions, and selecting software together with training programs to establish new skill, organization, and procedures for:

- Configuration
 - Hardware and software inventory and asset management
 - Software distribution, installation or reinstallation management
 - Storage management
 - Network management
- Operations
 - Measurement and reporting of service level metrics
 - Change management and problem management
 - Performance management and capacity planning
- Data management
 - Back up/restore and archive/retrieval solutions
 - Database management

- Batch/file transfer
- Availability
 - High availability cluster management
 - User and security management
 - User support/help desk
 - System monitoring: Error reporting and problem determination
 - Automated operations and event management

4.1.8.7 Business relationship

IBM Business Partners are trained and accredited to deliver the skills that implement IBM's systems efficiently and effectively. The service provider program allows partners to deliver IGS offerings that are developed and maintained on a world-wide basis.

4.2 Business solution assessment

The Business Solution Assessment is a platform migration analysis offering. BSA is used to assist customers who are considering porting code to a new platform as part of a server consolidation project. The BSA will assist in identifying a pilot, assessing risk, and estimating the level of effort to complete a port. The BSA typically takes 3-5 calendar weeks, during which 2-5 days are spent on-site interviewing technical people. Data is gathered throughout the entire BSA process through a series of questionnaires. Outputs from the BSA include a place to start, a high-level project plan, and estimates of cost and risk.

The BSA helps to determine the effort in migrating code to a new, consolidated platform. This presupposes that the customer has access to the source code. A porting assessment includes:

- How the code is written
- How standard is the code
- The scalability of the code

If the customer is running packaged software from an Independent Software Vendor (ISV), the nature of a BSA changes dramatically. A BSA can still be done, but in these cases, it would usually be done with the ISV being the "customer." If all the applications are ISV software, and the ISV is not interested in changing platforms or the applications already run on the target

IBM platforms, then server consolidation can still be pursued. The BSA can be an outcome of an ALIGN analysis or a stand-alone project.

BSA includes conducting a risk assessment and availability impact analysis for recommended solutions. In addition, the methodology utilizes knowledge based online tools and surveys to:

- Analyze IT environments at a high level
- Understand database migration and application support prerequisites
- Evaluate estimated costs

4.2.1 Overview

BSA is a repeatable business process. Setting expectations, understanding customer's business and technical requirements, providing recommendations, identifying first project candidates, developing a high level project plan, and suggesting next step activities are the objectives of BSA.

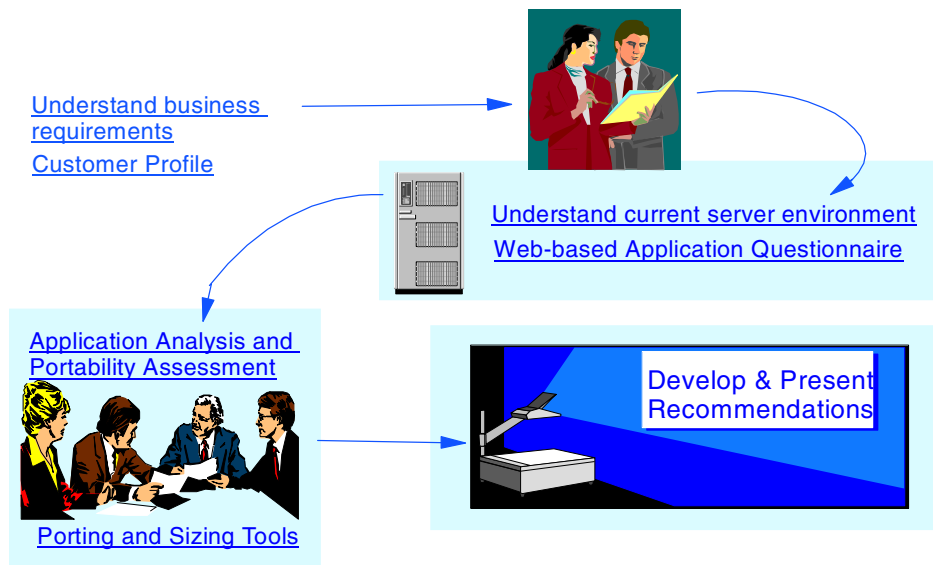


Figure 21. BSA overview

The BSA provides a consistent and methodical approach that ensures that proper expectations are set and that the customer's business objectives are met with minimum risk. This methodology partners the customer, various IBM resources, and IBM business associates to develop a high-level plan, recommendations, and next step activities. Since BSA is a flexible

methodology, recommendations can support customer requirements for a non-bias solution or to validate a predetermined solution.

Through the BSA process, surveys will be used to gather customer environmental information. Surveys will contain an inventory of hardware, software, applications, and databases. The information obtained will be analyzed and used to:

- Understand assessment personnel requirements
- Construct structured interview sessions
- Define assessment scope
- Develop interview session objectives

The BSA methodology consist of three phases: pre-assessment, Assessment and post-assessment.

Business Solution Assessment Process Phases

Phase one: Pre-assessment

The first phase of the BSA methodology is to identify the business sponsors and understand, in detail, their goals for the project. While most project initiatives are undertaken to reduce operational costs, other improvements in data access, service levels, standards, and capacity utilization are also possible. In addition, it is also important to identify all sponsors and affected parties within a client's organization. It is important to win the support for each project under consideration.

Conducting an inventory of existing applications and the current associated infrastructure enables the subsequent analysis phases of the process. Inventories, therefore, must be as complete as possible and efficiently conducted. Depending on the characteristics of the project, a variety of surveys are used to conduct thorough information gathering.

Once the inventory of hardware, software, applications, and associated databases are completed, the analysis and assessment project planning activities can take place. Personnel requirements, assessment objectives, the project scope, and schedules for interview sessions can be determined.

Phase two: Assessment

The second phase of the BSA methodology involves conducting a more detailed application analysis on-site with application owners and other personnel to validate the project effort. In addition to the information required to properly size the target hardware and software configuration, application changes and porting considerations are also estimated. Determine portability

of the application and the possible difficulty of the port. Sizing and performance tools from various IBM product divisions, labs, and business partners are used in this phase of the project.

Phase three: Post-Assessment

The third phase of the BSA methodology includes developing a high-level project plan and determining the proper resources needed. Since a primary reason for most customer I/T projects is cost reduction, a cost/benefit analysis, incorporating all aspects of a project, should be conducted to justify the project. The assessment results are reviewed with all parties involved. The standard documented results of the assessment include:

- Determine portability of the application
- Define recommended solution
- Adjust initial sizing
- Provide strategic recommendation
- What and how much to port?
- Recommend follow-on items
- Document alternatives

Customer benefits

- Understand how and where to begin
- Teamwork with IBM and IBM Business Partners for a solution assessment based on IT strategic direction
- Recommendations developed using knowledge-based tools and experiences from IBM and IBM Business Partners
- Identification of initial projects, project plans, and any associated risk
- Documented costs and business justification for investments
- Accelerated return on investment by reducing the time needed to place the solution into production

4.2.2 BSA case study

Company X, Inc.

This scenario represents the output of the Business Solution Assessment (BSA) that was performed for four major components of the APPLICATION 24 at Company X. The purpose of the assessment was to examine the characteristics of the specific Application 24 and determine the viability of moving the application to an IBM platform.

Current Environment

APPLICATION 24 was jointly developed by Company X and Vendor A. It was designed to maximize Company X's billing revenue in a five states region. APPLICATION 24 is focused on sorting and correcting the billing data that is incomplete and sending it back to the mainframe for proper billing. Based upon the interviews, Company X is satisfied with the application function; however, the application is faced with continuous growth and cannot handle the projected growth on the existing platform. Company X is planning to deploy this application in states where larger volumes are expected. Growth and volumes are the primary reasons that Company X is evaluating a new platform for APPLICATION 24. Today, two components of APPLICATION 24 are experiencing 12-15 outages annually. However, availability was not mentioned as a primary motivation for re-hosting the applications.

APPLICATION 24 has been in production since 1994 and currently supports five states with plans to grow to seven states in January of 1999. APPLICATION 24 processes 2.6 billion AMA records a month (across all RAOs). The addition of State 1 and State 2 will add approximately 4 billion records monthly because of the need to capture more detailed records for the billing regulations. Company X is currently running production sites on six NILE 150 processors and is evaluating upgrade. APPLICATION 24 is connected to the Company X mainframe in order to provide transfer to and from the billing system.

Current Challenges

Since deploying the applications in production, Company X's challenges have been focused on growth and increased volumes due to increase in the detail being captured. The application is growing at the rate of approximately five percent per month. The following represents IBM's understanding of the challenges related to these applications.

Immediate Growth concerns: The APPLICATION 24 team has the immediate need to upgrade three locations to alleviate the constraints. CPUs will be at maximum capacity within two, five, and nine months at each of the three locations, respectively. By accelerating the porting initiatives, it may be possible to avoid these upgrades.

State 2 Deployment: The APPLICATION 24 team has to respond to the increased scale of data required in State 1 and State 2. This deployment increases the average file size from 2.6 billion to 4 billion records. The current platform may not be able to handle this increased number of records.

Maintaining the ongoing projects (for example, Year 2000 initiatives): The APPLICATION 24 team needs to respond to ongoing and pending

projects as directed by Company X and government regulations. To reduce the cost of computing of mid-range platforms, the APPLICATION 24 team will leverage the increased buying power and support structure offered from IBM through the new mid-range agreement.

Recommendations

Based upon the information gathered, IBM recommends moving the non-MVS major component of APPLICATION 24 to the existing S/390 systems and moving three of the remaining components of APPLICATION 24 to an RS/6000 and 7133 environments. The IBM team is encouraged by our findings to date and looks forward to a more complete evaluation as we finish the review of the Code Generator portion of the application.

Our intent is to move the three components of APPLICATION 24, with minimal changes, to the RS/6000 AIX environment. This system was selected after close analysis of the facts learned during the BSA, evaluation of the proposed Pyramid platform, and the performance analysis completed by our technical staff. We feel the recommended solution will provide the capacity that Company X needs along with a strong platform for continued growth. The application will remain on Oracle 7.2.3 and utilize Oracle Forms 3.0. We are recommending the use of COBOL and C for the application environment.

IBMs intent is to deliver both the code generator and run-time in tandem; however, if this is not reasonable, it is IBMs recommendation to deliver the run-time portion of the application first in order to deliver value faster and a more cost effective production environment and then go back and port the generator to complete the effort. IBM, Company X, and Vendor B will port the existing code and run-time to the recommended IBM platforms.

IBM will include an educational assessment for training for the APPLICATION 24 and system administration personnel and provide an IBM Smooth Start as a component of this deployment. IBM will provide access to training through the use of the IBM education cards.

IBM and Vendor B have developed a good understanding of Company X's requirements through the BSA process. We understand the necessity of a clean and positive porting outcome and the important role the APPLICATION 24 plays at Company X. We believe our challenges to be in understanding the code generator application and see that as an outstanding risk, but feel that with closer analysis we will define this and present a complete and comprehensive statement of work. We strongly feel the challenges will be outweighed by the advantages of the IBM platforms and look forward to a successful project with the APPLICATION 24 team.

Hardware and Software Recommendations

IBM recommends the non-MVS components of APPLICATION 24 Main be ported to the existing S/390 systems. IBM also recommends that the following hardware configurations be installed to support the remaining components of APPLICATION 24 as currently deployed. This configuration was developed by taking into consideration the existing performance, capacity data, and future growth expectations provided by Company X and working with IBM technical resources to determine the solution. It is our intention to place an S70 in the geographical locations specified by Company X to provide coverage of the seven states operating environment. It is our understanding that we will place six production systems and one test system for this set of APPLICATION 24 components.

Additionally, we researched the proposed Pyramid solution for future upgrades to ensure that the IBM configurations will provide excess capacity and performance beyond the specifications presented to IBM by Company X.

Chapter 5. Customer consolidation scenarios

There are many factors that drive customers toward considering server consolidation. Out of these, the two most important are the reduction of the total cost of ownership and the enhancement of functionality. This chapter presents various scenarios of customers, some real and some hypothetical, who have successfully consolidated their systems to RS/6000.

Note

As of the writing of this redbook, some hardwares or solutions have been withdrawn.

5.1 IBM Yasu lab

Our first scenario is that of the IBM Yasu lab, located in Japan, that was able to achieve a 49 percent reduction in system costs by consolidating 53 of their servers onto an RS/6000 SP.

The following business needs were taken into account for the implementation of the consolidation:

- Technical need
 - Yasu lab needed higher density to design and develop LSI chips
 - Demand for shorter turn around time
 - UNI/SMP technology limit for computing power
 - 3D simulation needed for designing semi-conductor chips
 - Need for more CPU power for calculations
- Management/operation need
 - Need for more system administrators
 - Need for load balance between all facilities
 - Less difficulty in viewing the usage of all the systems
 - Effective usage of invested resources from the view point of whole center
 - Rapid growth in cost to manage and operate systems

The solutions for these needs were as follows:

- Consolidate 53 servers to three RS/6000 SPs with 21 nodes

- Server integration
- Slim clients
- High speed networks

The overall benefits to the lab (shown in Figure 22) were as follows:

- Reduced hardware cost to 66 percent
- Reduced license cost
 - VHDL simulator reduced 38 percent
 - Logic synthesis reduced 25 percent
 - Test pattern generator reduced 20 percent
 - Layout reduced 20 percent
- Server maintenance hardware cost reduced 63 percent
- Improved CPU utilization up to 35 percent
- Consolidated space of servers reduced 76 percent
- Improved total turn around time

IBM JAPAN - YASU

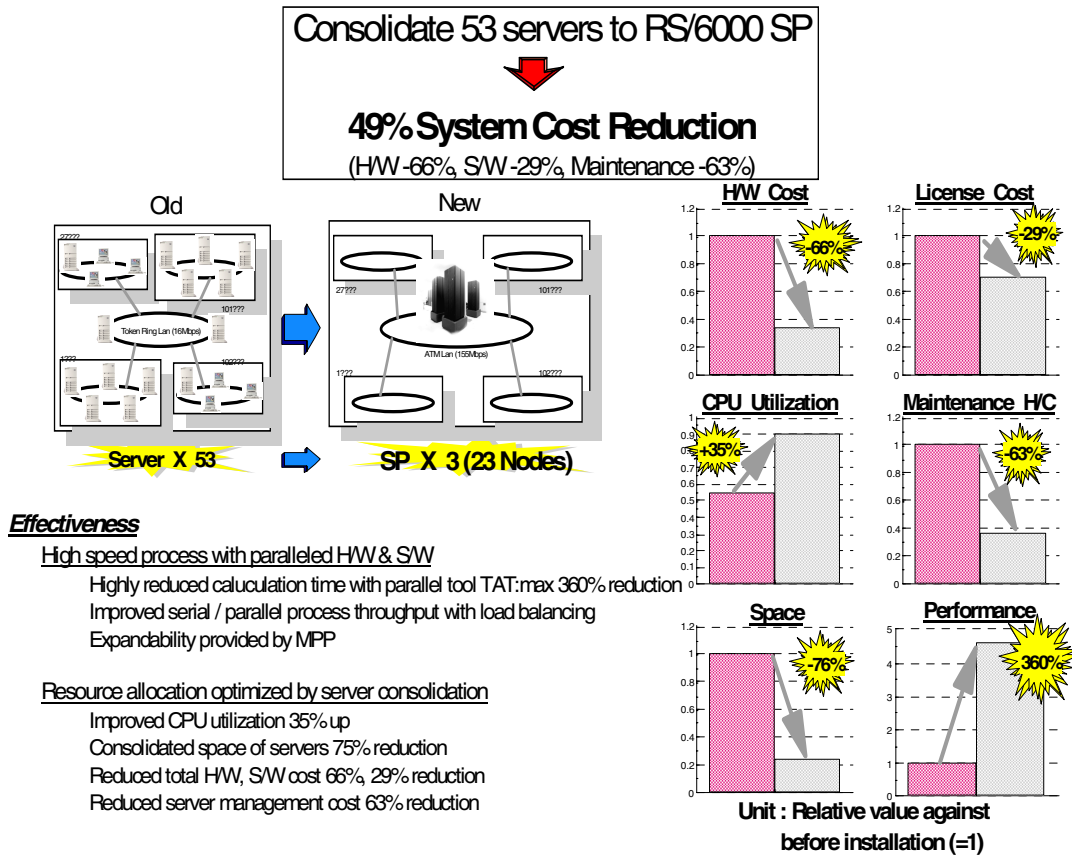


Figure 22. IBM Yasu lab

5.2 An investment services company

This scenario involves an investment services company whose main focus is in the rating of business and municipal debt. By consolidating their existing system onto an RS/6000, this company experienced significant improvements in systems management and scalability.

The following business needs were taken into account for the implementation of the consolidation:

- Proliferation of UNIX workstations resulted in systems management issues:
 - Performance planning
 - Capacity planning
 - Unmanageable backup/recovery
- Reduce the number of other servers
- Improve systems management capabilities
- Future scalability

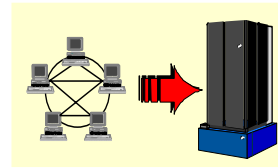
The solutions for these needs were as follows:

- Consolidate 15 servers onto a 5-node RS/6000 SP and five RS/6000 servers and workstations, all financed through IBM Credit Corporation

The overall benefits to the institution (shown in Figure 23) were as follows:

- Improved access to critical information
- The ability to consolidate Sybase servers into a flexible, manageable, scalable environment
- Systems administration costs reduced 50 percent
- System backup time cut 75 percent
- Database consistency check time cut 55 percent

- Debt Rating and Financial Information



- Phase I

- Multiple Sybase System 10 database servers consolidated on to single SP system
- ADSM storage management on SP for enterprise backup
- Partitioning by business area

- Phase II

- Additional SMP nodes
- Exploit high availability infrastructure
- Parallel database and query

- Result

- System administration cost cut to 1/2
- DB replication time cut to 1/8th
- System backup time cut to 1/4th
- DB consistency check time cut to 1/3rd

- Result

- Cost performance improvement
- Improved access to critical information
- More complex models and Analysis

Figure 23. Investment services company

5.3 An automobile manufacturer

This scenario is that of an international automobile manufacturer. This company is under pressure to reduce expenditures due to sales, general, and administrative costs and is also needing to upgrade its technology infrastructure to compete more effectively in the automotive marketplace.

Business Need

This company had two main concerns with their e-mail system: The international connectivity was erratic, and the response time was poor and unreliable. The CEO requested that its e-mail be fixed.

Description of Solution

The choice was made to use an IBM RS/6000 SP as the consolidation platform for SAP, PeopleSoft, and data warehouse. Their new e-mail system replaced the OfficeVision MVS (OV/MVS), which was several releases old. The fix e-mail mandate necessitated a change to the entire network and workstation infrastructure for this company to support the selected e-mail

package, which was Microsoft Exchange/Schedule+. The IBM content of the solution was program office and project management, technical and educational consulting, and IT application and infrastructure architectural design and development. All hardware and software was from third parties and directly contracted to the company.

IBM managed the design, development, and implementation for an upgraded network infrastructure and client/server WIN95 environment for this company. IBM directed the work efforts of more than 90 client and contractor personnel during a period of six months to complete the following across 32 national sites:

- Develop/implement a state-of-the-art ATM/fiber and intelligent hub network infrastructure for eight local sites
- Deploy Frame Relay across 32 national sites
- Upgrade/replace 52 older technology print/file servers
- Integrate more than 352 client applications into WIN95 and develop associated tools and WIN95 applications for application installation and upgrade
- Customize MS office suite and e-mail applications per client requirements
- Upgrade and replace more than 3,500 client workstations (3 desktop/4 laptop configurations) with standardized WIN95 software image
- Develop/implement new hardware "break/fix" processes nationally
- Train more than 2,900 clients across 32 national sites on new hardware, software, and processes

Key Functions of Program Office:

- Develop and manage overall integrated Program Office project plan and schedule, thus, enforcing close linkage to individual project plans.
- Develop and manage key infrastructure processes (communication, budget management, resource requirement, and issue management) across teams and up through the Chief Information Officer.
- Discuss and report on projects' status, budget, resource, and issues to Executive Steering Committee on a weekly basis.
- Drive weekly analysis and management of individual project managers' progress, issues, and resource requirements.
- Drive daily issues escalation/resolution meetings with PM team.

- Develop and manage Quality Assurance process for all program/project deliverables.

Benefits/Value to Customer

What was implemented provided a client/server based solution that produced better reliability, responsiveness, and access to the automobile manufacturer's international environment. A considerable sum of the annual budget was saved by reducing support staff requirements by 50 percent. Also, ADSM and SP Switch provided faster, more reliable backups.

5.4 A medical facility

This scenario is that of a hospital that provides an acute care, non-profit community medical center that has served its community for almost fifty years.

This facility combines high-caliber, skilled medical and nursing staffs with a friendly, helpful, sensitive, and supportive approach to patient care.

Business Need

This hospital wanted to consolidate its 13 stand-alone UNIX servers onto a single platform to increase manageability and reduce costs and floor space.

Description of Solution

It was decided that the best solution was to replace 13 stand-alone servers with an IBM RS/6000 SP system. The SP solution was used as an integrated application and database server for clinical management and back office financial systems. A second phase of the project included plans to consolidate additional servers and applications onto SP nodes.

RS/6000 Details:

- Type and number of RS/6000 installed: Two RS/6000 9076-550 SP frames with an SP Switch Router
- Type of applications being run on the RS/6000: Clinical and financial patient billing and accounting applications
- Number of users: 3,000
- Topology: FDDI network with plans to switch to an ATM
- Version of AIX being used: V4.3.1

Benefits/Value to Customer

The RS/6000 SP implementation enabled this facility to improve its IS infrastructure 300 percent without hiring additional UNIX administrators or

network administrators. The hospital has implemented new patient care and business applications without increasing floor space. The solution cut the hospital's overall IS budget costs by 30 percent.

5.5 A life insurance company

This scenario involves an international life insurance company that has \$10 billion worth of assets under management in the UK alone and makes payments of over \$1.4 billion to policyholders and beneficiaries each year.

It employs 1,500 office workers and has a field force of 850 in the UK. Its multiple lines of business offer its customers a range of financial services, including individual and group life insurance policies, pensions, and savings/investment plans.

Business Needs

This company has grown and expanded at a rapid pace over the past few years. Early in 1997, it became clear they were going to need a fresh business environment to cope with future growth.

This company had a number of business critical applications, including Lotus Notes, already running on small IBM RS/6000 servers. Lotus Notes had improved the company's communications and operations, and has become mission critical in that this company can not function effectively without it. An increasing dependency on Notes and a number of other applications, coupled with the desire for consolidation, led this company to an IBM RS/6000 SP.

The company had more than ten business-critical applications, all of which were reaching their peak. The choice was either to buy an appropriately-sized server for each application, or to look at it slightly differently - with a degree of foresight - and opt for server consolidation instead.

With an eye toward the future, the company evaluated what was the most sensible solution to solve their business needs. The answer wasn't to create another server farm, it was to have the application put onto one manageable and scalable platform.

Description of Solution

IBM RS/6000 SP: Consolidation and Upgradability

A great deal of time and effort was put into choosing and deciding upon the best business environment. The strength of IBM as a company and the security, reliability and availability of the RS/6000 SP were the main factors involved in the final decision.

Also, the scalability of the RS/6000 SP meant that this company would have no problem dealing with growth or expansion within the company.

Their RS/6000 SP was upgraded from one frame containing six thin and two high nodes to an additional frame with eight silver wide nodes. The upgrade provided a Lotus Notes fallback node and extended the functional use and number of Notes users to over 2,000 while, at the same time, providing more processing power for other applications.

The company went from zero to a frame that was three-quarters full in ten months and has filled one frame up inside a year. That would never have happened if the consolidation wasn't working well. The company now has fourteen applications, at least ten of which are business-critical, running on the RS/6000 SP, which includes everything from the securities and investment administration system to IT management. Looking ahead, this company is planning on installing some image processing software as well.

The concept of a single point of management has made all the difference to the business. It has meant improvement in information sharing, customer service and quality, and a reduction in overall costs.

Rather than fourteen separate smaller RS/6000 systems, each one in need of its own management, this company now has them all in one manageable environment. They now have two frames, and with the business growth they foresee, they should be looking for a third frame early next year.

5.6 An insurance company

This scenario is that of an insurance company, which is one of the largest insurance companies in eastern Europe. Their business portfolio consists of standard insurance, such as life, pension, and car insurance as well as related financial options.

Business Need

In 1995, they realized that its existing distributed IT environment was not able to accommodate increased user demands, and the system scalability and options for adding new applications was limited. The management of the distributed server farm was also causing the customer serious problems. The customer's servers were installed throughout the country's regions, and there was no online communication between them and the central server. Data from the various regional servers was collected off-line and processed as batch jobs, thus, causing a two to three week processing delay. Consequently, management decided to renew its IT infrastructure by consolidating the server farm.

The customer sought to replace the following systems:

- One Pyramid Nile 150
- One Sun 2000
- 16 ICL DRS/6000s
- Intel based PC servers

Description of Solution

This company decided to consolidate its old server farm onto an IBM RS/6000 SP system to enable the redesign and introduction of new business applications for the customer's changing business demands. The new environment ensured that the customer could consolidate databases from its different business units, therefore, making compiled data easier to manage and more consistent. Additionally, users would now have online access to applications.

Since the customer's applications are mission critical, they configured the SP nodes in a HACMP cluster. The nodes are grouped in pairs ensuring that another node can take over if one fails. The HACMP cluster is configured in "mutual-takeover" mode, which means that during normal operations, all nodes are active. If a node fails, the surviving node acquires the failed node's resources and continues to provide the failed node's critical services. Performance of the system is affected when this occurs; however, the service of the application will not stop, which is an essential requirement for this customer.

1.8 TB of 7133 Serial Storage Architecture (SSA) is connected to SP nodes for use as database servers. Each database is physically connected to two nodes in a twin-tail configuration for high availability purposes. 3590 Tape Subsystems are used on a database server node for backup, and the backup/restore functions are managed by ADSM for AIX, which utilizes the high-speed SP Switch to gather data from other nodes for backup.

RS/6000 Details:

- Type and number of RS/6000 installed: RS/6000 SP (9076-306) with three frames including eight high nodes, two wide nodes, and two wide Silver nodes
- Type of applications being run on the RS/6000: Oracle Financial V10.7 and several self-developed Oracle based applications
- Number of users: 6,000 (1,500 concurrently)
- Topology: LAN (FDDI, Ethernet); WAN (FDDI backbone, leased line)
- Protocol: TCP/IP

- Version of AIX: V4.2, V4.3

ADSM Details:

- Server: ADSM for AIX
- Version of ADSM: V3.1
- How many clients and of what type: Eight AIX clients
- Backing up: 6-700GB backed up to tape (3494, 3590)
- Database: ORACLE RDBMS V7.3
- Database backup occurs online with ADSM Connect Agent for Oracle and Oracle EBU

7133 SSA Details:

- Description of capacity/configuration: 1.8 TB (24 x 7133-020 with 16 x 4.5 GB disks, and 22 x SSA adapters)
- Mirroring: Yes
- Multi Host: Yes
- Processor Type(s)/Model(s): SP
- Operating System(s): AIX
- Main application/workload and database stored on the 7133s: ORACLE RDBMS v7.3
- Storage Management Application: ADSM for AIX V3.1

3590/3494 Details:

- Number of 3494 Tape Libraries (if present): 1
- 3494 Library Cartridge Capacity (if present): 200 Cartridges
- 3494 Library Tape Drives: 2x 3590
- Tape Drive Attachments: SCSI
- Control Programs: AIX V4.2
- Key Applications: ADSM

Benefits/Value to Customer

Although they have not quantified the business benefits of the solution, this company has acknowledged that operational cost have decreased. This company is committed to use IBMs RS/6000 SP system as a standard, and its future IT strategy will be based on the SP. The company feels that the system has offered an excellent return on its investment.

The factors that contributed most to the success of this company's consolidation include:

- Best solution for the price/performance ratio
- The company was convinced that IBM's SP system was the most scalable, powerful system capable of serving a mission critical environment.
- IBM established a good relationship with the company, which strengthened long-term confidence in IBM.

5.7 A university

This scenario involves a state university located in the United States. Throughout its 135-year history, this university has served the people, the region, the nation, and the world through extensive, multi-purpose programs encompassing instruction, research, and public service.

This university is designated a Research University I, the top category of the Carnegie Foundation's ranking of research institutions. This places them in the top 2 percent of the nation's colleges and universities. The Research I designation is shared by only 45 public and 25 private institutions.

One of only 25 universities nationwide holding both land-grant and sea-grant status, they are actively pursuing space-grant status. Their research is supported by its nationally known library and by a supercomputer, which places this university among the top 20 universities in the country in terms of computer capacity.

Business Need

This university had a modem pool that could no longer support the level of traffic required. The modem pool was always busy and the university was looking for an ISP to help solve this problem.

Description of Solution

The university rolled out the IBM Internet Connection for Education (ICE) offering to more than 500 users. These users were a combination of corporate dial users and credit card dial users. The users now have access to the university's homepage and can collaborate with peers and professors via the IBM Global Network.

Benefits/Value to Customer

This university is now able to spend its time on mission critical and education matters without being concerned over the specifics of running the modem pool.

5.8 An aircraft manufacturer

This scenario involves an aircraft manufacturer that is an European multi-national consortium with a worldwide reputation for setting the standards in modern, efficient transport aircraft. Created in 1970, this company has developed a complete family of short, medium, and long-haul aircraft with seating capacities ranging from 124 to approximately 400 seats.

In its 28 years of operation, they have booked a total of more than 3,200 firm orders and delivered over 1,890 aircraft to its international customers.

Business Need

Due to strong growth, the airline industry is a market that presents multiple unique challenges to the IT departments of aircraft manufacturers. This aircraft manufacturer was constantly adding new projects, and each time a new project was added, additional computer resources were needed. The new projects required new computer resources, and over the years, the customer had acquired a wide variety of UNIX servers and workstations, which made for a very distributed environment. To stop this distributed system approach, IBM proposed a consolidation of their UNIX servers onto an RS/6000 SP system.

This server consolidation project involved approximately 50 IBM and non-IBM UNIX servers. Most of the applications being run on the servers were specific applications including business, customer support, and technical documentation applications.

Description of Solution

IBM built a strategy with this customer in an effort to better manage costs through an optimization of existing computing resources with centralized management. The customer's old environment included many complex client/server applications. There were generally one or two applications per server, consequently, the high operating costs were proportional to the number of applications. IBM and the customer decided server consolidation was the best way to address these problems.

The initial RS/6000 SP implementation launched at for this aircraft manufacturer had a small configuration (four nodes) and was meant to

demonstrate the technical capabilities of the RS/6000 SP and its compatibility with the installed AIX base.

The customer's success conditions included:

- Accordance with their processes -- Applications needed to be tested by moving them from development to validation and production modes.
- Full staff training with the understanding that all tests needed to be completed before pursuing operations on the SP.
- Strong support from IBM engineers.

The trial was a success, and their RS/6000 SP system now includes five frames and 26 nodes, four of which are configured with HACMP for mission critical applications. Six nodes are used for SAP applications, two for Tivoli, and 18 nodes are used as application nodes. The SP is managed with standard SP tools (PSSP), and the customer is validating Performance Toolbox.

The IBM local team worked closely with the project managers throughout the implementation. Because a critical factor in the implementation was to demonstrate the SP to the customer in production mode, close collaboration with the project managers was important. The aircraft manufacturer's staff has taken part in various IBM meetings and marketing presentations, such as SP World, each year for further SP training and knowledge.

RS/6000 Details:

- Type of applications being run on the RS/6000: This includes a wide variety of applications on the different nodes including: SAP R/3, Oracle V7 (V8 planned), MQSeries, Tivoli, Lotus Domino, Communication Server, Loadleveler, and HR Access (developed by IGS). Some of their specific applications include: SB-COMP (technical documentation), ACMM Airport, BFEMS, EMS, and SCODA
- Number of users: All applications are client/server, and more than 2,500 users can access the SP. For example, SAP has 300 connected users and 7,500 customers.
- Topology: TokenRing and Ethernet.

SAP R/3 Details:

- ISV software application and release: SAP R/3 3.1.H
- Platform:
 - DB SAP Production Node: Wide 332 Mhz

- AS SAP Production Node: Wide 332 Mhz
- AS SAP Production Node: Thin 332 Mhz
- DB SAP Validation Node: Thin 332 Mhz
- AS SAP Validation Node: Thin 332 Mhz
- SAP Development Node: Thin 332 Mhz
- New SAP Integration Node: 332 Mhz
- New SAP Test Node: 332 Mhz
- Database: Oracle V7 (V8 planned)
- Disk by GB: <250GB SSA Disks
- Number of current active users: 300 connected users (7,500 total customers)
- IT Infrastructure software and technology: HACMP
- e-business, collaboration, and messaging software: Lotus Domino, MQSeries, Tivoli, and HR Access were added to SAPR/3 for human resources management

Benefits/Value to Customer

The most important benefits the customer has received from the SP consolidation include high availability, improved space management, and unlimited scalability. By moving away from the distributed architecture, the customer estimates that IT costs have been cut in half. Additionally, although costs related to software remain unchanged, it is now easier for the customer to immediately acquire hardware with the right software levels when needed, and SP resources are easily dispatched according to project requirements. The SP system enables the customer to better organize its UNIX information system both physically and logically. The customer has also realized savings in hardware and software maintenance through IBM CEMS contract, which includes services.

Additional key benefits for the customer include:

- Centralized administration
- Ease of applying norms and standards
- Possibility to distribute applications and software products from a reference node
- Flexibility to deploy different versions of the operating system
- Graphical monitoring of the complete SP
- Internal links between SP nodes

- Ability to optimize between nodes and frame (flexible switching backplane)
- Hardware reliability
- Professional hot line and support for the complete system

5.9 A financial services company

This scenario is that of a financial services company that offers a wide range of financial services to hundreds of thousands of retail customers from more than 350 offices throughout northern Italy. Striving for continuous growth through acquisition, they have managed to double their business volume over the last several years. Founded in the 1800s, the bank has won some of the highest international ratings in the United States.

To improve employee efficiency and reduce operating costs in its brokerage and finance trading department, the bank consolidated UNIX servers on an RS/6000 system and equipped the new server with IBM high-availability software.

Business benefits

Fewer data transfer delays, quicker time to market, savings from concentrated storage management and lower maintenance costs. Consolidation also allows for a flexible system to deal with growth strategy.

Business solution

Managing the assets of the bank or of clients is a dynamic activity and a strategic one as competition grows ever sharper. It is imperative that this company invest in the sophisticated solutions that would help them operate more effectively in this area.

Central to the state-of-the-art technology installed by the bank was a top-of-the-line RS/6000 SP server, which handles the work formerly run on 12 separate systems, each having its own characteristics and operating system versions. According to an engineer in the bank's data processing department, these discrepancies created huge management problems. For example, every time operators wanted to access a different service, they had to work from a different keyboard and monitor.

In addition, their former technology infrastructure could no longer efficiently process the volumes of work involved. The result: Reduced availability and higher cost of managing the systems. Trading delays and system failure increased the risk of serious economic consequences.

A team from the Financial Services Company researched possible solutions and visited other banks to study their operations. This team tried to foresee the evolution of the products involved and to evaluate the reliability of the providers. In the end, they decided to award the entire project, including its management, to IBM.

To create a more efficient trading environment, the IBM team began by stripping two rooms to the bare walls. New cabling, fire protection and a security system were installed. For greater flexibility, they relocated the customer's workstation CPUs centrally elsewhere, leaving only the keyboards and monitors on the trading floor. With this, users can access any available application or monitor from any one keyboard. The installation team linked the floor's 70 IBM PCs via a LAN, connecting them to the RS/6000 SP and to the bank's host mainframe as well. As prime contractor, IBM was responsible for the entire turnkey operation, even down to the choice of acoustics, lighting, and furniture.

A more efficient trading floor

Five months after the opening of their new financial management rooms, the innovations adopted have made their work much more productive and, most of all, eliminated the risk of unpleasant, expensive problems.

The new technological infrastructure - based on the SP system - allows this firm to face growing work volumes with high performance, availability, and reliability, up to their most ambitious expectations.

Today, their SP's multiprocessor nodes run their database and its attendant functions as well as all financial applications ranging from monitoring exchange rate positions to portfolio management. The SPs also control the attached PCs and handle file server, print server, and communication server tasks. The two nodes are connected via High Availability Cluster Multi-Processing (HACMP), which protects them from unscheduled outages. HACMP is a control application that can link multiple servers and nodes, thus, enabling parallel access to the same data and providing the redundancy and fault resilience required for business critical applications.

For safety's sake, in the event of a problem on the trading floor that might interfere with the flow of business, this company has a backup room in another building with 10 standby workstations that can be used under any special circumstances. In addition, the bank has installed ADSTAR Storage Management (ADSM) for AIX, a software product that integrates unattended network backup and archiving from as many as 25 multi-vendor platforms, with disaster recover planning and hierarchical storage management.

Benefits of server consolidation

According to the Organizational Department Director, the success of the project has meant optimization of hardware resources and lower licensing fees along with better manageability and performance on top of dealer productivity. Server consolidation has also eased the data backup process, doing away with LAN bottlenecks and allowing storage resource management to be concentrated in a single location. Consolidation will also mean great future savings, mostly in the systems management and maintenance areas. The new technological infrastructure - based on the SP system - allows this company to face growing work volumes with high performance, availability, and reliability.

5.10 A casino's success: Betting on RS/6000 technology

This scenario is on a casino that is the third largest in the world and boasts 150,000 feet of gaming space. The 20,000 guests who visit each day can also enjoy any of 20 restaurants for formal and informal dining as well as entertainment and retail shopping. Even before it opened, this casino won acclaim for uniquely incorporating legendary Native American themes in its architecture.

Business Needs

The owners of this casino wanted to establish an environment that would allow users a quick and efficient way to access all casino data - both financial and gaming related. The driving concept was to achieve a high level of integration between the different applications and take casino network design to a new level.

That requirement led to the selection of the Progress database, which then suggested a UNIX environment. The casino performed a detailed evaluation of products from two vendors, and IBMs RS/6000 came out the winner for several reasons - mainly, the scalability of the product line and a clear migration path for both hardware and operating system.

Business Solution

Twin RS/6000 G40s run the financials human resources and payroll application, another pair run this company's training, testing, and development environment while twin R40s run the patron management and gaming applications.

All three pairs operate under the High Availability Cluster Multi-Processing for AIX (HACMP) control program. HACMP enables up to eight RS/6000 servers

to access the same data in parallel, thus, providing facilities for application recovery/restart and fault resilience for these mission-critical applications.

Focus on scalability

The RS/6000 Model G40 symmetric multiprocessor (SMP) modular design makes it easy to plug in new cards and disk drives to accommodate up to four-way processing. The Model R40 SMP server features up to eight-way processing. The latest release of AIX - IBM's version of the UNIX operating system - is known for its special strengths in systems management and user environment.

Using a high-end RS/6000 SP, the casino runs AIX connections for print and user services on one node, Lotus Notes on another, with room for expansion. The customer plans on installing additional SP nodes for its data warehouse implementation. In addition, two IBM PC Server 520 file servers monitor, maintain, and administer the OLTP-intensive slot machine application, which tracks the drop of every coin in 3,000 machines.

Around 80 people on-site use Lotus Notes for scheduling, task assignment, document task routing, e-mail distribution, calendar and appointment scheduling, and not too far in the future, the casino plans to be a totally paper-less environment, this means all administrative procedures, documents, help center, and more.

Monitoring and administration of three separate networks is handled by NetView running on RS/6000 43Ps, while two more work together with the G40s to complete the development and test platforms. Meanwhile, software from a number of IBM Business Partners are in play at the Casino.

A Game Technology company supplies the Casino and more than 100 other establishments with systems solutions for both patron management and casino accounting including reservations, table games/slot machine activity analysis, player tracking, promotions, and credit verification and authorization. A Gaming Hospitality company provides its several dozen casino clients a fully integrated suite of back-office financial applications. In addition, the company offers a wide range of services including systems design, on-site support, database administration, and custom integration work.

The Gaming Hospitality company's scope of work for this customer included input throughout the design process from server hardware selection to cable plant design. A very practical example of this is the modification of the human resource application, integrating information across a variety of software systems so that thousands of employees can now use a single card for

building security access, time clock functions, and employee cafeteria meal tracking.

In addition, an IBM Business Partner provided network design and implementation services with continued networking system support. A main company focus is architecting and deploying high-speed ATM networks and integrating application suites for RS/6000 systems.

Counting up the benefits

IBMs state-of-the-art computer systems is helping provide outstanding guest service. With IBMs help, the casino is increasing their information flow to guests and providing them with a customer-friendly environment to play in.

The casino's computing environment helps the operation monitor and analyze gambling trends, down to pinpointing how much revenue is generated by each table and each slot machine. The level of patron rewards and hospitality is maximized by tracking customer use of a personal magnetic card in each gaming, shopping, and dining situation. Casino data is stored for later analysis on IBM 7133 Serial Storage Architecture Disk Subsystems.

With this, scalability is more than a buzz word. As more and more people flock to this casino, the ability to grow their computer systems to handle success confirms the strategic RS/6000-AIX decision made.

Part 2. Server consolidation solutions for RS/6000

Chapter 6. System management

The success of a server consolidation solution depends on the strength and quality of its system management. There are many UNIX tools offered by IBM and other vendors to assist with day-to-day management of both stand-alone and consolidated solutions.

Each of these tools can reduce the amount of system administration required in maintaining a system. They can also enhance the stability, security, and operability of servers, particularly when the environment is integrated or complex.

Areas that require strong monitoring and control are: Backup and recovery, problem and event management, and performance management.

Areas that typically require large amounts of system administration resource are: User management, print service management and disk space management.

We will review how various solution providers address these problems with their products. There are two types of consolidated environments that we will be looking at solutions for:

1. Large environments that have been consolidated but still have multiple nodes or are still heterogeneous.
2. Smaller environments that have been consolidated to a few systems, typically an RS/6000 SP or SMP platform.

6.1 Software installation management

Traditionally, the operating system and software for a UNIX server are installed through a tape device or CD ROM drive attached directly to the server being installed.

Server consolidation often means consolidating more than just processors. Generally, input/output devices will also be consolidated. As a result, it is not necessary to have a tape device or CD ROM drive for every server in the environment. In many cases, just one tape library and one CD ROM drive can be shared (directly or indirectly) between all servers.

6.1.1 Software installation management solutions

In the absence of local tape and CD ROM devices, an alternative method of operating system and software installation must be used. In a consolidated

environment, such as the RS/6000 SP systems, network based tools are used to install/update the operating system and software.

6.1.1.1 Network Install Manager (NIM)

NIM is client/server software that enables a client to obtain over the network a boot program and operating system from a server. The server, referred to as a boot/install server or NIM master, is configured to respond to the boot request from a system defined to the consolidated environment

The Parallel System Support Program (PSSP) is packaged with NIM, which is configured during the installation process. NIM is the standard tool used to install/update the nodes in an SP. By default, in a single frame SP, the control work station will be the NIM master. In a multiple frame SP, the first node in each frame becomes a NIM master for that frame by default.

6.1.1.2 Network File System (NFS)

NFS is currently the default network file system sharing utility over TCP/IP in the UNIX environment. NFS is provided with the operating system and can be configured through the System Management Interface Tool (SMIT) by the system administrator.

The purpose of NFS is to provide the facility to share a directory structure to other systems over the network. As such, the files that reside on an NFS server can be viewed as if they were local files on a client system in the network.

6.1.2 Operating system installation and updates

To install or update the operating system of a node in the consolidated environment, a number of steps must be performed.

NIM operating system installs:

1. Copy the operating system or update image to a designated area on the NIM master. These will be made available to the client being installed. The primary NIM master in such an environment will typically have a tape drive and CD ROM drive.
2. Notify the NIM master of the node(s) to be installed/updated.
3. Boot the node(s) to be installed. This step will initiate the installation process.

Using NIM to manage the operating system installs and updates provides a number of benefits to the consolidated server environment.

- Cost savings on input/output devices, such as tape and CD ROM drives

- Multiple systems in the environment may be installed simultaneously from the same image.
- Version consistency is easily maintained. Systems are installed from the same NIM master images.

NFS operating system updates

NFS can be used for minor operating system updates within the current release level of the operating system.

There are a number of steps required to perform an operating system install using NFS.

1. Write the software installation image to a designated area on the NFS server. Any system supporting NFS can be an NFS server, but generally, a single system is used for consistency, for example, the control work station in an SP system.
2. Define on the NFS server which systems may access the area containing the installation image.
3. Mount the software installation area on the client as a network file system. The installation images will appear as local files and may be installed as per the software installation instructions.

6.1.3 Other software installation products

The installation of third party software does not utilize NIM. The software to be installed must be made available across the network so that servers without tape or CD-ROM drives can access the software installation images.

In this section, we will be considering other software installation products from the following vendor:

- Tivoli

6.1.3.1 Tivoli Software Distribution

As its name implies, Tivoli Software Distribution provides facilities for the distribution and installation of software to managed systems in a Tivoli environment.

Tivoli Software Distribution uses the facilities provided by the Tivoli Management Framework to distribute file packages in an efficient manner. Administrators use the profile paradigm used by most other Tivoli applications to define file packages to be distributed. These file packages can include any files, such as executable programs, data files, and so on, and scripts that will

be executed before and after the distribution for a proper installation of the files on the target system.

By defining an appropriate repeater hierarchy for your network environment, large file packages will only be moved once across links but will still reach multiple target systems.

Table 2. Features and benefits of Tivoli Software Distribution

| Management service | What it does | What it means for you |
|---|---|---|
| Push and pull interfaces | Provides increased flexibility in defining and initiating updates | Enables you to initiate secure and timely distributions to intermittently connected end users, such as mobile computer users, or users missed during a scheduled distribution |
| Tight integration with Tivoli Inventory | Uses Tivoli Inventory query results to find the proper target of the distribution | Automatically determines target servers and desktops for software distributions |
| WAN-smart capabilities | Increases efficiency of network bandwidth usage | Enables you to use network properly without creating bottleneck |
| Tivoli's Application Management specification support | Enables compliant applications to automatically connect to Tivoli Software Distribution | Automatically generates distribution packages and dependency scripts based on information from these applications |
| AutoPack scriptless installation for PCs | Provides flexibility to install shrink-wrapped and custom applications | Eliminates the need to write scripts to package and distribute PC-based software |
| Transaction-based service | Enables client/server software to be deployed as a single unit | Simultaneously activates heterogeneous components for smooth cut-overs to new versions of applications |

6.2 User management

Although existing solutions from vendors of hardware, operating systems, relational databases, and applications address specific management objectives, they have not reduced the overall complexity of managing a distributed environment, particularly in the area of user account management.

As a result, customers must hire more and more system administrators as they add vendors for servers, databases, and applications. Unfortunately, because these are often cumbersome and require a significant manual entry of repetitive information, mistakes are common, and processes become extremely time-consuming.

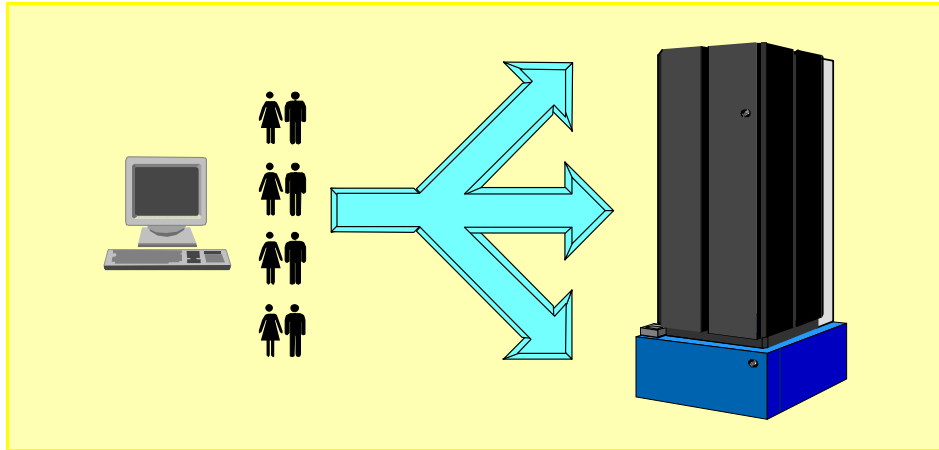


Figure 24. User management on a multi-node RS/6000 SP

One of the challenges for the consolidated server administrator is managing the users in the system. Should a user be allowed to access to one node, some nodes, or all nodes? The SP consolidated server can be viewed as one logical unit; therefore, users need to be defined across all nodes with the same login name, the same login password, the same group characteristics, the same home directory, and so on. This is only achieved by sharing the user database across the consolidated server.

A system administrator, therefore, has the responsibility of maintaining consistent copies of files, such as `/etc/passwd`, `/etc/group`, and `/etc/security/passwd` across all nodes in a server.

6.2.1 User management solutions

There are four methods available to maintain a consistent user database:

- Manage each user individually over each node in a consolidated server. This is a time consuming effort because the system administrator has to be aware of any changes, such as a password change, to the user database on every node. Once a change is implemented on one node, the system administrator has to then update every node in the system.

- For an IBM RS/6000 SP, you can use the SP File Collections facility provided with the Parallel System Support Programs (PSSP) software. File Collections defines and maintains sets of common files that are retrieved by the nodes at regular intervals.
- Use the Network Information System (NIS) available with AIX. NIS is a three-tiered system of clients, servers, and domains in which any changes to a set of defined common files, such as /etc/passwd and /etc/group, are automatically propagated throughout the system.
- Use an optional software solution from an independent software vendor.

6.2.1.1 Managing user databases on an SP

We can have two types of users residing within an SP consolidated server:

1. Those users that are created through AIX on an individual node and reside only on that node.
2. Those users created through SP user management and can have access to every node.

It is possible to have both types of users on a given node. This makes it difficult, if not impossible, to use file collections and NIS to manage the user database because the user database on each node is different from the other nodes. File collections and NIS are designed to manage one consistent copy of the user database across the system.

It is recommended that system administrators choose SP File Collections or NIS to maintain a user database when they have users that need to access two or more of the nodes or systems in their consolidated environment.

But, it is also possible that the system administrator may have some or all users who only need to access one system. In this case, using file collections and NIS is not possible, and other option must be used.

Both options, SP File Collections or NIS, provide the system administrator the ability to manage the user database from a single point of control, for example, the control workstation in an SP environment. In an SP system with a large number of users and/or a large number of nodes, SP File Collections or NIS may be the only feasible choices. Setting up individual users manually requires a large manual effort for users who need access to multiple nodes, while the other options make user management much more automated.

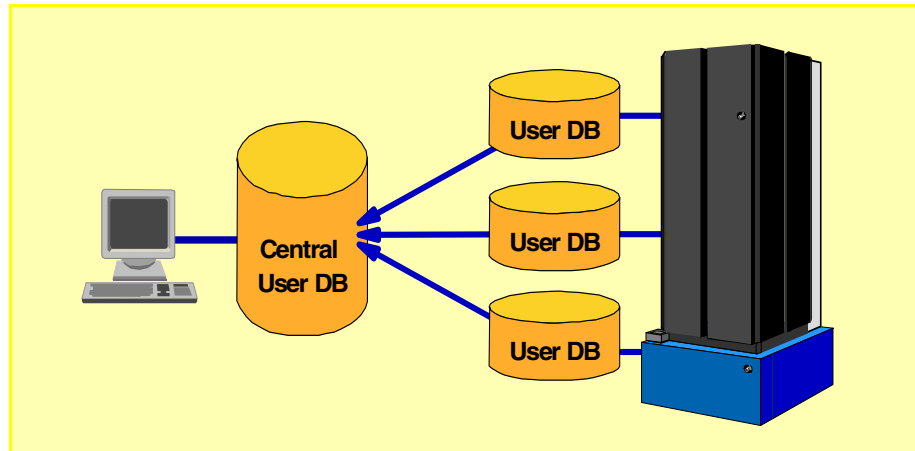


Figure 25. Managing a central user database on an RS/6000 SP

The purpose of user management is to maintain a user database across all nodes in an SP consolidated server. This enables the users to use the SP as one logical machine despite its makeup of multiple, individual RS/6000s. User management tasks include:

- Adding user accounts
- Deleting user accounts
- Changing user account information including the login password
- Listing user account information
- Controlling user logins

There are distinct sets of commands for managing AIX and SP users. To manage AIX users, use the AIX commands `mkuser`, `rmuser`, `chuser`, and `lsuser`.

To manage SP users, the PSSP software provides the SP user management commands. There are four SP user management commands: `spmuser`, `sprmuser`, `spchuser`, and `splsuser`. The two sets of command perform similar functions; the difference is the type of users they manage.

The PSSP software provides SP user management commands to enable an SP administrator to manage SP users. SP user management commands add and delete SP users, change account information, and set defaults for SP users' home directories.

6.2.1.2 SP access control

You may want to run either file collections or NIS to maintain a user database but have a requirement to restrict SP user access on nodes. Or, if your SP system is used for a parallel application, there may be a need to restrict user login on the processing nodes to improve performance. In either case, you can use SP access control. At the heart of SP access control is the command `spacs_cntrl`, which is run by the SP system administrator or by a job submission program on nodes where login control is required.

SP access control makes use of login and rlogin attributes in the `/etc/security/user` file to control a users' ability to log into particular nodes. It sets the attributes to either true or false depending on whether you want users to login (true) or not (false). The login attribute is used to control user local logins including using serial lines, while rlogin controls remote (network) logins.

6.2.1.3 SP file collections

The SP file collections facility is included with PSSP to group files and directories on multiple nodes into sets known as file collections. This simplifies their maintenance and control across an SP system. File collections is installed by default on an SP system.

A file collection can be one of two types: Primary or secondary. A primary file collection can contain a group of files or a secondary file collection. When a primary file collection is installed, the files in this collection are written to a node for usage. What if you want to use a node to serve a file collection to other nodes? This is made possible by using a secondary file collection.

When a secondary file collection is installed on a node, its files do not get executed on the node, rather they are stored ready to be served out to other nodes.

File collections uses a hierarchical structure to distribute files to nodes. The control workstation is defined, by default, to be the Master server, which is the machine where the master copy of a file is kept. It is only at this location that a file may be changed. Files are then distributed from the master server to all the defined boot/install servers to be distributed out to the nodes. If no boot/install servers are defined, file collections are served by the control workstation since the control workstation itself is a boot/install server.

You may change this hierarchy and use a boot/install server as a master server for one or some of the file collections. This way, you can maintain different copies of files on different groups of nodes. To implement this, you run the `supper offline` command on a boot/install server against a file

collection. This prevents that file collection from being updated by the control workstation. Changes specific to the group of nodes served by the boot/install server can now be made on the boot/install server.

6.2.1.4 Network Information System (NIS)

The main purpose of NIS is to centralize administration of files, such as `/etc/passwd`, within a network environment.

NIS separates a network into three components: Domain servers, and clients.

A NIS domain defines the boundary within which file administration is carried out. In a large network, it is possible to define several NIS domains to break the machines up into smaller groups. This way, files meant to be shared among, for example, five machines, stay within a domain that includes the five machines and not all the machines on the network.

A NIS server is a machine that provides the system files to be read by other machines on the network. There are two types of servers: Master and slave.

There are four basic daemons that NIS uses: `ypserv`, `yplibd`, `yppasswdd` and `ypupdated`. NIS was initially called yellow pages; hence, the prefix `yp` is used for the daemons.

The `yppasswdd` daemon makes it possible for users to change their login passwords anywhere on the network. When NIS is configured, the `/bin/passwd` command is linked to the `/usr/bin/yppasswd` command on the nodes. The `yppasswd` command sends any password changes over the network to the `yppasswdd` daemon on the master server. The master server changes the appropriate files and propagates this change to the slave servers using the `ypupdated` daemon.

6.2.1.5 The automounter

The automounter is provided for an SP system administrator to manage the mounting of file systems across an SP system. The convenience of using the automounter is best illustrated with an example.

Consider an SP system with four nodes. You, as the system administrator, want to let all users have the capability to access all four nodes. You have chosen to use file collections to handle the management of the user database. What about the users' home directories? As with the user database, it is preferable to allow users to maintain one home directory across all four nodes.

The automounter can take care of all of this for you. When the mounting of a file system is under automounter control, the automounter transparently mounts the required file system. When there is no activity to that file system for a prescribed period of time, the file system is unmounted.

The automounter uses map files to determine which file systems it needs to control. There is typically one map file per file system that you want to manage. It is also possible to use NIS maps instead of automounter maps to specify which file systems the automounter is to handle.

6.2.2 Other user management solutions

Besides SP solutions for user management across a multi-node RS/6000 SP server, we will be considering other user management products from the following vendor:

- Tivoli

6.2.2.1 Tivoli User Administration

Tivoli User Administration, which is one of Tivoli's core applications for security management discipline, provides an automated, secure way to manage user attributes and user services across a consolidated environment. It is more suitable for large, complex environments that have a complicated mix of users and user setup requirements. It is designed to give system administrators centralized control.

Features and benefits

Tivoli User Administration provides a network computing environment with the following features:

- Centralized and GUI-based control of user administration tasks
- Consistent user administrative policy definition
- Automated repetitive user administration tasks
- Parallel operations performed on many users and systems
- Delegation of administrative tasks to other administrators
- Configuration error reduction via profile-based methodology
- Single-action user management to synchronize logins and passwords

These capabilities provide a high level of control and security and greatly improve the productivity of system administrators. With Tivoli User Administration, you can manage user and group accounts on different operating system platforms from one single location.

Example

If a new sales person joins the company and requires access to three nodes on an RS/6000 SP server, the administrator can create a single Salesperson template that automatically pushes the account information to the appropriate servers in a single action rather than manually adding user accounts on all the different resources one at a time. Tivoli's default policy automatically generates values for user attributes across account types.

Similarly, if an employee leaves the company, the administrator can work from a single template, therefore, deleting all of the user's accounts in a single action.

Hints and tips

- The Tivoli User Administration OnePassword utility enables users and administrators to change user passwords in the Tivoli database and distribute the updated passwords to the appropriate target systems. Users can access a common Password Manager Web site, thus, enabling them to change their own passwords, and administrators can access a restricted Web site to assign new passwords for users. It utilizes the Secure Sockets Layer (SSL) for security.
- The Tivoli User Administration LDAP Connection provides a method for managing user accounts on any system accessible via the Lightweight Directory Access Protocol (LDAP). LDAP is a directory service protocol that runs over TCP/IP. This product adds new user profile attributes and a default policy script to all user profiles.
- Some Tivoli Managers for databases, such as Tivoli Manager for Oracle, provide a user management component to manage the specific database users.

6.2.2.2 Tivoli Global Sign-On

Tivoli Global Sign-On provides a secure, single point of entry to computing resources, which enables organizations to connect disparate networked systems.

Features and benefits

- Increased security - Supports password, fingerprint, or smartcard authentication to confirm authorized users.
- Easier administration - Simplifies setting up and managing passwords and IDs for users.
- Increased productivity - Reduces the time required to complete logons and simplifies system and user management.

Tivoli Global Sign-On is easy to use and manage because it eliminates the need for multiple passwords and IDs. It strengthens existing enterprise security by preventing security exposures generated by multiple passwords, which can lead to users writing down passwords to remember them or using trivialized passwords (simple words that are easy to decode). It also permits only authorized users to store, transmit, and distribute passwords.

Tivoli Global Sign-On integrates with existing operating systems and is consistent with, and across, the leading operating system platforms for truly global, single sign-on capabilities. It is easily extended to allow maximum flexibility to add new applications and minimizes the programming effort needed to support them.

Hints and tips

- To easily manage secure environments, Tivoli Global Sign-On is fully integrated with Tivoli User Administration and the Tivoli Management Framework. In this environment, a designated Tivoli administrator manages Tivoli Global Sign-On users and targets from a centralized console. Role-based administration is also available with this support. A Tivoli Global Sign-On Plus Module enables monitoring the Tivoli Global Sign-On resources.

6.2.2.3 Tivoli Security Management

Tivoli Security Management saves a tremendous amount of time and effort for security administrators when implementing the security policy across the enterprise. It also helps align systems more rapidly within the organization. With Tivoli Security Management, you can manage system security on managed nodes as well as Tivoli Management Agent endpoints. This product provides centralized role-based security administration across distributed computing platforms, such as UNIX, Windows NT, OS/390, AS/400, NetWare, or OS/2.

Features

- Access limitations to resources based on job functions
- System-wide security policies for passwords, audits, and logins
- Common interface to different security systems, thus, reducing the amount of training
- Audit capabilities
- Integration with existing Tivoli applications and Framework services

6.3 Performance management

Performance management is an important component of system administration in any environment and even more critical in the consolidated environment. Proper performance management will enable an administrator to maintain the optimal performance of isolated and distributed systems. This section examines three basic sets of performance management tools.

- Standard UNIX performance tools

These tools are native to the UNIX operating system; however, they are limited in that they will only monitor the local system. These tools will not centrally monitor multiple servers in a consolidated environment.

- Performance Toolbox

The Performance toolbox provides a centralized tool for managing the performance multiple server from a single graphical user interface.

- Performance Toolbox Parallel Extensions

This utility provides extensions to the performance toolbox package along with enhancements for an SP consolidated environment.

6.3.1 Performance measurements

System performance may be divided into four basic categories:

- CPU

Relative workload on the processor(s)

- Memory

Utilization of the system wide memory resource

- Input/Output

Data throughput and bandwidth between the processor and external storage devices, such as disk and tape subsystems

- Network

Data throughput and bandwidth of a system through available network connections

There are a number of ways in which these areas of performance may be influenced:

- Modification of the hardware environment

For example, adding physical memory, adding CPUs, changing disk or network technology, and so on.

- Modification of application profile

Many applications may be tuned to control the way in which the application utilizes the system resources. This may include running jobs in batch or modifying the way in which an application allocates resources.

- Modification of user behavior

User attributes and access may be controlled in order to restrict the unnecessary abuse of system resources.

- Modification of system tuning parameters

It is possible to tune the manner in which operating system resources are allocated to improve system performance. This includes device attributes, storage allocation, memory allocation, CPU allocation, network parameters, and so on.

Modifying the performance attributes of a system will not necessarily improve performance unless the appropriate attributes are modified.

For example, if a system is consistently I/O bound (potentially a disk throughput issue), adding additional CPUs will probably not improve performance.

6.3.2 Performance management solutions

To enable accurate monitoring and management of performance, a suite of tools have been provided to monitor and record system performance. These tools are discussed in the remainder of this section.

6.3.2.1 Standard UNIX performance tools

Standard UNIX performance tools provide the facilities to monitor system resources, such as CPU utilization, memory, I/O throughput, and disks. These tools will generally monitor only the local server and not remote systems. The following list represents some the common UNIX performance tools.

- vmstat (Virtual Memory Statistics)
- iostat (Input / Output Statistics)
- sar (System Activity Reporting)
- ps (Process Status)
- netstat (Network Statistics)
- nfsstat (Network File System Statistics)
- no (Network Options)
- nfso (Network File System Options)

- prof, gprof (program profiling tools)
- rmss (Reduced-Memory System Simulator)
- filemon (File system Monitor)
- fileplace (File Placement)
- netpmon (Network Performance Monitor)

6.3.2.2 Performance Toolbox (PTX)

PTX is a comprehensive tool for monitoring and tuning system performance. PTX uses the client/server model to monitor local and remote system performance with several graphical windows that are fully user configured. It is a Motif-based toolbox that includes 2D and 3D views of performance statistics. Analysis and tuning facilities incorporate existing performance tools into a menu-driven environment.

This product is especially useful in a consolidated environment consisting of multiple servers. When properly customized, this allows the administrator to monitor desired performance statistics from all servers in the environment from a single point of control.

Features

The client/server implementation of PTX is an excellent aid to monitor and tune the performance of various UNIX systems in a network environment from a single graphics workstation. There are a number of important PTX features.

- With a fully configurable graphical interface, PTX gives the user the ability to concurrently visualize live performance characteristics and pinpoint either local, distributed, or network bottlenecks.
- PTX allows the user to define conditions and appropriate responses, which include alerting a specific administrator to initiating a corrective action without any human response.
- Separately installable manager and agent components that allow a manager to monitor multiple agents and an agent to supply data to multiple managers.
- HP-UX (V9.03), SunOS (V4.1.3), and Solaris (V2.3, V2.4, and V2.5) agents to allow monitoring of performance data on OEM machines.
- Application Programming Interfaces (API) that allow programmers to access local or remote data as well as register custom data with the local agent.
- Ability to respond to SNMP requests and to send traps to an SNMP manager (AIX agents only).

- Support for RS/6000 SP systems using the Performance Toolbox Parallel Extensions Feature of the PSSP.

Benefits

There are numerous benefits derived from using PTX.

- **Monitoring System Performance**

PTX/6000 provides easy monitoring of both local and remote systems including the performance of the network. In network client/server environments, the performance of system groups working collectively can be as important as the performance of an individual system. Likewise, the performance of multiple applications working together can be as important as that of an individual application. Therefore, it is very important to be able to get the big picture by graphically viewing many correlated parameters concurrently across multiple nodes in a network. PTX/6000 allows a user to concurrently visualize the (near real time) performance characteristics of the clients and server applications across the network.

- **Analysis and Control of System Performance**

By providing an umbrella for tools that can be used to analyze performance data and control system resources, the manager program, xmperf, assists the system administrator in keeping track of available tools and appropriately applying them. The menus of xmperf are preconfigured to include most of the performance tools shipped as part of the tools option of the agent component.

All performance-related tools already available in AIX can be accessed through this interface. In addition, the ability to record load scenarios and play them back in graphical windows at any desired speed provides ways of analyzing a performance problem. Features for analyzing a recording of performance data are provided by the azizo program and its support programs. Finally, using the agent component filter, filtld, you can define conditions that, when met, could trigger any action you deem appropriate, therefore, initiating corrective action without human intervention. This facility is entirely configurable so that alarms and actions can be customized to your installation.

- **Capacity Planning**

If you can make your system simulate a future load scenario, xmperf can be used to visualize the resulting performance of your system. By simulating the load scenario on systems with more resources, such as more memory or more disks, the result of increasing the resources can be demonstrated.

- **Network Operation**

The xmservd data supplier daemon can provide consumers of performance statistics with a stream of data. Frequency and contents of each packet of performance data are determined by the consumer program. Any consumer program can access performance data from the local host and one or more remote hosts. Any data supplier daemon can supply data to multiple hosts.

- **SNMP Interface**

By entering a single keyword in a configuration file, the data supplier daemon can be told to export all its statistics to a local snmpd SNMP agent. Users of an SNMP manager, such as IBM NetView, see the exported statistical data as an extension of the set of data already available from snmpd.

6.3.2.3 Performance Toolbox Parallel Extensions

Performance Toolbox Parallel Extensions for AIX (PTPE) was developed to integrate performance monitoring on RS/6000 SP systems with PTX. When installed on an SP, PTPE provides specific SP performance statistics for software and hardware utilized by an SP installation.

- LoadLeveler
- IBM Virtual Shared Disk (VSD)
- High Speed Switch

The PTPE software provides additional functionality to a consolidated environment utilizing SP technology in which PTX will be used for performance management. PTPE may also be used to interface directly into user written applications without using PTX. This is made possible through an API designed for the purpose.

PTPE uses a distributed approach to allocate workload and responsibilities across a number of nodes. Through this architecture, large SP complexes can be efficiently monitored without excessive burden on any one node. This distributed hierarchy has three major elements:

- A data sampler on every node configured as a reporter

In a typical hierarchy, every node has a data sampler daemon (reporting role). This daemon extracts performance-related information from the hardware and software components on that particular node based on collection criteria.

- At least one data manager on selected node(s) configured as a manager(s)

At least one node in the SP has a data management role. This node manages one or more reporting nodes. This additional task includes relaying instructions to its assigned reporter nodes as well as collecting and manipulating collected performance information. The manager node can also provide additional data reduction, such as preparing summary performance figures for the nodes they manage.

- Exactly one data coordinator on the node configured as a coordinator

The central coordinator node administers all the manager nodes and provides a single point of control for the PTPE hierarchy. This reduces the number of individual hosts that must be contacted to gather information and allows for a central process to manage all API requests. The central coordinator calculates SP-wide statistical averages of the performance summaries prepared by all the data manager nodes to provide an overall view of system activity.

You can easily start, stop, and control the operation of PTPE through the use of simple AIX commands or by using the performance monitoring facility included in the PSSP SP Perspectives Graphical User Interface. Much of the PTPE setup and configuration can be performed automatically.

PTPE can assist in maximizing your SP's computing power. PTPE can be used to monitor node performance, assist in SP workload balancing, and tune your configuration for maximum throughput. PTPE can provide easy access to all the performance data available on your SP system.

Features and benefits

There are numerous benefits to implementing PTPE in an SP environment:

- Extends the Performance Toolbox for AIX products to the SP environment.
- Lightens administrative loads.
- Provides run-time monitoring.
- Records performance data and saves it for later analysis, as with PTX.
- Familiar run-time displays show how RS/6000 SP nodes are performing.
- Monitor as many aspects of system performance as you want, on as many SP nodes as you want, with minimal performance impact.
- Group nodes for meaningful summary statistics and display the results now or archives them for later analysis.
- Distribute data management across multiple nodes, therefore, eliminating bottlenecks and preventing your performance analysis effort from affecting system performance.

- Collect and display statistical data for SP hardware (nodes and switches) and IBM software (LPPs), such as LoadLeveler and PSSP.

6.3.3 Other performance management solutions

In this section, we will be considering other performance management products from the following vendors:

- BMC Software
- Compuware

6.3.3.1 BEST/1 from BMC Software

The BEST/1 product is designed for IT professionals who need to understand and manage performance across Unix, Windows NT, OS/390, Parallel Sysplex, VM, and AS/400 computing environments. To meet their needs, the BEST/1 product provides the ability to:

- Comprehend normal performance and analyze deviations
- Report resource consumption with a business perspective
- Track long-term performance trends to understand demand
- Predict the impact of change on response times
- Identify hardware requirements prior to deployment
- Forecast the need for additional computing resources

Features and benefits

BEST/1 is an application management product designed to support IT service level objectives.

- BEST/1 has performance data collectors for Oracle, Sybase, Informix, Microsoft Exchange, and R/3 performance data, as well as the ability to create workloads to represent other applications. These performance data collectors provide the foundation for the analysis, reporting, and predictive capabilities of BEST/1.
- From the BEST/1 Performance Console, a user can browse the BEST/1 managed Windows NT and Unix nodes and retrieve key performance indicators every ten seconds for analysis and diagnostics.
- The BEST/1-Visualizer product, the common graphical reporting tool for all BEST/1 supported platforms, is a PC-based tool that allows users to graphically diagnose, display, and report the performance of applications, databases, and systems while using a common interface. The collected data is stored in a performance warehouse where it is viewed through a series of graphs that can be published to the Web.

- BEST/1 uses analytical queuing theory to predict the impact of an IT environment change. Whether the change is an increase in users or a server consolidation, BEST/1 models the change and provides scientific information used for planning. When using the predictive capabilities of BEST/1 for server consolidation, the workloads of multiple servers can be combined onto a single server to understand the impact of the consolidation before it takes place. Whether the impending bottleneck is I/O, memory, or CPU, the modeling capabilities of BEST/1 would proactively identify the situation and identify the appropriate steps to be taken. The BEST/1 hardware table, the source for hardware alternatives, has over 1000 entries that can be *test driven* through BEST/1 modeling.
- The planning capabilities of BEST/1 allow users to apply incremental growth to existing configurations and determine the life expectancy of resources. This feature results in just-in-time purchases and accurate hardware budgeting.
- BEST/1's automation reduces human intervention to a minimum, thus, maximizing the success of the performance management process. BEST/1 has automation for handling data collection, data aggregation, analysis, modeling, reporting, and Web publishing.

Hints and tips

- The event monitoring capabilities of Patrol combined with BEST/1's ability to analyze and predict performance provides a strong arsenal for business availability.

6.3.3.2 BEST/1 Performance Module from BMC Software

The BEST/1 Performance Modules are plug-ins to the BEST/1 for Distributed Systems products. These modules are currently available for Oracle, Sybase, Informix, SAP R/3, and Microsoft Exchange. These modules provide the specific information required to effectively manage the performance of databases and applications.

Workload Response Time Detail

Workload Financialsapp@TARGET in BMC_PLAN on 2/10/99

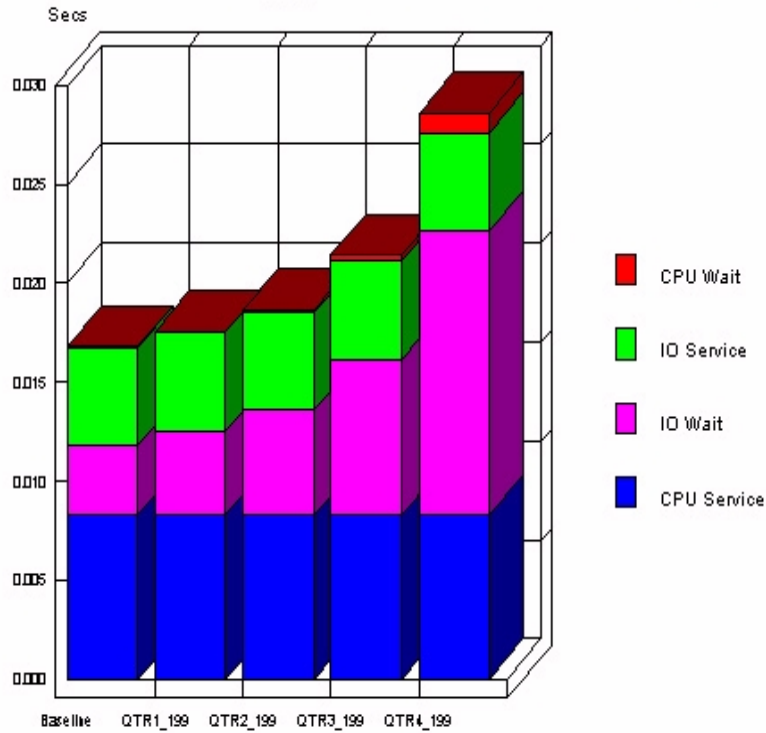


Figure 26. Workload response time of BEST/1

Features and benefits

The BEST/1 Performance Module for Oracle helps to manage and improve database performance across Oracle environments. It also:

- Provides performance analysis to let you diagnose database metrics and immediately uncover and resolve problems.
- Provides historical analysis and graphical reporting to give insight into what is normal for your database and identify deviations.
- Provides *what-if* performance modeling and planning to identify and prevent database problems and allows you to plan for database growth.

6.3.3.3 EcoTOOLS from Compuware

EcoTOOLS brings mainframe-like control and availability to the UNIX environment by delivering functionality crucial to managing client/server, internal intranet, and Web electronic commerce applications.

Features and benefits

EcoTOOLS can provide the following UNIX event management and performance monitoring:

- Detects excessive CPU usage and identifies top CPU resources
- Detect excessive paging and swapping and identifies top memory consumers
- Identifies I/O and assists in balancing across disks
- Detects running out of space
- Detects shortages of critical resources, such as process entries, semaphores, and inodes
- Detects runaway processes
- Monitors key process and automatic restart
- Detects hung printers or peripheral devices

EcoTOOLS can also provide the following network event management and performance monitoring:

- Monitors network traffic/packet activity
- Detects network congestion, packet errors, and collisions
- Detects and reconnects lost or broken connections
- Monitors NFS calls and client time-outs
- Monitors SNMP devices
- Shares events and data with SNMP-based products

Managing application service levels requires new approaches to planning, monitoring, and managing applications. Service level management also requires two new types of measurements:

- Executive management must be able to measure whether recent service level commitments made by IT operations to business users are met on an overall basis as well as historical trends experienced by those service level measures.

- IT operational level management requires detailed metrics to examine and understand un-met service level expectations as well as background information on prioritized problem areas.

EcoSYSTEMS service level management reports provide both kinds of measurement of application service levels. EcoTOOLS for UNIX, for example, provides reports by application indicating application and application component availability for the specified time period. Reports are provided at both the executive summary and detailed operations levels allowing a truly complete view of the availability of business-critical applications.

EcoTOOLS provides the following automated operations:

- Automatic discovery of system and database health and status
- Run corrective and preventative actions automatically when an event occurs
- Initiate actions based on multiple thresholds
- Specify periods when monitoring of computers and databases are unavailable

It also has the ability to assist in capacity planning with the following functions:

- Predict capacity needs by analyzing real-time and historical data presented in reports
- Build repository or actual resource utilization over time
- Log events and data directly to ASCII files for easy reporting to other products

EcoTOOLS security monitoring is designed to protect your environment from unauthorized access with features, such as:

- Notify and take actions resulting from unauthorized attempts to access resources
- Monitor for modifications to system files and directories
- Check export files and UUCP for global/world access
- Verify user passwords are installed and user directories do not have global access
- Enforce Access Control List for systems management functions

Report Subject: SAP R/3 availability report by component XYZ Corporation

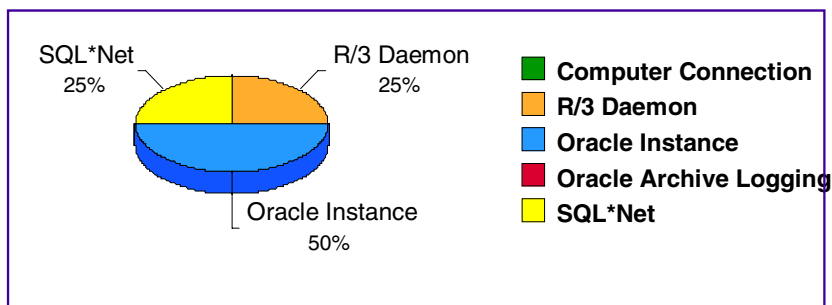
Duration: 10/1/1997 to 8/31/1999

Prepared for: David Kim, IT Management
 Prepared by: Joan Munger, IT Operations
 Date: 10/26/1998
 Location: San Jose, California

Number of Servers: 3
 Avg concurrent users: 100

SAP R/3 System Failure by Component

Failure Distribution by Component



Individual Statistics

| Component | Failure rate (per day) | # of failures | MTBF(hrs) |
|---------------------|------------------------|---------------|-----------|
| Computer Connection | 0 | 0 | N/A |
| R/3 Daemon | 0.01 | 1 | 21548 |
| Oracle Instance | 0.02 | 2 | 1077 |
| Oracle Archives | 0 | 0 | 0 |
| SQL*Net | 0.01 | 1 | 21549 |

If #of failures is zero, MTBF is undefined.

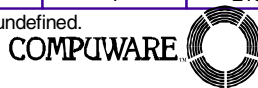


Figure 27. Sample EcoTOOLS service level management report

6.4 Disk space management

Disk space management in a consolidated environment requires additional consideration for stand-alone systems. It may be a requirement of a multiple node consolidated environment that disk space be shared and accessible between the nodes.

6.4.1 Disk space management solutions

In a consolidated system of servers, it can be a requirement that these servers have access to the same data through sharing disk space, while at the same time, optimizing performance. There are a number of tools that enable disk sharing between servers:

- IBM Shared Disk
- General Parallel File System (GPFS)
- Parallel I/O File System (PIOFS)

These tools, discussed below, make it possible to share disk space and file systems between servers while providing superior performance over network tools, such as NFS.

6.4.1.1 IBM Shared Disk

There are three components within the IBM PSSP to assist with the sharing of disk space, such as raw logical volumes, between applications executing on different servers in a SP system.

- Virtual Shared Disk (VSD)
- Hashed Shared Disk (HSD)
- Recoverable Virtual Shared Disk (RVSD)

Virtual Shared Disk (VSD)

VSD is an IBM subsystem that allows application programs executing on different nodes of an SP system access to a raw logical volume as if it were local at each of the nodes. Each virtual shared disk corresponds to a logical volume that is local at only one of the nodes (the server node). The VSD subsystem routes I/O requests from the other nodes (client nodes) to the server node and returns the results to the client nodes.

The I/O routing is done by the VSD device driver that interacts with the AIX Logical Volume Manager (LVM). The device driver is loaded as a kernel extension on each node. Thus, raw logical volumes can be made globally accessible.

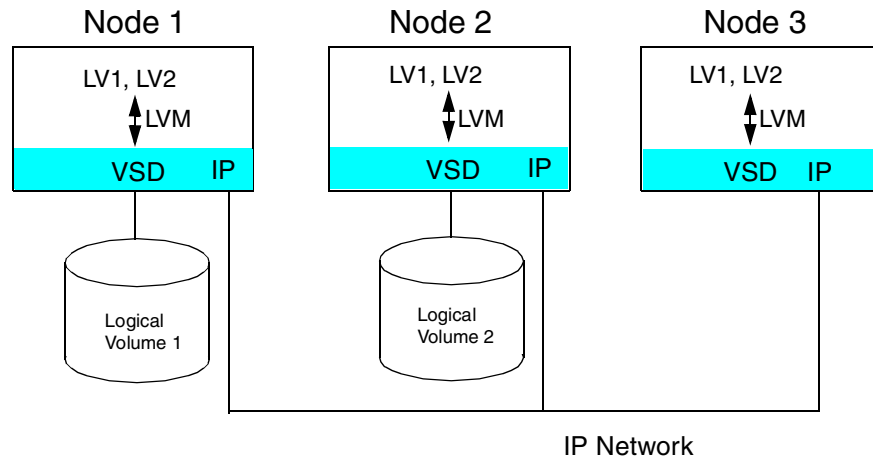


Figure 28. Example: Virtual Shared Disk configuration

In the above example (Figure 28), there are three nodes connect by an IP network. Node 1 has raw Logical Volume 1 (LV1) locally attached, and Node 2 has raw Logical Volume 2 (LV2) attached. The LVM and VSD device drivers are integrated in such a way that LV1 and LV2 appear to be local to users on all three nodes.

Hashed Shared Disk (HSD)

The HSD component of disk management in the SP environment has a data striping device driver that distributes data across multiple nodes and multiple virtual shared disks.

The benefit this provides is a reduction in I/O and potential performance improvements. Instead of writing all the data from an application program I/O request onto a single virtual shared disk at a specific location, the data striping device driver writes blocks of the data on each of the several separate virtual shared disks.

Recoverable Virtual Shared Disk (RVSD)

The VSD function allows all nodes in the system partition access a given disk, even though that specific disk is physically attached to only one node. If the server node to which this disk is attached should fail, access to the disk is lost until the server node is recovered.

IBM RVSD overcomes this problem. RVSD, in combination with twin-tailed disks or disk arrays, allows a secondary node to take over the server function from the primary node when certain types of failure occur. This means that, in

the case where a server running RVSD fails, the disks locally attached may be managed from another server until the failed node is available again.

A twin-tailed disk is a disk or group of disks that are attached to two nodes of an SP. For recoverability purposes, only one of these nodes serves the disks at any given time. The secondary, or backup, node provides access to the disks if the primary node fails, is powered off, or if you need to change the server node temporarily for administrative reasons.

The RVSD component automatically manages your virtual shared disks by detecting error conditions, such as node failures, adapter failures, and disk failures (I/O errors), and then switching access to the disk from the primary node to the secondary node so that your application can continue to operate normally. The RVSD component also allows you to cut off access to virtual shared disks from certain nodes and to dynamically change the server node.

When your applications exploit the RVSD component, you can recover more easily from node failures and have continuous access to the data on the twin-tailed disks.

For more detailed information on these functions, see the redbook *Managing Shared Disks*, SA22-7349.

6.4.1.2 General Parallel File System (GPFS)

GPFS provides global access to files in the RS/6000 SP. Files created in GPFS can be accessed from every node that runs GPFS code. For those nodes not running GPFS, files can still be accessed through NFS.

GPFS is implemented as a standard AIX Virtual File System (VFS), which means that applications using standard AIX VFS calls, such as standard Journaled File System or JFS calls, will run over GPFS without modification. It also provides a byte-range locking mechanism allowing parallel applications to access non-overlapping blocks of a file with minimal contention.

GPFS offers high scalability and high availability by allowing multiple servers and multiple disks to serve the same file system. If a server fails, the file system will still be available as long as another server has access to the disks containing the data and a network path to the client. If a disk fails, GPFS will continue providing access to the file system as long as the data contained in the failed disk is not file system metadata but user data, or as long as it has been replicated. User and file system data can be replicated through mirroring to provide an even more highly available environment.

The implementation of GPFS on the RS/6000 SP is based on two key components:

- PSSP High Availability Infrastructure
- Virtual Shared Disks.

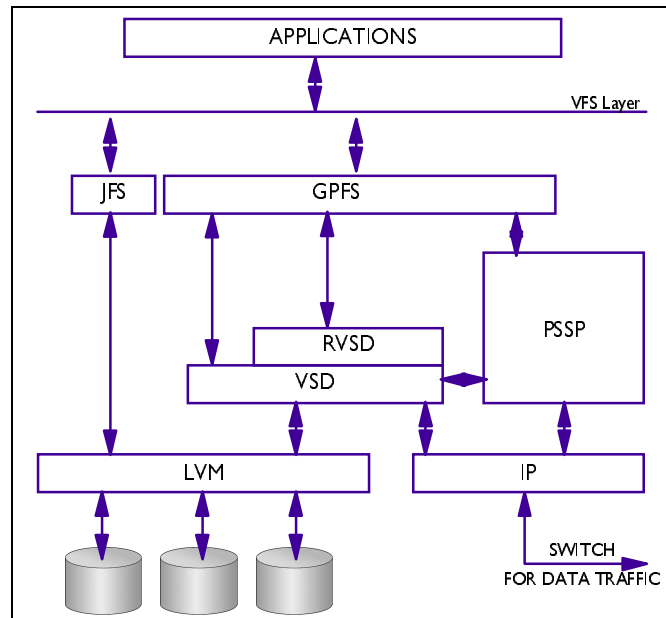


Figure 29. GPFS overview

Why use GPFS?

A parallel file system not only offers performance advantages by eliminating the limitation of a single server for file services but also a great deal of flexibility.

With a parallel file system, since all nodes have access to the file system, it is easier to move applications from one node to another. This is especially valid for high availability solutions where the application is moved if unrecoverable errors occur in the main server.

Sharing the same file system among several nodes has the benefit of increasing the maximum I/O bandwidth that otherwise would be limited by the maximum local I/O bandwidth of the single server.

Benefits

GPFS provides numerous benefits to the consolidated system.

- Improved performance for file access
Multiple node access of the same file system will potentially increase I/O bandwidth and file access performance.
- Increased data availability
GPFS is a logging file system that creates separate logs for each SP node. These logs aid fast recovery and consistency of data in the event of node failure, even when a node fails while modifying file data.
- Enhanced system flexibility
Disks can be added or deleted while the file system is mounted. When appropriate, file systems may be rebalanced across currently configured disks.
- Ease of administration
All GPFS administration tasks can be performed from any node in the SP configuration. The GPFS commands operate on multiple nodes. A single GPFS multi-node command can perform a file system function across the entire SP system. GPFS administration tasks can also be executed from SMIT menus.

6.4.1.3 Parallel I/O File System (PIOFS)

Loading data in large Decision Support or Data Warehouse environments may present problems when handling large file sizes. The PIOFS offers the capability to handle files in sizes much larger than the current limit of the AIX operating system.

PIOFS for the RS/6000 SP is designed for serial or parallel applications that require large temporary files and high I/O bandwidth.

PIOFS allows the creation of files as large as 128 terabytes, which may span multiple server nodes. A PIOFS file can be treated as a single, regular AIX file or logically partitioned into subfiles, each containing a portion of the file's data. Each subfile can then be processed in parallel by a separate task.

PIOFS allows parallelized access to your data without the inconvenience and administrative overhead of maintaining multiple data files. PIOFS files can be dynamically partitioned into subfiles many different ways, all without altering or moving the contents of the file.

PIOFS is designed as a client/server application. Server nodes provide file space for the PIOFS clients over the network. PIOFS kernel extensions are loaded into the AIX kernel within the Virtual File System (VFS) layer. PIOFS is a TCP/IP based application and can be configured and used over any TCP/IP connection. However, the best results will be achieved by using the PIOFS together with the Switch network of the RS/6000 SP. The High Performance Switch connects a PIOFS client to the PIOFS servers over point-to-point links providing the full bandwidth from each client to each server.

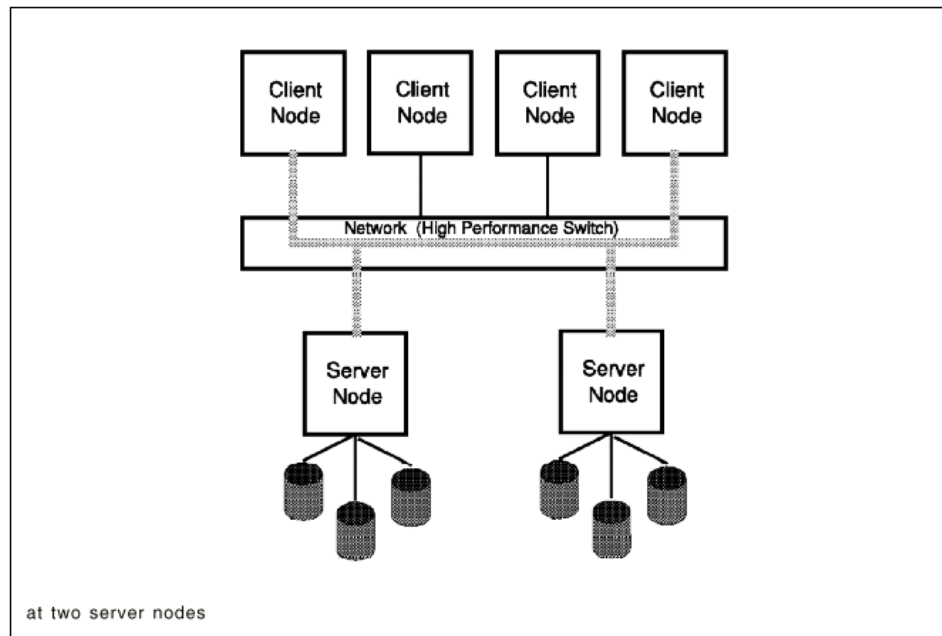


Figure 30. PIOFS supports the simultaneous access

For more detail on the PIOFS, refer to *IBM AIX Parallel I/O File System: Installation, Administration, and Use*, SH34-6065 and *IBM Systems Journal*, SJ34-2, G321-0120.

Why use PIOFS?

The PIOFS facility will prove to be the greatest benefit for sites implementing large data handling facilities with requirements for:

- Very large data file handling
- Parallel access to large data facilities for flexibility and performance

- Availability of data and minimization of down-time in large database environments

Benefits

- PIOFS supports parallelism in two ways, physically and logically.
 - A file may be divided physically over multiple disks and servers.
 - A file may be divided logically into multiple subfiles.
- A key benefit of PIOFS is its scalability. Significant parallelism for applications can be achieved by using PIOFS to spread files across multiple disks and servers without using file partitioning.
- Files significantly larger than the AIX file size limit may be created up to 128 terabytes. The PIOFS is architected to support a file size of up to 2E63 bytes. The current implemented limit, however, is a file size of (512 x 256 GB = 128 TB).
- Achieves moderate parallelism and, hence, better I/O performance through spreading files across multiple server nodes.
- PIOFS provides the ability to capture and save a snapshot of files during processing. If the application fails, the checkpointed version of the file can be used to restart the program where it was last checkpointed. File checkpointing may also be used to save a particular version of a PIOFS file.

Hints and tips

- Files 2 GB or larger must be defined in a profile to indicate that the files should be accessed using the PIOFS internal file offset.
- PIOFS files can be treated as standard AIX files with the additional benefit that they may grow to the greater size supported by the Parallel I/O File System.
- Files smaller than 2 GB can be copied between a Parallel I/O File System and another file system. You can also create new files.

6.5 Automation

Using automation is important for helping the operational environment of a consolidated server. Today, there are automation products in the marketplace that help customers in managing the production tasks.

The purpose of automation is to:

- Help the operator manage repetitive tasks
- Avoid wrong decisions during sensitive situations

- Improve system availability (24 hours a day and 7 days a week is the objective)

6.5.1 Automation solutions

UNIX automation and system administration tools are now offered by many vendors. The tools are diverse, but they mainly fall into these categories:

- Job Schedulers
- Distributed system management tools
- Network management tools, such as Tivoli NetView

Tivoli Enterprise software consist of an underlying infrastructure known as the Tivoli Management Framework and a growing set of Tivoli and third party management applications that can utilize this framework to manage heterogeneous systems and applications in a consistent manner. Because it use an object technology, the distributed object framework performs the same functions in the same way on different operating systems, such as UNIX, NT, OS/390, Netware, and so on. The Tivoli Framework serves as a single point of integration for the Tivoli and third-party applications.

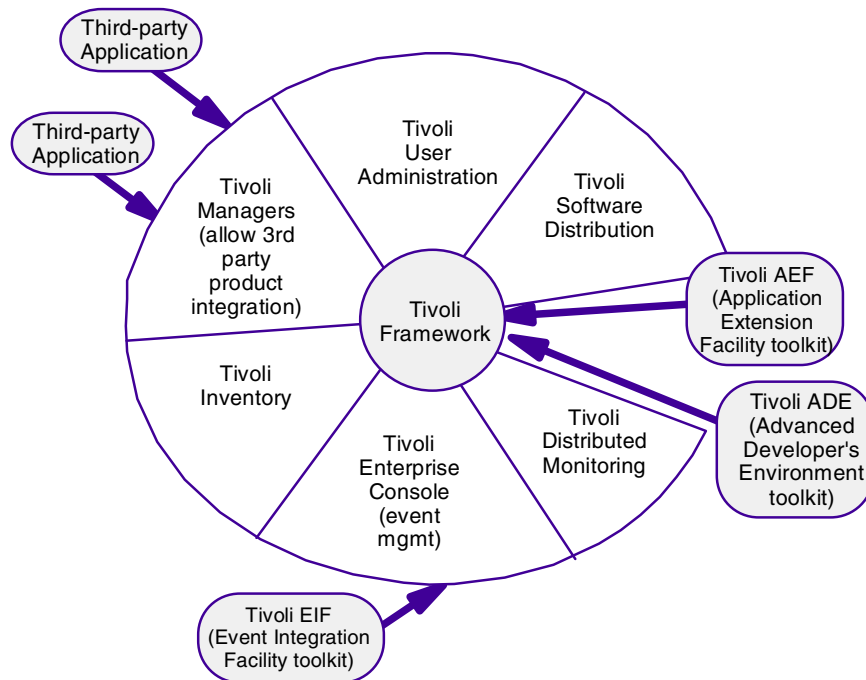


Figure 31. Tivoli Enterprise software components

6.5.1.1 Tivoli Global Enterprise Manager (GEM)

Tivoli GEM allows you view the overall health of enterprise applications, wherever they reside, within the enterprise network and graphically displays the relationships and data flow among the interconnected applications. This allows management, from a business perspective, to quickly determine the business impact of a component failure.

Tivoli GEM can help you understand how all the individual entities collectively come together to form a single business system. It enables you to perform distributed management even when a business system spans multiple platforms.

In brief, Tivoli GEM can help you manage:

- Business components that communicate across multiple systems
- Distributed business components

6.5.1.2 Tivoli NetView for UNIX

This product does the following:

- Monitors various network components (SNMP and SNA devices)
- Collects messages (events) and alerts from these components
- Performs some event correlation and automation on these platforms
- Notifies Tivoli Enterprise Console of the critical events if they cannot be resolved

6.5.1.3 Tivoli Decision Support

This is a tool that allows systems administrators and managers to make business decisions based on historical data. Decision Support can also help resolve problems. For example, the Decision Support module for server Performance Management can show you how well your servers are performing and also show you where you can allocate resources to improve service levels, such as additional servers, more memory, and so on.

6.5.2 Other automation solutions

In this section, we will be considering other automation products from the following vendors:

- Candle Corporation
- BMC Software

6.5.2.1 Candle Command Center

Candle Command Center is a suite of products. You can implement any number of these products to monitor the availability and performance of all the systems in your enterprise from one or several designated workstations. Command Center also allows you to monitor and manage diverse systems on diverse platforms throughout your network.

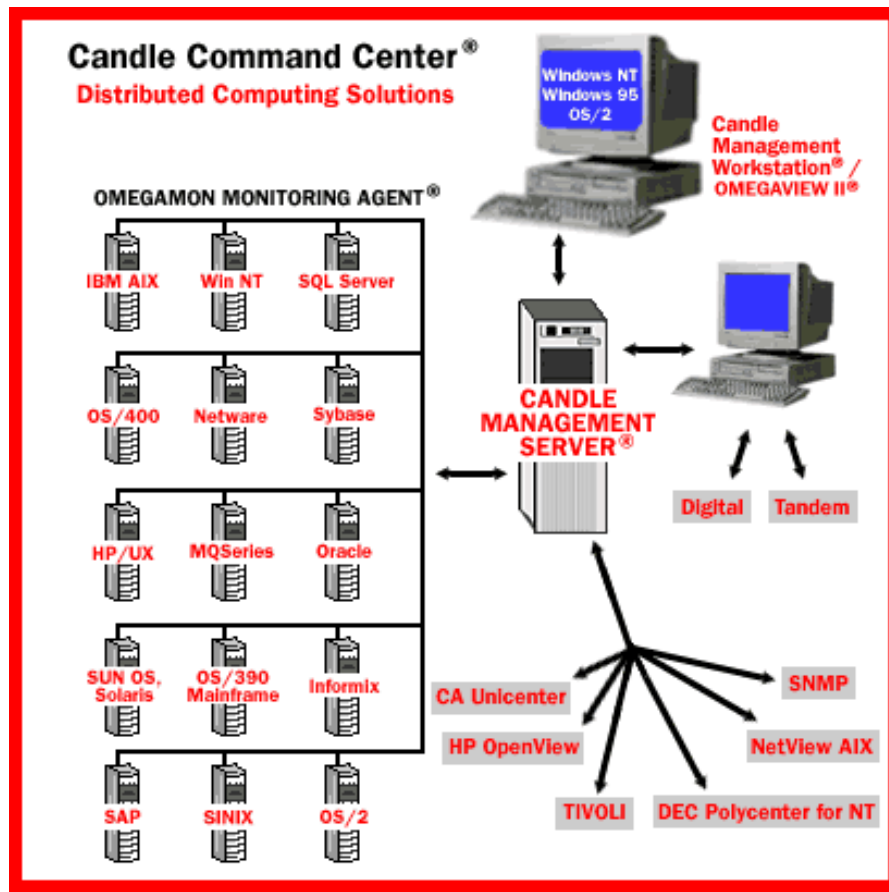


Figure 32. Candle Command Center overview

Features and benefits

With Command Center, you can:

- Establish your own performance thresholds
- Create situations, which are conditions to monitor
- Create policies, which are a collection of activities that provide the capability of automating responses to events or routine operator tasks
- Monitor for alerts on the systems and platforms you are managing
- Trace the causes leading up to an alert

You can divide monitoring by business applications, geographical locations, time zones, and types of systems, subsystems, or databases.

Example

Candle's Intelligent Remote Agents alert you when conditions in your environment meet threshold-based conditions so that you can limit and control network traffic. Agents provide key performance metrics, which are contained in detailed reports:

- System identification and activity
- CPU usage, system virtual memory, and load average
- Disk utilization, performance, and disk i-nodes
- Network activity
- Processes activity

Hints and tips

- You can also program your Command Center solution to take automatic action through Candle's Localized Automation capability. Localized Automation automatically corrects recurring availability and performance problems, as they occur, eliminating the need for manual intervention. It also provides problem identification and analysis and integrates your existing scripts and commands into the Command Center database. It also allows you to call third-party applications, such as pagers, to develop customized automation solutions.

6.5.2.2 COMMAND/POST from BMC Software

COMMAND/POST enables users to monitor and manage all technologies, including systems, networks, frameworks, and applications, from a single console. COMMAND/POST can create a tightly integrated, automated, advanced service management solution overseeing events across individual platforms throughout multiple geographical locations.

COMMAND/POST integrates with other tools, such as PATROL and third-party management solutions, to provide an enterprise-wide view of your network.

Features and benefits

- A central point for collecting and integrating management data from all applications, computers, workstations, LANs, WANs, and communications devices.
- COMMAND/POST Enterprise Server

The central server that exists on UNIX platforms and provides the core engine for integration, automation, and correlation.

- COMMAND/POST Explorer

The *cockpit of the enterprise* for viewing and managing business processes and services.

- COMMAND/POST PhonePoint

A notification system that enables events to be escalated to telephones, pagers, faxes, and so on, and actions to be taken directly from the telephone keypad.

- COMMAND/POST connect PATROL

This provides bi-directional integration with frameworks and other products, including PATROL, Tivoli TME, HP OpenView, and any SNMP compliant devices.

6.5.3 Batch processing

In this section, we will be considering batch processing products from the following vendors:

- IBM
- Tivoli
- BMC Software

6.5.3.1 Batch processing overview

Batch applications are generally used to satisfy large-volume processing. The out-of-the-box, standard UNIX has very little support for batch application since its natural environment is the interactive process. Over the years, many batch support tools have appeared on UNIX. They include:

- Batch submission packages, mainly used for creating and executing job chains
- Load balancing tools used to balance the batch workload over a cluster of UNIX systems

Under UNIX, job completion reports have traditionally been sent through e-mail to their owner. This mechanism is clearly insufficient, especially in the case of distributed tasks belonging to the same batch chain. Each vendor has developed its own job reporting method, which results in little or no interoperability between batch management packages.

6.5.3.2 IBM Job Scheduler

IBM Job Scheduler for AIX allows the administrator to define a workload (chains of jobs) and set rules for its execution (machines, time of the day, maximum run time, valid return codes, and so on). It allows administrators to define dependencies between jobs and files and to set up corrective actions that are automatically triggered in the case of errors.

6.5.3.3 Tivoli Workload Scheduler

Tivoli Workload Scheduler is a production automation solution for managing workloads in a distributed computing environment. It schedules, coordinates, and automates mission-critical application execution across an enterprise, thus, ensuring consistent and reliable operations.

Features and benefits

- From a single point of control, it supports multiple platforms such as UNIX, Windows NT, OS/390, and MPE.
- It integrates with applications, such as SAP R/3, Oracle Applications, PeopleSoft, and Baan.
- It is fully integrated with OPC, a host-based workload management solution.
- Job scheduling can be correlated with other enterprise management events, and the Tivoli Enterprise Console can manage it all from a single point of control.

6.5.3.4 CONTROL-M from BMC Software

CONTROL-M provides production control and scheduling over a multitude of platforms from a single point of management.

Features and benefits

- Comprehensive Scheduling and Production Control
With a Consolidated Active Environment, you get a real-time graphical view of all connected systems and platforms. Enterprise Controlstation provides a focal point of control for managing and automating complex cross-platform environments. A common user interface for all platforms ensures that production problems are detected and solved through a consistent platform-independent management tool.
- Reliability, Scalability, Availability, and Flexibility
Control-M solution works in both centralized and distributed environments using various combinations of client/server technology. It offers maximum scalability and flexibility regardless of hardware configuration or physical

layout. Communication failover problems are overcome through the use of mirror image databases.

- Interoperability and Management by Exception

A user notification facility detects exception situations on the network immediately and displays them in an alert window. The alert window allows users to drill-down immediately into the job and its environment, to do research, and to solve the problem.

Hints and tips

- CONTROL-M integrates with numerous third-party products and applications, such as SAP R/3, Oracle Applications, PeopleSoft, and Baan.
- Tight integration is also available for PATROL and COMMAND/POST and for management frameworks, such as Tivoli, HP OpenView, and CA-Unicenter.
- Enterprise Production Management also integrates with problem tracking and management applications, such as Remedy, Vantive, and Info/Management.

6.6 Event management

A consolidated system will often consist of multiple servers in close interaction or relatively complex configurations. In such an environment, it is important to have close control over, and a timely response to, events as they occur.

An event by definition is a matter worthy of remark. In this discussion, we consider an event to be an occurrence within our server environment that warrants attention. For example, there may be a hardware failure, a file system may be approaching capacity, memory may be over-utilized, and so on.

6.6.1 Event management solutions

In a complex system environment, such as a consolidated server configuration, an administrator cannot be monitoring all aspects of system activity all of the time. In most cases, it is impractical to do so. There are numerous software products available to assist in monitoring events. Some of these are:

- IBM PSSP
- Tivoli

- Independent Software Vendors solutions (ISV)

These products will monitor for defined events. When an event occurs, the software will react in a predetermined manner. This may involve sending a page message to the administrator, sending a mail message, or even executing a command. Different events may be handled differently within the same system.

6.6.1.1 IBM Parallel System Support Programs (PSSP)

The IBM PSSP used by the RS/6000 SP contains a component called Event Management (EM). EM is a member of the PSSP Group Services that maintains common state and membership information across all nodes in the SP complex.

EM provides a monitoring service by means of an application programming interface called the Event Manager Application Programming Interface (EMAPI). Through EMAPI, applications (called EM clients) can request the EM subsystem to monitor specific conditions. EM monitors those conditions, and it notifies its clients when those conditions are met.

The function of the EM subsystem is to match information about the state of system resources with information about resource conditions that are of interest to EM clients. A client may be an application, subsystem, or other hardware/software components.

The Event Manager daemon receives the status of system resources from Resource Monitors. Resource Monitors are utilities that observe the state of specific system resources. The Event Manager daemon applies expressions, specified by EM clients, to each resource instance being monitored. If the expression is true, an event is generated and sent to the appropriated EM client.

For example, a resource being monitored may be free space in a file system, and the client expression may be if free space is less than 15 percent. In this case, the client that defined the expression will be notified in the event of free filesystem space dropping below 15 percent of the total file system space.

The following represents a list of the Resource Monitors available within PSSP:

- IBM.PSSP.harmlid
Supplies information for the SP Switch, Virtual Shared Disk, and LoadLeveler subsystems.
- IBM.PSSP.harmpd

Provides information for the number of processes executing a particular program and is used to determine whether a particular system daemon is running.

- IBM.PSSP.hmrmd

Provides information from the PSSP hardware monitoring subsystem (hardmon).

- IBM.PSSP.pmanrmd

Supplies information from the PSSP Problem Management (PMAN) subsystem.

- aixos

Provides information from the AIX operating system.

- IBM.PSSP.CSSLogMon

Supplies information representing the state of SP Switch error log entries.

- IBM.PSSP.SDR

Provides information representing the modification state of System Data Repository (SDR) classes.

- Membership

Supplies the Host Membership and Adapter Membership states.

- Response

Provides information from the host_responds and switch_responds SDR classes.

Considerations

The PSSP software is typically only found on SP installations, and, as such, the PSSP Event Manager will only be available to systems within an SP complex. There are a number of advantages in using the Event Manager in an SP environment.

- The PSSP Event Manager is a component of the PSSP software and is available on any installed SP at no additional licensing cost.
- Events monitored through the Event Manager may be configured and monitored through the SP management interface *Perspectives*.
- The Event Manager Application Programming Interface (EMAPI) provides the ability to write applications that can take advantage of the existing resource monitors.

- Event Manager can be configured to communicate with monitoring tools, such as Tivoli and independent software vendor products, through PSSP Problem Management (PMAN)

For additional information on the PSSP Event Manager, refer to the *PSSP Administration Guide*, SA22-7348.

6.6.1.2 Tivoli Solutions

There are components within the Tivoli Enterprise software. There are two Tivoli products, based on the Tivoli Framework, capable of monitoring system resources and activities.

- Tivoli Distributed Monitoring (DM)
- Tivoli Enterprise Console (TEC)

Both TEC and DM have the capacity to monitor system resources and activities.

TEC is for asynchronous monitoring, and DM is for synchronous monitoring. TEC correlates events reported from different sources. TEC also maintains a history, whereas DM does not. TEC has the capacity to monitor a greater variety of events than DM can. Generally speaking, DM is used for local monitoring and is combined with TEC for further analysis.

6.6.1.3 Tivoli Distributed Monitoring

Tivoli Distributed Monitoring (DM) is an application that allows status monitoring of a wide range of hardware from different vendors running different operating systems, including resources that are not part of the Tivoli Environment.

DM possesses the ability to monitor and control specific aspects of a resource (percentage of disk space, status of a print queue, database process status, load average of a system, and network collisions). Its definition contains threshold values and various response actions triggered upon reaching a threshold (in essence, an event).

Tivoli Distributed Monitoring provides your network computing environment with the following features:

- Centralized monitoring of remote resources
- Predefined monitors for almost every resource
- Strong mechanism to generate events and alarms
- Automated decisions and actions in response to alarms or events

- Various responses (email, triggering a program)
- Custom scripts for monitoring specific applications
- Full integration with the Tivoli Enterprise Console event server
- Data collection for statistical analysis and capacity

6.6.1.4 Tivoli Enterprise Console

Tivoli Enterprise Console (TEC) provides a centralized point of integration and control for enterprise client/server environments. TEC provides functionality for administrators to monitor information about the environments for which they are responsible. TEC is at the center of the Tivoli availability solution.

TEC has been designed to detect potential problems before they cause outages. When problems are detected, TEC responds immediately. TEC can be configured to prevent administrators from being flooded with unnecessary data that masks the real problems.

For example, TEC can perform automatic actions or filter out duplicate messages. By maintaining a comprehensive history of reported conditions, TEC allows handling only serious problem that happen in a particular time frame or in the context of other previously received events.

Features

- Collects events from Tivoli Distributed Monitoring or any other resource.
- Correlates these events and helps define rules to determine the root cause of a problem. For instance, if 500 clients fail, and then the server fails, you can develop such a rule: If you get a client alarm, and then the corresponding server sends an alarm, you want to mark the client alarms as harmless, pointing out to the administrator only for critical server problems.
- Automates responses to solve the problems.

Considerations

Tivoli Enterprise software represents a large, comprehensive product suite with many management tools that belong to one of the four categories: Deployment management, availability management, security management, and operations management. To support and integrate these products, Tivoli uses the Tivoli Framework software as a foundation. The complexity of this solution makes it best suited for larger environments with 200 or more servers.

6.6.2 Other event management solutions

There are numerous, third-party solutions available for event management on the heterogeneous system, including the RS/6000 platform. In this section, we will be considering other event management products from the following vendor:

- BMC Software

6.6.2.1 COMMAND/POST from BMC Software

COMMAND/POST enables users to monitor and manage systems, networks, frameworks, and applications from a single console. It also provides an integrated, automated, service management solution overseeing events across platforms at multiple geographical locations.

COMMAND/POST integrates with other BMC tools, such as PATROL and third-party management solutions, to provide an enterprise-wide view of one or multiple subscribing clients.

Features and benefits

COMMAND/POST offers the following features:

- Client/Server monitoring and management structure.
- PhonePoint, which is an advanced notification system enabling events to be escalated to telephones, pagers, faxes, and so on.
- Bi-directional integration with other products including PATROL, Tivoli TME, HP OpenView, and any SNMP compliant devices.
- Management for OS/390 environments, enabling you to monitor, control, and IPL your systems.

The information on this application was researched, with permission, from BMC Software's Web site. See Appendix A for more information.

6.6.2.2 PATROL from BMC Software

BMC Software's PATROL product has become the industry standard for monitoring and managing applications in distributed computing environments. Only PATROL can administer all of the system components, including applications, databases, middleware, operating systems, and underlying technologies.

How PATROL works

PATROL uses libraries of expertise called PATROL Knowledge Modules (KMs) that provide information about each system component to a stand-alone, independent PATROL Agent. The PATROL Agent sends the

status information about critical health metrics gathered by the PATROL KMs to a central console for viewing by an administrator.

PATROL uses an interactive GUI that graphically depicts the monitored objects to show the complete health of the enterprise, therefore, enabling administrators to find and correct problems on the system *before* they impact users. It will also automatically notify administrators about situations requiring attention, and it can speed up and automate database administration including backups, recoveries, and reorganizations.

Features and benefits

- Provides centralized storage, versioning, and deployment for all PATROL Knowledge Modules (KMs)
- Allows centralized, remote installation of PATROL products and components
- Enhances security through the use of externalized security interfaces

BMC PATROL software can be practically utilized in small, as well as large, server environments.

6.7 Print services management

The consolidated server environment can face an unprecedented deluge of output created by a wide range of distributed applications and systems. Coupled with a variety of output and devices, the distribution of print services can present a challenge to the consolidated server administrator.

This output must be delivered in a reliable and timely way to end users working in a wide range of operating environments and who require output in varied formats. But, companies often waste expensive print and administrator resources on inefficient output processes. Users frequently receive written information late, at the wrong location, or with incorrect data.

The challenge is to effectively manage the growing complexity of distributed output while ensuring the optimal and most cost-effective use of printing and other output resources.

6.7.1 Print services management solutions

Tivoli has a comprehensive solution to the problem of print service management. In the next section, we examine some of the benefits of Infoprint Manager for AIX and Tivoli Output Manager.

6.7.1.1 Infoprint Manager for AIX

IBM Infoprint Manager is a comprehensive software solution for managing digital printing. It combines award-winning print management technology with enhanced file management and spooling capabilities to address the requirements of a variety of print markets. Using Infoprint Manager, you can submit diverse file types to a single system to be managed, printed, stored, and reprinted quickly and efficiently.

Infoprint Manager

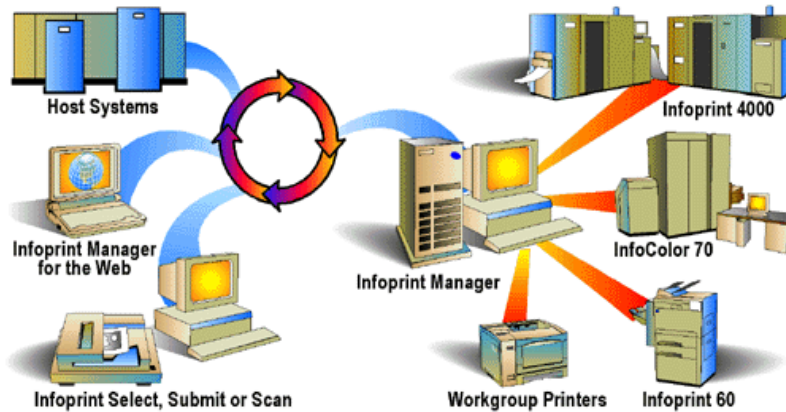


Figure 33. Infoprint Manager

Combining superior print technology with built-in Advanced Function Presentation™ (AFPTM) and Adobe PostScript support, Infoprint Manager delivers production-volume, single-copy integrity with PostScript quality. Infoprint Manager simplifies the print process and reduces your operating expenses, whether you are printing books on a high speed printer or managing all the printing for your distributed enterprise. It offers a complete system that meets your diverse printing needs.

Features

Centralized Output Management

- Configure, manage, and monitor from remote locations.
- Get your jobs to the printers, fax machines, and other destinations.
- Operators get notified when events happen.

- Balance printer workload.
- Enhance printing with imposition and output finishing.
- Support a wide range of printers.

Job Submission and Print Management Applications

- Submit jobs from a range of platforms.
- Manage jobs and printers from various platforms.

Hard Copy Scanning

- Take advantage of the IBM Infoprint Scan integrated solution, which combines hardware and software in a single system.
- Dial hardware and software services with a single phone number.
- Cut and Paste, Clean up, align, and annotate scanned images for printing.
- Save scanned files as PDF source for viewing, printing, and distribution on the Web.

Variable Data

- Incorporate a range of variable items, from database entries to rotated graphics.
- Create and manage a variable data job like any other Infoprint job.
- Tuned for performance efficiency, in color or black and white.

Benefits

- Provides secure, scalable enterprise printing support.
- Provides reliability for mission critical applications, such as SAP/R3.
- Manage and print to printers, fax machines, and more.
- Supports a wide range of printers.
- Provides hardware, software, and services available.
- Include printing in your Tivoli System Management solution.

6.7.1.2 Tivoli Output Manager

Tivoli Output Manager controls the distribution of information in many forms: Print output, e-mail, FAX, Web pages, and others. It provides an automated information delivery and a secured and centralized administration of output. It is based on a Fault-Tolerant architecture, and it affords powerful tools for end users.

Features

- Tivoli Output Manager supports encrypted and compressed output file distributions over any network.
- All users of Tivoli Output Manager must be authorized to use the various features of the product. Security is managed by role definitions. A user is assigned to one or multiple roles.

Benefits for the system administrators

Tivoli Output Manager provides consolidated, enterprise-wide views of output management resources including users, domains, nodes, devices, and output status. The results of this are:

- Greatly reduced incidents of lost reports
- Automation of complex report distribution schemes
- Efficient use of all printer and output device resources
- Centralized spool file management for networked systems
- Easy tracking and auditing of spool file processing
- A significant reduction in paper costs

Benefits for the end users

Tivoli Output Manager empowers the end users by making available to them all authorized output resources directly from the desktop. It also allows them to create and store output automation rules directly from their desktop. End users benefit from:

- Expanded access to authorized devices
- Automatic report distribution
- The ability to print to multiple devices

6.8 Configuration management

One of the challenges in a consolidated server environment is keeping track of the hardware and software installed on each machine. System configuration information is vital to managing a consolidated environment. For example:

- An administrator preparing to distribute a software upgrade automatically generates a list of systems worldwide that meet the new version's hardware prerequisites.

- A help desk technician quickly analyzes recent configuration changes to an end user's malfunctioning system and determines that a critical file has been inadvertently deleted.
- A new accounting application is planned for company-wide roll out. A list of all obsolete servers is automatically generated to ensure that they are replaced before the application is distributed.

6.8.1 Configuration management solutions

6.8.1.1 Tivoli Inventory

The Tivoli Deployment solution consists of two products:

- Tivoli Inventory - Creates a repository of all the configuration data of your systems in your enterprise.
- Tivoli Software Distribution - Efficiently delivers new applications on your systems.

These two applications rely upon the Tivoli Framework for the delivery services.

Tivoli Inventory addresses this problem by providing the means to gather hardware and software information related to each system and then storing that information in a relational database. Queries and reports can be run to display the information in this database.

Features and benefits

- Tivoli Inventory stores inventory information in a Relational Database Management System (RDBMS) and, therefore, allows any non-Tivoli applications that can access SQL data to share the inventory information. Moreover, it benefits from the advanced features of an RDBMS system, such as scalability and performance.
- Tivoli Inventory has close links with Tivoli Software Distribution. As its name implies, Tivoli Software Distribution provides facilities for the distribution and installation of software to managed systems in a Tivoli environment.

6.8.1.2 Tivoli Asset Management

Tivoli Asset Management is a component of Tivoli Service Desk. While Tivoli Problem Management and Tivoli Change Management address the operational aspects of delivery IT services, Tivoli Asset Management deals with the assets that are used to deliver those services.

Features

- Tivoli Inventory provides functions to track assets from acquisition through maintenance to disposition.
- The costs associated with the maintenance of an asset are recorded as part of the history log.
- Tivoli Inventory provides a number of schemes called Data Manager Hierarchies for organizing and identifying assets.
- Tivoli Inventory also incorporates a reporting function for generating reports for different purposes.

6.9 Change management

Most organizations change every day. Any change, whether it involves your computer network, the software you are using, system administration, or security could have an impact on your company's efficiency and performance as well as the ability of your service desk to manage service levels across your organization. As the volume, speed, and complexity of change increases, your enterprise needs to manage change more efficiently and reduce the associated risks. Change management is the process through which these alterations to the server environment are introduced, executed, and recorded.

A change management solution needs to address the need to efficiently and safely execute changes and manage approvals and planning while avoiding *sneaker net* for software updates. It also needs to address model and version control and currency across large enterprises. Change management needs to quickly implement changes for maximum business impact to help keep your business competitive, help reduce the complexity of applications, and facilitate large numbers and varying types of users.

6.9.1 Change management solutions

Various independent software vendors provide some sort of configuration management. For example, EcoTOOLS from Compuware has a configuration management function for integrating systems. Tivoli has an extensive solution, which will now be described.

6.9.1.1 Tivoli Change Management (TCM)

TCM is a component of the Tivoli Service Desk. Tivoli's approach to change management acknowledges that change must be factored into the overall business objectives, must be automated as much as possible, and must minimize the end-user impact.

TCM gives you complete control over the process involved in making any alterations to the consolidated server infrastructure.

Examples

- Changes to a system can include software upgrades, hardware location moves, or network modifications that are underway. With Tivoli Change Management, you will have critical information about any change including an analysis of the impact it may have so that your service desk can manage all stages of the process.

Tivoli Change Management allows you to plan better by leveraging past experience and understanding the impact of changes before initiating them.

Features and benefits

With Tivoli Change Management, your organization can:

- Embed business rules into the change management process. Policies unique to an organization are made part of the process. You can ensure that your best-practices approach is followed during execution. For instance, the system might require that specific actions occur, approvals are obtained, notifications are delivered, and costs are associated with specific change categories.
- Interpret the impact of a change. By analyzing past changes and any associated problems or assets, it is easier to estimate who or what may be affected by a proposed change so that you can develop a plan that minimizes the impact.
- Escalate activities to keep the process moving. Many times, a change takes longer to implement than planned. Now, you can raise the awareness of critical path activities, delayed tasks, outstanding approvals, and uncompleted changes to keep the change process on schedule.
- Completely manage every step in the process. Change management often fails because one or more of the critical steps are disregarded. Tivoli Change Management gives you the power to ensure that every step is followed: Assigning priorities, analyzing for impact, establishing tasks, assigning responsibilities, and routing them electronically for review and approval, all of which are based on the business rules that the organization has defined.
- Create templates to institutionalize change. By defining and working with models of commonly occurring changes, it is easy to create templates for future change requests. Now, you can free yourself from the laborious task of specifying procedures for repetitive changes. You simply define the process once, improve it over time, and ensure consistent execution.

- Notify change participants automatically. Users of the system can notify anyone necessary at specific times in the change process. Half the battle in executing any smooth change is apprising the many people involved in the status and their responsibilities. Now, you can automatically notify those affected by a change, those executing it, and those managing it.
- Continuously analyze and evaluate the change process. Through the continuous collection of change data, reports can be generated, and you always have access to comprehensive statistical information about the overall change process.

6.10 System recovery

Failure due to hard disk malfunction, operating system corruption, irrecoverable data errors and so forth can cause much distress to an organization whose business depends on the availability of the system. No administrator will disagree on the importance of having good backups. Not only do they prevent data loss, they also reduce recovery time in the event of system failures.

There are various aspects of backup that need to be considered, namely, the operating system, user data, and configurations.

No doubt when we deal with the SP system, we are dealing with a cluster of RS/6000 systems. The strategies involved in system recovery can prove to be of utmost importance to an administrator dealing with hundreds of nodes.

6.10.1 Centralized back up and recovery

Back up and recovery of software, data, and the operating system is the same for UNIX systems of any type and flavor. The AIX operating system has special backup commands that allow you to make copies of the currently installed operating system and configuration information. There are also many software packages that allow you to back up software and data to archive media, such as tape and read/writable CD-ROM.

With server consolidation, we can introduce the concept of a centralized backup and recovery solution. This allows us to implement a uniform and consistent level of backup with the ability to recover easily.

Poor management in this area can drastically increase backup time and result in unnecessary waste of archival space. More importantly, it could compromise your recovery position. Always ask yourself when you need to back up and what you need to back up.

A quick guide as to when to back up the operating system would be when major changes are made. This includes installation of new software, upgrades, patches, changes to configurations, hardware changes, especially to system planar, hard disks belonging to rootvg, power supplies, and so forth.

The timing and content of data backups is business driven to minimize the point of recovery and maximize data integrity.

6.10.1.1 Tape Management

Another benefit of server consolidation is the ability to centralize tape management and resources; however, it is only possible if the system or application software provides the necessary functions. Such functions include:

- Tape volume control (inventory, ownership, and access rights)
- Allocation of tape drives
- Centralized (manual) mount interface for operators
- Remote access to tape drives or tape data

All of the above functions are missing in UNIX operating systems and, therefore, must be provided by additional software products, such as NetBackup or ADSM, which have appeared on the market in the past few years.

You probably have manual, stand-alone, internal and external tape devices as well as automated tape libraries in your UNIX environment. Single, manual drives require tape cartridges to be loaded and unloaded one at a time by hand. The tape libraries have one or more drives, contain a number of tape cartridges stored in slots in a removable cassette, and use a robotics accessor mechanism to mount (load) and dismount (unload) tapes in the drives.

Automated tape libraries clearly have a distinct advantages over manual tape devices. To begin with, you can leave automated tape libraries unattended for long periods of time because their robotics load and unload tapes in the drives. In addition, you can physically lock the access door, therefore, making the tapes relatively secure. You can use an automated tape library for more than one application at a time because there is usually more than one drive in the library. Automated tape libraries are well suited for use with an application like ADSM.

6.10.2 Centralized backup and recovery solutions

In this section, we will be considering some of the backup and restore products on the market that assist the back up and recovery of data and software including the following vendors:

- Sysback/6000 from IBM
- NetTAPE Tape Management from IBM
- Tivoli Storage Manager (previously called ADSM)
- NetBackup from VERITAS
- NetWorker from Legato
- SmartMedia from Legato
- ARCserve/*T* from Computer Associates
- AMASS Offline Media Manager from ADIC

6.10.2.1 Sysback/6000 from IBM

Sysback/6000 provides an easy-to-use way to back up and restore Oracle datafiles. You can use SMIT screens to do both local and remote backups of logical volumes.

Sysback/6000 is a flexible backup tool that handles the recovery and replication of AIX systems. Sysback is a single product for backing up and restoring user data, applications, and the operating system.

Sysback/6000 is a single product for backing up and restoring user data, applications, and your operating system. Sysback/6000 performs back up, listing, verification, and restoration from various backup types including:

- A full system backup (including multiple volume groups)
- Select volume groups
- Select file systems
- Select files or directories
- Raw logical volumes

Features and benefits

- Provides easy back up and recovery of either individual files or your entire system
- Replicates your AIX operating system
- Improves performance and capacity of your backups
- Utilizes the AIX interface, System Management Interface Tool (SMIT)

- Provides backup and recovery from local and remote devices

6.10.2.2 NetTAPE Tape Management from IBM

NetTAPE provides enhanced tape capabilities in the AIX environment. NetTAPE offers a single-system image of tape drives and libraries for operations and end-user access. All of the tape and library devices that NetTAPE manages appear as a set of resource pools to operators, applications, and end users, thus, providing a single-system image.

Features and benefits

- Central operation - Tape drives can be managed from one or more operator stations through command lines or a Motif-based graphical interface.
- Remote tape access - An application can access a tape drive connected to a remote system.
- Device allocation - NetTAPE manages the selection and allocation of devices to requesting applications and end users. Therefore, tape drives can be shared between applications running on either the same or different systems.
- Exchange of tape volumes - NetTAPE's support of tape labels enables you to exchange tape volume data between AIX and mainframe applications.
- ADSM support - NetTAPE provides an interface to ADSM so that tape and library devices can be managed by NetTAPE instead of being dedicated to ADSM. Therefore, drives can be shared between ADSM servers and/or other applications.
- ADSM cannot use NetTAPE to access remote devices.

6.10.2.3 Tivoli Storage Manager

Tivoli Storage Manager manages distributed storage. Using a client/server architecture, it enables a variety of client nodes to send data to a central server machine, the Tivoli Storage Manager server. This server is responsible for managing the storage devices to which the client data is sent and stored, and it uses a database to keep track of the location of every piece of data. Clients can get their data back when they ask for it.

Because UNIX operating systems do not provide the necessary functions for managing tapes, this server provides those functions for its own use.

If you are in a situation where it can satisfy your needs for data storage, you can use it to manage your entire tape and automated library environment to provide your distributed clients with:

- User home directory backup, archive, and space management
- Backup and archive of distributed databases
- Lotus Notes backup and archive
- Backup and archive for specialized applications, such as image and print-on-demand

Features and benefits

Using Tivoli Storage Manager to support tapes enables you to get away from a model where a user owns a tape volume. When you use it, user data is managed in a way that makes tape volumes transparent to the end user.

Using it has the following advantages:

- Improved tape volume utilization
Tivoli Storage Manager tape volume utilization will remain consistently high because tape volumes are filled to full capacity, and, when data is deleted from these volumes, the space is reclaimed by consolidating several partially filled volumes onto a full volume.
- Improved tape drive utilization with disk caching
It provides the ability to use disk space as a write cache in front of your tape drives. Disk caching avoids the problem of users holding on to drives for many hours for little usage and, therefore, provides better utilization of your tape drives.
- Management of the tape volume inventory
It manages its tape volume inventory so that you do not have to manage the inventory yourself.
- Improved data recovery
It enables you to selectively create duplicate copies of your backed up data (whether the backed up data is stored on tape or disk) and automatically uses the duplicate copies if the physical volume containing the primary copies is lost or damaged.
- Automated library support
Because it supports and manages automated libraries, it is easy for you to implement tape automation.
- Disaster Recovery

It includes functions for disaster recovery that enable you to reconstruct your entire production environment at a recovery site. The disaster recovery feature can manage an off-site rotation strategy for Tivoli Storage Manager tapes, automatically creates a disaster recovery plan for the this server, and stores client recovery information in the Tivoli Storage Manager database.

The information on this Tivoli application was researched from Tivoli's Web site and from Tivoli redbooks. See Appendix A for more information.

6.10.2.4 NetBackup from VERITAS

NetBackup ensures continuous data availability. From a single management interface, VERITAS NetBackup automates enterprise backup operations for thousands of users across multiple servers and consolidates management of all storage devices (stand-alone, departmental) and those in the data center. It improves data reliability for organizations that store critical information on heterogeneous systems and distributes it around the world.

When a disaster occurs, it can be as simple as a disk array crash or as big as the computer room being flooded. NetBackup not only has the ability to perform full or partial recovery from a primary backup but can be used to recover applications or complete servers in an off-site scenario.

NetBackup provides the ability to automatically create copies of the primary backups. These secondary tapes can then be sent off-site for storage. However, there is more to the story than just copying tapes. First of all, NetBackup de-multiplexes tapes so that data is *co-located* on tapes. The reason for this is that most installations have business critical applications that must come up first followed by secondary and tertiary applications.

The process of performing a selective restore is much faster if the data is co-located. Very rarely does an organization choose to restore a complete server at a hot-site location. Also, the backup copies that NetBackup creates are TAR compatible. While NetBackup uses its own method for moving data and writing data to tape to ensure reliability, it provides the capability for these tapes to be read by basic UNIX utilities.

For complete disaster recovery automation, NetBackup provides an option for complete vault management. This includes everything from ejection of the backup copies to the I/O bin in a tape library to pick/pull reports written in a variety of formats including Arcus and Datasafe. Additionally, tapes are automatically rotated to and from the off-site vault.

Features and benefits

- Heterogeneous Support - Coverage for all major UNIX platforms. Support for Windows NT, Novell NetWare, and database and application options including all leading databases and applications: Oracle, Microsoft SQL Server, Sybase, Informix, Microsoft Exchange, and SAP R/3, with more in development.
- Multiplexed Backup and Restore - Ability to write and read multiple data streams to one or more tapes from one or more clients/servers in parallel for optimum performance.
- Scalable Image and Media Catalogs - Distributed, small footprint, catalogs that track backups and tape media based on a fast access, segmented structure that can easily be backed up, restored, or replicated.
- Non-proprietary Tape Format - Ability to create *TAR compatible* tapes.
- True Image Restore - The intelligence to re-create data based on the current allocations, negating the recovery of obsolete data.
- Data Encryption Option - 56-bit encryption available for customers in the United States and Canada; 40-bit encryption available for all domestic and international customers.
- Remote GUI Administration - Full backup and restore capabilities from any location including dial-up networks. Easily define schedules, set backup windows, and identify backups with meaningful names.
- User Initiated Backup and Restore - Easy-to-use interfaces available for end-users that reduce system administrator intervention.
- Network Bandwidth Throttling - Option to control NetBackup network utilization.
- Job Prioritization - Ability to set priorities for backups based on importance.

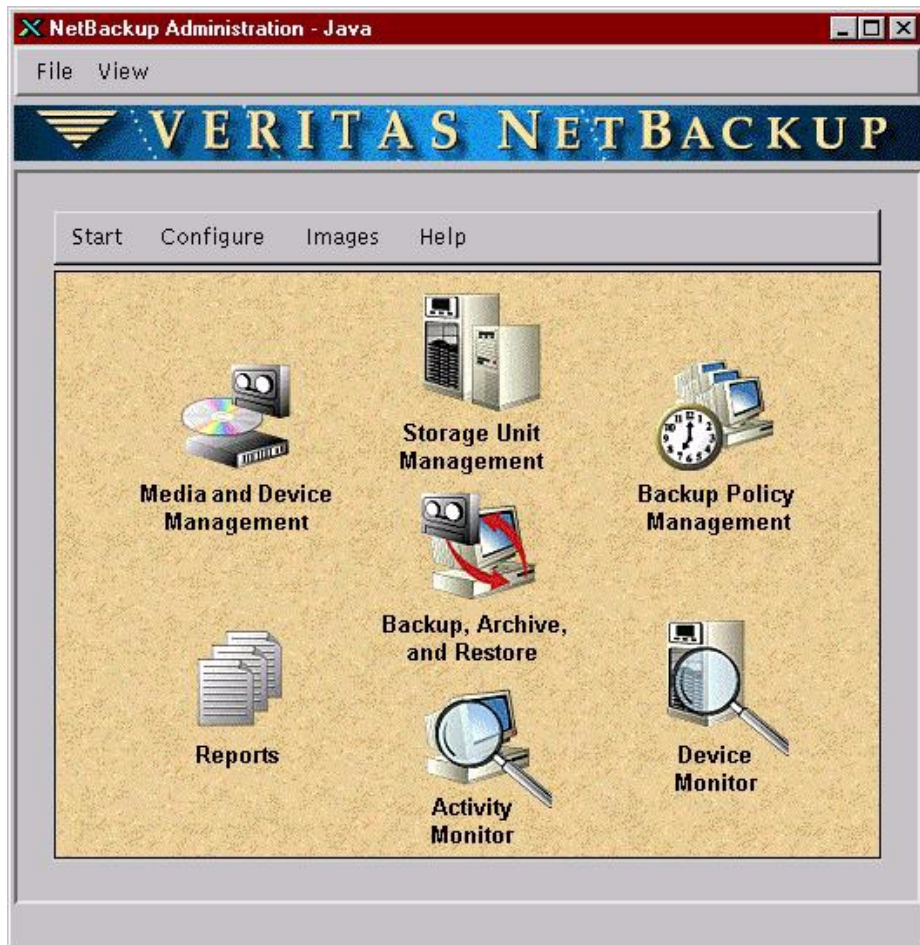


Figure 34. NetBackup Java interface

- Tape library sharing - Lower tape robotic costs by sharing tape library with multiple NetBackup servers.
- Complete Tape Management - Manage the entire tape life cycle.

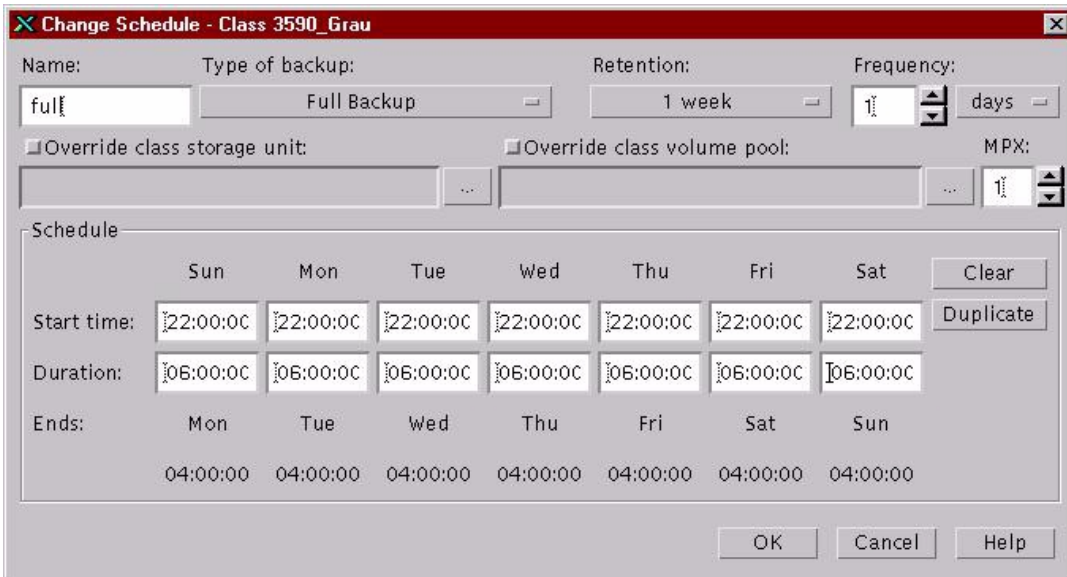


Figure 35. NetBackup GUI Scheduler interface

6.10.2.5 NetWorker from Legato

Legato NetWorker helps organizations to easily protect vast amounts of critical distributed data by offering heterogeneous platform support, automated media handling, interoperable tape format, data stream parallelism, remote tape management, and automatic integration with the popular storage management frameworks in the network environment. Legato NetWorker takes these capabilities to a new level for state-of-the-art data protection, making it the most flexible and scalable solution available.

Legato NetWorker is based on the client/server architecture and comprises three distinct components: Client, Storage Node, and Server. This three-tier architecture provides the flexibility and performance required to protect and manage data on the most complex networks.

The data protection process is encompassed in a data zone, which is a collection of Clients, Storage Nodes, and data protected by a single Legato NetWorker Server. Legato NetWorker data zone can be collectively administered with a single set of data protection policies and procedures. Legato NetWorker excels as a data zone manager, delivering both centralized and distributed storage management.

Features and benefits

Comprehensive Data Protection

- Protects all local and remote data running on Windows NT, UNIX, and NetWare servers and all major desktops including security information, user profiles, access control lists, and Windows registries and event logs
- Add-on modules protect open files and popular databases and applications
- Full, incremental and differential backup
- Data encryption and password protection
- Automatic cloning provides additional data integrity and availability
- Add-on Archive and HSM applications enable manual and policy-driven archive, retrieval, migration, and recall of data
- Failover capability based on Storage Node priority list increases backup reliability and availability

Superior Performance

- Fast performance for both backup and restore
- Multi-tier storage management with optional staging of storage management to intermediate high speed disk
- Local device support using Storage Nodes optimizes backup/recovery performance and reduces network traffic
- Client-side parallelism and data compression accelerate throughput while minimizing server load and network traffic
- Built-in support for parallelism and data interleaving of 32 data streams-per-device ensures devices are driven at maximum speeds
- Legato NetWorker Power Edition provides built-in support for 64 parallel data streams for fastest backup for VLDB and data warehousing environments Year 2000 Compliant
- Designed, developed, and tested for post-2000 environments

Easy to Manage and Use

- Windows GUI provides speed icons, online help, and single point of administration to configure, schedule, or monitor network-wide backups
- Scheduled, manual, or ad-hoc backup and restore
- Support for Sun DiskSuite, AIX LVM, and Veritas file managers

- Desktop users can perform ad-hoc backups and browse and recover own files from a graphical, online index
- Automatic event notification via pager, e-mail, or SNMP

Advanced Media Management

- Broad support for 4 mm, 8 mm, DLT, and other devices
- Power Edition provides built-in support for high-speed devices and concurrent operation of 32 devices per Storage Node
- Advanced media management automates media handling, cleaning cartridge support, and media verification
- Support for electronic labeling, bar codes, and cartridge access ports reduces media inventory requirements

Enterprise Scalability

- Ability to dynamically expand capacity by adding client connections, storage devices, and Storage Nodes via software enablers
- Fully interoperable across Windows NT, UNIX, NetWare, Windows 95/98 and 3.1, DOS, and MacOS
- Interchangeable tapes between Legato NetWorker servers for Windows NT, UNIX, and NetWare
- Legato NetWorker Workgroup Edition protects smaller networks across the enterprise
- Legato NetWorker Power Edition provides unsurpassed performance for very large file systems and databases
- Support for unlimited client index size

6.10.2.6 SmartMedia from Legato

Legato SmartMedia is an open media management application that provides standard interfaces for applications, robotic library control, drive control, and administration, thereby enabling enterprises to better manage their growing base of removable media and devices distributed across a heterogeneous storage environment.

Whether deployed within or across data zones, Legato SmartMedia acts as a gateway for drive and library sharing. It controls application requests for media from a central location and allows all Legato SmartMedia aware applications, such as Legato NetWorker, to access and share all Legato SmartMedia attached devices.

Features and benefits

- Device sharing
Leverage all available devices across participating applications, increasing performance through dynamic device sharing
- Access control
Protects media and devices from errors that can occur with device sharing
- Advanced device integration
Rapid support for, and compatibility with, new and updated devices
- Off-site storage management
Quickly locate media stored outside of the library
- Distributed architecture
Flexible implementation to maximize network bandwidth
- Single media database
Leverage limited System Administrator resources through central media management

6.10.2.7 ARCserve/IT from Computer Associates

ARCserve/IT combines both native Windows NT and UNIX based support with single console management featuring cross platform support. The combined solution provides the widest platform, application, and technology support for heterogeneous IT environments. ARCserve/IT features: Unicenter TNG integration, virus scanning, data compression, virtual library support, automatic media management, and parallel streaming to multiple devices. ARCserve/IT offers market-leading functionality with reliability and performance. ARCserve/IT now offers powerful new add-ons such as fibre channel SAN support, shared library support, AS/400, delta backup to OS/390, Non-Stop SQL/MX support, and a Java based console for remote Web based management on UNIX. ARCserve/IT also contains integrated RAID support and add-ons, such as Image backup, Open file protection, online application backup, and multi-platform client agents. ARCserve/IT continues to set the standard for high performance enterprise managed storage.

Features and benefits

- ARCserve/IT automates every traditional storage management task, can lower operating expenses, eliminates errors, and allows you to easily perform essential data protection routines.

- ARCserve/IT has comprehensive data protection functionality and reliably secures critical data. In the event of a disaster, ARCserve/IT's automated and remote disaster recovery functionality can quickly and easily recover your vital information to keep your business running.
- ARCserve/IT provides high performance, automated storage management for Windows NT and UNIX server platforms, and extends protection to virtually any client including Windows NT, Novell NetWare, Macintosh, OS/2, and a variety of UNIX platforms.
- ARCserve/IT's remote installation feature can remotely install its options and agents throughout the enterprise. This eliminates time consuming local installation of systems on each agent.
- ARCserve/IT's Intelligent Data Compression feature increases effective device capacity, decreases network traffic, and increases backup and restore performance.
- ARCserve/IT features centralized, cross-platform management. This helps manage multiple ARCserve/IT jobs, devices, and database with one single view. It also enables centralized reporting across Windows NT and UNIX platforms.
- ARCserve/IT maximizes performance by simultaneously backing up, storing, or copying data to or from multiple devices. To enhance performance for remote client data, ARCserve/IT uses *push* client agent technology. The data is pre-packaged at the client site and passed to ARCserve/IT, therefore, saving server CPU cycles and maximizing throughput.
- ARCserve/IT's mainframe-class rotation scheme improves administrator productivity and eliminates accidental overwrites. Media rotation automates the process of monitoring media usage and recycling. Grandfather-Father-Son or customized rotation is automated to provide *lights out* backup/copy operations.

6.10.2.8 AMASS Off-line Media Manager from ADIC

AMASS is a archiving software that makes a tape or optical library look like magnetic disk. AMASS-OMM is an option of AMASS. OMM tracks media outside the library when AMASS archives the files. AMASS has an HSM option to handle disk management called DataMgr.

AMASS Off-Line Media Manager (AMASS-OMM) software provides easy access to data contained on media that has been removed from an automated optical or tape library. Access is achieved by using a compatible stand-alone drive or multiple drives attached to the AMASS library server. By

using this hardware configuration and AMASS-OMM, access to off-line data on archived media is now simplified.

In standard storage configurations, older or infrequently accessed media is usually removed from the automated library and stored off-line. This opens up slots in the library for more active media. However, at various times, the data stored on this *shelved* media will need to be located quickly and then reaccessed. To complete this task in a timely manner, an operator must free up slots within the library by temporarily removing active media. Unfortunately, this could temporarily prevent other users from accessing active data on active media.

By using AMASS-OMM, the time-consuming process of locating and then accessing off-line data has been streamlined through the use of a compatible stand-alone drive or multiple drives attached to the AMASS library server.

The AMASS Index, located on magnetic disk, tracks the files in the library file system and tells the user whether the sought-after media is located in the library or is on the shelf. If the media is resident in the automated library, the read or write operation is completed transparently. If the media is located off-line, the operator receives a load request message in a sysop window on the AMASS server console. The load request specifies the media identification number and side A or B if optical media is being used. After the operator loads the desired media into a stand-alone drive, the user's read request is completed transparently.

6.10.3 Software recovery

Within the RS/6000 environment, two basic types of software error will be encountered:

- System errors
- Application errors

Through the reporting of software errors, we may facilitate software recovery. Proper reporting and monitoring of software errors will enable timely recovery from software related problems. With many server consolidation tools, it is also possible to automate actions based on reported errors.

6.10.3.1 System errors

System errors are errors relating to the operating system and associated software, such as AIX, PSSP, HACMP, Tivoli, and so on. Within the AIX operating system, there are three basic methods for reporting system errors:

- syslog

- errorlog
- text log files

Syslog is a BSD error logging facility also available with AIX. The errors logged by the syslog daemon are defined in the file `/etc/syslog.conf`. All errors that are to be reported by the syslog daemon must be defined in the `syslogd.conf` file.

The errorlog facility is enabled for all AIX operating systems and reports on detected software errors. Any AIX user may generate a report of errors that have been logged by the errorlog subsystem.

Many of the AIX subsystems and associated software products will write errors to designated log files on disk. This occurs with products, such as PSSP, HACMP, ADSM, and so on. These files must be viewed to see errors that have occurred.

6.10.3.2 Application errors

The reporting of application errors is dependent on the application and may be different for each application. An application may report errors through a number of methods:

- Writing directly to the user session
- Writing the error to a log file
- Utilizing a native error reporting facility, such as syslog or the error log
- Writing the error internally to application logging

To determine the error reporting methods used, the application documentation should be consulted.

6.10.3.3 Consolidated software errors and recovery

Server consolidation products and tools provide the ability to monitor errors on all servers in the environment from a central point of control or console. Many tools also provide the ability to define an action to be taken based on reported errors.

By comparison, in a server environment that is not consolidated, it would be necessary to monitor each server in the environment separately. This has a higher management overhead and increases the possibility of overlooking errors.

There are a number of tools available that enable centralized software error management:

- PSSP (native to the RS/6000 SP environment)
- Tivoli
- ISV Products

6.10.4 Hardware recovery

In the business environment of today, system outages can result in a significant loss of revenue. Hardware recovery has been a focus of RS/6000 design in an effort to reduce system outages for business critical applications. In addition to standard RS/6000 features, the SP environment has been designed with hardware availability as a major objective.

6.10.4.1 RS/6000 hardware recovery

Hardware recovery on the RS/6000 of today is achieved through a number of mechanisms.

- Error Checking and Correction (ECC) on storage

Most recent motherboards use ECC to detect and correct memory errors. The ECC provides detection of double-bit errors and correction of single-bit errors.
- Parity checking on internal circuit

The POWER processor has imbedded error detection circuitry that prevents incorrect data from being processed.
- Disk Drive Predictive Failure Analysis (PFA)

Some IBM disk drives, such as the ones used in the RS/6000 Model F50, incorporate PFA technology. The built-in PFA circuitry continuously monitors drive performance and can determine an early tendency toward disk drive failure long before it actually happens. When this condition is detected, an error status byte is returned to the AIX device driver, which results in an error log entry in the AIX error log facility. Automatic Error Log Analysis, if enabled, sends a warning message to the system administrator. Preventive measures can then be taken to preserve the disk data and perform a repair action before a serious impact to customer operations occurs.
- Other hardware improvements

Some UNIX machines now include fault-tolerant hardware. For example, IBM RS/6000 Model F50 incorporates seven new Reliability, Availability, and Serviceability (RAS) design features:

 - Method and system for reboot recovery
 - Environmental and power error handling extension and analysis

- Automatic processor surveillance
- Machine check handling for fault isolation in a computer system
- Method and system for check stop error handling
- Error collection coordination for software-readable and non-software-readable fault isolation registers in a computer system
- A method and system for fault isolation for PCI bus errors
- Processor redundancy

Some of the current UNIX systems can run on either SMP machines or loosely-coupled parallel systems and use some form of take-over redundancy.

For example, the High Availability Cluster Multi-Processing (HACMP/ES) facility for AIX is a control application that can link up to 32 RS/6000 servers or SP nodes into highly available clusters. Clustering servers or nodes enables parallel access to data, which can provide the redundancy and fault resilience required for business critical applications.

HACMP automatically detects system failures and recovers users, applications, and data on backup systems, thus, minimizing downtime to minutes or seconds. In addition, using HACMP for AIX minimizes planned outages since users, applications, and data can be moved to backup systems during scheduled system maintenance. With the addition of the new Dynamic Reconfiguration facility, customers will no longer have to stop their cluster to add or remove processors, adapters, or networks. Using administrative interfaces, these activities can now be performed on an active HACMP cluster. Dynamic Reconfiguration can also be used to change the failover behavior of the cluster without disrupting current operations, therefore, helping the customer approach non-stop cluster operations.

HACMP software detects and recovers from failures of disks, disk adapters, networks, network adapters, and processors. If a node fails, nominal recovery time is approximately 30 to 300 seconds. Actual recovery time is a function of the system configuration, the application configuration, the size of the user's databases, and the user's recovery script (if any).

For more information on HACMP/ES, see Chapter 8, "High availability" on page 233.

6.10.4.2 Additional SP hardware recovery

In addition to the RS/6000 fault tolerance, there are some additional features specific to the SP environment that assist in hardware recovery.

- First Failure Data Capture (FFDC) diagnostics

These are the same problem determination techniques as used on mainframe systems and have been designed into the latest 332 Mhz SP nodes. The implementation of FFDC diagnostics also means that a long boot is no longer necessary, therefore, reducing the boot time of these nodes to 6-8 minutes.

- Redundant power supplies

The SP frame and switch power systems are designed with N+1 power so that a power supply failure does not cause a system failure. In addition, a failed power supply can be replaced concurrently without stopping the system.

- Hardware fault tolerance

A service processor failure does not stop the system. A failing frame supervisor card may be hot-plug replaced without powering down the system. The SP Switch has been designed to continue operating, where possible, even when certain hardware failures occur. The SP Switch has a multiple path design such that switch errors may also be tolerated. Data transmission across the switch is protected by Cyclic Redundancy Check (CRC) code so that any transmission errors can be detected and recovered from.

- Processor modularity

If a node fails, the remaining nodes in the SP system can continue to operate. Maintenance can be performed on the failed node without disruption to the other nodes.

- Service director

Service Director is now shipped with every SP system and automatically initiates a call to IBM when a system problem is detected.

For end users, a *service* is a single computational resource representing a single virtual server. Interactive Service Support (ISS) balances the workload across the various systems in the *service* to provide optimum productivity.

6.10.5 Disaster recovery

Until the advent of UNIX-based business solutions, there was little incentive to consider disaster recovery requirements. Only recently have disaster recovery solutions appeared for UNIX machines. Since UNIX evolved on inexpensive systems, the philosophy was that reliability could be achieved through redundancy. Adding more boxes was a relatively cheap way of guaranteeing continuous operations.

6.10.5.1 Outsourcing

Several IT vendors offer UNIX recovery solutions based on a duplication of the customer's data on a separate, geographically distant machine. IBM, for example, offers its Business Recovery Service for RS/6000 SP systems.

6.10.5.2 Backup site

A backup site solution generally involves setting up a WAN telecommunication line and using it to duplicate data (preferably in real time) between the two sites. Should the main site fail, the backup site would take over with minimal delay. This generally means that the customer's database has to be migrated to a backup site and updated in real time. Update transactions should be sent separately to both sites.

6.10.5.3 High Availability Geographic Cluster (HAGEO)

IBM HAGEO for AIX is a good example of this kind of solution. HAGEO provides a flexible, reliable solution for automatic real-time disaster recovery. HAGEO extends the loosely-coupled clustering technology of RS/6000 SP to encompass two physically separate data centers.

Each data center maintains an updated copy of essential data and runs key applications ensuring that mission critical computing resources remain continuously available at a geographically separate site should a planned (maintenance) or unplanned (disaster) event disable an entire site.

An HAGEO cluster consists of two geographically separated sites, each capable of supporting up to four high-availability cluster system nodes. There are four modes of disaster protection and one mode of recovery: Remote hot backup, remote mutual takeover, concurrent access, and remote system recovery. Chapter 8 provides more information on HAGEO.

6.10.5.4 High Availability Cluster Multi-Processing (HACMP)

IBM HACMP for AIX is a software product that allows customers to automatically detect system failures and recover users, applications, and data on backup systems, therefore, minimizing downtime to minutes or seconds. Using HACMP virtually eliminates planned outages as users, applications, and data can be made available using the backup systems in a cluster during scheduled system maintenance.

With features such as Cluster Single Point of Control (CSPOC) and Dynamic Reconfiguration, the systems administrator can add users, files, new hardwares, and other security functions without needing to halt mission critical resources when making changes to the cluster. Chapter 8 provides more information on HACMP.

Chapter 7. Workload management

Workload management encompasses the tools and concepts used to manage the distribution of workload across available resources. The goal of workload management is to assign resources to components of work in the most productive manner.

7.1 Introduction

In this chapter, we will consider two basic forms of workload management:

- Allocation of resources within individual servers
- Workload distribution across multiple servers

One form of server consolidation is the migration of multiple applications executing on multiple servers to a single server. In such a situation, the newly combined applications will compete for system resources, such as CPU time and memory. To assist system administrators in controlling server resource allocation, we will look at the *AIX Workload Manager*.

A consolidated environment may consist of multiple servers combined as a single unit to share the workload, such as a group of Web servers serving the same Web site. In this type of environment, we want to control the distribution of work across the servers to get the most from our investment. We will discuss two tools that assist in this form of load balancing, *Loadleveler*, and *SecureWay Network Dispatcher*.

7.2 AIX Workload Manager

AIX Workload Manager (WLM) is an operating system feature introduced in AIX Version 4.3.3. It is provided as an integrated part of the operating system kernel at no additional charge.

7.2.1 Overview

WLM is designed to give the system administrator greater control over how the scheduler and Virtual Memory Manager (VMM) allocate CPU and physical memory resources to processes. This can be used to prevent different classes of jobs from interfering with each other and to allocate resources based on the requirements of different groups of users.

The major use of WLM is expected to be for large SMP systems, typically used for server consolidation, where workloads from many different server

systems, such as print, database, general user, transaction processing systems, and so on, are combined. These workloads often compete for resources and have differing goals and service level agreements. At the same time, WLM can be used in uni-processor workstations to improve responsiveness of interactive work by reserving physical memory. WLM can also be used to manage individual SP nodes.

Another use of WLM is to provide a buffer between user communities with very different system behaviors. WLM can help prevent effective starvation of workloads with certain behaviors, such as interactive or low CPU usage jobs, from workloads with other behaviors, such as batch or high CPU usage.

WLM gives the system administrator the ability to create different classes of service and to specify attributes for those classes. The system administrator has the ability to classify jobs automatically to classes based upon the user, group, and/or pathname of the application.

WLM configuration is performed through a text editor, AIX commands, or through the AIX administration tools SMIT or a Web-based System Manager graphical user interface.

7.2.2 Concepts and configuration

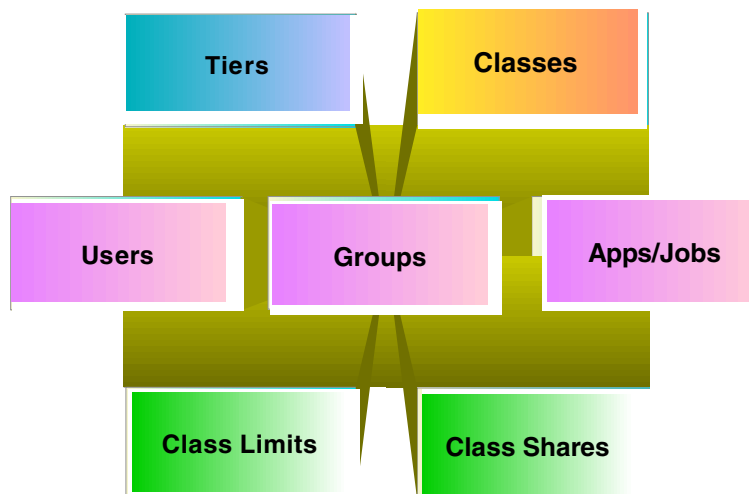


Figure 36. Basic WLM elements

WLM monitors and regulates the CPU utilization of threads and the physical memory consumption of processes active on the system. The manner in

which the resources are regulated is dependent on the WLM configuration defined by the system administrator.

There are a number of controlling variables in WLM that facilitate managing classes of jobs to achieve the automatic application of the resource entitlement policy you define. The primary concept to remember is that classes are what you manage in WLM, and that there are three job attributes available for process identification: Users, groups, and applications pathnames. Class resource shares and class resource limits allow you to define resource entitlements for each class, and tiers allow you to group the classes you are managing.

7.2.2.1 Job Attributes

Users

The user name (as specified in the password file) of the user owning a process can be used to determine the class to which the process belongs. Users can be excluded by using an exclamation point (!) prefix. Applications that use the `setuid` permission to change the effective user ID they run under are still classified according to the user that invoked them.

Groups

The group name (as specified in the group file) of a process can be used to determine the class to which the process belongs. Groups can be excluded by using an exclamation point (!) prefix. Applications that use the `setgid` permission to change the effective group ID they run under are still classified according to the group that invoked them.

Application pathname

The full pathname of the application for a process can be used to determine the class to which a process belongs. The full pathname can include pattern matching with the same style as in the Korn shell. Application pathnames can be excluded by using an exclamation point (!) prefix.

Combinations of user, group, and application pathname attributes may be used to determine the class.

7.2.2.2 Categories

Classes

Up to 27 classes can be defined by the system administrator. In addition, two classes and two pseudo-classes are automatically created as follows.

There are two pre-defined classes that cannot be removed.

- Default class

The default class is named *Default* and is always defined. All non-root processes that are not automatically assigned to a specific class will be assigned to the Default class. You can also assign other processes to the Default class.

- System class

This class, named *System*, will have all privileged (root) processes assigned to it, if not assigned by rules to a specific class, plus the pages belonging to all system memory segments kernel processes and kernel threads. You can also assign other processes to the System class. The default for this class is to have a memory minimum limit of 1 percent.

And, in the first release of WLM, there are two pseudo-classes named *Shared* and *Unclassified*. No classification rules, resource limits, or resource shares can be specified for these classes. Pseudo-classes are outside of WLM control and, therefore, fall under default AIX resource allocation control. Under default control, if the demands of unclassified processes are great enough, they could potentially starve processes under WLM management.

- Shared pseudo-class

This pseudo-class receives all memory that is used by processes in more than one class. This includes pages in shared memory regions and pages in files that are used by processes in more than one class. Shared memory and files that are used by multiple processes that all belong to a single class are associated with that class. It is only when a process from a different class accesses the shared memory region or file that the pages are placed in the Shared pseudo-class.

- Unclassified pseudo-class

This pseudo-class receives all processes that cannot be classified. Processes that have called the `plock` system call will be in the Unclassified pseudo-class. Note that in the first release of WLM, processes that are already in existence at the time that WLM is initialized will be in the Unclassified pseudo-class.

The attributes that must be defined for a class are:

- Class Name (up to 16 characters in length)
- Tier (zero is the default)
- Resource shares
- Resource limits

When an application is started, it will be classified into one of the defined classes based on the class assignment rules pertaining to the user, group, and/or application pathname.

Tiers

Tiers are based on the importance of a class relative to other classes in WLM. There are 10 available tiers from 0 through to 9. Tier value 0 is the most important; the value 9 is the least important. As a result, classes belonging to tier 0 will get resource allocation priority over classes in tier 1; classes in tier 1 will have priority over classes in tier 2, and so on. In the first release of WLM, tier resource enforcement is primarily directed at CPU utilization.

7.2.2.3 Resources

Class assignment rules

After a class has been defined, class assignment rules should be created. The assignment rules are used by WLM to assign a process to a class based on the user, group, application pathname, or a combination of these three attributes. The exclamation point (!) as the lead character represents an exclusion. The exclamation point (!) can be used with all attributes: User, group, and application pathname.

Note

In the first release of WLM, a process that is already running when WLM is initialized will not be classified and will be placed in the Unclassified pseudo-class.

Legend: dash = all, exclamation point = exclusion

| class name | user name | group name | application |
|------------|--------------|------------|-------------------|
| system | root | - | - |
| support | tech1, tech2 | - | - |
| marketing | - | - | /bin/analysis |
| skilled | - | webmasters | /bin/emacs |
| promoted | sally | staff | /bin/ksh, /bin/sh |
| games | !bob, !ted | !managers | /bin/solitaire |
| bad | hacker | - | - |

Figure 37. Example of classes and class assignment rules

The class assignment is done by WLM, when a process invokes the system call `exec`, by comparing the process attribute's user ID, group ID, and application file it is going to execute against the values specified in the rules file for these attributes. When doing so, WLM takes the rules in the order in which they appear in the assignment rules file and classifies the process in the class corresponding to the first rule for which a match is found.

There are two "default" rules that are always defined (that are "hardwired" in WLM). These are the default rules to assign all processes started by the user root to the System class and all the other processes to the Default classes. These are the only rules in the "standard" configuration installed with AIX. These rules can safely be omitted in the assignment rules file. In this case, the class assignment will work as if they were at the bottom of the file, with the rule for System class first.

In the above example, the rule for Default class is omitted. The rule for System class is explicit and has been put first in the file. This is done voluntarily so that all processes started by root will be assigned to the System class. The system administrator could have chosen to assign to System class only the root processes that would not be classified in another class, for example, because of the application executed, by putting System class farther down in the file. In the above example, with the rule for System class on top, if root executes `/bin/analysis`, the process will be classified as System class. If the rule for the System class were after the rule for the Marketing class, the same process would be classified as Marketing class.

These examples show that the order of the rules in the assignment rules file is very important. The more specific assignment rules should appear first in the rules file, and the more general rules should appear last. An extreme example would be putting the default assignment rule for the Default class, for which every process is a match, first in the rules file. This would have the effect of classifying every process in the Default class. The other rules would, in effect, be ignored.

You can define multiple assignment rules for any given class. You can also define your own specific assignment rules for the System and/or Default classes. The "default" rules for these classes would still be applied to processes that would not be classified using any of the explicit rules.

The way the WLM classification works has some consequences that can sometimes surprise system administrators:

- The classification is done based on *real* (not *effective*) user and/or group IDs. As a consequence, applications having the `setuid` or `setgid` bit set in the application's file permission bits will be classified according to the user

ID/group ID of the process, which calls `exec` (*real* IDs), and not according to the user ID/group ID of the file (*effective*).

- A process calling the `setuid/setgid` subroutines is not reclassified. This is sometimes misleading because a `ps` command will show a process with a user ID that would cause it to be assigned to a user defined class while the process is shown as belonging to the System class. This is often the case for processes spawned by daemons, such as `ftpd`, `telnetd`, and so on.
- The classification is done at the process level. This means that all the threads of a multi-threaded process will be in the same class.
- There is no way in the first release to assign different instances of the same application to different classes if they run with the same user ID and group ID.

Class resource shares

The number of shares of a resource for a class determine the proportion of a system resource that is allocated to the processes assigned to the class. In simple terms, the resource shares are specified as relative amounts of usage between different classes.

- The default share value is 1.
- The allowed share values for one resource for one class are integers from 1 to 65535.
- A class is active if it has at least one process assigned to it.

System resources are only allocated to a class with active processes; so, resource percentages are calculated based on the total number of shares requested by all active classes. The reason for using resource shares rather than percentages is that the desired amount is automatically recalculated when new resource shares are added (when a new class is created) or when resource shares are unused (because a class has no processes currently assigned to it). The calculation of entitlement is a percentage equal to the class resource shares divided by the total shares of all active classes in the same tier.

However, WLM makes sure that the calculated percentage goal for a resource remains compatible with the minimum and maximum limits for the resource. If the calculated percentage is below the minimum, WLM uses the minimum as the entitlement. If the percentage is above the maximum range, WLM uses the maximum as the entitlement. If the percentage is between the minimum and maximum, WLM uses the calculated percentage value (based on shares) as the entitlement.

Resource shares are specified in the resource shares file. The resource shares are listed by resource type within stanzas for each class.

The following example displays resource allocation before and after a new class is activated. Initially, there are three active classes that have been allocated: 5, 7, and 2 resource shares, respectively. These resource shares in combination are allocated 100 percent of the resource in accordance with their relative share values. When the new class, which has three resource shares, is activated, there are four active classes with resource shares of 5, 7, 2, and three with the total active resource shares equal to 17. As a result, when all four classes are active, the class with five resource shares will be allocated 5 of the total of 17 shares, or 29 percent of the system resource (29.4 percent will be rounded down to 29 percent).

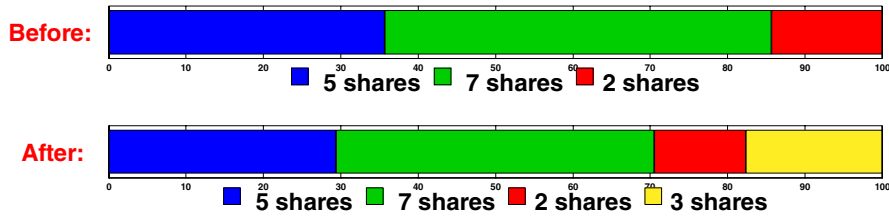


Figure 38. Example of share distribution automatically adjusting resources

Class resource limits

The class resource limits define the minimum and maximum amount of a resource (in the first release of WLM, for CPU and memory) that may be allocated to a class as a percentage of the total system resources.

Instead of having the minimum limit reserve a percentage of the resource for a given class, a higher resource allocation priority is given to processes in classes that are getting less than their minimum limit of the resource. This makes it more likely that these classes will get the resource if they try to use it. For CPU and memory, if a class is not using the resource at least up to its minimum, the rest is allocated to other classes. WLM cannot guarantee that processes actually reach their minimum limit. This depends on how the processes use their resources and on other limits that may be in effect. For example, a class may not be able to reach its minimum CPU limit because it cannot get enough memory. The possible values for the minimum limit are integers from 0 to 100. The default value is 0.

A maximum limit restricts the amount of resource that may be utilized by a given class. The possible values for the maximum limit are integers from 1 to 100. The default value is 100.

Resource limit values are specified in the resource limit file by resource type within stanzas for each class. The resource limits are specified as a minimum to maximum range, separated by a hyphen (-) with whitespace ignored. Each resource limit value is followed by a percent sign (%).

Class resource limits follow some basic rules:

- Resource limits take precedence over class share values.
- The minimum limit must be less than, or equal to, the maximum limit.
- The sum of all minimum limits for a resource for classes in the same tier cannot exceed 100 percent.
- For CPU, a class use of CPU cycles may exceed the maximum limit if the CPU cycles would otherwise be wasted (for example, given to the `waitproc`).

7.2.2.4 Resource handling for the Unclassified class

The Unclassified class' resources cannot be explicitly controlled. However, WLM does have to make choices about resources that are used by the Unclassified class.

For CPU usage, WLM does not adjust the scheduling priority of Unclassified work. This work has its scheduling priority modified by the scheduler based upon the CPU usage of each thread, exactly as happens when WLM is not invoked. CPU usage by the Unclassified class is not counted toward the total amount of CPU usage available to other classes.

For physical memory usage, the Unclassified class is treated the same as a class that is below its desired resource share value even though it contributes no shares to the total. Thus, pages will tend to be taken away from classes that are over their desired resource share value and given to classes that are below their desired resource share value and, thus, also to the unclassified work. Thus, Unclassified work is somewhat favored for memory usage relative to other classes.

7.2.2.5 Interaction with other scheduler control mechanisms

The `nice` command causes a process to have its CPU usage selectively favored or penalized with respect to other processes in the system. This effect will also work in a WLM environment. The `nice` command will cause a process to have its CPU usage selectively favored or penalized with respect

to other processes in the same class as the process. The `nice` command will not affect the CPU utilization of processes in other classes, because WLM will work to have the class' resources meet the requested number of resource shares and resource limits.

The CPU utilization of fixed priority processes is not managed by WLM. This is because setting fixed priority for a process (which can only be done by root) is designed to prevent the scheduler from changing the priority of a process. This is generally done because the process is so important that it must be fixed at a priority higher than the other processes in the system can achieve.

The `schedtune` command can be used to modify the behavior of the scheduler. All options to `schedtune` continue to work in a WLM environment. The use of `schedtune` options will not significantly impact the ability of WLM to manage CPU usage.

7.2.2.6 Interaction with other physical memory control mechanisms

The `vmtune` command can be used to modify the behavior of the Virtual Memory Manager (VMM). All options to `vmtune` continue to work in a WLM environment. Some of the options to `vmtune`, particularly `minperm` and `maxperm`, can hamper WLM's ability to achieve the specified physical memory usage goals.

7.2.3 Administration

WLM can be administered using three different methods:

- Web-based System Manager graphical user interface initiated with the AIX command: `wsm wlm`
- System Management Interface Tool (SMIT) initiated with the AIX command: `smit wlm`
- Command line and file editing

7.2.3.1 Property files

The Web-based System Manager and SMIT interfaces record the information in the same flat text files. These files are called the "WLM property files" and the files are named *classes*, *description*, *rules*, *limits*, and *shares*, respectively. The WLM property files can only be loaded by the root user.

Also, multiple sets of property files can be defined. These configurations are located in the subdirectories of `/etc/wlm`. A symbolic link, `/etc/wlm/current`, points to the directory containing the current configuration files. For example, the current running rules file is stored in a file `/etc/wlm/current/rules`. This link

is updated by the `wlmcntrl` command when WLM starts with a specified set of configuration files. The sample configuration files shipped with AIX are in the `/etc/wlm/standard` directory.

For example, you can create a subdirectory named `sample_config`. The file, `/etc/wlm/sample_config/description`, contains a character string describing the WLM configuration in subdirectory `sample_config`. This string appears in the WLM Manage Configurations menu in the Web-based System Manager. The only whitespace that is significant in these files is the carriage return. Begin comment lines with an asterisk.

Example /etc/wlm/sample_config/description file

My sample configuration

Example /etc/wlm/sample_config/classes file

Default:

```
description="The WLM default class"
tier = 0
```

System:

```
description="The WLM system class"
tier = 0
```

student:

```
description="The WLM student class"
tier = 1
```

Example /etc/wlm/sample_config/limits file

Default:

```
CPU = 0% - 100%
memory = 0% - 100%
```

System:

```
CPU = 10% - 100%
memory = 20% - 100%
```

student:

```
CPU = 10% - 100%
memory = 20% - 100%
```

Example /etc/wlm/sample_config/shares file

Default:

```
CPU = 20
memory = 20
```

System:

```
CPU = 20
memory = 20
```

student:

```
CPU = 10
memory = 20
```

Example /etc/wlm/sample_config/rules File

| * class | reserved | user | group | application |
|---------|----------|-------|---------|-------------|
| * | _____ | _____ | _____ | _____ |
| System | - | root | - | - |
| student | - | - | student | !/bin/ksh |
| Default | - | - | - | - |

7.2.3.2 Initiating AIX Workload Manager

By default, WLM is not enabled at system install and must be activated by the system administrator.

This may be performed from the command line with the command `wlmcntrl`; however, this only causes WLM to be initialized at that moment, not on every system boot. To configure the WLM to start automatically at system boot, initiate WLM using the SMIT or Web-based System Manager graphical user interface. If you use the SMIT, use the command `smit wlm`, then select **Start/Stop/Update WLM -> Start Workload Manager** and specify **Start Workload Manager** to be **Both**. WLM may also be stopped using these same menus.

Instead of using SMIT or Web-based System Manager, you can directly change the `/etc/inittab` file to start WLM at every system boot. To classify the maximum number of processes, the following line should be placed in the `/etc/inittab` file:

```
wlm:2:once:/usr/sbin/wlmcntrl > /dev/console 2>&1 #Start WLM
```

The `wlmcntrl` command does some very important processing of the WLM property files before passing the configuration information down to AIX. In particular:

- It converts all the user and group names into numerical user IDs and group IDs.
- It expands the wild cards (if applicable) in the application pathnames in the rules file and accesses all the target application files to transform the ID and inode number.

The `wlmcntrl` command will issue an error message and will not start WLM if it cannot translate a user or group name or cannot access an application file.

Thus, there are two main constraints to keep in mind when starting WLM:

- All processes that exist when WLM is started remained Unclassified (and, thus, are not controlled by WLM). This means that in order for WLM to be able to control as many of the processes as possible, it must be started very early in the boot process.

- On the other hand, WLM will not start if all of the application files cannot be accessed. This means that WLM cannot be started from the inittab before all the file systems on which the applications listed in the rules file reside are mounted. In case of remote file systems, such as NFS, this also means that TCP/IP and NFS are started.

These two constraints are somewhat contradictory, and there is no way for the system administration tools SMIT and WebSM to determine the optimal place for the line to start WLM in the inittab. These tools will just add the line at the end. It is recommended that system administrators edit the /etc/inittab file and move the WLM start command to best suit their system.

The options of the `wlmcntrl` command are:

- `-d Config_dir`: Use /etc/wlm/Config_dir as the directory to use for the classes, resource limits, resource shares, and rules files, and to make /etc/wlm/Config_dir the current configuration.
- `-u`: Update request to change the resource limits or resource shares of the running classes. The tier value for a class can also be changed in this way. Classes cannot be added or removed in this way.
- `-o`: Stops WLM
- `-q`: Query WLM state. It returns 0 if WLM is running, if not, it returns 1.

In the first stage, WLM allows system administrators to dynamically modify the shares, limits, and tier numbers using the `-u` option of WLM. It is also possible to use the `-d` option in conjunction with `-u`.

A system administrator has the option of modifying the shares, limits and/or tier numbers of the current configuration files and make the changes active using `"wlmcntrl -u"`. He/she also has the option to create a new configuration, with the same classes but different shares, limits, and/or tier numbers, and make this new configuration active by using `"wlmcntrl -u -d <new_config>"`.

This second option is interesting since it allows administrators to create different configurations with the same classes and different shares, limits, and/or tier numbers, for example, a `day_config` and a `night_config`, and flip from one to the other at given times using the AIX `cron` facility.

In the first stage, WLM also allows to update the assignment rules using the `-u` option (with or without `-d`), with the caveat that the processes that existed before the update will remain classified according to the "old" rules, and all the new processes will be classified using the new rules. There may be cases where this is perfectly acceptable.

For changes involving adding new classes or deleting existing classes, it is required in this first stage to stop and restart WLM. This has the inconvenience, already mentioned, that the processes that exist when WLM is restarted will be Unclassified. In practice, this kind of change will also require stopping and restarting the main applications. Hopefully, this should be a rare occurrence on production systems once the initial WLM set up has been completed.

7.2.3.3 Adding a Class *Web-based System Manager*

Initiate the Web-based System Manager graphical user interface with the command `wsm wlm`.

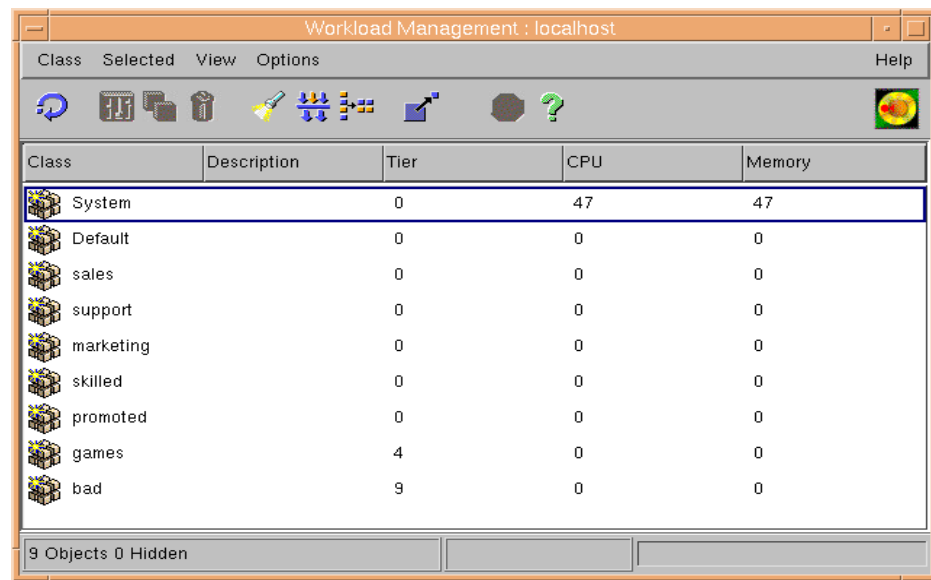


Figure 39. Web-based System Manager: Main menu for WLM

Select **New Class** from the Class menu of the top-level container. This displays a Create Class (or Class Properties) dialog:

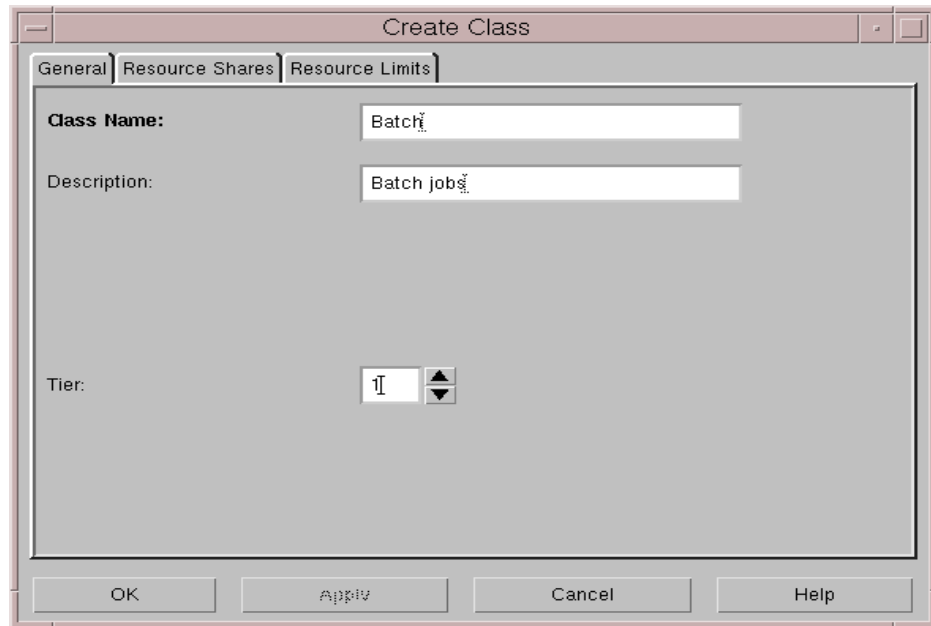


Figure 40. Web-based System Manager: Create Class dialog

Use the Create Class dialog to specify the name, description, and properties for a new WLM class. A similar dialog is used to edit or view properties of an existing class.

The next dialog illustrates how you specify resource shares and resource limits from a subdialog of the Create Class dialog.

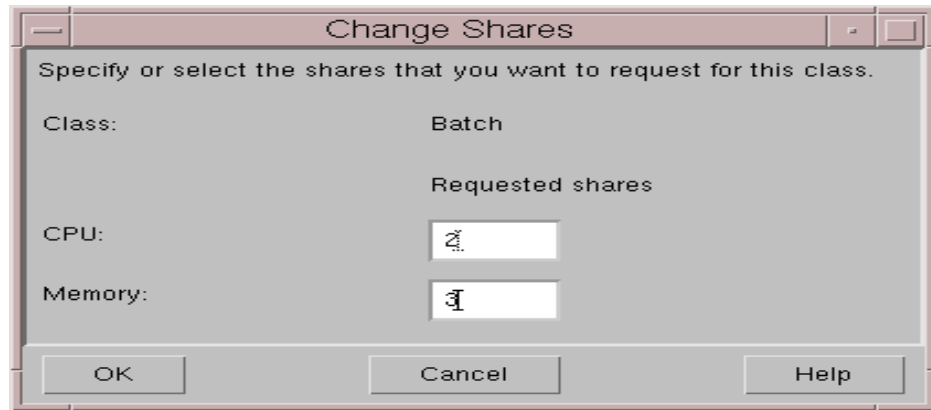


Figure 41. Web-based System Manager: Change Shares dialog

The Change Shares dialog is invoked from the Create Class dialog. Use it to specify the resource share values for memory and CPU. A similar dialog is used to specify minimum and maximum limits, if desired.

To edit the class assignment rules file, use the Class Assignment Rules dialog.

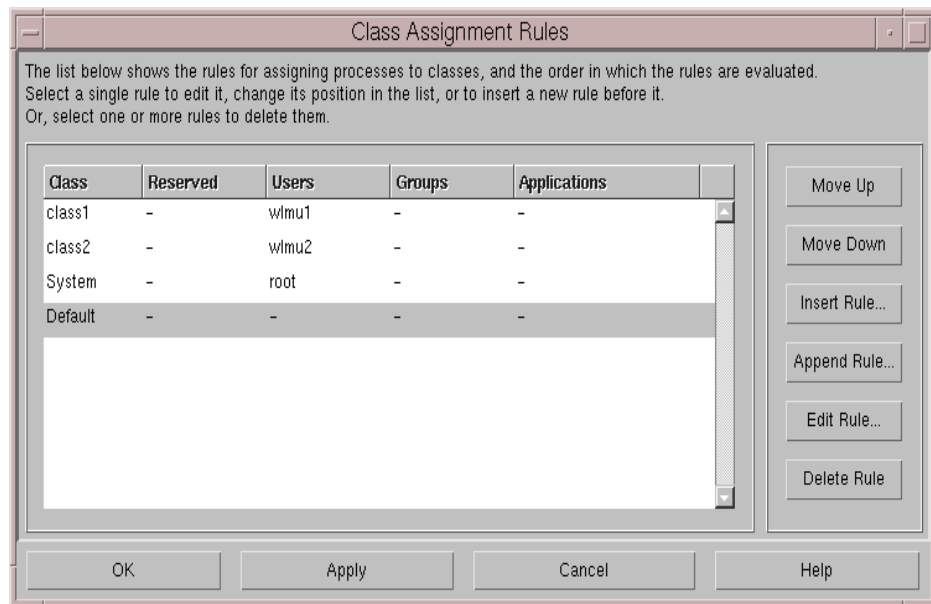


Figure 42. Web-based System Manager: Class Assignment Rules dialog

Use the Class Assignment Rules dialog to create new class assignment rules, edit or delete existing rules, or change the order of rules in the class assignment rules file.

Processes may be automatically classified based on several process attributes:

- User
- Group
- Application pathname

Selecting values for these attributes can be used to create assignment rules, which determine which class a process will be assigned to.

If you select **Edit** or **Append Rule**, the New Class Assignment Rule dialog is displayed.

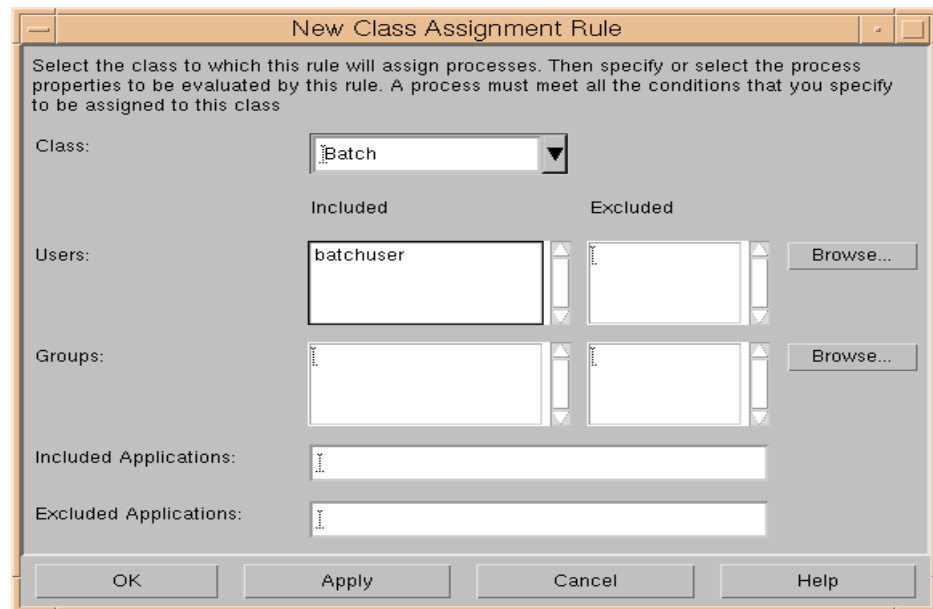


Figure 43. Web-based System Manager: New Class Assignment Rule dialog

Use the New Class Assignment Rule dialog to create new class assignment rules. A similar dialog is used to edit existing rules.

To work with WLM configurations other than the current configuration, select **Manage Configurations** from the Class menu to display the following dialog:

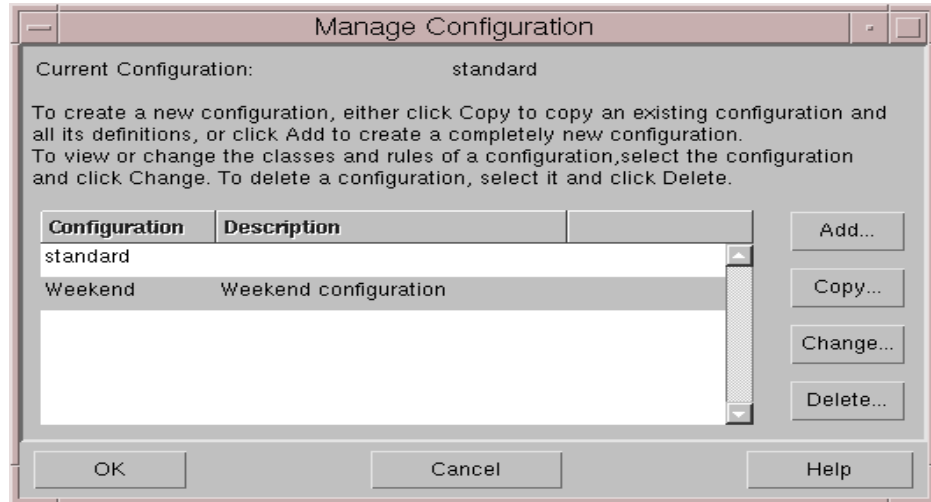


Figure 44. Web-based System Manager: Manage Configuration dialog

Use the Manage Configuration dialog to work with configurations other than the current configuration. From this dialog, you can create alternate configurations and work with classes and rules for those configurations without affecting the ongoing operation of your system.

Finally, to edit an alternate configuration, use the Change Configuration dialog:

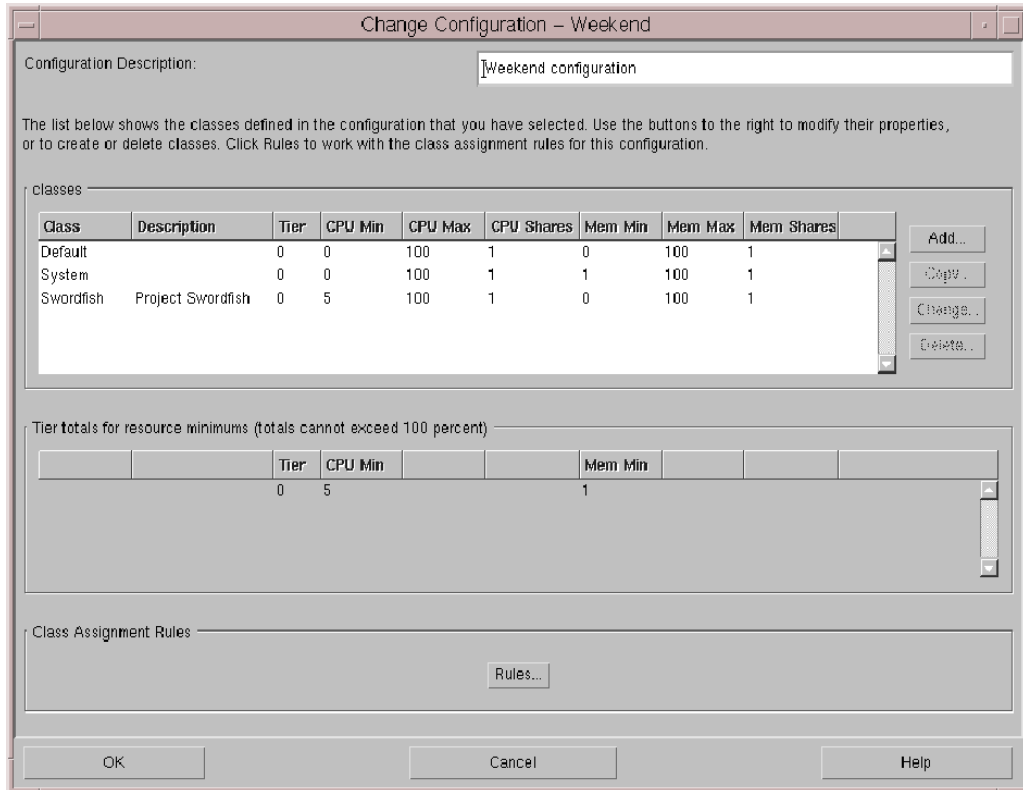


Figure 45. Web-based System Manager: Change Configuration dialog

Use the Change Configuration dialog to view or edit alternate WLM configurations. The first table lists all the defined classes for the selected configuration along with their tier values, resource shares, and resource limits. You can select a class to edit its properties. The second table shows the sum of minimum limits for classes within tiers. The sum of the minimum limits for classes in the same tier cannot exceed 100 percent. To work with the class assignment rules for this configuration, click the **Rules** button.

SMIT

Initiate the SMIT interface with the command `smit wlm`. In the SMIT menu, select **Add a Class**. In the menu presented, the class name must be defined and must not be the same as that of any other class. An optional description may be entered to explain the purpose of the class. The tier value can be defined here. This is also where you define the resource limits and shares for CPU and memory usage.

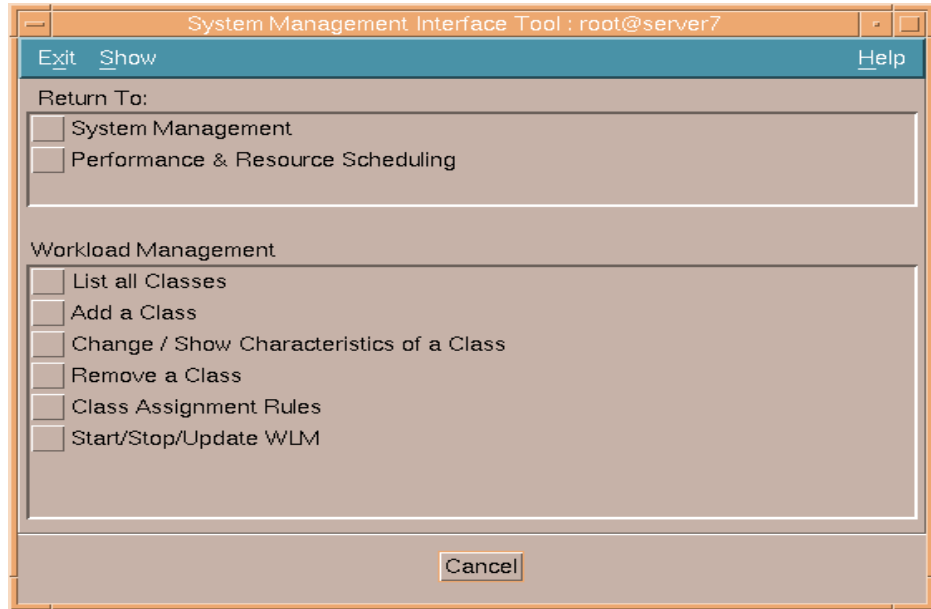


Figure 46. SMIT: Main menu for WLM

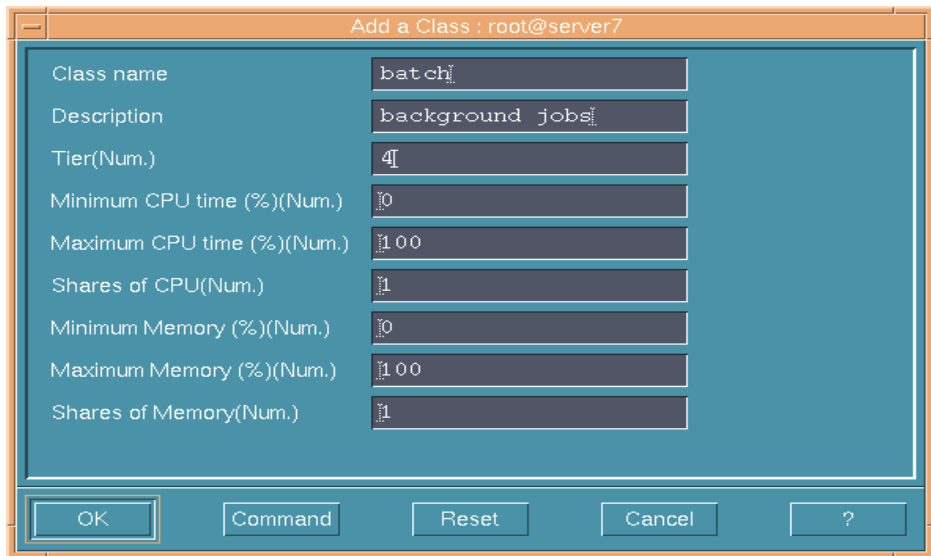


Figure 47. SMIT: Adding a class

To define an assignment rule for a class that has been created, initiate SMIT with `smit wlm`, select **Class Assignment Rules** and **Create New Rule** to

define a new class rule, or **Change / Show Characteristics of a Rule** to modify an existing class rule.

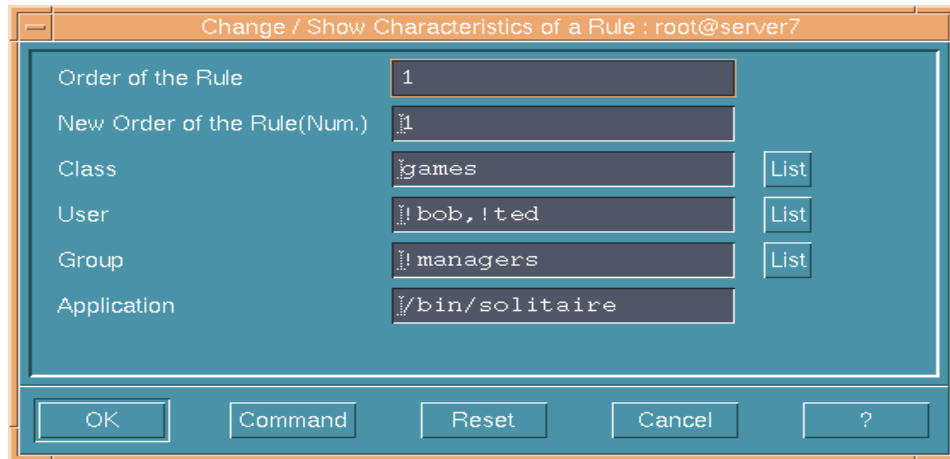


Figure 48. SMIT: Assign process attribute values

In the dialog provided, specify the order of the rule (this determines which rules will be assessed first), and the name of the class for which the rule is to apply. Also, set the criteria that will classify a process as a member of the class (this may be an AIX user list, group list, application pathname, or combination of the three). In the user, group, and application pathname lists, the exclamation point (!) as the lead character represents an exclusion (for example, "!bob" means "exclude bob"). Multiple process attribute values may be excluded.

7.2.3.4 Monitor AIX Workload Manager

wlmstat

To monitor the current status of WLM and resources, use the command `wlmstat`. It shows the amount of resources currently being used by each class. The syntax is:

```
wlmstat [-l class] [-c | -m] [interval] [count]
```

The options are:

- -l class: Show statistics for the specified class name. If not specified, all classes are displayed along with a summary for appropriate fields.
- -c: Show only CPU statistics.
- -m: Show only memory statistics.

- interval: Specifies an interval in seconds (defaults to 1).
- count: Specifies how many times `wlmstat` will print a report (defaults to 1).

If a count is specified, `wlmstat` displays the statistics that many times and sleeps the specified number of seconds after set of statistics is displayed. A summary is also displayed after each set of statistics.

ps

The `ps` command may be used to view the class of each process currently in the system. The `ps` command is useful in verifying that processes actually belong to the classes intended.

To view just the processes in a single class or set of classes, use the "`-c classlist`" option with the `ps` command. This option takes a list of class names as a parameter. The list of class names must be separated with commas only (no blanks).

Use the `ps` command with the `-o class` option displays the current class assignment for each process. The normal output from the `ps` command is unchanged without this option. For example, this `ps` command will provide the following output:

```
ps -ae -o pid,user,class,pcpu,thcount,vsz,wchan,args
```

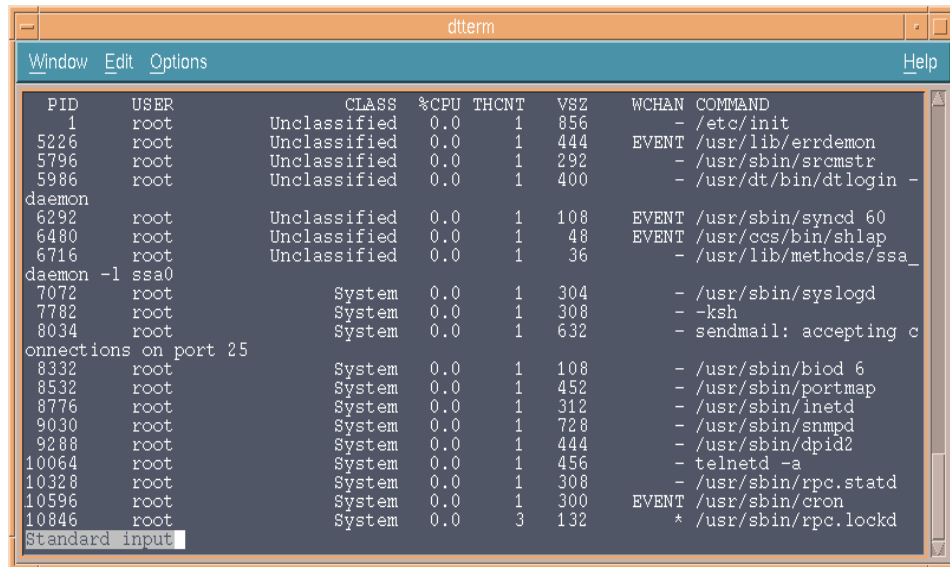


Figure 49. Output of the `ps` command demonstrating a class column printed

svmon

`svmon` is one of the tools in the Performance Toolbox. This tool generates snapshots of a system's virtual memory. This tool has been enhanced with usability, scalability, and speed improvements on the largest enterprise server systems. In addition, it has been enhanced to generate reports on users, commands, and WLM classes to support WLM functions.

7.2.4 Benefits

WLM allows application separation for server consolidation.

WLM can be used to cause applications to remain in memory for more predictable performance results.

When properly used, this tool will allow a system administrator far greater control of resource allocation within an RS/6000 server. The result being better distribution of resources across the workload and more resources available for critical applications. This is particularly beneficial in consolidated environments where multiple applications have been combined on the same server.

Note

Because WLM is a part of the kernel and influences thread priority, WLM is a very powerful capability. If WLM is incorrectly configured, it may result in degraded performance of important processes and applications. To help achieve the results you desire, be very careful in all phases of WLM use from application behavior analysis to workload anticipation, planning, and ongoing use monitoring.

7.2.5 Hints and Tips

The following points represent a compilation of hints and tips to assist in configuring and managing the WLM.

- Always study and anticipate the behaviors of your applications before beginning to use WLM.
- When configuring WLM, know your users and applications. It is important to understand the user base and their computing needs. It is also important to have an understanding of the resources required by all applications in the system.
- Before starting WLM for the first time, make sure you have appropriately set up the resource share values for the Default and System classes. The

default setup for WLM equally divides resources between the Default and System classes.

- Keep it very simple at first; start with just a few application, then build.
- The overhead of WLM will increase based on the number and complexity of the rules.
- Use resource shares rather than limits to start with. Resource shares are goals given to WLM to achieve, and this allows greater system flexibility than hard limits. If the resource shares set by an system administrator are not optimal, the system should still be able to balance the load reasonably well. With hard limits, WLM can do little to prevent applications being starved of resources. For example, if the maximum memory limit is set smaller than the average working set of the application, the application may incur significant performance degradation. In summary, it is better to wait to assign limits until after experience is gained with the results from setting resource shares.
- When setting resource limits, start by setting just the minimum.
- When configuring WLM on a server, follow the steps below:
 1. Balance the load using only shares and monitor WLM and the system for a reasonable period of time to assess application performance and tune these values if necessary.
 2. Set minimum limits for the applications that do not appear to be allocated their share of resources.
 3. Prioritize workloads using tiers, if necessary, to promote a ranking among jobs. For greater tier impact, increase the separation of tiers. For example, the impact of a tier 1 and 7 separation will be greater than the impact of a tier 1 and 4 separation.
 4. Set maximum limits only if absolutely necessary to control poorly behaving applications.

Note

WLM configuration should be tested in a non-production environment to avoid possible disruption to users and applications.

If any undesirable behavior occurs when WLM is running, WLM can be stopped by using the command `wlmcntl -o`. Stopping WLM will turn off all WLM management of resources, and the system behavior will quickly return to the normal state.

- Start WLM as early in the system startup as possible but not before mounting all file systems. When WLM is started, the application files in all rules are examined. In the first release of WLM, WLM will not start if any of these files cannot be accessed. This can happen, for instance, if the file resides on an unmounted file system or if the permissions of any component of the path name does not grant root access.
- When a class' memory working set is larger than its maximum limit for memory, performance of processes in the class may significantly degrade. It is suggested that memory minimums for *other* classes be used before imposing a memory maximum for any class.
- When creating class assignment rules, the order number of the rule must be considered. This number is set relative to the order of all other rules. When assigning a process to a class, the class rules will be examined by WLM, and the first rule to match the process will determine the class assignment (starting at rule 1). If a process matches more than one assignment rule, only the first rule will be used to classify the process. The order should be determined by whether the rule has exceptions or not. If a rule has no exceptions, it should go earlier in the list.
- Tiers, resource limits, and resource shares can be modified while WLM is running. While in the first release of WLM, adding or removing classes requires stopping and restarting WLM.
- If a tailored management policy is required for different times of day or different days, define several WLM configurations and potentially activate different configurations at different times of the days of the week, if you need. The process for accomplishing this in the first release of WLM is to use `cron` and the root crontab to load different configurations at the appropriate dates and times.

7.2.6 Examples

The following are examples of WLM designed to demonstrate the behavior of this utility.

7.2.6.1 Example 1

There are two classes, *alpha* and *beta*, with the following attributes:

| Class | Tier | Min CPU | Max CPU | CPU shares | Min Memory | Max Memory | Memory Shares |
|-------|------|---------|---------|------------|------------|------------|---------------|
| alpha | 0 | 0 % | 100 % | 3 | 0 % | 100 % | 1 |

| Class | Tier | Min CPU | Max CPU | CPU shares | Min Memory | Max Memory | Memory Shares |
|-------|------|---------|---------|------------|------------|------------|---------------|
| beta | 0 | 0 % | 100 % | 2 | 0 % | 100 % | 1 |

Here, the system administrator has only imposed shares on CPU usage while all other attributes of the classes have been left at the default values. As a result, WLM will not force hard restrictions on system resources. It will attempt to allocate 60 percent of CPU resources to class *alpha* and 40 percent to class *beta* when both classes are active.

Initially, only class *alpha* is active. Assuming there are no other application constraints, such as memory, class *alpha* will acquire 100 percent of the CPU.

If class *beta* becomes active, class *alpha* CPU utilization will progressively decrease to 60 percent while class *beta* CPU utilization will increase from 0 to 40 percent. The system will stabilize with the two classes at 60 percent and 40 percent CPU utilization, respectively, within a matter of seconds.

Note

This is a very simplified example and assumes resources are not required by processes outside these two classes. On a real running system, there may be other processes in the *Default* and *System* classes, or in the *Unclassified* or *Shared* pseudo-classes.

7.2.6.2 Example 2

Basically, this example is similar to Example 1. The test machine is a 64 GB, 24-way Symmetric Multiprocessor (SMP) machine. In this example, three classes are defined: class1, class2, and class3. These classes are loaded with CPU-bound processes with little or no memory consumption. CPU maximum limits are not defined. Class1 has one CPU share, class2 has two CPU shares, and class3 has three CPU shares.

WLM is started, and processes are started in class1, with class2 and class3 being inactive. Then processes in class2 are started, with class3 remaining inactive, and, finally, processes are started in class3. After some time, all the processes are stopped in reverse order: First class3 and then class2. In each phase, the classes quickly stabilize at or near the appropriate percent usage (which varies according to the number of active classes and their number of shares).

The following graph shows the CPU usage results.

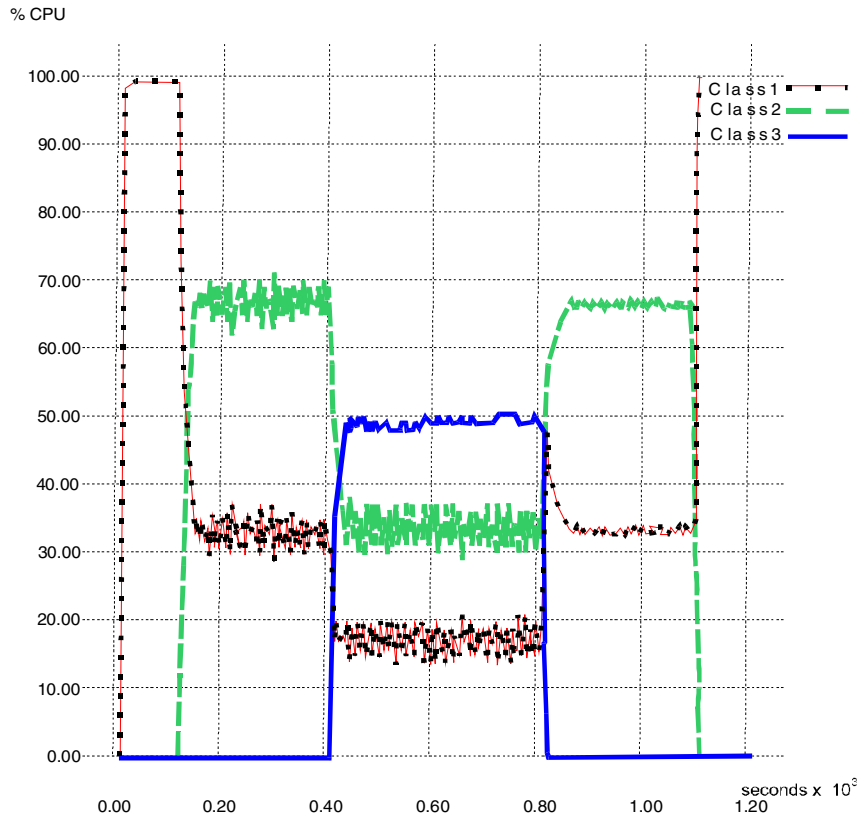


Figure 50. CPU usage results of example 2

7.2.6.3 Example 3

This example is an extension of example 2. In this example, WLM is started with one CPU share each for class1, class2, and class3. After some time, the CPU shares are changed to one, two, and three, respectively, using the `wlmcntrl -u` command. The CPU consumption of each class quickly stabilizes around its new CPU shares. There are six total shares, and class 1 gets 1/6, class 2 gets 2/6 (or 1/3), and class 3 gets 3/6 (or 1/2).

The following graph shows the CPU usage results with a SMP machine.

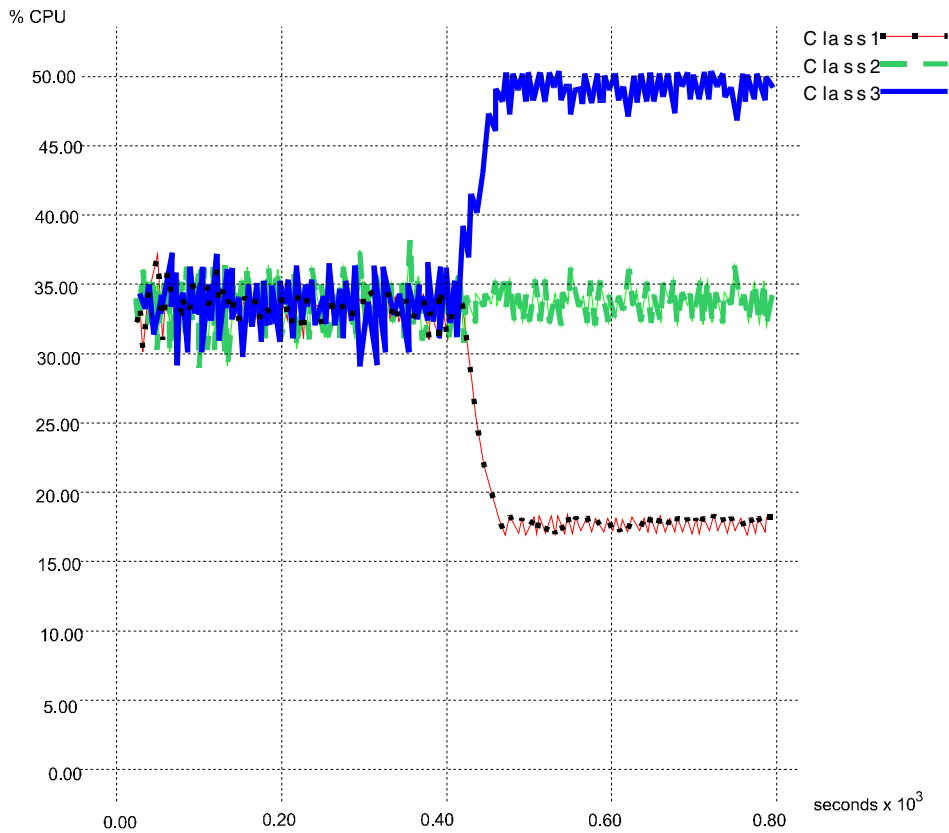


Figure 51. CPU usage results of example 3

7.2.6.4 Example 4

There are three classes defined, *App1*, *Batch*, and *Night*.

| Class | Tier | Min CPU | Max CPU | CPU shares | Min Memory | Max Memory | Memory Shares |
|-------|------|---------|---------|------------|------------|------------|---------------|
| App1 | 1 | 0 % | 100 % | 1 | 20 % | 100 % | 50 |
| Batch | 3 | 0 % | 100 % | 1 | 50 % | 100 % | 50 |
| Night | 2 | 0 % | 100 % | 15 | 40 % | 100 % | 50 |

This system runs an application, the processes of which are assigned to the *App1* class. The application is in use 24 hours a day. It has been discovered through performance monitoring that for users to receive an acceptable

response time, the application requires a minimum of 20 percent of the system memory.

Processes classified in the *Batch* class represent background jobs that are continuously running day and night. The processes in this class are of the lowest priority with respect to resource satisfaction and have, as such, been placed in a higher-numbered tier.

The *Night* class represents a nightly job of relatively high importance that runs each night and must complete as quickly as reasonably possible. Because of the importance of *Night* processes completing, it has been assigned a lower-numbered tier than regular *Batch* processes and is allocated a minimum memory resource of 40 percent. This will ensure that during the evening process it will be given a higher priority and more memory than the less critical *Batch* processes. To increase the importance of *Night* over *Batch*, move *Batch* to higher tier numbers, for example, from tier 3 to tier 5.

During daytime processing, the *App1* and *Batch* classes will have a total memory minimum limit of 70 percent (at this time, *Night* processes will not be active). As a result, both classes will be able to acquire their reserved memory. When the *Night* class becomes active, the total minimum memory required by all classes totals 110 percent (actually, the total is 111 percent because the System class has a memory minimum of one percent). WLM will ignore minimum requirements for lowest tiers first; hence, *Night* will steal memory from the *Batch* class' 50 percent reserve. Because *App1* is of a higher tier, *Night* is unable to take from the *App1* memory reserve. The *Batch* class processes will be starved of memory during the processing of the *Night* class processes. Once *Night* has completed, the memory allocation to *Batch* will return to normal. During this time, the memory requirements of *App1* will always be honored, ensuring users experience desired response times regardless of the background processing taking place.

The following is an actual test of the above situation in the first release of WLM. The applications in all three classes are represented by a test program attempting to consume a specified amount of memory and CPU time (percentages passed as arguments). WLM is started with one copy of the test program in class *App1* to consume 25 percent of CPU time and 25 percent of memory and one copy of the test program in class *Batch* to consume 50 percent of CPU time and 55 percent of memory. This simulates the load during the day. Everything is fine since the system is not overcommitted, and both classes get what they need (CPU and memory). When everything has stabilized, another copy of the test program in class *Night* is started, and it is instructed to request 50 percent of CPU time and 60 percent of memory.

When the job in *Night* starts (about 500 seconds into the test), *App1* has its requested 25 percent and *Batch* has its requested 55 percent. At this point, there is still memory available, and *Night* starts to ramp up its memory usage without affecting the other two classes. At about 600 seconds, the system runs out of free memory. From then on, *Night* steals pages from *Batch* (in a lower-priority tier) while *App1* (in a higher-priority tier) is pretty much unaffected. Ultimately, *Night* will satisfy its memory requirements (60 percent) at the expense of the lower-priority tier while *Batch* gets whatever memory is left. The stabilization of memory usage is progressive since it takes time to page out the stolen pages. The graph below shows the memory usage results.

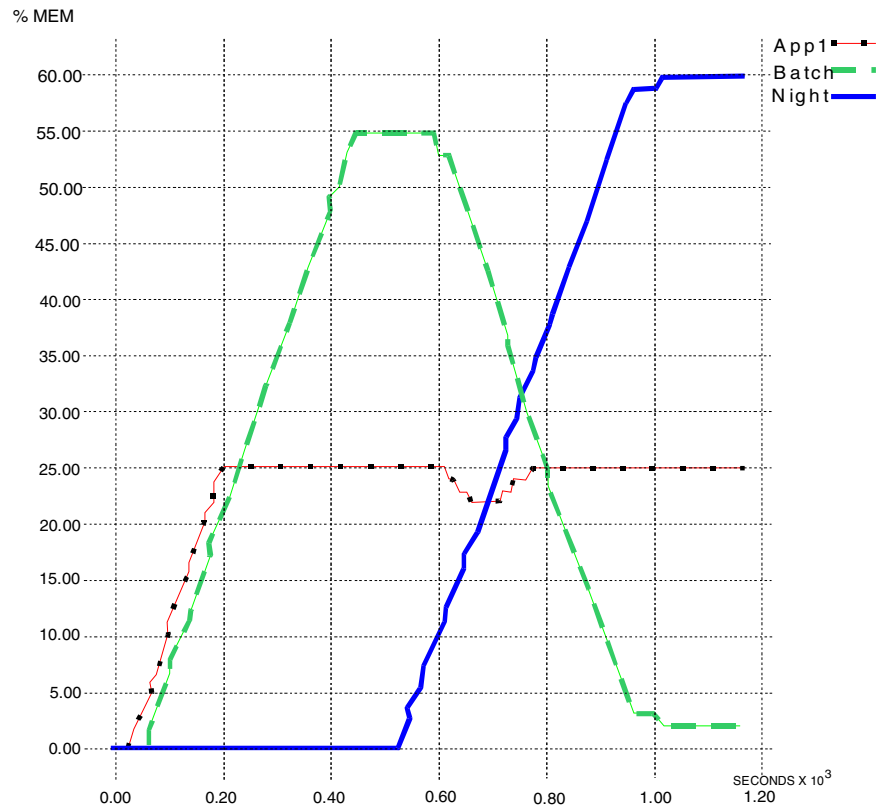


Figure 52. Memory usage results of example 4

7.3 LoadLeveler

LoadLeveler is an IBM software product designed to automate workload management.

7.3.1 Overview

Essentially, LoadLeveler is a scheduler that also has facilities to build, submit, and manage jobs. The jobs can be processed by any one of a number of machines, which together are referred to as the LoadLeveler cluster. Any stand-alone RS/6000 may be part of a cluster although LoadLeveler is most often run in the RS/6000 SP environment. The proper use of Loadleveler enables a server administrator to automatically distribute the workload over multiple servers, thus, making best use of available resources.

There are a number of basic terms to understand with respect to Loadleveler.

- Cluster
A group of machines that are able to run LoadLeveler jobs. Each member of the cluster has the LoadLeveler software installed.
- Job
A unit of execution processed by Loadleveler. A serial job runs on a single machine. A parallel job is run on several machines simultaneously and must be written using a parallel language Application Programming Interface (API). As LoadLeveler processes a job, the job moves in to various job states, such as *Pending*, *Running* and *Completed*.
- Job Command File
A formal description of a job written using LoadLeveler statements and variables. The command file is submitted to LoadLeveler for scheduling of the job.
- Job Step
A job command file specifies one or more executable programs to be run. The executable and the conditions under which it is run are defined in a single job step. The job step consists of several LoadLeveler command statements.

7.3.2 Concepts and configuration

There are three important functional machine types in LoadLeveler.

- Scheduling server

When a job is submitted to LoadLeveler, it gets placed in a queue, which is managed by the scheduling server. This server then submits a request to the central manager to find a machine in the cluster that can process the job.

- Central manager server

This server evaluates the resources required by the job specified in the job command file and selects a machine capable of running it. The central manager is also called the negotiator.

- Executing server

Machines that are assigned and execute jobs.

When a job is submitted in a Loadleveler cluster, there are four basic steps to be performed.

1. A job is submitted to the LoadLeveler scheduling server.
2. The scheduling server contacts the central manager to inform it that a job has been submitted and to determine if there is a server available capable of fulfilling the job requirements.
3. Once an executing server with the required capacity is found, the central manager informs the scheduling server which machine is available.
4. The scheduling server contacts the executing server and submits the job information and executable program. The executing server sends job status information to the scheduling server and notifies it when the job has completed.

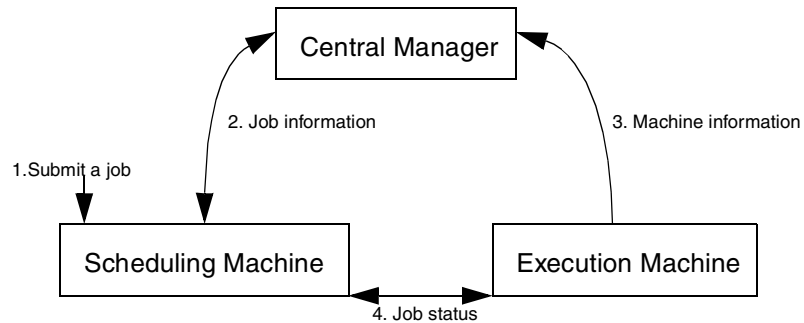


Figure 53. Loadleveler job status flow

There is an additional type of server known as a submit-only server. This type of machine may only submit jobs although it is also able to query and cancel them.

The negotiator calculates a priority value for each job's System Priority (SYSPRIO) that determines when the job will run. Jobs with a high SYSPRIO value will run before those with a low value. There are several different parameters used to calculate SYSPRIO.

- How many other jobs does the user already have running.
- When the job was submitted.
- What is the priority the user has assigned to it.

The priority of a job in the LoadLeveler queue is completely separate and must be distinguished from the AIX `nice` value, which is the priority of the process the executable program is given by AIX.

LoadLeveler also supports the concept of job classes. These are defined by the system administrator and are used to classify particular types of jobs. For example, we can define two classes of jobs that run in the clusters called *night* jobs and *day* jobs. We might specify that *executing machine A*, which is very busy during the day because it supports a lot of interactive users, should only run jobs in the night class. However, *machine B*, which has a low workload in the day, could run both. LoadLeveler can be configured to also take job class into account when it calculates SYSPRIO for a job.

As SYSPRIO is used for prioritizing jobs, LoadLeveler also has a way of prioritizing the executing machines. It calculates a value called MACHPRIO for each machine in the cluster. The system administrator can specify several different parameters that are used to calculate MACHPRIO.

- Load average
- Number of CPUs
- The relative speed of the machine
- Free disk space
- The amount of memory

Machines may be classified by LoadLeveler into pools. Machines with similar resources, for example, a fast CPU, might be grouped together in the same pool so that they could be allocated CPU-intensive jobs. A job can specify as one of its requirements that it run on a particular pool of machines. In this way, the right machines can be allocated the right jobs.

7.3.3 Why use LoadLeveler?

In the computing environments of today, parallel job assignment is becoming increasingly important. Larger and more complex job processing across multiple servers is becoming more common place, thus, generating a need for multiple server job control.

LoadLeveler allows a group of servers to function as a single job processing system. Through the monitoring of servers in such a group, Loadleveler facilitates the best assignment of jobs to the most appropriate system. This gives the ability to dynamically assign jobs to the most appropriate server based on job requirements, such as memory size, machine architecture, programs, classes of service, and available resources.

In an environment of multiple servers sharing job processing, LoadLeveler provides the means to make the optimum use of available resources, hence, maximizing investment and performance.

7.3.4 Benefits

LoadLeveler allows users to run more jobs in less time by matching their processing needs to available resources. LoadLeveler distributes workload across multiple servers, maximizing on server investment.

This presents numerous benefits to the consolidated environment.

- Scalable job management from the RS/6000 SP system down to workstations on end-user desktops.
- Administration of machines and jobs can be accomplished from a single point of control.
- Job requirements, such as memory size, machine architecture, programs, and classes of service are all easily specified and matched with available resources.
- LoadLeveler's Central Manager can automatically recover on other designated machines with no loss of job information.

7.4 SecureWay Network Dispatcher

SecureWay Network Dispatcher is load distribution software from IBM.

SecureWay Network Dispatcher is a scalable, highly-available load-balancing software solution for HTTP, FTP or other TCP-based servers. It balances the load of Internet servers on a variety of operating systems, such as AIX, Windows NT, and Sun Solaris.

7.4.1 Overview

SecureWay Network Dispatcher boosts the performance of servers by routing TCP/IP session requests to different servers, thereby, balancing the client requests to systems in the environment. This routing is transparent to users and other applications, such as e-mail servers, World Wide Web servers, distributed parallel database queries, and other TCP/IP applications.

IBM SecureWay Network Dispatcher provides customers with advanced functions to meet their site's scalability and availability needs. It consists of three components:

- Interactive Session Support (ISS) provides the same DNS interface as before for your clients. It provides a least-disruptive migration path for applications that are already deployed using Round Robin DNS.
- Dispatcher provides an advanced IP level load balancing mechanism that you install instead of Round Robin DNS. Once installed, the Dispatcher remains completely invisible to clients but can deliver superior load balancing, management, and availability function.
- Content-based Routing (CBR) provides full-function load balancing based on information in the HTTP data stream, such as URLs, paths, cookies, and so on.

These three components can be deployed separately or together in various configurations to suit a wide variety of customer application requirements.

7.4.2 Interactive Session Support

Interactive Session Support (ISS) is the DNS-based component. It provides a load-monitoring daemon that can be installed on each of the servers that form part of your installation. This group of daemons is referred to as a cell. One of the members of the cell becomes the "spokesman" for the load-monitoring service. You can use standard memory or processor utilization figures to measure load with the ISS daemon. Alternatively, you can provide your own set of criteria, for example, with an application-aware executable module or a simple shell-script, and ISS can be configured to use it. This is referred to as a custom metric.

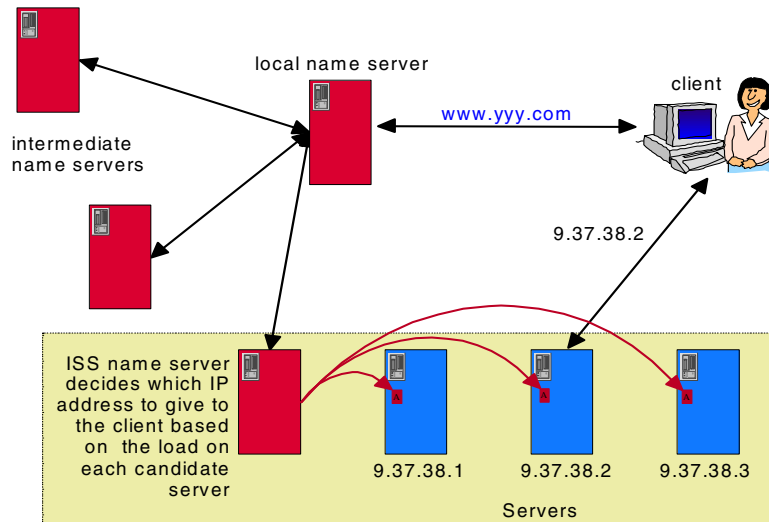


Figure 54. Interactive Session Support (ISS)

ISS provides an "observer" interface to enable other applications to use the load-monitoring service. Observers watch the cell and initiate actions based on the load. The observer applications provided in SecureWay Network Dispatcher are of two kinds: Name server observer and dispatcher observer. As shown in Figure 54, the name server observer allows the load monitoring service to provide an intelligent DNS name resolution service that returns selected IP addresses to your clients in response to a `gethostbyname()` call based on feedback from the ISS daemons running on each server machine rather than the strict rotation of round-robin. Also, if a server is down, its IP address will not be returned.

The dispatcher observer is used in conjunction with the Dispatcher component. ISS can also be configured in "ping triangulation" mode where the ISS daemons respond based on the ping time from each of them to the client. This allows DNS-based load balancing based on network topology.

7.4.3 Dispatcher

The Dispatcher is an IP-level load balancer. It uses a fundamentally different approach to load balancing based on patented technology from IBM's Research Division. Dispatcher does not use DNS in any way although normal static DNS will still usually be used in front of the Dispatcher. Once installed and configured, the Dispatcher actually becomes the site IP address to which your clients send all packets. This externally advertised address is referred to

as the cluster address. You can define as many cluster addresses as you need. You then define the ports you want to support inside each cluster and then the actual servers that will provide the service on each of those ports. If you want, you can also conceal from your clients the real IP addresses of the servers in the cluster by filtering them at the gateway router. This object-oriented cluster-port-server structure provides a simple configuration interface that can be created and modified dynamically, thus, permitting true 24 x 7 operation.

The core function of the Dispatcher is the Executor. It is a kernel-level function (or device driver) that examines only the header of each packet and decides whether the packet belongs to an existing connection or represents a new connection request. It uses a simple connection table stored in memory to achieve this. Note that the connection is never actually set up on the Dispatcher machine, it is between the client and the server, just as it would be if the Dispatcher were not installed, but the connection table records its existence and the address of the server to which the connection was sent. If the connection already exists, which means it has an existing entry in the in-memory connection table, then, without further processing, the packet is rapidly forwarded to the same server chosen on the initial connection request. Since most of the packets that flow are of this type, the overhead of the whole load balancing process is kept to a minimum. This is one of the reasons why the Dispatcher is so superior to its competition in performance and scalability. If the packet is a new connection request, the Executor will look at the configuration to see which servers can support a request on the port requested by the client on the requested cluster address. Then it uses stored weights for each such server to determine the “right” server to which the connection will be forwarded. An entry mentioning this server is made in the connection table, therefore, ensuring that subsequent packets for this connection are correctly forwarded to the chosen server.

Note that the “right” server is not always the “best” server since it is desirable for all eligible servers to process their share of the load. Even the “worst” server needs to shoulder some of the burden. If traffic is only ever forwarded to the best server, it can be guaranteed that it will rapidly cease to be the best, and the load on the servers will swing backwards and forwards. Dispatcher's patented algorithm for choosing the right server and its advanced smoothing techniques achieve optimal balance in the shortest possible time and maintain that balance.

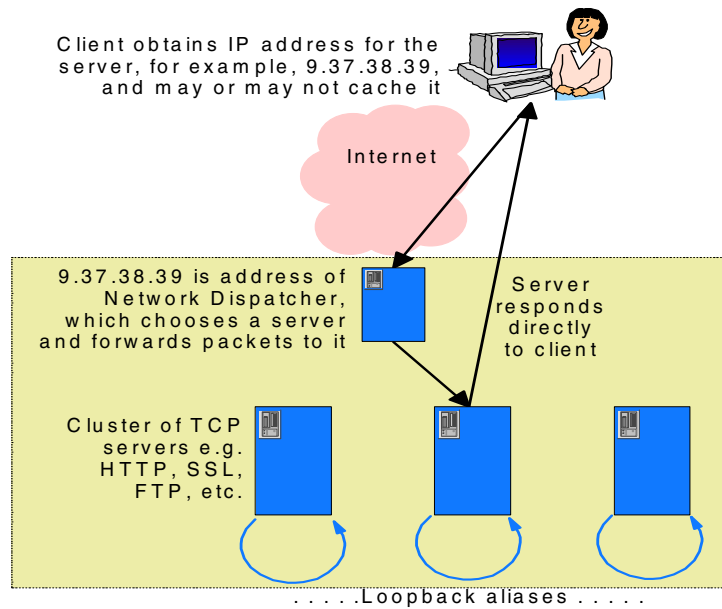


Figure 55. Dispatcher

An overview of Dispatcher packet flow is shown in Figure 55. The Executor does not modify the client's IP packet when forwarding it. Because the Dispatcher is on the same subnet as its clustered servers, it simply forwards the packet explicitly to the IP address of the chosen server, just like any ordinary IP packet. The Dispatcher's TCP/IP stack re-addresses only the MAC frame containing the packet, in the operating system approved manner, and sends it to the chosen server. To allow the TCP/IP stack on that server to accept the unmodified packet from the Dispatcher and pass it to the chosen port for normal application processing, the IP address of the Dispatcher machine is also installed as a non-advertising alias on each of the clustered servers. This is achieved by configuring the alias on the loopback adapter.

The server's TCP then establishes the server-to-client half of the connection according to standard TCP semantics by simply swapping the source and target addresses as supplied by the client rather than determining them from its own basic configuration. This means that it replies to the client with the cluster IP address (that is, the IP address of the Dispatcher). As a direct result, the balancing function is invisible both to the client and the clustered servers. This invisibility means that the Dispatcher is not dependent upon server platforms, provided they implement standard TCP/IP protocols.

Another key performance and scalability benefit to the Dispatcher customer is that the application server returns the response to the client's request directly to the client without passing back through the Dispatcher. Indeed, there is no need even to return using the original physical path; a separate high-bandwidth connection can be used. In many cases, the volume of outbound server-to-client traffic is substantially greater than the inbound traffic. For example, Web page HTML and imbedded images sent from the server are typically at least 10 times the size of the client URLs that request them. Because the Dispatcher is a truly generic TCP/IP application, its functions can be applied not only to HTTP or FTP traffic but also to other standards-compliant types of TCP and UDP traffic.

7.4.3.1 Dispatcher load balancing with weights

The Dispatcher has a Manager function, which sets the weights that the Executor obeys. As shown in Figure 56 on page 224, the Manager can use four metrics to set these weights:

- Active connection counts for each server on each port. This count is held inside the Executor.
- New connection counts for each server on each port. This count is held inside the Executor.
- A check that each server is up-and-running. The Advisor performs this function.
- A check that each server has "displaceable capacity," which means that it can realistically process the work. ISS performs this function when used in the Dispatcher Observer mode. This fourth metric is optional and usually only needs to be deployed for custom applications.

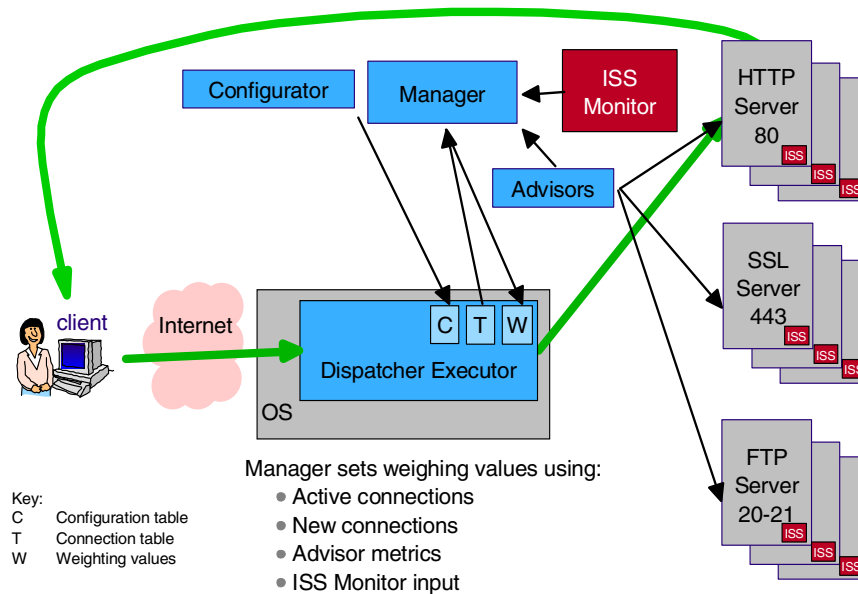


Figure 56. Dispatcher component

The Advisor is a lightweight client that runs as a part of the Dispatcher but actually passes real commands to each application server. It only executes a trivial command, but it needs to be more than a "ping," which will only verify that the TCP/IP protocol stack is up. If this request is successful, the server is deemed to be up. If it fails, the Advisor informs the Manager, and the Manager sets the weight of that server to zero until the server responds again. If the application server has the ISS Load Monitor daemon installed (with or without a custom metric), periodic load reports can be passed to the Dispatcher via the Dispatcher observer. The results of these reports can be factored into the process of setting the individual server weights for the Dispatcher. In most cases, you won't need to use all four of the metrics defined above. The first three will be adequate for typical applications. This means that in practice the server-resident load monitor is an optional item when using Dispatcher. Typically, it is recommended to start out with only the three metrics in use and to evaluate over time if the fourth metric is required. Typically, this will be required only if, in normal use, the application has dramatic shifts of resource utilization, perhaps by factors of 10 or more. The majority of customers do not use the fourth metric with Dispatcher.

7.4.3.2 Dispatcher high availability

The Dispatcher already provides high availability because one of its basic functions is to avoid choosing a failed server clustered behind it. The Dispatcher can also be configured to eliminate the load balancer itself as a single point of failure as part of a comprehensive hardware and software implementation of high availability. The Dispatcher can optionally be configured with a secondary/standby machine on the same subnet that listens for a heartbeat from the primary/active machine and synchronizes its state, including that of its connection table, with that of the primary/active machine. If the standby machine detects that the heartbeat from the active machine is no longer being received, it becomes active, takes over the cluster IP addresses that are being served by the site, advertises the addresses by means of “gratuitous arp” so the other devices on the network are immediately aware of the change, and takes over the role of forwarding packets.

Typically, failover occurs in five seconds or less, therefore, minimizing the number of connection attempts that might fail while the recovery is in progress. Other failover approaches that rely on the age-out of arp entries can take as much as a minute to complete. Additional customer-defined “reachability” criteria can also be specified as part of the criteria for failover, such as access to gateway routers across duplicated adapters and networks, and so on. In the event of a failure, the connection table on the standby machine is closely synchronized with that of the now-failed primary machine so that the great majority of the existing connections in flight will survive the failure. The newly active machine still knows where to send all packets that it receives, and TCP automatically resends any individual packets that were lost during the actual failover. The most likely connections to fail are those that are just in the process of either opening or closing. In the case of opening requests, the client TCP/IP stack or the application may be able to retry and be successful without the client being aware of the failure. In the case of closing requests, it is likely that there will be no loss of data.

7.4.3.3 Dispatcher IP rules-based balancing

IP rules-based balancing introduces the concept of rules and a set of servers among which to load balance if the rule is obeyed. The following rules are available:

- Client IP address
- Client port
- Time of day
- Connections per second for a port

- Active connections for a port

This allows the site to take account of its traffic and the identity of the clients that access it when setting up the load balancing policy. A range can be specified where required, so, for example, all client IP addresses in a particular subnet can be forwarded to a particular set of servers, or between the hours of 8am to 5pm, use this set of ten servers, otherwise, use that set of five servers. This provides a simple method of implementing Quality of Service (QoS) guidelines for individuals, groups of clients, or time-of-day, without imposing unique or proprietary additional semantics or protocols on the client-server relationship.

7.4.3.4 Wide area network Dispatcher

The basic mode of the Dispatcher requires that all clustered servers be on the same subnet as the Dispatcher. But, if you need to, you can configure remote servers as far away as you like, either inside your private network or even across the Internet, to provide a site that is geographically distributed over a few miles or across the globe.

Another Dispatcher must be installed at the remote location. Using this option, a truly distributed wide area network site can be configured. Any client data can be forwarded to any of the configured servers at any location. The choice can be based on server load, or rules, or a combination of the two. The choice of which site should receive the client's packets first can be achieved in several ways including the "which site is closest" mode provided by ISS running in ping triangulation mode (see 7.4.3.5, "ISS and Dispatcher together"). The transmission of packets to remote sites is achieved by encapsulating the unmodified client packets at the originating Dispatcher and then un-encapsulating them at the receiving Dispatcher. The server to client data flow goes direct, as it does with the local area network Dispatcher, thus, keeping the overhead to a minimum.

7.4.3.5 ISS and Dispatcher together

ISS and Dispatcher can be deployed together to support a geographically distributed site. As shown in Figure 57 on page 227, in response to a standard name resolution request from an application program, the ISS name server returns to the client the IP address of the chosen site based on feedback from the ISS daemons running on the Dispatcher machines running at the individual sites. The ISS daemons can be configured to return results in one of two ways:

- Based on system load, which will direct the client to the "least loaded" site. The ISS daemon can use a custom metric if needed.

- Based on "ping triangulation" that will direct the client to the "closest" site from a network topology point of view.

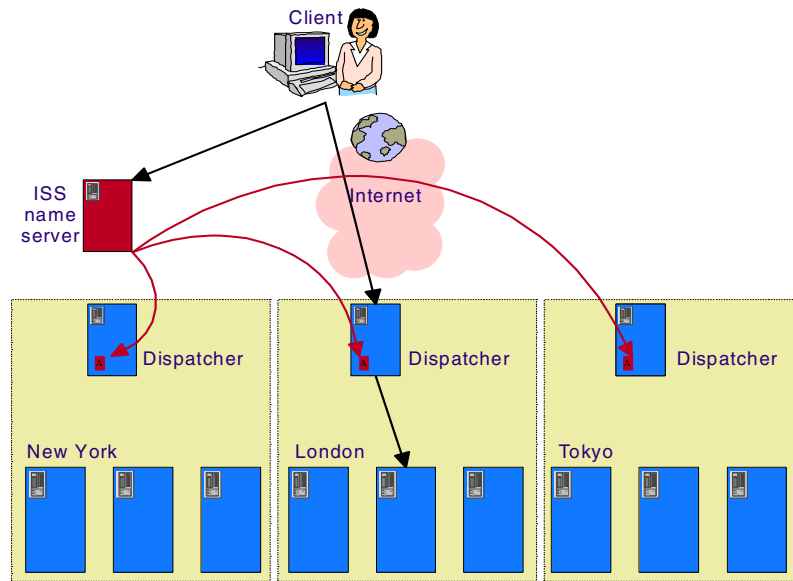


Figure 57. ISS and Dispatcher together

The client then sends its connection request to the selected Dispatcher machine at the chosen location, which will, in turn, select the individual server from the cluster at that site based on its own set of weights. Note that this configuration can also include the wide area network Dispatcher support described above.

7.4.3.6 Dispatcher server affinity

This function is sometimes referred to as the "sticky" option. It is provided to allow load balancing for those applications that preserve some kind of persistent state between (across) separate connections on behalf of clients without losing the state when the client reconnects. When a client originally contacts the site, it is associated with a chosen server in the normal way. When you configure the sticky option, the difference is that any subsequent connections sent in by the client will be dispatched to the same server until a configurable time-out value expires. The Dispatcher lets you configure this sticky capability on a per-port basis.

The sticky option should be used with care since it is keyed off the source IP address coming from the client. There are known limitations to using the source IP address in this way, particularly when many of the clients are

reaching the Internet by means of proxies of various kinds. Some proxies collapse large numbers of users to a very small number of IP addresses, therefore, causing users who are not part of the session to be routed to the same server simply because they are coming through the same proxy. This can cause unwanted hot spots. Other proxies do somewhat the opposite using a pool of IP addresses at random, thus, completely invalidating the affinity. Typically, it is safer to use source IP address affinity in an intranet environment than on the Internet.

7.4.3.7 Dispatcher Server-Directed Affinity API

You may already feel that, from the description in the previous section, affinity is something that should be under the control of your application instead of being chosen by a load balancer's algorithm. The Server-Directed Affinity (SDA) API lets you do exactly that. Your application can now decide where users need to be directed and tell Dispatcher to set up affinity relationships before they ever contact the site. Using this API requires you to write code on your application server to invoke it. A complete working sample program in the Java language is provided.

7.4.3.8 Dispatcher Advisors and Custom Advisors

The Dispatcher supplies standard Advisors for HTTP, HTTPS (SSL), FTP, NNTP, SMTP, POP3, and Telnet. In addition, a specific advisor for the IBM Web Traffic Express (WTE) caching proxy server, a component of IBM WebSphere Performance pack called the WTE advisor, is provided which, in addition to normal availability checking, provides specific awareness of activity on the WTE node, such as garbage collection or cache refresh. A customer may wish to use other standard or private protocols with the Dispatcher or to implement specific extensions to the standard Advisors. For this purpose a Custom Advisor facility has been provided with well-documented sample source code. Custom Advisors must be written in Java at release 1.1.2 or higher of the Java Development Kit, which is not provided with Dispatcher. In fact, the WTE advisor previously mentioned is an example of a Custom Advisor.

This powerful extension capability can be coupled with custom code written to run on the application server machine to provide a high degree of synergy between the Dispatcher and the applications it balances. One example of this use would be a link between a custom advisor and a Java servlet running inside a Web server. The servlet could be coded to extract in-depth performance data from each server and return it to the Dispatcher, thus, allowing the Dispatcher to benefit from results more precisely tailored to the customer's real application environment. This Custom Advisor capability, coupled with the Server-directed affinity API described earlier, provides

customers with application integration and manageability that are quite simply unmatched in the industry.

7.4.3.9 Dispatcher collocation option

If you are starting from a small site, but you have plans to grow rapidly, you have some simple options to keep costs down in the early stages of your growth while still benefiting from the high availability and scalability options of SecureWay Network Dispatcher, therefore, positioning you for that growth when it occurs. As your site's load increases to the point where you need more than one server to handle the traffic, you can add the Dispatcher to your existing network infrastructure with a minimum of hardware and software investment by installing the Dispatcher software on one of the machines where the application servers reside. You will not have to change your existing network configuration in any way. In the simplest case, all you need to do is configure the Dispatcher and either replicate your application server content on to another server or share it with a shared file system. Both file content replication and sharing are provided by another component of IBM WebSphere Performance Pack, the IBM AFS enterprise shared file system.

The collocation option of the Dispatcher allows you to start quickly with minimum cost and evolve to a stand-alone Dispatcher or to a high-availability configuration when the volume of traffic requires it. Even when you have deployed a dedicated Dispatcher, you may still choose to collocate your standby Dispatcher with one of the application servers. Collocation is supported on AIX and Solaris. It is not supported on Windows NT.

7.4.3.10 Wildcard cluster

You can define a wildcard cluster to catch all IP addresses that are not explicitly defined. This allows you to perform completely transparent load balancing as well as providing a mechanism to catch misdirected or malicious packets and log them.

7.4.3.11 Wildcard port

You can define a wildcard port to catch all ports on a particular cluster that are not explicitly defined. This provides a further mechanism to trap and log misdirected or malicious packets.

7.4.3.12 ISP configuration

If you are an Internet Service Provider (ISP) with a large backbone network, or are hosting a busy search engine, you can address your scaling and availability requirements by coupling SecureWay Network Dispatcher in high availability mode with high-capacity caching proxy servers and an enterprise shared file system. These components have been combined into the IBM

WebSphere Performance Pack. By combining the remote cache access and file sharing capabilities of the caching proxy server together with the load balancing capabilities of SecureWay Network Dispatcher, high-availability peer caching can be achieved without incurring the costs normally associated with redundant caching of content.

An efficiently managed cache can significantly reduce backbone network congestion, which allows you to deliver faster response times to your clients and better customer service while reducing your traffic and storage costs. The additional components in the IBM WebSphere Performance Pack are the IBM Web Traffic Express caching proxy server and the IBM AFS enterprise shared file system.

7.4.3.13 Dispatcher command line

The command-line interface of SecureWay Network Dispatcher is the lowest-level administrator interface. It can be imbedded in scripts to permit you to implement specific customized functions.

7.4.3.14 Dispatcher graphical user interface

A fully functional graphical user interface is provided for managing both Dispatcher and ISS. The current configuration is represented on the left hand side of the GUI as a graphical tree. High-level objects are represented at root level, such as the Dispatcher itself. Under it are clusters, ports, and servers, along with the Manager and Advisors. Each cluster port and server can be selected, as can the manager and advisors. A context menu is presented. A different user dialog is presented on the right hand side of the GUI according to the entity that has been selected in the tree on the left hand side. Appropriate dialogs are presented in the normal way when options are selected. A status display area at the bottom of the GUI and full online help via a browser bean are provided. This graphical tree includes the capability to show and manage multiple Dispatcher, ISS, and/or CBR configurations on a single GUI.

7.4.3.15 Dispatcher SNMP support

A Simple Network Management Protocol (SNMP) Management Information Base (MIB) for Dispatcher data is provided. This MIB contains a comprehensive set of values associated with the state and current performance of the Dispatcher. A SNMP-enabled management tool, such as Tivoli NetView or any other similarly equipped tool, can access this MIB. In addition, SNMP traps are also generated if a clustered server fails or if the Dispatcher itself fails over to its standby.

7.4.4 Content-based routing

In some cases, the information upon which the load-balancing decision is to be made is not to be found in the TCP/IP packet headers but is contained in the application-defined data that flows between the client and the server. This implies the use of a proxy or a proxy-like function that actually terminates the client's connection, reads the application data, and creates another connection to the chosen server. This implies significantly higher overhead than the IP-level packet forwarding of the base Dispatcher, but, in many cases, it is required.

The Content-based Routing (CBR) function of SecureWay Network Dispatcher combines the powerful load balancing functions of the Dispatcher with the IBM Web Traffic Express caching proxy server to permit load balancing based on the content of HTTP requests at the application level. Content-based routing can only be used with the HTTP protocol. The same rules-based paradigm is used as for the packet-level, Rules-based routing capability. An example of the kind of data that can be specified in rules and parsed is the name of the object being requested, for example, an HTTP Uniform Resource Locator (URL) contained in a request from a Web Browser. In fact, many of the MIME headers that flow in an HTTP request can be parsed and specified in rules.

The most powerful option is the ability to parse for a Network Dispatcher cookie in the HTTP data stream. This can be inserted in the outbound flow from the server to the client so that when the client contacts the site again, the same cookie will flow. This, in turn, delivers a powerful new means of establishing and maintaining affinity at the individual client level, bypassing the well known "collapsed proxy" limitation when using the source IP address to establish session affinity.

7.4.5 Remote administration

The administration of SecureWay Network Dispatcher can now be performed remotely using both the command line and the GUI. This will allow you to administer multiple Dispatcher, ISS, and CBR configurations from a single management station. The communications used for this function are secured and authenticated by a centrally generated key file.

7.4.6 Why use SecureWay Network Dispatcher?

In the business world of today, more and more computers are involved with integration and communication. The Internet is the world's largest network, and it has become essential in organizations, such as government, academia, and commercial enterprises. Transactions over the Internet are becoming

more common especially in the commercial arena. The information that organizations have on their traditional or legacy business applications may now be published and accessible to a wide audience. Making this information or service available for their customers is a competitive advantage for any organization. The value of an Internet solution, however, is greatly reduced if the information cannot be accessed in a reasonable response time.

To improve performance and, hence, response time to the customer, Internet services may be distributed among multiple servers. Instead of a service belonging to a single server, it belongs to a group of servers to which additional systems may be added or removed. This provides the ability to dynamically modify the service capability to maintain the desired service levels to the customer. In such an environment, load balancing is required to ensure that the workload is appropriately shared between servers in the group providing optimum performance. The IBM SecureWay Network Dispatcher software has been designed specifically to address these types of load-balancing issues, thus, providing optimum performance for customers and maximization of system investment.

7.4.7 Benefits

The benefit provided by SecureWay Network Dispatcher is a well balanced flexible environment allowing multiple servers to operate as a single virtual entity in a TCP/IP serving capacity. This makes the Dispatcher ideal for managing environments, such as large Web servers, with numerous advantages.

- Load balancing across servers for optimum performance, maximizing server investment
- The ability to route service request around problem servers, increasing service availability
- The ability to add or remove servers in the environment without disrupting the current work flow

Chapter 8. High availability

In a consolidated server environment, access to mission critical data is paramount. This chapter will concentrate on the issues involved in retaining availability of critical, online customer resources. We will cover the major products and operating system facilities required to retain application access during system failure. A combination of High Availability Cluster Multi Processing (HACMP) licensed program products, the AIX Logical Volume Manager (LVM), and disk management will be utilized to provide this rapid and seamless recovery.

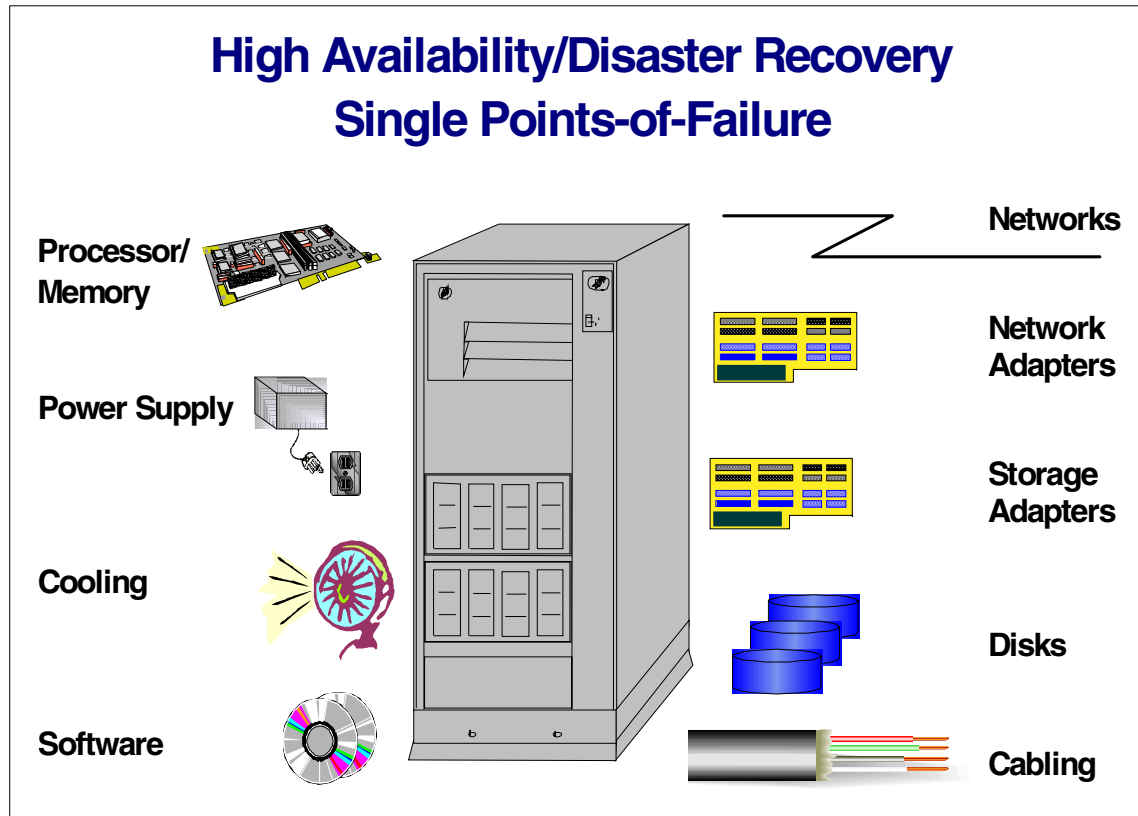


Figure 58. High availability overview

8.1 Overview

The focus of this chapter is to illustrate how the AIX operating system, in conjunction with HACMP, can improve your application service level through achieving higher availability of data.

HACMP is a software product that allows customers to automatically detect system failures and recover users, applications, and data on backup systems, therefore, minimizing downtime to minutes or seconds. Using HACMP virtually eliminates planned outages as users, applications, and data can be made available using the backup systems in a cluster during scheduled system maintenance. With features, such as Cluster Single Point of Control (CSPOC) and Dynamic Reconfiguration, the systems administrator can add users, files, new hardware, and other security functions, without needing to halt mission critical resources when making changes to the cluster.

The AIX Logical Volume Manager, among other functions, provides the tools to manipulate physical and logical volumes and physical and logical partitions, as well as volume groups. Through the capacity to assign multiple physical partitions to each logical partition, copies of vital information can be maintained on separate physical volumes for additional security.

Utilizing RAID technology on both SCSI disk devices and serial storage architecture (SSA) creates a further level of data security and availability.

8.2 High availability solutions

IBM has various offerings for high availability. This section is intended as an overview of the major offerings within the HA suite of products. We will cover HACMP, HAGEO, HACWS, and HACMP/ES in this section.

8.2.1 High Availability Cluster Multi Processing (HACMP)

The HACMP licensed program product is based on the concept of high availability. High availability is often compared to, and sometimes equated to, fault tolerance. However, it is important to understand that high availability is not fault tolerance.

Fault tolerant systems rely on specialized hardware and software to detect hardware failures and instantly continue workload on redundant components built into the system. The fault tolerant system is seen as one logical system entity. There is one system running one instance of operating system software running one instance of any application. Within that one system, there is a complete set of redundant hardware components: Redundant

memory, processors, buses, disks, network adapters, and so on. If any component fails, its redundant partner continues. The active processes are preserved and continue so that there is no perceived impact to the users. While this lack of impact is an attractive advantage, there is a price to be paid. Since all redundant components are included in a single system entity, the redundant components cannot be used for any independent workload in non-failure conditions. There is another drawback to fault tolerant systems. Since they are constructed of specialized hardware and software, they are too expensive for many businesses to justify. This cost becomes even higher if you consider it in price/performance terms because the redundant hardware is not providing any application performance benefits.

Highly available systems are made up of off-the-shelf components. An HACMP cluster is an example of a highly available system. A cluster is made up of several independent systems that share resources, such as disks and application access; so, if a component fails, a redundant component can take over its function. If a component fails, clients will lose access to system resources until the cluster recovers. However, recovery time can be relatively short. Many factors impact recovery time, but resources are typically available within minutes. When a system is used to run a business or used to support customer interactions, the system's availability is one of the most important criteria when selecting a server platform. Loss of system availability means loss of revenue to the business in many situations.

In classic HACMP, anywhere from two to eight nodes can be configured in a cluster. Utilizing High Availability Cluster Multi Processing Enhanced Scalability (HACMP/ES), up to 32 nodes can be included in the cluster. Clearly, from the standpoint of maximizing system output by optimizing resource utilization, HACMP, in conjunction with a consolidated server environment on such hardware platforms as IBMs SP systems, provides a cost effective and robust solution to a multitude of business requirements.

When considered in the context of consolidating a server environment, it is intended as a tool for selecting the appropriate features for a customer's business critical applications, data, and hardware requirements. When moving entire environments, be they development or production, into a consolidated hardware platform, data integrity and availability becomes a principal concern. With careful planning and proper implementation, HACMP

reduces system downtime to a minimum in the event of single or multiple server loss in a cluster.

Table 3. Benefits of highly-available consolidated server systems

| Highly available consolidated server clusters | Traditional open distributed systems |
|---|--|
| Redundant Components Automated takeover of applications between CPU's Excellent price/performance Operating system failure protection Single point of control Eliminate downtime due to planned outages Minimize downtime due to system failure Centralized administration and maintenance | Flexible Non proprietary Good performance Many single points of failure High administrative costs Hardware and software maintenance overhead due to geographic separation |

8.2.1.1 Application and database protection

Ultimately, there are three defining factors in designing any highly available consolidated server environment:

- Access to data
- Integrity of data
- Cost effectiveness of solution

These components will have differing importance depending upon your situation's individual demands. With regard to application and database access protection, we are primarily concerned with the first two points here.

Most important is the matter of data integrity. Most applications and databases provide their own logging mechanisms to ensure data reliability after a system failure. In an HACMP cluster, this is not compromised during a system failure, as multiple start and stop scripts for the applications and databases ensure their continued functionality during and after system failure. The following three diagrams illustrate the basic cluster layouts for the most commonly used non-concurrent cluster configurations:

- Hot Standby
- Mutual takeover
- Rotating Standby

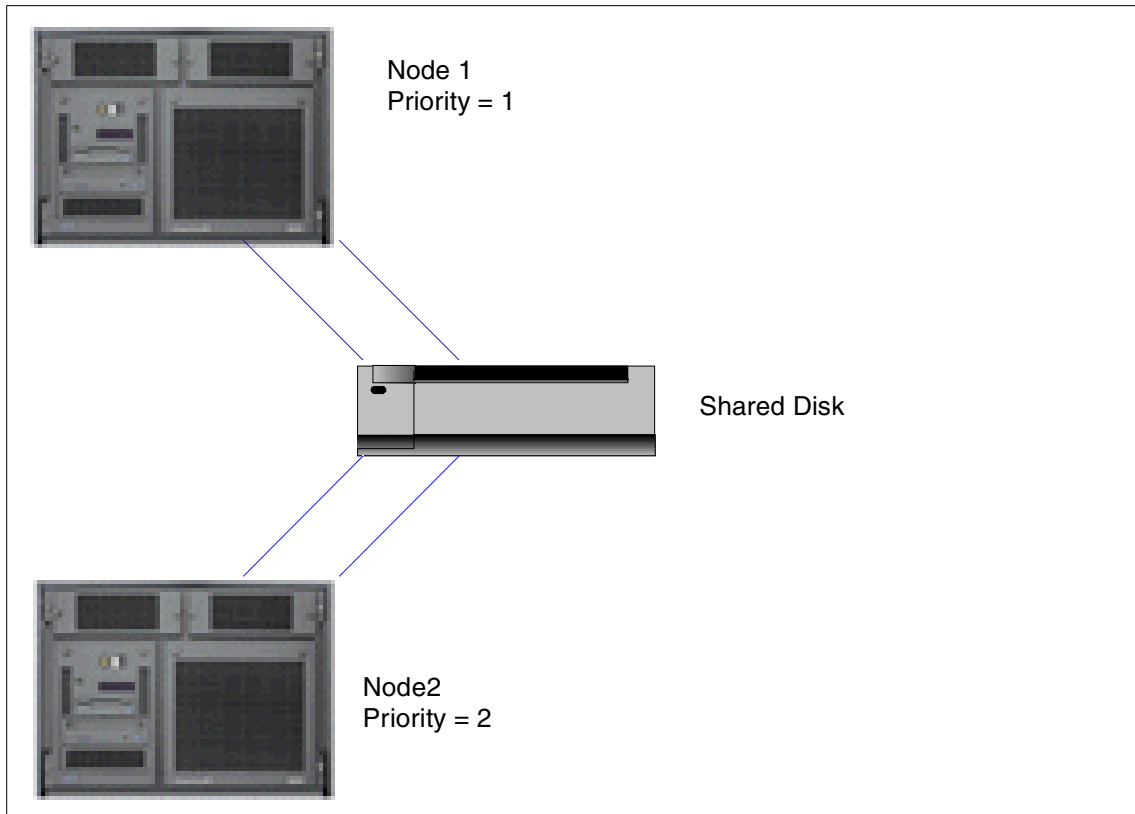


Figure 59. A two node Hot-Standby cluster configuration

- Hot-Standby configuration

In this configuration, there is one cascading resource group consisting of the SSA disk sub-system and its constituent volume groups and file systems. Node 1 has a priority of 1 for this resource group, while Node 2 has a priority of 2. During normal operations, Node 1 provides all critical services to end users. Node 2 may be idle or may be providing non-critical services and, hence, is referred to as a Hot-Standby node. When Node 1 fails or has to leave the cluster for a scheduled outage, Node 2 acquires the resource group and starts providing the critical services.

- Rotating Standby Configuration

This configuration is the same as the previous configuration except that the resource groups used are rotating resource groups.

In the Hot-Standby configuration, when Node 1 reintegrates into the cluster, it takes back the resource group since it has the highest priority for it. This implies a break in service to the end users during reintegration.

If the cluster is using rotating resource groups, reintegrating nodes do not reacquire any of the resource groups. A failed node that recovers and rejoins the cluster becomes a standby node. You must choose a rotating standby configuration if you do not want a break in service during reintegration. Since takeover nodes continue providing services until they have to leave the cluster, you should configure your cluster with nodes of equal power. While more expensive in terms of CPU hardware, a rotating standby configuration gives you better availability and performance than a Hot-Standby configuration.

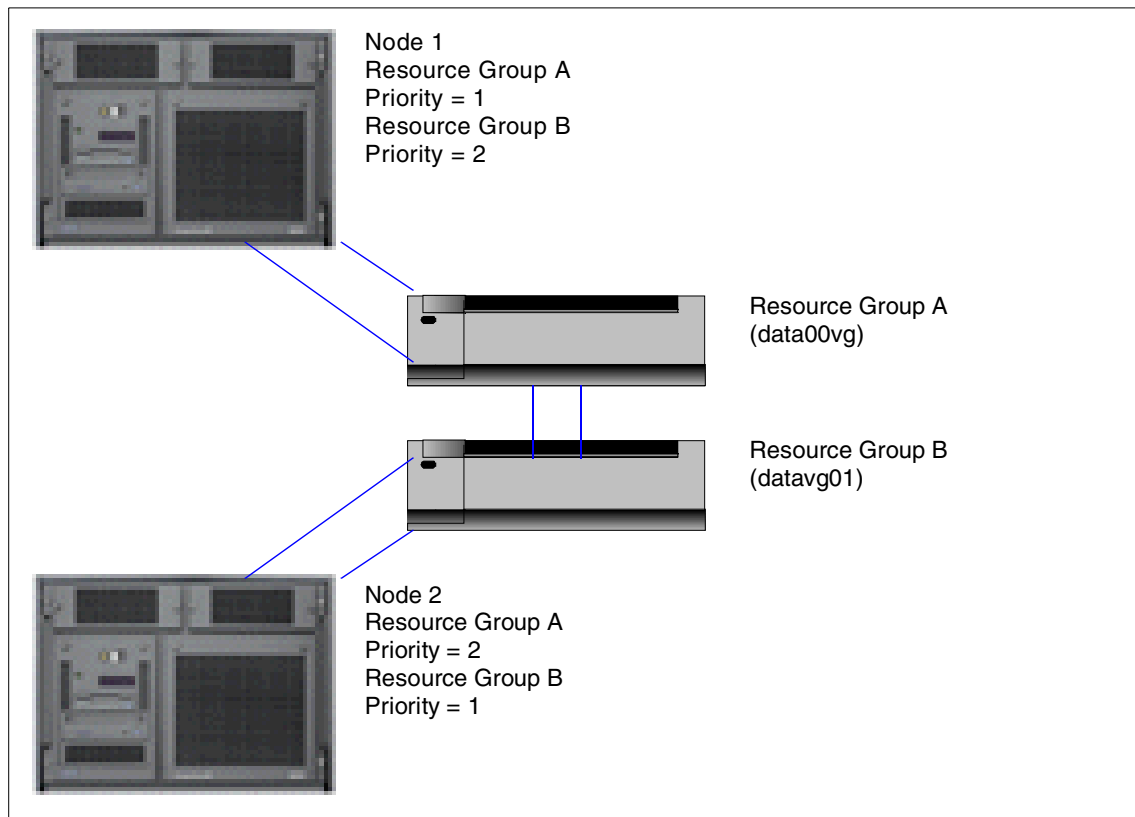


Figure 60. Mutual takeover configuration

- Mutual Takeover configuration

Figure 60 on page 238 illustrates a two node cluster in a mutual takeover configuration. In this configuration, there are two cascading resource groups: A and B. Resource group A consists of multiple SSA disks and one volume group, datavg00. Resource group B consists of multiple SSA disks and one volume group, datavg01. Node 1 has priorities of 1 and 2 for resource groups A and B, respectively, while Node 2 has priorities of 1 and 2 for resource groups B and A, respectively. During normal operations, nodes 1 and 2 have control of resource groups A and B, respectively, and both provide critical services to end users. If either node 1 or node 2 fails, or has to leave the cluster for a scheduled outage, the surviving node acquires the failed node's resource groups and continues to provide the failed node's critical services.

When a failed node reintegrates into the cluster, it takes back the resource group for which it has the highest priority. Therefore, even in this configuration, there is a break in service during reintegration. Of course, if you look at it from the point of view of performance, this is the best thing to do since you have one node doing the work of two when any one of the nodes is down.

Later in this chapter, we will give an overview of the ClusterProven licensing program. The ClusterProven licensing program introduces new high availability requirements that may be applied to any combination of operating system, middleware, or end-user application. A solution that satisfies a platform's technical criteria may be validated with IBM and licensed to be marketed with IBM's ClusterProven trademark.

8.2.1.2 IP Address protection (adapter and node failure)

The mechanism of IP Address Takeover (IPAT) is a fundamental part of HACMP. Hardware address takeover can also be incorporated into this.

The basic configuration of each node in the cluster will include a minimum of two network adapters. One adapter is configured as a service adapter, and the other as a standby. If the service adapter of one node fails, the standby adapter will be reconfigured by the cluster manager to take over that adapter's service IP address. If a node fails, the standby adapter in the surviving node will be reconfigured to take over the failed node's service IP address. In combination with the other facilities of HACMP that have been configured for the environment, the resources that have been failed over to the standby node should now be available to the system users.

8.2.1.3 Network failure protection

As an independent, layered component of AIX, the HACMP software works with most TCP/IP-based networks. HACMP has been tested with standard ethernet

interfaces, Token-Ring and Fiber Distributed Data Interchange (FDDI) networks, with IBM Serial Optical Channel Converter (SOCC), Serial Line Internet Protocol (SLIP), and Asynchronous Transfer Mode (ATM) point-to-point connections. The HACMP for AIX software supports a maximum of 32 networks per cluster and 24 TCP/IP network adapters on each node. These numbers provide a great deal of flexibility in designing a network configuration. The network design affects the degree of system availability in that the more communication paths that connect clustered nodes and clients, the greater the degree of network availability.

8.2.1.4 Candle Command Center for High Availability

By monitoring the key components that effect your systems and applications, the Candle Command Center for High Availability Systems keeps you informed of all potential problems and performance issues. In essence, the Candle Command Center for High Availability Systems is your early warning system for business critical applications. This solution:

- Delivers systems management capabilities
- Displays performance and availability data
- Sets and displays alert or threshold values
- Makes implementation and management easier
- Provides end-to-end response time for SAP and PeopleSoft

8.2.2 High Availability Geographic Cluster (HAGEO)

HAGEO for AIX is a product that extends the capabilities of IBMs HACMP to deal with complete site disasters. HAGEO mirrors disk contents over long distance networks to a backup site. It automates recovery procedures to facilitate a server installation in withstanding a failure or disaster that disables an entire location.

In the unlikely event of losing an entire location, HAGEO, in combination with the various application and database recovery aids, can be used to rebuild the system environment and user data.

If you require more information regarding HAGEO, refer to the redbook: *Disaster Recovery with HAGEO: An Installer's Companion*, SG24-2018.

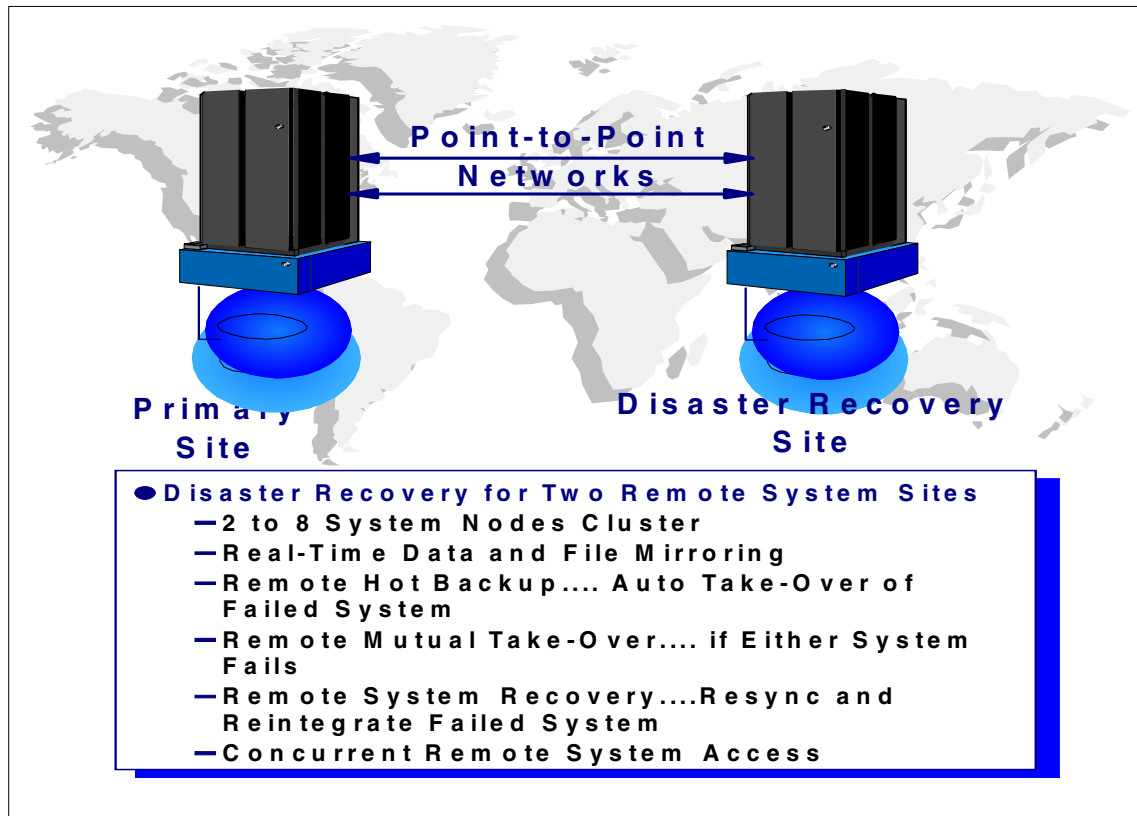


Figure 61. Typical HAGEO configuration

8.2.3 High Availability Control WorkStation (HACWS)

HACWS is a two node cluster consisting of a primary control work station with a backup control workstation configured as a rotating cluster.

HACWS provides a high availability solution on the control workstation, which serves as a single point of control for managing and maintaining the RS/6000 SP nodes using Parallel System Support Program (PSSP). Depending on your system environment and the type of applications that are running on your control workstation, the impact of a failure in the CWS may not affect the operation of the RS/6000 SP.

If your needs are such that you choose to run critical applications on the CWS, then you should consider protecting it with HACWS. Users who do not run any critical application, for example, Name Server, LoadLeveler, or DB2 Parallel Edition, on the control workstation or do not care about spending the

time to restore system backup and rebuild the control workstation after its failure, do not require HACWS. For such users, however, HACWS still provides minimum down time for maintenance of the control workstation.

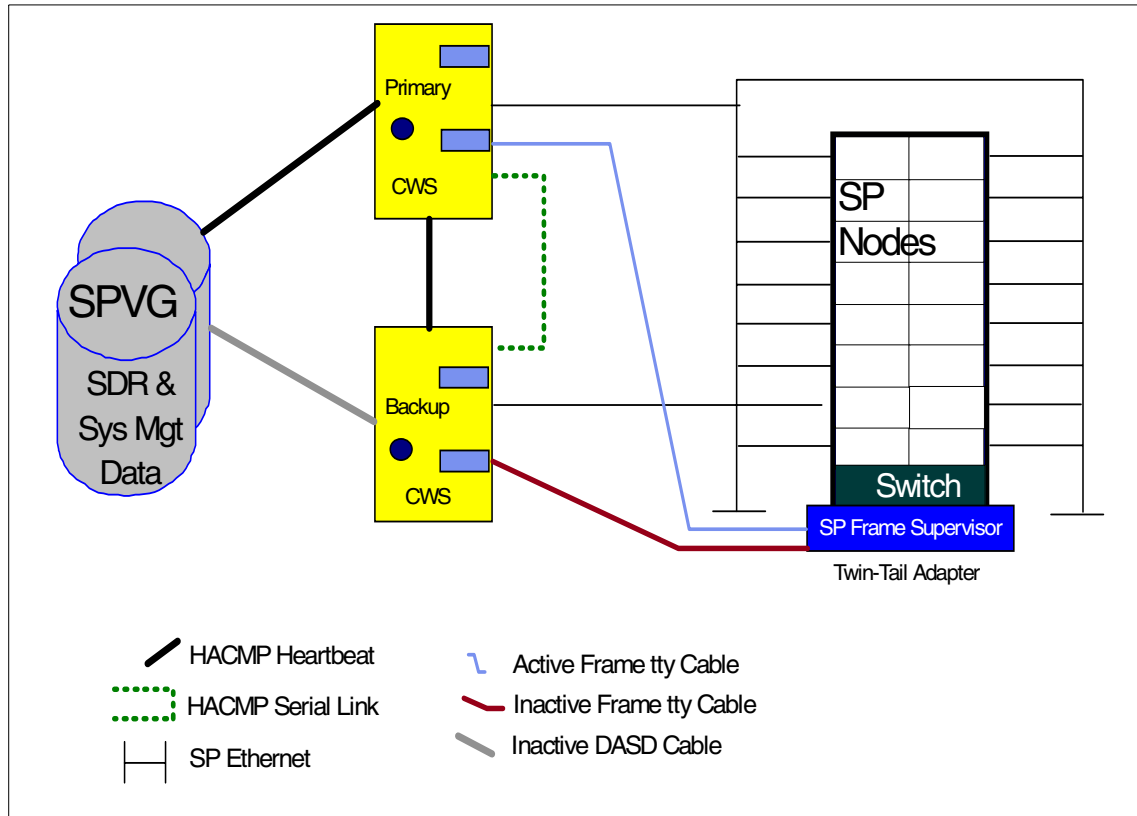


Figure 62. Typical HACWS configuration

The loss or failure of the control workstation will have the following effects on the management of the RS/6000 SP system but should have little or no impact on the active jobs on RS/6000 SP nodes:

- It will not be possible to control the RS/6000 SP hardware.
- System Data Repository will be unavailable.
- Existing jobs will continue to completion, but new parallel jobs cannot be started.
- No configuration changes can be made.
- Software installations cannot be done from the CWS.

- Should a switch fault occur, reset processing cannot be completed
- Error logging of alerts raised by RS/6000 SP nodes will be lost (though the information will still be logged on the individual nodes)
- While the control workstation is unavailable, some administrative tasks that use PSSP will not be able to proceed

For further information regarding HACWS, refer to the following redbook: *Implementing High Availability on RISC/6000 SP*, SG24-4742.

8.2.4 HACMP/ES

HACMP Enhanced Scalability 4.3 brings a new component to the business of server consolidation. It achieves this through adding a greater degree of flexibility to considerations of availability in many possible solutions. HACMP/ES 4.3 allows you to amalgamate any combination of stand-alone RS/6000s in clusters with RS/6000 SP nodes.

The RS/6000 High Availability Infrastructure (HAI) was introduced as part of PSSP Version 2.2. With Version 3.1, the RS/6000 SP term HAI has been replaced by the newly introduced official term IBM RS/6000 Cluster Technology (RSCT), which refers to the following three key distributed subsystem components:

- Topology Services
- Group Services
- Event Management

If you install HACMP/ES on an SP node, you need AIX, PSSP, and HACMP/ES. If you install HACMP/ES on a RS/6000, you will require AIX and HACMP/ES. RSCT is included in both PSSP 3.1, and HACMP/ES 4.3 software.

8.2.4.1 HACMP/ES V4.3

We will now briefly discuss the new or enhanced functions provided with Version 4.3 of the HACMP/ES product:

- Support for the RS/6000 family
- 32-node support
- Topology Dynamic Automatic Reconfiguration Event (DARE)
- Packaging
- Concurrent Access
- Improved snapshot facility

- ATM support
- SDR independency
- Heartbeat tunable on a network basis

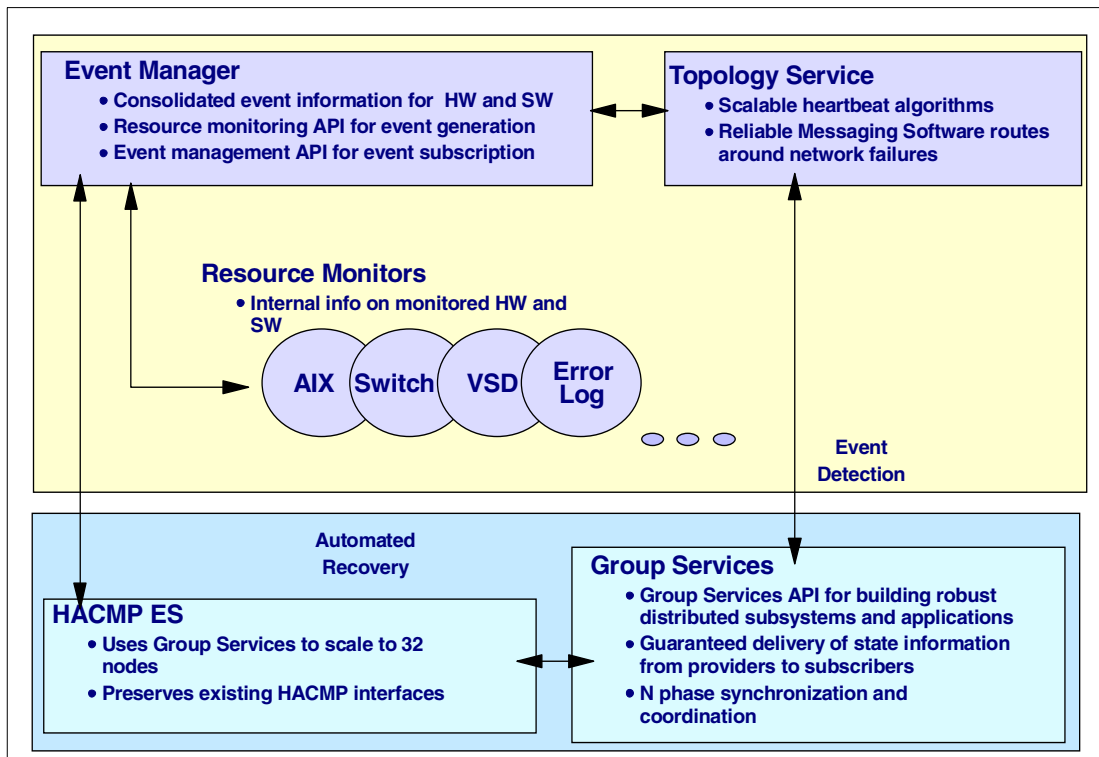


Figure 63. HACMP/ES Version 4.3

- Support for the RS/6000 family

One major limitation of HACMP/ES V4.2.2 is that it is only supported on the RS/6000 SP node. HACMP/ES V4.3 can now be installed on both RS/6000 SP nodes and RS/6000s. This means it is now possible to have an HACMP/ES V4.3 cluster made up of:

- RS/6000s only
- RS/6000 SP nodes only
- A mixture of RS/6000s and RS/6000 SP nodes
- RS/6000 SP nodes belonging to different RS/6000 SP systems

In case of an HACMP/ES V4.3 cluster composed of RS/6000 SP nodes only, these nodes can belong to different PSSP partitions.

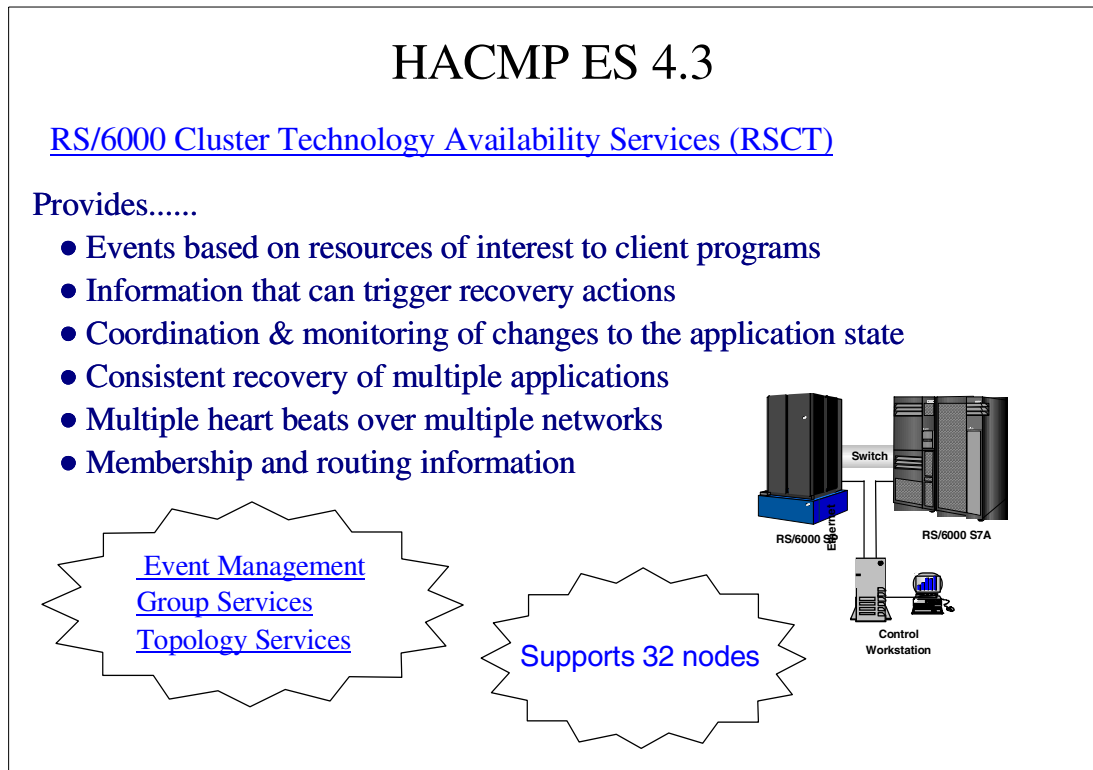


Figure 64. RSCT overview

8.2.4.2 32-node support

HACMP/ES V4.2.2 supports a maximum cluster size of 16 RS/6000 SP nodes without PTF, while HACMP/ES V4.3 supports clusters of up to 32 RS/6000 SP nodes. Like all previous versions, HACMP for AIX V4.3 is still limited to 8-node clusters.

8.2.4.3 Topology DARE

HACMP/ES V4.3 now supports DARE of the Cluster Topology. DARE allows the system administrator to perform changes to the active cluster configuration without having to stop and restart HACMP/ES, hence, increasing the availability of customer applications and reducing downtime. HACMP/ES V4.2.2 supports DARE only on the Cluster Resources, not the Cluster Topology.

8.2.4.4 Packaging

In order to install the HACMP/ES V4.3 software, it is now required to install the RSCT filesets first. These filesets include the support on the Topology Services, Group Services, and Event Management components. Before V4.3, these components were part of PSSP. Now, having extended the support of HACMP/ES to the RS/6000 family, these components are part of a new package called RSCT, which no longer depends on PSSP. The fileset names are rsct.basic and rsct.clients.

8.2.4.5 Concurrent access

HACMP/ES V4.3 now supports concurrent access clusters, hence, offering the opportunity to use the IBM-provided lock manager, the cclockd daemon. This means all the cluster nodes are able to read/write the data on the external shared disks concurrently.

8.2.4.6 Improved snapshot facility

HACMP/ES V4.3 implements an improved snapshot facility that preserves user-defined events. When applying a cluster snapshot, user-defined events are restored in the /usr/sbin/cluster/events/rules.hacmprd file. With HACMP/ES V4.2.2, it is a customer's responsibility to save user-defined events when taking a snapshot and recreate them when applying the snapshot.

Also, HACMP/ES V4.3 allows you to define your own commands to those that are issued by the standard snapshot. This allows you to add extra information, such as application related data, to your snapshot.

8.2.4.7 ATM support

HACMP/ES V4.3 now supports Asynchronous Transfer Mode (ATM) networks.

8.2.4.8 SDR independency

HACMP/ES V4.3 no longer has a dependency on the System Data Repository (SDR). With HACMP/ES V4.2.2, daemons were used to query the SDR to collect information about the cluster configuration, hence, creating a dependency on the Control Workstation where the SDR resides. Now the HACMP/ES V4.3 daemons gather the necessary information from the Global ODM (GODM). This change became necessary since HACMP/ES V4.3 can now also run on a cluster composed of only RS/6000s, while the SDR is a peculiarity of the RS/6000 SP system.

8.2.4.9 Heartbeat rate tunable on network basis

HACMP/ES V4.3 now allows the system administrator to configure a different heartbeat rate for every network type, thus, providing the same granularity the classic HACMP for AIX product has. HACMP/ES V4.2.2 only allows one heartbeat rate that is valid for all networks.

8.3 High availability options at the operating system level

Configuring a cluster for high availability is more than simply implementing HACMP or HACMP/ES software. A singularly important factor in obtaining maximum system accessibility is disk redundancy. We will cover both disk mirroring and RAIDant arrays in this section.

8.3.1 Disk mirroring/logical volume mirroring

We will not be going into any great detail regarding mirroring, and, as such, we'll be taking a fairly high overview of the concepts involved.

Mirroring allows either your system or your data to remain available in the event of a disk failure. It allows the capacity to boot from an alternate disk in the eventuality of losing an operating system boot disk.

Up to three copies of a disk/logical volume/volume group are catered for with the AIX operating system. The AIX Systems Management Interface Tool (SMIT) makes the creation of logical volume/volume group mirrors relatively straightforward. Alternately, the mirrors can be created from the command line.

Mirroring adds another layer of security to availability when consolidating an environment. Be it operating system or data integrity, constant access to information will be facilitated through mirroring.

8.3.2 Redundant Array of Independent Disks (RAID)

We will briefly describe the basic elements of RAID technology and provide an overview of the various RAID levels.

RAID is an acronym for Redundant Array of Independent Disks. Disk arrays are groups of disk drives that work together to achieve higher data-transfer and I/O rates than those provided by single, large drives. An array is a set of multiple disk drives plus a specialized controller (an array controller) that keeps track of how data is distributed across the drives. Data for a particular file is written in segments to the different drives in the array rather than being written to a single drive. By using multiple drives, the array can provide higher

data-transfer rates and higher I/O rates when compared to a single large drive; this is achieved through the consequent ability to schedule reads and writes to the disks in parallel.

Arrays can also provide data redundancy so that no data is lost if a single drive (physical disk) in the array should fail. Depending on the RAID level, data is either mirrored or striped. Striping involves splitting a data file into multiple blocks and writing a sequential set of blocks to each available drive in parallel and repeating this process until all blocks have been written. Mirroring describes the situation where data written to one disk is also copied exactly to another disk, thereby, providing a backup copy. The RAID levels currently supported by the 7135 RAID Array subsystem are: RAID 0, 1, 3, and 5. RAID 1, 3, and 5 offer data redundancy or protection of the data in the event that a drive should fail. Subarrays are contained within an array subsystem. Depending on how it is configured, an array subsystem can contain one or more sub-arrays, also referred to as Logical Units (LUNs). Each LUN has its own characteristics, for example, RAID level, logical block size, and logical unit size. From AIX, each subarray is seen as a single hdisk with its own unique name.

8.4 Integrating an existing cluster into a consolidated environment

Many of the issues involved in moving your existing HACMP cluster to a consolidated environment will have been dealt with elsewhere in this publication. There are a few issues, particular to HACMP clusters, that should be considered here. Only some of the issues mentioned below will be relevant in any particular scenario.

8.4.1 Physical hardware and software relocation

There are several factors to be taken into account when physically moving the location of a cluster, along with the costs, both in fiscal and human terms:

- The cost of relocating the hardware itself
- Future growth of the cluster and its processing requirements
- Making available personnel with the appropriate HACMP expertise at the new site
- Initially making available personnel with knowledge of the previous site's database, application, user issues, security issues, overall system administration, and so on.

Though most of these issues cannot be quantified here, they will require careful planning when deciding upon an overall strategy. For example, when

calculating future processing requirements for a cluster, it may well be the case where migrating to new hardware would financially show greater viability than physically relocating the existing cluster hardware.

8.4.2 Migrating a cluster to new hardware

Some issues to be considered when migrating a cluster to new hardware, over and above those of simply relocating the cluster, are:

- The operating system level must be compatible with the new hardware.
- Some of the tailored failover scripts may require modification due to new hardware, such as disk subsystems, disk adapters, TCP/IP adapters, being configured.
- Heartbeat rate(s) may require adjustment due to potential throughput improvement.

8.5 Planning and integrating a cluster from conception

The area of cluster planning is a large one. It includes planning for the types of hardware, such as CPUs, networks, and disks to be used in the cluster. It also includes resource planning, that is, planning the desired behavior of the cluster in failure situations along with other aspects. Resource planning must take into account application loads and characteristics as well as priorities.

Once the decision has been made to migrate an application into a consolidated cluster, a number of factors will influence the final desired environment. These are covered in detail in the Cluster Planning and Implementation guides included with the HACMP or HACMP/ES licensed program product. In this section we will give a brief description of these planning considerations

8.5.1 Cluster nodes

One of HACMP's key design strengths is its ability to provide support across the entire range of RS/6000 products. Because of this built-in flexibility and the facility to mix and match RS/6000 products, the effort required to design a highly available cluster is significantly reduced.

The following table gives you an overview of the currently supported RS/6000 models in an HACMP cluster.

8.5.2 CPU options

HACMP is designed to execute with RS/6000 uniprocessors, Symmetric Multi-Processor (SMP) servers, and the RS/6000 SP. The minimum configuration and sizing of each system CPU is highly dependent on the user's application and data requirements. Almost any model of the RS/6000 POWERserver family can be included in an HACMP environment. The following table gives you an overview of the currently supported RS/6000 models as nodes in an HACMP or HACMP/ES cluster.

| HACMP version | 4.2 | 4.3 | 4.2/ES | 4.3/ES |
|------------------------|-----|-----|--------|--------|
| 7009 Model Cxx | yes | yes | no | yes |
| 7011 Model 2xx | yes | yes | no | yes |
| 7012 Model 3xx and Gxx | yes | yes | no | yes |
| 7013 Model 5xx and Jxx | yes | yes | no | yes |
| 7015 Model 9xx and Rxx | yes | yes | no | yes |
| 7017 Model S7x | yes | yes | no | yes |
| 7024 Model Exx | yes | yes | no | yes |
| 7025 Model Fxx | yes | yes | no | yes |
| 7026 Model Hxx | yes | yes | no | yes |
| 7043 Model 43P, 260 | yes | yes | no | yes |
| 9076 RS/6000 SP | yes | yes | yes | yes |

Table 4. Supported RS/6000 models as cluster nodes

8.5.3 Node considerations

Much of the decision on choosing nodes centers around the following areas:

- Processor capacity
- Application requirements
- Anticipated growth requirements
- I/O slot requirements

When designing a cluster, you must consider the requirements of the cluster as a total entity. This includes understanding system capacity requirements of other nodes in the cluster beyond the requirements of each system's

prescribed normal load. You must consider the required performance of the solution, during and after failover, when a surviving node has to add the workload of a failed node to its own workload.

8.5.4 Cluster networks

HACMP differentiates between two major types of networks: TCP/IP networks and non-TCP/IP networks. HACMP utilizes both of them for exchanging heartbeats. HACMP uses these heartbeats to diagnose failures in the cluster. Non-TCP/IP networks are used to distinguish an actual hardware failure from the failure of the TCP/IP software. If there were only TCP/IP networks being used, and the TCP/IP software failed, therefore, causing heartbeats to stop, HACMP could falsely diagnose a node failure when the node was actually still functioning. Since a non-TCP/IP network would continue working in this event, the correct diagnosis could be made by HACMP. In general, all networks are also used for verification, synchronization, communication, and triggering events between nodes. Of course, TCP/IP networks are used for communication with client machines as well.

The following lists the supported TCP/IP network types and network considerations:

- Generic IP
- ATM
- Ethernet
- FCS
- FDDI
- SP Switch
- SLIP
- SOCC
- Token-Ring

As an independent, layered component of AIX, HACMP software works with most TCP/IP-based networks.

8.5.4.1 Non-TCP/IP Networks

Non-TCP/IP networks in HACMP are used as an independent path for exchanging messages or heartbeats between cluster nodes. In case of an IP subsystem failure, HACMP can still differentiate between a network failure and a node failure when an independent path is available and functional.

Currently, HACMP supports the following types of networks for non-TCP/IP heartbeat exchange between cluster nodes:

- Serial (RS232)
- Target-mode SCSI
- Target-mode SSA

8.5.5 Cluster disks

There are two types of disk subsystems supported under HACMP:

- SSA
- SCSI

8.5.5.1 SSA Disks

SSA is a high-performance, serial interconnect technology used to connect disk devices and host adapters. SSA is an open standard, and SSA specifications have been approved by the SSA Industry Association and has also been accepted as an ANSI standard through the ANSI X3T10.1 subcommittee.

For a full description of SSA and its functionality, see the redbook *Monitoring and Managing IBM SSA Disk Subsystems*, SG24-5251.

8.5.5.2 SCSI Disks

After the announcement of the 7133 SSA Disk Subsystems, the SCSI Disk subsystems became less common in HACMP clusters. However, the 7135 RAIDiant Array (Model 110 and 210) and other SCSI Subsystems are still in use at some customer sites.

8.5.6 Resource planning

HACMP provides a highly available environment for resources essential to mission critical processing. It further defines the relationships between the nodes in the cluster, thus, ensuring appropriate failover behavior. When a cluster node fails, the Cluster Manager distributes its resources according to how the *Resource Groups* are categorized.

HACMP considers the following as resource types:

- Volume groups
- Disks
- File systems
- File systems to be NFS mounted

- File Systems to be NFS exported
- Service IP addresses
- Applications

8.5.6.1 Resource group options

Each resource in a cluster is defined as part of a resource group. This allows you to combine related resources that need to be together to provide a particular service. A resource group also includes the list of nodes that can acquire those resources and serve them to clients.

A resource group is defined as one of three types:

- Cascading - The active node with the highest priority controls the resource group.
- Rotating - The node with the rotating resource group's service IP address controls the resource group.
- Concurrent - All active nodes have access to the resource group.

Each of these types describes a different set of relationships between nodes in the cluster and a different set of behaviors upon nodes entering and leaving the cluster. For further information, refer to IBM Redbook Web site:

<http://www.redbooks.ibm.com/>

8.5.7 Application planning

The central purpose for combining nodes in a cluster is to provide a highly available environment for mission-critical applications. In many organizations, these applications must remain available at all times.

In an HACMP for AIX cluster, these critical applications can be a single point of failure. To ensure the availability of these applications, the node configured to take over the resources of the node leaving the cluster should also restart these applications so that they remain available to client processes.

For more information about creating application server resources, see the *HACMP Version 4.3 AIX: Installation Guide*, SC23-4278.

8.5.7.1 Performance requirements

There are a number of possible states that an HACMP cluster can be in at any point in time. Under normal circumstances, the load of any one application is generally being serviced by the cluster node that was designed to deal with it. In the case of a failover, however, an alternate node in the cluster will have to take over this workload. The performance requirements of

any cluster application needs to be understood in order to have the computing capacity available to support mission critical applications in all possible cluster states.

The Application Planning Worksheets can be found in Appendix A of the *HACMP Version 4.3 AIX Planning Guide*, SC23-4277. These should be filled out when in the initial stages of cluster design.

8.5.8 Other cluster consolidation issues

When conceptualizing a cluster in a consolidated environment, there are a number of other considerations to be taken into account that are worthy of mention at this point.

- Application start-up and shutdown routines, licensing methods, co-existence, and prioritisation
- Cluster event customisation and error notification
- Cluster user and group IDs, passwords, and home directories

8.6 ClusterProven and Advanced ClusterProven

In an ongoing, strategic effort to assist customers on their way to continuous system availability, IBM has introduced the ClusterProven Program. This initiative demonstrates the commitment IBM and its solution developers have to high availability and scalability.



Figure 65. ClusterProven logo

High availability of a computer system cannot be achieved by the servers alone. Whatever high figures for uptime its hardware and operating system has, the whole chain is as weak as its weakest part. If middleware or applications fail, the whole system loses its availability for the end-user. Only those solutions which have all of their components tested to meet high

availability requirements provide a real value to customers. IBM has designed the ClusterProven licensing program to encourage solution developers to bring full HA solutions to the market by utilizing the high availability features of IBM servers.

The ClusterProven licensing program introduces new high availability requirements that may be applied to any combination of operating system, middleware, or end-user applications. A solution that satisfies a platform's technical criteria may be validated with IBM and licensed to be marketed with IBM's ClusterProven trademark.

The program offers two levels of ClusterProven status that a solution developer may pursue. One status is ClusterProven, which applies to applications that exploit the basic availability features of an IBM server platform, the other level being Advanced ClusterProven, which may be granted to applications that go beyond that level of exploitation to realize increased efficiency in system management, data protection, and downtime reduction.

In order to help customers to recognize solutions that go beyond basic availability exploitation to the increased efficiency in system management, data protection, and downtime reduction, IBM is also introducing the Advanced ClusterProven status. The Advanced ClusterProven status is intended to denote applications that provide significant benefits above those bearing the ClusterProven status and that move customers closer to continuous operations. This status implies a higher level of application integration and the delivery of superior high availability and scalability.

Examples of Advanced ClusterProven characteristics include:

- Failure recovery with minimal impact to application availability at the end-user level.
- Application recovery with no loss of in flight data or transactions.
- Further reduction of costly downtime for planned upgrades.
- Automatic corrective action when application or system conditions become marginal.
- Efficiency of operations and management.
- Increased throughput with scalable incremental computing growth.

The criteria of ClusterProven and Advanced ClusterProven Solutions are set forth in general terms across the server brands. Each of the IBM server brand teams provides an explanation of what is required from an application to

qualify as ClusterProven or an Advanced ClusterProven solution on this server brand.

8.6.1 ClusterProven verification process

The intent of IBMs ClusterProven program is to demonstrate that an application can take advantage of an RS/6000 clustered environment to attain greater levels of availability and/or scalability than is possible on a stand-alone server.

To be verified as ClusterProven, an application must demonstrate recovery from local hardware failures with no corruption of data. It is recognized, however, that not all downtime is caused by hardware failure. Therefore, an application that has been verified as ClusterProven, and accepted by IBM, may seek a further level of verification for *Advanced ClusterProven*. To be Advanced Cluster Proven, an application must demonstrate recovery from, or immunity to, a defined set of software failures. This set of failures includes events directly under the control of the application as well as a set of related services in the AIX operating system.

8.6.2 ClusterProven verification criteria

To demonstrate a product is ClusterProven, a solution developer must design, implement, and test its application in an RS/6000 clustered environment and submit documentation of the successfully tested cluster to IBM, along with the solution developer's verification form, for approval. Where appropriate, the solution developer will make the sample scripts, sample AIX, and application setup information used in setting up the cluster available to customers and service providers through a Web site or other easily accessible methods.

8.6.3 Cluster design guidelines

- The cluster should be representative of a typical implementation.
- If more than one application is required for the test (for example, an application server and a database server), then each application should be run on a separate node to demonstrate greater robustness and scalability.
- Any clients used in the test environment should be *outside* of the cluster.
- The cluster should be designed to eliminate any single point of failure
- The cluster must demonstrate:
 - Recovery from a complete node failure.
 - Recovery from a network adapter failure.

- Client able to reconnect to same IP address after recovery from either of the above failures.
- No corruption of data during any failure.

The solution developer may demonstrate adherence to these guidelines in one of the following ways:

- HACMP. Implement and test the application in an HACMP-based cluster.
- Network Dispatcher. Implement and test the application in a Network Dispatcher-based cluster.
- Custom Implementation. In some instances, an application has cluster awareness built into it, thus, providing functions similar to HACMP or Network Dispatcher. To verify the application as ClusterProven through this method, an application must be tested against the same set of failure events as HACMP or Network Dispatcher.

8.6.4 Advanced ClusterProven overview

To be considered Advanced ClusterProven, an application must, in addition to being verified as ClusterProven, demonstrate the capability to monitor its critical software environment and to take preventative and/or corrective action based on the information gathered. In addition to monitoring the health of its own application, the solution developer must ensure that the critical applications, or operating system services upon which it is dependent, are monitored as well.

An application that monitors and measures this environment and takes actions designed to improve availability and performance can be verified as Advanced ClusterProven. Given the complexity of this task, IBM is identifying enablers for Advanced ClusterProven that will allow an application to meet this criteria.

By using an enabler for Advanced ClusterProven, a solution developer can then focus on how to determine the health of its own application. IBM will provide collateral in the form of a list of approved advanced enablers and guidelines for providing monitoring hooks and actions to be used by the advanced enabler. By providing information in a standard format, an application can be verified as Advanced ClusterProven by testing with any Advanced Enabler that will use the information. The customer who installs and uses this application is then free to implement an environment with the enabler that the application was certified with or with any other enabler supplying the same function.

8.6.4.1 Advanced ClusterProven verification criteria

- An application must first meet all criteria for ClusterProven status. The application must be verified as ClusterProven, and the verification must be accepted by IBM.
- Identify application failure modes and recovery action as discussed in IBM supplied documentation *RS/6000 Advanced ClusterProven Guidelines*, which can be found at: <http://www.ibm.com/servers/clusters>
- Test cluster for recovery from, and prevention of, operating system and application failures.
- The solution developer must submit the following documentation to IBM along with its verification form:
 - A description of the application and environment tested
 - The successfully executed test plan
 - Documentation of failures and events monitored and actions taken

Note

In all cases, IBM will be the sole arbiter as to whether submission material meets the criteria for ClusterProven and Advanced ClusterProven and must accept all verification submitted before use of any trademarks are permitted under the trademark license.

Regardless of whether an application is verified as meeting the criteria, or IBM has accepted such verification, no right to use the trademarks is granted until both parties have signed the ClusterProven Trademark Agreement.

8.7 Enterprise Storage Server

The Enterprise Storage Server (ESS) is the latest IBM storage product to be developed using IBM's Seascape architecture. It provides all the open systems functions of the Versatile Storage Server, and now with System/390 attachment, it provides all the functions of the 3990 Storage Control.

IBM has some exciting new functions for both the S/390 customer and the open systems customer. For S/390, IBM has some new features that significantly enhance performance. These features, which together with OS/390 software deliver a new world to the S/390 storage environment, are probably the biggest change since disk caching was introduced. For the

non-mainframe customer, we have delivered features that were previously only available on mainframe storage, functions that will enable you to better manage your data. For customers, we have the capacity and performance to meet the largest of your requirements.

The Seascope architecture is the key to the development of IBMs storage products, both now and in the future. Seascope allows IBM to take the best of the technologies developed by the many IBM laboratories and integrate them to produce flexible and upgradable storage solutions. Technologies, such as the PowerPC, the Magstar Tape drives, and IBMs award-winning Ultrastar disk drives. Serial interconnect technology, SSA, now delivering a 160 MB/sec data rate, is used within the Seascope architecture to connect the processors and internal disks to ensure high-performance and availability. Seascope also allows you to adapt to new technologies quickly, such as Fibre Channel and S/390 FICON.

IBM has already delivered Seascope solutions with the Virtual Tape Server (VTS), the Network Storage Manager (NSM), and the Web Cache Manager. The Versatile Storage Server was the first of the Seascope disk servers; now we have announced the Enterprise Storage Server, the follow-on to the VSS, utilizing the latest technology and taking advantage of the Seascope architecture's flexibility to upgrade and replace components easily and quickly. For detailed information regarding the Enterprise storage Server, please refer to the redbook *Implementing the Enterprise Storage Server in Your Environment*, SG24-5420. The following session is a brief overview extracted from that publication.

8.7.1 Positioning

The Enterprise Storage Server is the natural successor to the IBM 3990. It provides all the functions that were available on the 3990 including peer-to-peer remote copy (PPRC), extended remote copy (XRC), and concurrent copy. For the many customers that have installed the RAMAC Virtual Array (RVA), with its revolutionary log structured file (LSF) architecture, we have protected their investment in this technology too. The Enterprise Storage Server will, in the future, implement an LSF architecture. You will even have the choice of how much capacity you have under LSF arrays and how much under the existing ESS design. If you have implemented SnapShot, then your investment in the function will be protected on ESS, initially by a function (FlashCopy) that will provide fast, real, time zero (T0) copies, and later when IBM has LSF by the ESS SnapShot equivalent. IBM has made the software interface transparent to exploit RVA SnapShot, ESS FlashCopy, or even Concurrent Copy without changing your procedures.

The Remote Copy functions available on both 3990 and RVA are available on the Enterprise Storage Server with the same operational interfaces, therefore, allowing you to mix, for example, PPRC on 3990, RVA, and ESS in the same installation.

The Enterprise Storage Server also supports TPF, thus, providing a very high-performance solution for airlines and other TPF users who need a high-performance, high availability solution to support their critical business applications.

For the open systems or AS/400 customer, IBM has delivered the next generation of VSS. The Enterprise Storage Server builds upon the VSS and adds more function, more capacity, and more attachment capability. If you have an existing VSS, it can be configured to attach to an Enterprise Storage Server and utilize all of its installed capacity. In addition, IBM has introduced a remote mirroring capability for disaster recovery. Utilizing ESCON technology, you can locate the remote site at distances of up to 103 kilometers from the primary location.

The IBM 7133 (the D40 and 020) can be attached through a VSS expansion rack to protect your current investments.

8.7.2 Benefits

ESS can help achieve business objectives in many areas; it provides a high-performance, high-availability subsystem with flexible storage that can be configured according to your requirements.

- **Storage Consolidation**

The Enterprise Storage Servers high performance, attachment flexibility, and large capacity enable you to consolidate your data from different platforms onto a single high-performance, high-availability box. Storage consolidation can be the first step towards server consolidation, thus, reducing the number of boxes you have to manage and allowing you the flexibility to add or assign capacity when and where it is needed. ESS supports all the major server platforms, from S/390 to AS/400, and Windows NT to many of the flavors of UNIX. With a capacity of up to 11 TB, and up to 32 host connections, an ESS can meet both your high-capacity requirements and your performance expectations.

- **Performance**

The Enterprise Storage Server is designed as a high-performance storage solution and takes advantage of IBMs leading technologies. In today's world, where your business can reach global markets through e-business,

you need business solutions that can deliver high levels of performance continuously every day, day after day. You also need a solution that can handle different workloads simultaneously so that you can run your Business Intelligence models, your large databases for Enterprise Resource Planning (ERP), and your online and Internet transactions alongside each other with minimal impact.

Some of the performance enhancing capabilities of ESS are: Remote Copy, PPRC, XRC, FlashCopy, Concurrent Copy, I/O Priority, Queuing, Parallel I/O, Multiple Allegiance, Parallel Access, Volumes, Sysplex Wide, and I/O Tuning.

There may be a concern about running S/390 and Open systems workloads together on an Enterprise Storage Server because of the often widely differing workload characteristics. Typically, S/390 workloads are cache-friendly and take advantage of large caches; whereas, the open systems workloads are often very cache-unfriendly. For the S/390 workload, we have sophisticated cache management algorithms and a large cache. For the workloads that cannot take advantage of cache, we have high performance disk arrays with fast disks and serial interconnect technology. Therefore, whatever the workload, even mixed workloads, ESS delivers high-performance.

Another example of why an Enterprise Storage Server can deliver performance is the RAID design; the RAID function is managed not by the main RISC processors in the ESS, but at the disk loop level. Up to 16 RAID functions can be performed simultaneously, therefore, delivering fast response times and high throughput.

- Disaster recovery and availability

The Enterprise Storage Server has been designed with no single points of failure. It is a fault tolerant storage subsystem, which can be maintained and upgraded concurrently with customer operation.

Some of the enhanced functions of the ESS are: ESCON, PPRC - Synchronous Remote Copy, XRC - Asynchronous Remote Copy (OS/390 only), Protect data on S/390, UNIX, and Windows NT.

The Peer-to-Peer Remote Copy function is now recognized as the future for disaster recovery in the S/390 Sysplex world by all the leading S/390 storage vendors. PPRC, together with enhancements to OS/390 and Geographically Dispersed Parallel Sysplex (GDPS), lead the industry in high availability solutions. Recent Gartner analysis shows a Parallel Sysplex solution as having, on average, less than 10 minutes outage per year. With GDPS and PPRC, IBM is bringing the recovery time following a disaster into minutes rather than days.

The PPRC solution is available with the Enterprise Storage Server for UNIX and Windows NT. Management of the PPRC setup is through the ESS Specialist Web interface. Now, we have disaster solutions for many platforms using a simple and easy-to-use interface.

Finally, XRC has been enhanced, and OS/390 has a disaster recovery solution that you can use over long distances. Enhancements to XRC delivered by the Enterprise Storage Server eliminate the need to recopy the data should the Data Mover function fail. Now, the ESS will keep track of changes on the primary storage, and the Data Mover will just copy the changed data to the secondary storage after it is restarted.

- Instant copy and your backups:

For all environments today, taking backups of data probably takes a long time. Even though we have high-availability storage that is fault tolerant and protected by RAID, backups need to be taken to protect data from logical errors and disasters. Backups are often taken outside prime shift because of the impact to normal operations. Databases must be closed to create consistency and data integrity, and the online systems are normally shut down.

To help reduce the impact of backups and other copy requirements, IBM introduced an instant copy function, called SnapShot, on S/390 and the RVA. SnapShot used the unique architecture of the RVA to be able to take a volume or dataset copy almost instantaneously. Then, you could take your backups from the copies in parallel with normal processing. This enabled RVA customers to save valuable time (a study has shown an average of four hours) out of the backup window, therefore, not only saving time, but also requiring almost no additional capacity to take the copies.

The Enterprise Storage Server, although it does not have the architecture that can perform instant copies without using disk capacity, does have an equivalent function called FlashCopy. This new function not only applies to the S/390 but also to all the other platforms.

- Data Sharing

Data Sharing is one of those areas where there has been much discussion, but little in the way of delivery. IBM has defined three levels of data sharing:

- Partitioned Storage:

Data is consolidated onto a common storage box, but the capacity is partitioned between the different attached hosts. An example of this is VSS and also ESS. The Enterprise Storage Server can attach to S/390, AS/400, Unix systems, and Windows NT. The advantage of ESS is that

storage can be dynamically added to any of the hosts or reallocated from one host to another.

- Data Copy Sharing:

Data is copied from one platform to another and, at the same time, may undergo some form of translation and reformatting so that the other platform can understand the data. An example of this is the IBM InfoSpeed, which transfers data at channel speed between OS/390 and UNIX or NT.

- True Data Sharing:

True data sharing between homogeneous hosts has been available for many years; for example, OS/390 IMS Datasharing has been available for more than 15 years. Similar data sharing capabilities exist on Compaq VMS and Sun systems.

In terms of datasharing with database access from homogeneous hosts today we have S/390 with DB2 and IMS/DL1. UNIX systems can use the Oracle Parallel Server to share data. But there are as yet no cases where one can share databases between heterogeneous systems. The issue to be resolved is not a hardware one (it is easy to physically share disks) but rather a software one. The software on each platform must be able to understand the format and content of the data and manage database integrity through common locks, logs, and recovery facilities.

Sharing of non-database data is easier providing both parties understand the content of the data. Agreeing on a common format will be a first step to true datasharing. The Seascope architecture of the Enterprise Storage Server will enable one, in the future, to include powerful new functions to start down the path towards true datasharing. ESS has the powerful, intelligent, UNIX-based RISC processors to enable one to achieve this.

- Storage Area Networks (SAN) Announcement Preview

For the open systems customer who is looking at Storage Area Networks and Fibre Channel, ESS supports a variety of Fibre Channel attachment options. Initially, support is provided by the IBM SAN Data gateway, which provides support for Fibre Channel attachment to ESS SCSI ports. As servers migrate from SCSI to Fibre Channel, the IBM gateway strategy allows ESS to support this migration while protecting customer investments.

IBM is previewing plans for the ESS to support native Fibre Channel, providing a basis for future development of full SAN exploitation in areas,

such as disk pooling, file pooling, and copy services. Up to 16 Fibre Channel ports will be available on an ESS. Each port will support point-to-point and fabric (switched) connections as well as FCAL. Fibre Channel ports also support FICON, the Fibre Channel interface for S/390 servers. Fibre Channel ports will be available as an upgrade option for installed ESSs.

IBM has more than eight years production experience with fiber, with its ESCON technology being based on early Fibre Channel developments.

The key to SANs is their management. In the ESCON SAN environment, the infrastructure is managed by using System Automation for S/390, and the data is managed with DFSMS/MVS. In the UNIX and NT environments, there is the Storwatch range of products that also manage the fiber infrastructure and the storage capacity as well as performance measurement information.

8.8 Customer example

This section provides a scenario using a manufacturing company in Sydney, Australia as an example. This company is one of Australia's largest food manufacturers, producing many of Australia's and New Zealand's most popular and well-known brands as well as products and ingredients for the food service, commercial, and industrial sectors.

Business Need: This company's multiple business units did not have common data processing capabilities. Some applications were not Y2K ready, and there was a duplication of common services, such as purchasing and distribution across multiple business units. The customer wanted to standardize on a common ERP system across its various business units and needed a new server system to run the new applications.

Description of Solution: The customer selected SAP R/3 as the common ERP applications to be used across the business units, therefore providing a central group IT capability. This company can now implement common, shared services across the multiple business units with improved efficiency and less overhead. With 3,000 users online to a single data center at a central location, this was a major change in the customer's IT direction and represented a significant investment in mission critical technology and infrastructure. The solution consists of two frames of RS/6000 SP thin and wide Silver nodes for application servers and three RS/6000 Model S7A servers. One S7A is used for a SAP production database; one is used as an application and HACMP failover server, and one is used for production

staging, thus, representing a duplicate of production staging environment for SAP software and performance testing.

Users log onto SAP from remote PC terminals running the SAP GUI. All SAP data processing occurs on the SP nodes using 7133 Serial Storage Architecture storage. ADSM for AIX is used to manage backup and restore functions, and 3590 Tape Subsystems and a 3494 Tape Library are used to back up data. Dual SP Switch routers and communication networks have been established for redundancy. This company has 1.3 TB of SSA spread across 16 drawers. Mirroring is used on the SAP production S7A system, and RAID 5 is used on the SP thin Silver node SAP development system and also on the ADSM server wide Silver node. The 3494 Tape Library has four drives.

IBM/SAP designed and installed all of the technical infrastructure including a performance guarantee. The equipment is now outsourced to GSA for ongoing management.

- RS/6000 Details
 - Type and number of RS/6000 installed: Two frames of RS/6000 SP with 14 thin Silver nodes and two wide Silver nodes (additional thin Silver node application servers will be added as additional business units are added), three RS/6000 S7As.
 - Type of applications being run on the RS/6000: SAP R/3 (development, test, and training for all of the production applications. Production SAP applications), Oracle Sales Analyzer, Manugistics, ADSM.
 - Number of users: The customer planned for 3,000 named users. Current production implementation is 600 named users.
 - Topology: TCP/IP SP Switch, Ethernet
 - Version of AIX being used: V4.3
- SAP R/3 Details:
 - ISV software application and release: SAP Version 3.1H
 - Platform: RS/6000 SP, three S7As
 - Operating Systems: AIX
 - Database: Oracle
 - Disk by GB: 1.2 TB of SSA
 - Tape facilities/drives: 3494 plus four 3590s.
 - Number of current active users: 600 named, 400 active.
 - Number of planned users: 3,000.

- IT Infrastructure software and technology: ADSM/Backint, Tivoli, HACMP, ADSM, SP Switch router.
- e-business collaboration and messaging software: MQSeries, Lotus Notes.
- ADSM Details:
 - Server: ADSM for AIX (SP wide node).
 - Version of ADSM: ADSM Version 3.1.
 - Disaster Recovery Manager (DRM): Yes.
 - Number and type of clients: 25 AIX clients, 10 Microsoft Windows NT clients.
 - Backing up: 1.3 TB of SSA disk backup via 300 GB SSA disk RAID-5 attached to ADSM server then backed up to four 3590s and 3494.
 - After data backup, does customer migrate data: Yes.
 - Database: Oracle V7.
 - Is database backup occurring: Off-line + Backint agent for online and off-line SAP backup.
 - Why did customer choose ADSM over the competition: ADSM was one element of total IT infrastructure solution.
- 3590/3494 Details:
 - Number of 3494 Tape Libraries: One library.
 - 3494 Library Cartridge Capacity (if present): 594.
 - 3494 Library Tape Drives: Four.
 - Tape Drive Attachments: SCSI.
 - Control Programs: AIX.
 - Key Applications: ADSM.
 - Tape Management System: ADSM.
 - Benefits: The environment is now managed by GSA as an outsource contract. As this is a new data center installation, there are no direct comparisons available.
- 7133 SSA Details:
 - Description of capacity/configuration: 1.3 TB (14 drawers of 16 x 4.5 GB, plus 2 drawers 16 x 9.1 GB. One SSA adapter per drawer.
 - Mirroring: Yes, on SAP production S7A system.

- RAID: Yes, on SP thin Silver node SAP development system and also ADSM server wide Silver node (with SSA adapter cache).
- RAID Level: RAID 5.
- Multi Host: Yes, for HACMP attached to two S7As.
- IBM 7190 or a VICOM product: No.
- Fibre-optic extenders: No.
- Processor Type(s)/Model(s): RS/6000 SP S7A and SP wide and thin Silver nodes.
- Operating System(s): AIX
- Main application/workload and database stored on the 7133(s): SAP development, test, and training SAP production, Oracle Sales Analyzer, Manugistics, ADSM.
- Storage Management Application: ADSM.
- Benefits/Value to Customer: Effective teaming and information sharing, enabling business transformation, improved customer service, improved quality, reduced costs.

This company went into live production with SAP R/3 with one business unit in March, 1999 as part of a staged roll out. Business benefits have yet to be quantified. IBM was on time with all equipment delivery and technical implementation.

Chapter 9. Server consolidation for key business applications

Server consolidation is a solution that helps customers deliver higher IT service levels in a more-cost effective fashion by optimizing both the quantity and distribution of servers and by integrating data and applications supporting their mission-critical IT functions. The concept of server consolidation has expanded to include a broader range of activities and strategies. Customers can benefit from server consolidation in every business area because server consolidation is an enabling technology for the cross industry applications and business processes, such as e-business, business intelligence, ERP, customer relationship management solutions, supply chain management solutions, and so on.

9.1 Overview

Inter/Intra Business Integration is a core business need. Server consolidation has expanded to cover the broad range of business areas. Server consolidation is an enabling core solution for business integration. Server consolidation topics related to business integration will be presented in the following discussions.

This chapter also presents cross industry solutions for business applications and the server consolidation for key business applications from the viewpoint of data/application integration.

9.1.1 Server consolidation for data/application integration

Server consolidation is broader than just physically combining assets. As presented in previous chapters, we have defined four general classes of server consolidation implementation strategies based on a review of existing and planned customer consolidation architectures:

- Centralization
- Physical consolidation
- Data integration
- Application integration

The last two area of consolidation, data/application integration, are the most complex. With application integration, we are attempting to break the paradigm of *one application to one server*. Server consolidation has evolved to deliver higher IT service levels in a more effective way by integrating data and applications.

Figure 66 on page 270 shows the server consolidation classification.

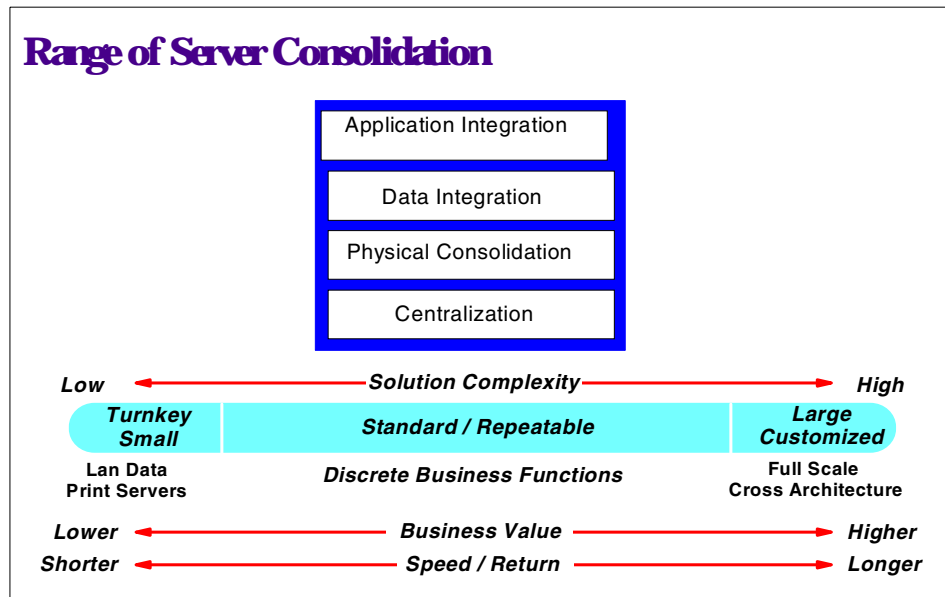


Figure 66. Server consolidation classification

From the viewpoint of end-users, applications are their major focus. From the viewpoint of IT managers, they focus on infrastructure as well as application. IT managers have tried to provide higher application services in more efficient infrastructures.

Server consolidation involves infrastructure optimization and architectural Integration to provide IT services in a more effective way.

At its core, server consolidation is an enabling solution for:

- Optimizing existing IT infrastructure
- Integrating existing architectures across applications/data
- Providing a foundation for new solution investment and implementation

9.1.2 Expanded server consolidation for business integration

Business integration is a new IT trend. Business integration involves data integration, application integration, and platform and architecture consolidation. It improves business effectiveness based on the best shared use of operational information. It provides IT capability to dynamically

manage changing business environments. Business integration will generate more server consolidation opportunities.

Server consolidation is an enabling background solution for business integration. It enables intra- and inter-enterprise business integration based on proven products, services, and methodologies.

Enterprises can benefit from server consolidation in the following business opportunities:

- Merger and acquisition integration
- Automating business processes
- Supply chain integration
- Customer relationship management integration
- Enterprise Resource Planning (ERP) application integration
- e-business application integration

Figure 67 shows the changing environment of business applications, where the business integration is the trend.

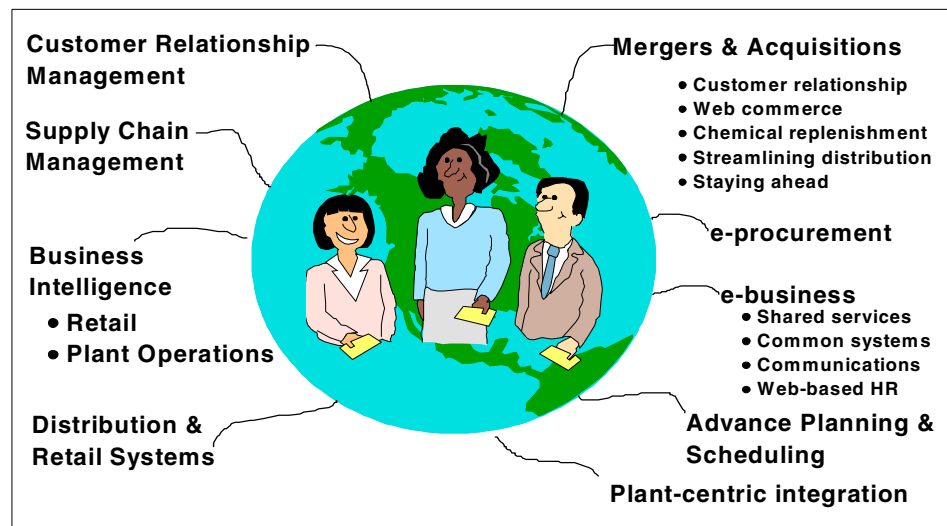


Figure 67. Consolidating business applications

9.2 Cross-industry solutions for business applications

Cross-industry solutions are the popular base softwares widely used for business applications.

- Database Management System (DBMS)

- Online Transaction Processing (OLTP)
- DSS - Business Intelligence Solutions
- Enterprise Resource Planning (ERP)
- Groupware - Lotus Notes
- e-business solutions

These cross-industry solutions are key software components for business applications. These solutions have been developed to deploy the client/server, distributed computing and Internet environment. In these environments, server consolidation has been widely adapted to implement these cross-industry solutions in a more effective way.

Figure 68 shows the RS/6000 SP system for server consolidation. RS/6000 SP system has been widely used as a flexible consolidation server.

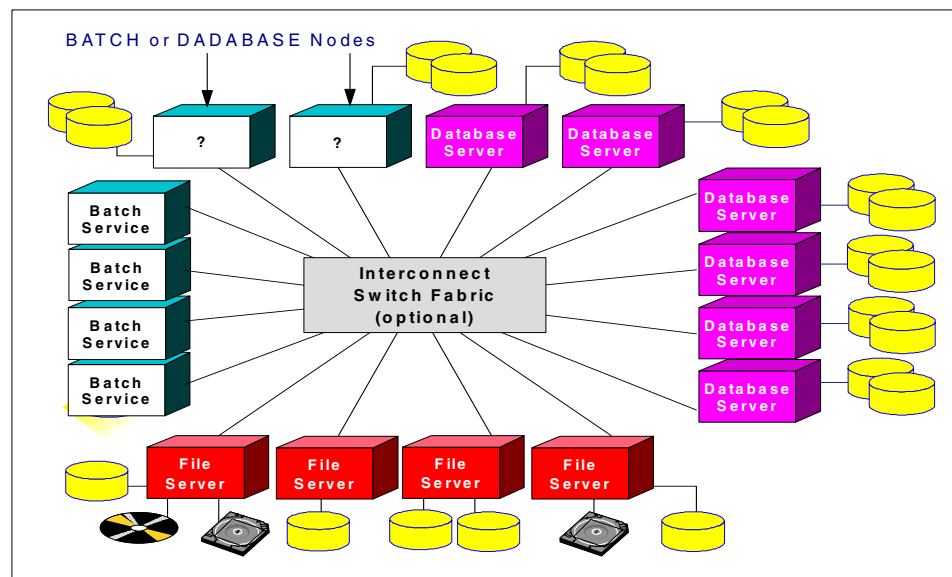


Figure 68. RS/6000 SP as a flexible consolidation server

9.2.1 DBMS

In this section, we will describe RS/6000 core databases for business applications. These consist of the following:

- DB2
 - DB2 Database Server

- DB2 Universal Database
- Oracle
 - Oracle 7.3, Oracle 8
 - Oracle 8i
 - Oracle Express
- Sybase
 - Sybase Adaptive Server
 - Sybase MPP
 - Sybase IQ
- Informix
 - Informix Universal Server
 - Informix Online Dynamic Server
 - Informix Online Extended Parallel Server
 - Red Brick

DB2 Universal Database is IBM's object-relational database for UNIX, OS/2, and Windows operating environments.

- DB2 Universal Database Personal Edition: Provides a single-user, object-relational database management system for your PC-based desktop that is ideal for mobile applications or the power-user.
- DB2 Universal Database Workgroup Edition: A multi-user, object-relational database for applications and data shared in a workgroup or department setting on PC-based LANs. Ideal for small businesses or departments.
- DB2 Universal Database Enterprise Edition: A multi-user, object-relational database for complex configurations and large database needs ranging from Intel to UNIX platforms and from uniprocessors to the largest SMPs. Ideal for midsize to large businesses and departments, particularly where Internet and/or enterprise connectivity is important.
- DB2 Universal Database Enterprise-Extended Edition: Provides a high performance mechanism to support large databases and offers greater scalability in Massively Parallel Processors (MPPs) or clustered servers. Ideal for applications requiring parallel processing, particularly data warehousing and data mining. This is the upgrade path from DB2 Parallel Edition.

Oracle8 is the universal data server from Oracle. It delivers advances over the characteristics of Oracle7 technology. Oracle8 is a key component of the Network Computing Architecture of Oracle.

Oracle8i is the Database for Internet Computing. Oracle8i includes several key features to develop Web-enabled applications for Internet and enterprise applications. This new release of the Oracle database server has a Java Virtual Machine built in to allow developers to write, store, and execute Java code within the database itself. Oracle8i is the mission-critical application platform specifically designed for Internet, intranet, and extranet applications. To keep Web applications running smoothly, Oracle8i leverages the power of Oracle Parallel Server to protect against system downtime. Oracle Enterprise Manager offers the management framework to support Internet environments.

Oracle8i product family:

- Oracle8i Enterprise Edition
- Oracle8i Personal Edition
- Java Products

Oracle8i Business Solutions:

- Web Information Management
- Electronic Commerce
- Data Warehousing
- Application Integration
- Manageability
- Mobile Computing
- Online Transaction Processing

Sybase Adaptive Server IQ12 is designed for high-performance data analysis with a combination of innovative query processing technologies, exclusive column-based indexing, and performance-optimized algorithms. It delivers fast, ad hoc query performance and maintains performance even as the number of users increases and users' questions change to meet new business requirements. Adaptive Server IQ12 does not require resource-intensive tuning to obtain excellent performance.

As part of Sybase's Adaptive Component Architecture, Sybase provides three data stores to fit users' needs:

- Adaptive Server Enterprise
- Adaptive Server Anywhere
- Adaptive Server IQ

Informix Dynamic Server.2000 is designed for the most demanding OLTP, e-commerce, and Web applications. Informix Dynamic Server.2000 is fully integrated with Informix's unique extensibility technology including support for DataBlade modules. Informix Dynamic Server/AD/XP meets the demands of large-scale enterprise decision support applications. Informix Red Brick Warehouse is optimized for use in non-IT environments. Informix also sells a number of other database server products aimed at specific end-user and OEM markets. This includes Informix SE for embedded applications and popular C-ISAM libraries.

Informix Extended Parallel Option provides:

- Unlimited Scalability
- VLDB Manageability
- Parallel SQL Functionality
- Partitioning of Data
- Transaction Logging
- Recovery

9.2.2 OLTP mission-critical applications

Online Transaction Processing (OLTP) is the solution for any high-end business where the systems are highly available and utilize UNIX.

9.2.2.1 Distributed System

A *distributed system* consists of multiple software components that run in separate independent processes on different machines in a network. Even though the software is distributed across a network, it can be accessed reliably by multiple users as if it were running on a single system. Bank representatives in diverse locations can process a debit or credit request, accessing the programs and data as if they were local.

A *transaction* is a set of related operations that must be executed as a unit (though each operation may run in a different process).

2-Tier and 3-Tier Concepts

One of the distributed architecture forms is the three-tiered client/server model. In this model, a client/server system organizes software into two separate parts, clients and servers. *Clients* interact with users and make requests to servers. *Servers* perform services, such as updating or retrieving data, in response to requests from clients.

Figure 69 on page 276 shows the typical 3-tier architecture in the distributed computing environment.

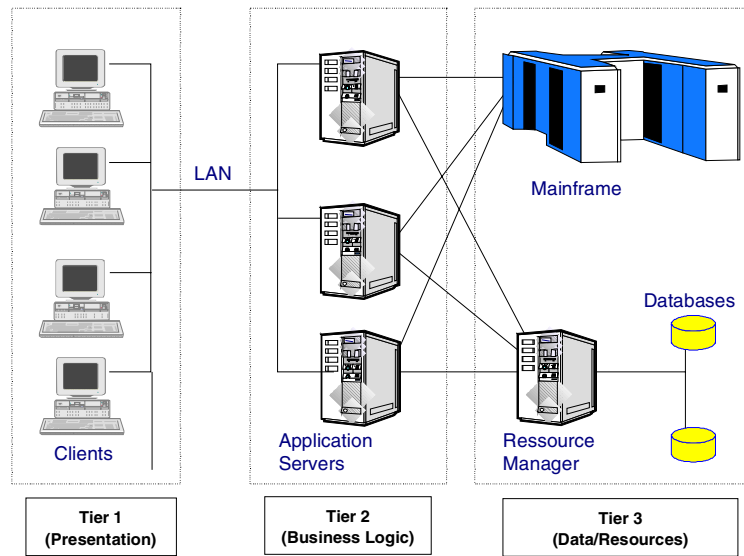


Figure 69. 3-tier architecture

Servers communicate with resource managers. A *resource manager* is an application that manages shared data, such as an Oracle or DB2 database used to hold account information.

The first tier contains presentation software, which consists of client applications that interact with users via screens or command-line interfaces. The client applications send requests to server applications in the second tier.

Second-tier applications contain the business logic, while the third tier contains data and resources, thus, separating them from processing logic.

For smaller applications, other solutions exist, such as the 2-tier approach, where the presentation and business logic are merged.

Figure 70 on page 277 shows the 2-tier and 3-tier architectural concepts.

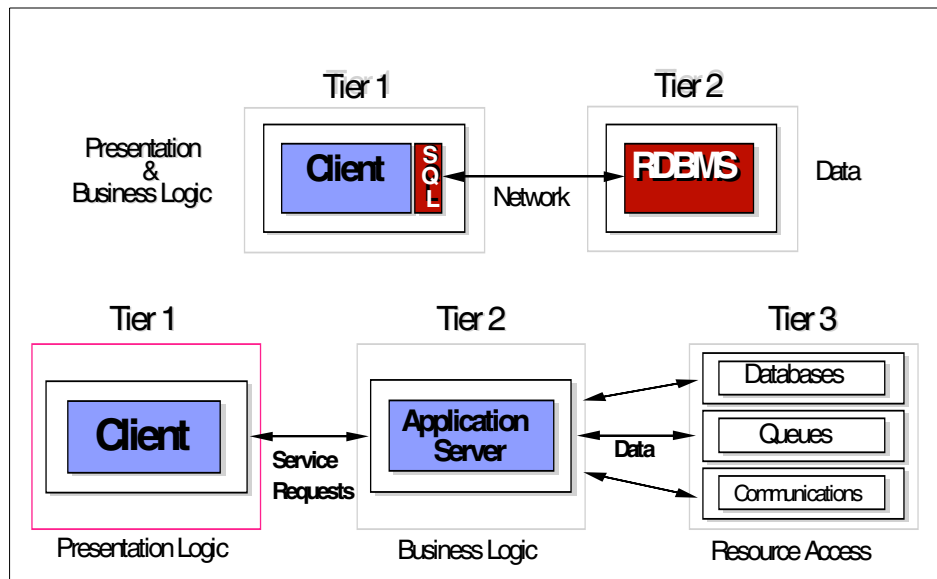


Figure 70. 2-tier and 3-tier architecture

The differences between these two architectures are as follows:

The 2-tier concept solution is database-oriented with a single source of data. This approach is ideal for:

- Applications supporting under approximately 100 users
- Applications with low security requirements

The 3-tier concept solution is ideal for domains with the following requirements:

- Flexibility
- Scalability with large numbers of users
- Multiple data sources support
- Transaction processing monitors for manageability, flexibility, and security

9.2.2.2 TP Monitors: CICS, Encina, and Tuxedo

The basic idea of transaction monitors is to allow different parts of an application to communicate. These products are used for developing, executing, and administering distributed transaction processing systems.

CICS

CICS is IBM's general-purpose, online transaction processing (OLTP) software. It is a powerful application server that runs on a range of operating systems ranging from the smallest desktop to the largest mainframe. CICS seamlessly integrates all the basic software services required by OLTP applications and provides a business application server to meet the information/processing needs of today and the future.

Each product in the CICS family is designed to run on a particular operating system and hardware platform, and each product has powerful functions to allow interproduct communication with other members of the CICS family. CICS provides a cost-effective and manageable transaction processing system, therefore, allowing you to write your own applications or to choose from many existing vendor-written business products.

A transaction management system (sometimes called a transaction monitor), such as CICS, performs the following:

- Handles the start, running, and completion of units of work for many concurrent users.
- Enables the application (when started by an end user) to run efficiently, to access a number of protected resources in a database or file system, and then to terminate, normally returning an output screen to the user.
- Isolates many concurrent users from each other so that two users cannot update the same resource at the same time.

CICS is a layer of middleware that shields applications from the need to take account of exactly what terminals and printers are being used while providing a rich set of resources and management services for those applications.

In particular, CICS provides an easy-to-use application programming interface (API), which allows a rich set of services (including file access and presentation services) to be used in the application and to be ported to and from a wide variety of SP environments.

Encina

Encina provides the infrastructure, development environment, and monitoring capabilities necessary for building and managing distributed transaction processing systems.

Core Distributed Services: The Encina Toolkit Executive and the Toolkit Server Core, together with Distributed Computing Environment

(DCE) services, provide an infrastructure that ensures security, recoverability, and reliability for the distributed system.

Transaction Processing Monitor: The Encina Monitor orchestrates activities in the distributed system by keeping server processes running, routing transactions across the system, and making sure that distributed work is completed successfully and accurately.

Application Development Environment: Encina provides a rich set of interfaces and language extensions for developing custom client and server applications. These include the Monitor Application Programming Interface (API), the TX interface, Encina++, and Transactional-C.

Recoverable Servers: The Recoverable Queueing Service (RQS) and Structured File Server (SFS) manage data stored in queues and record-oriented files, respectively. Both RQS and SFS are built using the Encina Toolkit services. The Toolkit Services can also be used to build custom-recoverable servers. Custom-recoverable servers can interact with Encina native resource managers (RQS and SFS) and with other resource managers. Encina also provides APIs for writing RQS and SFS clients.

Central Administration: The Enconsole graphical user interface (GUI) permits central administration and monitoring of Encina.

Monitor Systems: The Encina Control Program (enccp) is a command-line and scripting interface for administering Encina.

Monitor Cells: XA Distributed Transaction Processing (DTP) Integration. The Transaction Manager-XA (TM-XA) interface is used to develop applications that can interact with XA-compliant resource managers. XA-compliant resource managers follow the X/Open standard, which is a blueprint on how resource managers must behave in order to support transactional access. The TM-XA interface coordinates Encina's two-phase commit protocol with RDBMS.

Mainframe Connectivity: The Peer-to-Peer Communications (PPC) Services permit connectivity between Encina and mainframe systems using a standard Systems Network Architecture (SNA) LU 6.2 protocol. The PPC Services coordinate Encina's two-phase commit protocol with mainframe systems.

BEA Tuxedo

BEA TUXEDO provides a platform of services required for the development of mission-critical applications. With BEA TUXEDO, developers can focus on

providing the functionality that is important to the business and that can give the business a competitive edge.

Distributed Transaction Management: BEA TUXEDO offers services for cooperative processing that allow clients and servers to participate in a distributed transaction. BEA TUXEDO manages two-phase commit processing in a way that is transparent to the applications. BEA TUXEDO is based on the X/Open DTP model, which includes three basic elements: The application program, the transaction manager, and the resource managers.

Application to Transaction Manager Interface (ATMI): BEA TUXEDO uses an Application Programming Interface (API) known as ATMI. ATMI enables developers to write BEA TUXEDO applications regardless of the hardware on which the program will reside.

Dynamic Workload Balancing: BEA TUXEDO automatically generates and manages parallel copies (replicas) of applications and performs all the needed workload balancing among the copies to ensure that they are all evenly utilized. This is true whether the copies are on the same node or spread across nodes.

Transaction Queuing: BEA TUXEDO provides transaction queuing (/Q) to allow distributed applications to work together in an asynchronous, connection-less fashion. /Q is a modular store-and-forward capability that prioritizes queues based on message context, message content, and time of day.

Application Parallelization: By dynamically replicating distributed applications throughout the enterprise, BEA TUXEDO *parallelizes* without any special programming efforts or additional resources. This makes the system more efficient, therefore, allowing the same types of transactions to be processed simultaneously on different, distributed nodes.

Data Dependent Routing: BEA TUXEDO can route messages based on their content using a feature called Data Dependent Routing (DDR). This feature enables transactions to be processed where the data can be most efficiently utilized.

Automatic Recovery: BEA TUXEDO provides automatic recovery from application failures, transaction failures, network failures, and node failures. When an application component fails, the system monitor notifies the node manager and restarts the failed process. When an application program fails, the server manager recovers the failed program by rolling back the transaction that was in

progress. Node failures are also detected and recovered by the system monitor.

The RS/6000 SP system and BEA TUXEDO are ideal for each other. BEA Tuxedo was designed to allow business applications to be distributed, and the SP provides a great distributed hardware environment to host such a partitioned system.

Figure 71 shows the RS/6000 SP System for the 3-tier environment. The RS/6000 SP system has been widely used for the distributed OLTP environment as a flexible Server Consolidation platform.

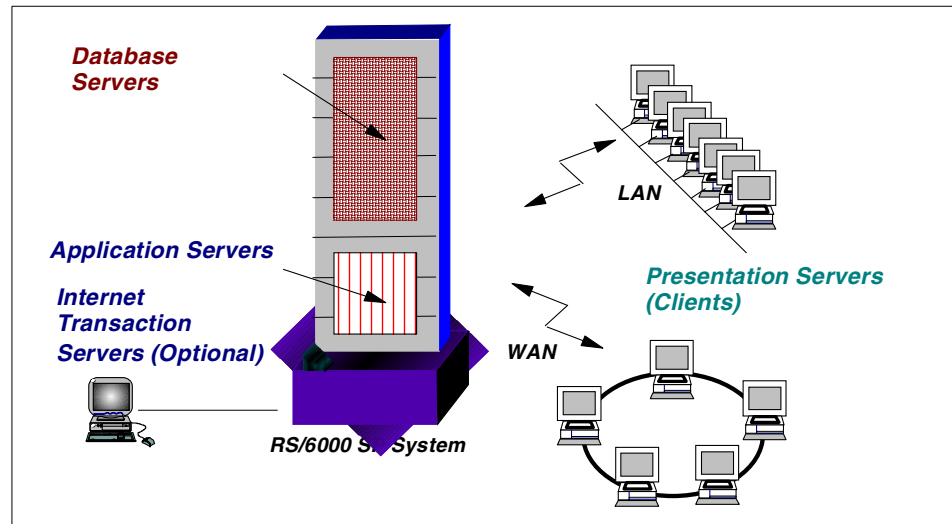


Figure 71. RS/6000 SP system for the typical OLTP 3-tier environment

9.2.3 Decision Support Systems (DSS)

Decision Support Systems (DSS) provide *business intelligence* to explore data and to enhance business decisions. Decision support systems include the following.

Data Warehouse: Defines a collection of data designed to support management decision making. The core database is a central repository for enterprise-wide data. A warehouse is normally a large relational database that comes with its own set of tools for loading, creating indexes, and performing other standard utility operations.

Data Marts: Are designed to focus on the information requirements of a single department, function, or group. Ideally, data marts are populated with data drawn from the data warehouse. Sometimes these marts employ special techniques to improve efficiency when dealing with queries in a single object context. These techniques include Online Analytical Processing (OLAP).

OLAP Tools: Enable users to analyze different dimensions of multidimensional data. For example, they provide time series and trend analysis views. Data access is provided by a mixed set of tools on the end user desktop. These tools range from personal databases to spreadsheets and desktop OLAP software and also extend to intranet or Internet access.

Data Mining: Is generally a batch process where computer algorithms shift through data to detect previously unknown patterns and correlations. Data can be discovered with data mining tools, such as Intelligent Miner (IM). IM is a suite of tool functions that support data mining operations and deploy a variety of techniques to:

- Create classification and prediction models.
- Discover associations and sequential patterns in large databases.
- Automatically segment databases into groups of related records.

Figure 72 on page 283 shows the Business Intelligence Framework. Typically, data must be transformed in some way as it travels from operational systems to the warehouse. A fully populated column in a warehouse table may come from several different operational systems, and logic needs to be maintained on what the selection criteria was from each source and what the transformation logic is to map that to the warehouse target. This mapping is stored by *metadata* tools.

RS/6000 and SP systems are the business computer systems widely used for Business Intelligence Solution areas.

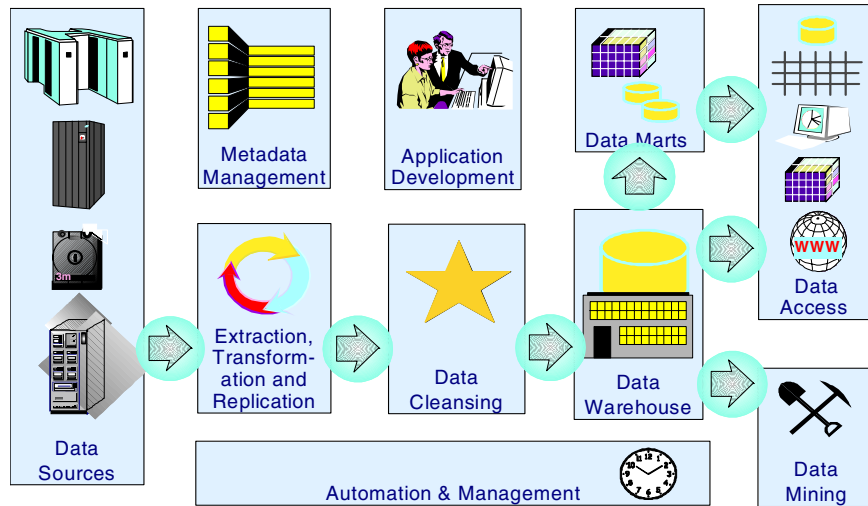


Figure 72. Business Intelligence framework

9.2.4 ERP

There are several ERP products available for SP. The leading companies in the international ERP marketplace are:

- SAP R/3
- Oracle Applications
- PeopleSoft Applications
- JD Edwards' OneWorld

SAP R/3

SAP is one of the largest standard application software companies in the world. Its R/3 system, a client/server set of applications for distributed enterprises based on a flexible, tiered architecture, is designed to integrate and consolidate the financial, sales, distribution, production, and other functions within a company.

SAP R/3's suite of client/server data processing products relies on combining all the business activities and technical processes of a company into a single integrated software solution. Users do not have to access different databases for data created from different divisions of a company. Instead, data from a wide variety of sources, such as financial, asset management, controlling, production planning, project system, quality assurance, and human

resources, can coexist in the same database and be accessed in real-time from a variety of applications.

SAP R/3 Design

The design of SAP R/3 is based on three categories of platform-independent software services (see Figure 73).

Presentation services: These consist of graphical user interfaces on PC, X-terminal, or workstation machines.

Application services: These services provide the application logic, and they may run on one or more UNIX or Windows NT-based systems. Application services may run in batch or interactive mode and are responsible for gathering monitoring data to be dispersed to the presentation services.

Database services: These are the services that a typical database engine would provide. SAP R/3 can use several database engines depending on the particular needs of the client. These services would typically run on a mainframe or on a cluster of UNIX or Windows NT machines.

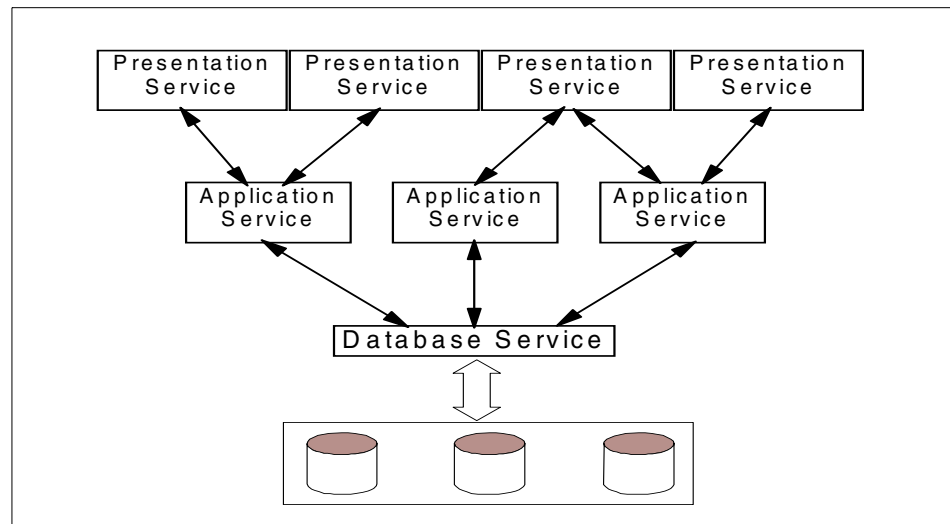


Figure 73. Conceptual model of SAP R/3 structure

An application service is designed and implemented in a layered way, therefore, isolating the SAP R/3 application logic from the operating system-dependent services. A middleware layer, called the *basis layer*,

communicates with the operating system and the network. Figure 73 on page 284 shows the layering of an application service.

Services are provided by administrative units called *instances* that group together components of SAP R/3. When first installed, the SAP R/3 system has a central instance, which has services, such as dialog, update, enqueue, batch, message, gateway, and spool. After installation, these services can be moved to other application servers in order to balance workloads.

Figure 74 shows the typical ERP configuration for Client/Server Computing environments. The Distributed Computing environment consists of many Servers working for Client/Server Computing.

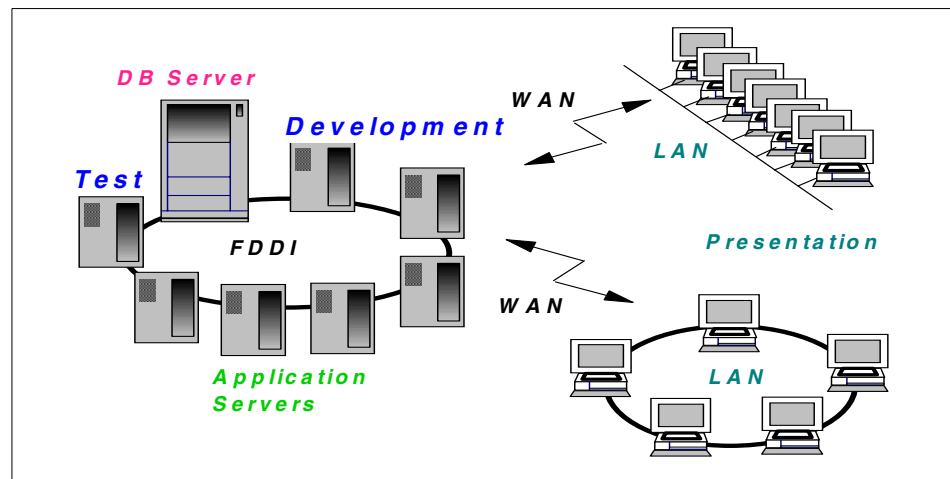


Figure 74. Typical ERP client/server configuration

RS/6000 SP implementation

Figure 75 on page 286 shows the RS/6000 SP Consolidated Configuration for the typical ERP environment. ERP users can benefit from Consolidated RS/6000 SP system Configuration for ERP solutions.

The SAP R/3 implementation on the RS/6000 SP follows a three-tier model. Contrary to most multi-tier configurations that use TCP/IP for network communication, the RS/6000 SP can take advantage of its internal fast switch to expedite network traffic and do so in a highly available manner. Because of its HACMP/ES capabilities, a cluster of nodes that acts as a database service would sense a node that fails and take appropriate steps to compensate for the loss.

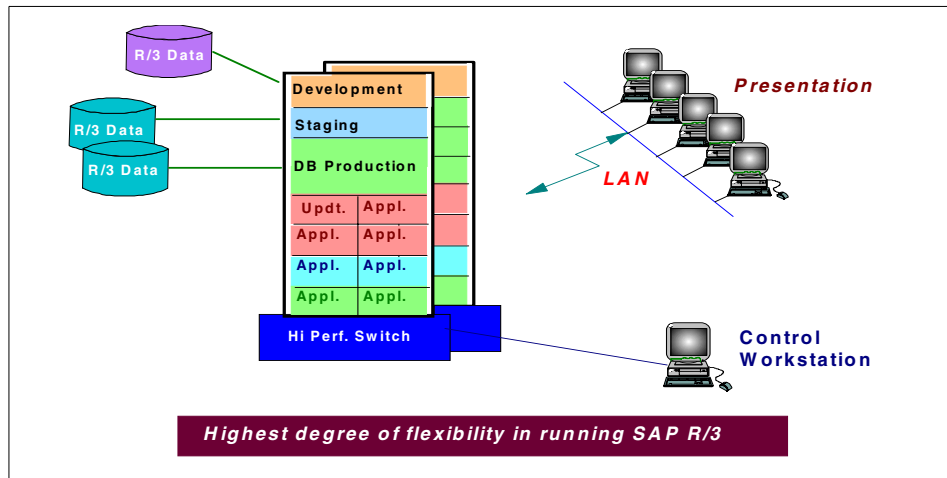


Figure 75. Typical ERP SP consolidated configuration

9.2.5 Groupware: Lotus Notes

Lotus Notes is a software application that promotes an effective workgroup computing environment. Lotus Notes provides the technology that allows people to collaborate regardless of software or hardware platforms, or technical, organizational, or geographical boundaries. There is a lot of literature in the market for implementing, configuring, and using Lotus Notes. For this, refer to the Lotus Notes official Web site at: <http://www.lotus.com>.

Groupware is used as a concept to characterize software whose objective is to improve the sharing of information among members of a group or a team. Lotus Notes incorporates a shared, distributed document database, which resides on a server, an easy-to-use graphical user interface, a built-in e-mail function, and built-in security to allow access control for critical information. Furthermore, Lotus Notes provides a rapid application development environment so that users can create custom applications that reside on top of Lotus Notes, therefore, enabling Lotus Notes to conform to particular business needs. The improved coordination that Lotus Notes offers provides a competitive advantage for organizations that choose to utilize the groupware capabilities of the program.

Figure 76 on page 287 shows the RS/6000 SP system for Lotus Notes Solution. The high speed of the SP Switch alleviates bottleneck problems and offers a Lotus Notes server with a single point of control. These benefits, along with the high availability infrastructure and scalability potential of the RS/6000 SP, make it an excellent choice for Lotus Notes implementations

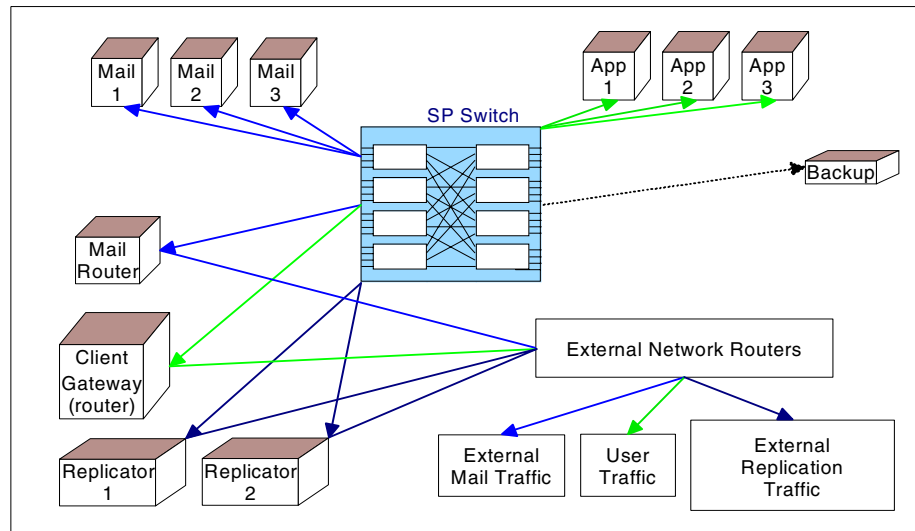


Figure 76. Logical model of Lotus Notes on the RS/6000 SP

9.2.6 e-business solutions

The RS/6000 Solution Series for e-business is currently comprised of the following offerings:

- Lotus Domino R5
- WebSphere Application Server - Standard Edition V2.0
- WebSphere Application Server - Advanced Edition V2.0
- WebSphere Application Server - Enterprise Edition V3.0

RS/6000 Solution Series for Lotus Domino

Domino is a server based product from Lotus Corporation of IBM that is positioned in the e-business marketplace in many segments since it provides solutions to a wide range of customer business problems. Domino today is marketed by Lotus as a messaging/groupware system with Internet/intranet dynamic application serving. The high-level functions Domino supports are e-mail, directory, calendaring, workflow, data/document storage, application development and delivery, with access to existing transaction systems and databases.

Lotus Domino is a messaging system. Lotus Domino is used for messaging in almost all implementations, and in many instances, Domino is installed only for messaging. Domino uses the Internet MIME and S/MIME formats for mail and has standard interfaces to SMTP and TCP/IP. Most all advanced

functions of e-mail are included in Domino, such as encryption, tracing, format translations, attachments, logging, and so on. Domino contains a full calendaring and scheduling function that provides free time searches, sending of meeting notices, time zone support, Internet access, security, and attachment capabilities.

Domino is a generic name of the server product. It comes in three levels: Domino Mail for mail, groupware, calendaring, and Internet support functions; Domino Server, which is mail plus application development capabilities; and Domino Enterprise, which adds clustering capabilities on top of Domino Server. The interface to Domino can be either through a Web browser or through what are called Notes Clients. Notes clients run on NT and Window 95 systems and provide a client/server aspect to the Domino server with a Windows look and feel. Figure 77 shows the Domino Configuration.

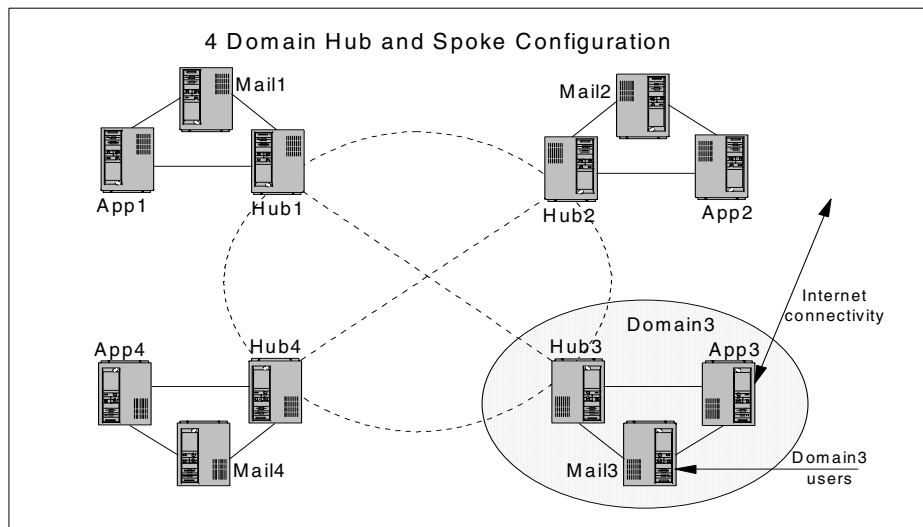


Figure 77. Enterprise-Wide Domino configuration

IBM WebSphere Application Server

IBM WebSphere Application Server, a member of the WebSphere Software product line, serves as the foundation for building e-business Web sites.

There are three editions of the WebSphere Application Server:

The Enterprise Edition for WebSphere Application Server, Version 3.0:

Provides the highest levels of security, performance, and availability. It offers comprehensive support for integrating existing and new IT resources for extensive enterprise and e-business connectivity. Combining distributed object and business process

features with world-class transaction processing features, Enterprise Edition can be used for the development, deployment, and management of heterogeneous business-critical applications. Enterprise Edition contains all of the features of the Advanced and Standard Editions.

The Advanced Edition, Version 2.02: Introduces server capabilities for applications built to Sun's Enterprise JavaBean specification. Deploying and managing JavaBean components provides a stronger CORBA implementation that maps to portable Java technologies. Advanced Edition contains all of the features of Standard Edition.

The Standard Edition, Version 2.02: Combines the control and portability of server-side business applications with the performance and manageability of Java technologies to offer a comprehensive Java-based Web application platform. It enables powerful interactions with enterprise databases and transaction systems.

In addition to the WebSphere Application Server, the WebSphere product line also includes:

WebSphere Performance Pack: Web-facilities management software that supports rapid growth of high-volume Web sites. It brings together, in a single package, caching, load balancing, and Web site replication.

WebSphere Studio: A set of integrated Web development tools that makes it easy to create dynamic content and Java-based applications for the WebSphere Application Server. It includes wizards, a workbench, and other hot development tools, such as IBM VisualAge for Java, NetObjects Fusion, NetObjects BeanBuilder, and NetObjects ScriptBuilder, for building new and powerful Web applications.

Figure 78 on page 290 shows the WebSphere Application Server Architecture. WebSphere is the Consolidation Solution for the Web Application Server.

WebSphere Application Server

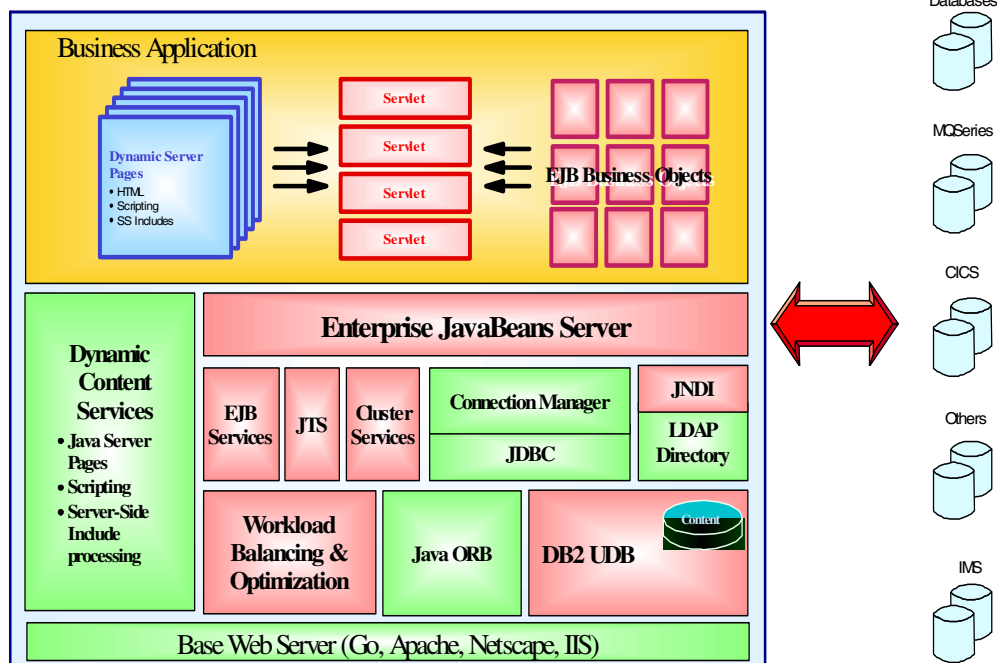


Figure 78. WebSphere Application Server architecture

9.3 Server consolidation for data integration

In this section, we present the Parallel DBMS, DB Connectivity Tool, Lotus Enterprise Data Integrator, and SAN (Storage Area Network) Solution for Data Integration.

9.3.1 VLDB using parallel DBMS for DB integration

Parallel DBMS is particularly well-suited for data-intensive tasks associated with data warehousing, decision support, and other very large database (VLDB) applications. Its unique parallel processing and management capabilities provide an order-of-magnitude improvement in performance by spreading out database operations across all available hardware resources.

Parallel DBMS provides a set of features directly associated with improving VLDB application performance, availability, and manageability, including

enhanced parallel SQL operations, high-availability capabilities, and a suite of systems management tools. It also provides support for relevant open systems standards. The result is a next generation parallel database architecture that delivers outstanding scalability, manageability and performance, minimal operation system overhead, and automatic distribution of the workload. Parallel DBMS can be used to support data/application integration, distributed operations, and mixed application workloads.

9.3.1.1 DB2 parallel architecture

DB2 Parallel Edition is a parallel database product that operates on AIX-based parallel processing systems, such as the IBM RS/6000 SP.

DB2 Parallel Edition supports a shared-nothing architecture and uses the function-shipping execution model for its parallel process flow.

The data is managed by a database manager. The database manager controls CPU, memory, disk, and communications resources. The database manager also provides users with the ability to store and access the data. The collection of data and system resources that is managed by a single database manager, together with its database manager software, is referred to collectively as a *node*. The node resides on a processor, and one or more nodes can be assigned to each processor. In DB2 Parallel Edition, a node is called a Parallel Database Node or PDB node.

Parallel database nodes

The mapping of logical database nodes to physical machines is accomplished by the node configuration file, `db2nodes.cfg`.

Figure 79 shows an example configuration with four parallel database nodes.

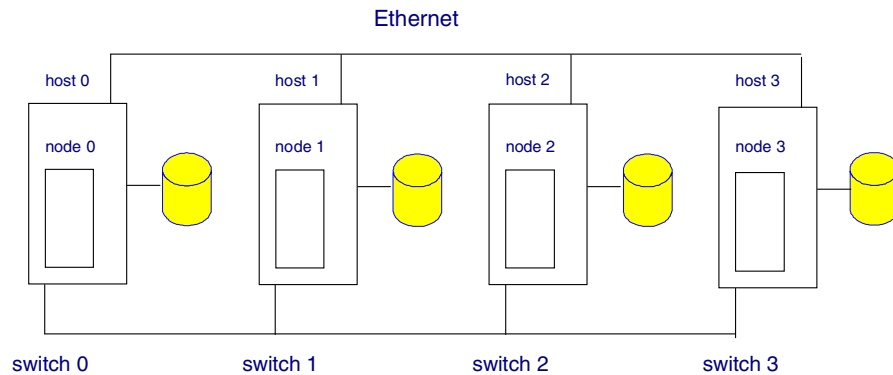


Figure 79. Four physical nodes using DB2 and the SP Switch

Database instance

A DB2 Parallel Edition instance is defined as a logical database manager environment. Every DB2 Parallel Edition instance is owned by an instance owner and is distinct from other instances. The AIX login name of the instance owner is also the name of the DB2 Parallel Edition instance. The instance owner must be unique for each DB2 Parallel Edition instance. Each instance can manage multiple databases; however, a single database can only belong to one instance.

Instance owner

The instance owner has complete control over the instance. The instance owner can start or stop the database manager, modify the database manager configuration, or modify parameters specific to a database.

Nodegroups

In DB2 Parallel Edition, data placement is one of the more challenging tasks. It determines the best placement strategy for all tables defined in a parallel database system. The rows of a table can be distributed across all the nodes (fully declustered) or through a subset of the nodes (partially declustered). Tables can be assigned to a particular set of nodes. Two tables in a parallel database system can share exactly the same set of nodes (fully overlapped), or at least one node (partially overlapped), or no common nodes (non-overlapped).

Data partitioning

DB2 Parallel Edition supports the hash partitioning technique to assign each row of a table to the node to which the row is hashed. You need to define a partitioning key before applying the hashing algorithm. The hashing algorithm uses the partitioning key as an input to generate a partition number. The partition number is then used as an index into the partitioning map. The partitioning map contains the node number(s) of the nodegroup.

Partitioning key

A partitioning key is a set of one or more columns of a given table. It is used by the hashing algorithm to determine on which node the row is placed.

Partitioning map

A partitioning map is an array of 4096 node numbers. Each nodegroup has a partitioning map. The content of the partitioning map consists of the node numbers that are defined for that nodegroup. The hashing algorithm uses the partitioning key on input to generate a partition number. The partition number has a value between 0 and 4095. It is then used as an index into the partitioning map for the nodegroup. The node number in the partitioning map

is used to indicate to which node the operation should be sent. Figure 80 on page 293 summarizes this concept.

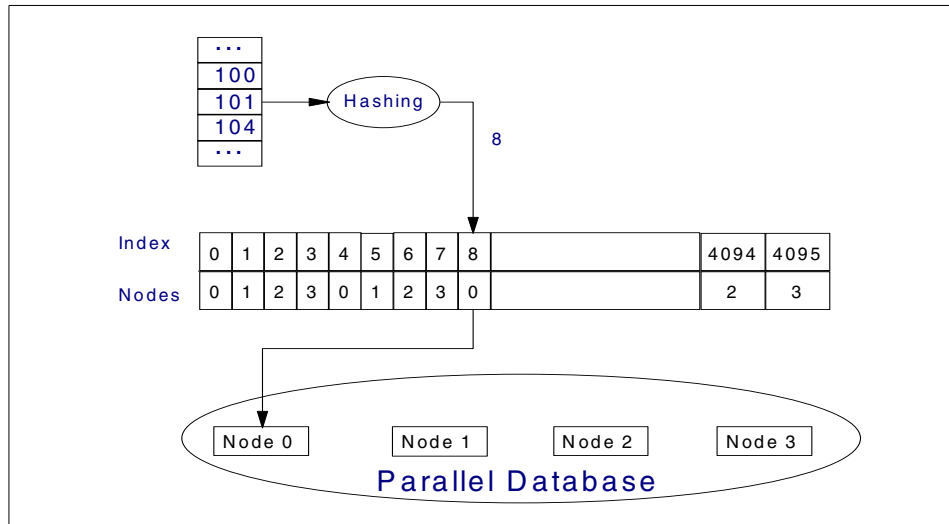


Figure 80. Partitioning map

DB2 and high-availability

Continuous access to data (disk recover) and continuous access to applications is done through HACMP. Figure 81 illustrates an example configuration with four nodes.

The nodes are made up of four clusters of two nodes. Each cluster is a mutual takeover cluster. After a failure, the backup node may vary on the volume groups of the failed node and start the failed DB2 PE partition.

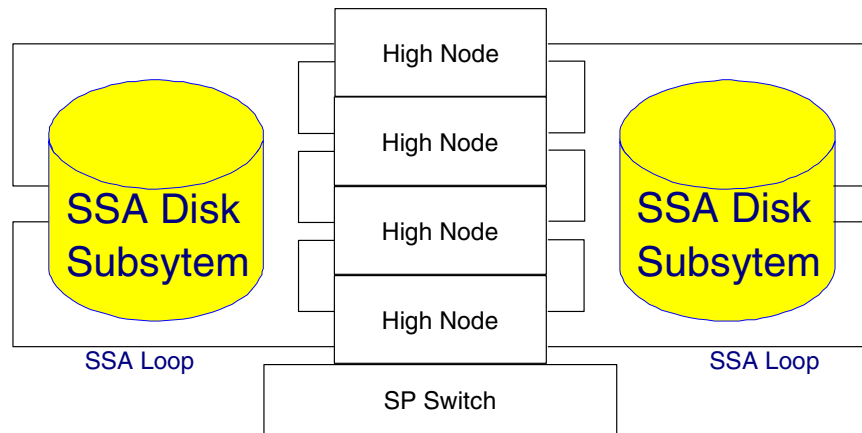


Figure 81. RS/6000 SP configuration with HACMP

9.3.1.2 Oracle Parallel Server

Oracle runs the parallel mode server (OPS) and the parallel query (OPQ) on multiple nodes. This software allows multiple instances of Oracle to run on different nodes simultaneously while maintaining the look of a single image to the end user. Oracle Parallel Server consists of the normal Oracle executables, a component called the Distributed Lock Manager (DLM), and special I/O routines that use IBM Virtual Shared Disk (VSD).

The distributed lock manager coordinates the resources between Oracle instances on nodes of the IBM RS/6000 SP. One instance of Oracle runs on each node of the RS/6000 SP.

In an MPP system, such as the IBM RS/6000 SP, each node has its own memory. Each node belonging to an Oracle Parallel Server system maintains its own cache in main memory. The cache holds the data blocks needed for transaction or query processing. The blocks are held in memory as long as possible to be available for the following transactions. If space is needed for new data blocks, old datablocks are removed from the cache according to a Least Recently Used (LRU) algorithm. In an MPP system, the contents of the caches on the different nodes must be kept consistent across node boundaries. This is the task of the DLM.

Oracle uses IBM's VSD software to allow instances to access disks that are physically attached to other nodes. Data that needs to be exchanged between nodes is passed through the RS/6000 SP's high-performance switch for increased system throughput.

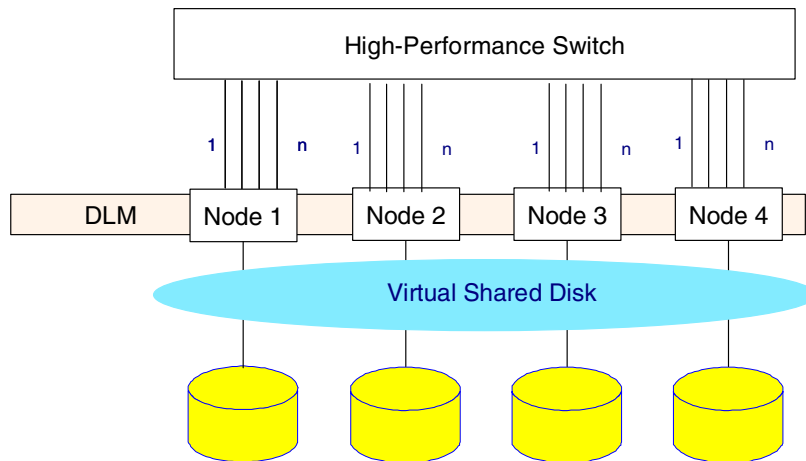


Figure 82. Oracle Parallel Server

Figure 82 on page 295 shows the Oracle Parallel Server Architecture on an RS/6000 SP system. On the IBM RS/6000 SP, the DLM uses the high-performance switch for communication. The DLM uses the high-performance and scalability of the interconnect to efficiently handle distributed cache management.

Parallel Queries: Oracle's parallel query option provides scalability for most common query operations. This dramatically improves response times for very large, complex queries. Scans, sorts, joins, distinct and group can all be performed in parallel.

Parallel Data Load: Parallel data load makes it possible to load large amounts of data onto an Oracle database in a very fast and efficient manner. For example, data located on different tapes can be read in parallel and then distributed over a large number of disks.

Parallel Create Index: Oracle Parallel Server exploits the benefits of parallel hardware by distributing the workload generated by a large index created among a large number of processors. The time needed to create a large index can be diminished linearly with an increasing number of processors.

Parallel Create Table As Select: This feature makes it possible to create aggregated tables in parallel. This is especially useful in large data warehousing environments where smaller data marts are created out of large amounts of data coming from the data warehouse.

Partitioned Views: This feature makes it possible to partition large tables that are often encountered in data warehousing environments, yet still accessing the data in a global way. Partitioned views improve the manageability and availability for large tables and can improve query performance dramatically.

Star Queries: A *star schema* is a natural representation for many data warehousing applications. A star schema contains one very large table (often referred to as a fact table) and multiple smaller tables (commonly called dimension tables).

Parallel Cost Model for the Optimizer: The parallel cost model allows the optimizer to avoid non-parallelizable execution plans. This feature exploits the parallel capabilities of the underlying hardware as much as possible.

Oracle and high-availability

Oracle on the RS/6000 SP offers advanced fault tolerance features to exploit the inherent redundancy of an MPP system in order to provide maximum reliability and availability.

In a fault-tolerant configuration, twin-tailed disks are used so that each disk is connected to two nodes. The twin-tailed connection is designed such that in the case of a node failure, the connection to the failed node is automatically deactivated, and the connection to the backup node is activated.

At the VSD level, there is a product called Recoverable VSD. Each VSD corresponds to one or more disks on a node and is mastered by the node the disks are attached to. If node failure is detected, all the VSDs mastered by the failed node are automatically remastered by the backup node. In this case, the backup node is the node that has the second connection to the disks belonging to the VSDs. The whole process is transparent to Oracle.

At a database level, the fault-tolerant lock manager plays a crucial role in providing fault tolerance. If a node fails, the instance on that node goes down, but all other instances in the system can still access the database. The lock manager is the component that detects the failure and alerts the other instances. The instance that first detects the failure performs instance recovery for the failed instance. The instance performing the recovery rolls uncompleted transactions back and rolls completed transactions forward.

This brings the database back to a consistent state. Once the database is in consistent state, it can be accessed as before. Since the additional workload can be evenly distributed across all nodes, the performance loss in such a case is only $1/n$, where n is the number of nodes.

HACMP software mode 1 (mutual access) and mode 2 (cascading access) are also supported on the SP to provide features, such as IP address swapping or takeover. The HACMP mode 3 (concurrent access) functionality is completely replaced by the Oracle FT DLM and the recoverable VSDs on the SP. Therefore, the traditional HACMP mode 3 is not supported by the Oracle Parallel Server on the SP.

9.3.1.3 Informix-OnLine eXtended Parallel Server (XPS) for RS/6000

Informix-OnLine eXtended Parallel Server (XPS) for RS/6000 SP leverages the shared-nothing nature of the SP architecture to provide a highly-available fault-tolerant database environment including automatic switch-hitter, dataists, and log and database mirroring capabilities.

- This product is designed for data-intensive tasks associated with data warehousing, decision support, and other very large database (VLDB) applications.
- It is able to break each database operation down into multiple tasks, which are spread across all available nodes for independent parallel execution.
- It provides a single system view, enabling database administrators to manage multiple database objects residing on multiple nodes from a single console.
- It is tightly integrated with Tivoli Management Environment (TME) to significantly ease database administration.
- It supports application transparency to facilitate migration of databases across a range of hardware environments without requiring recording.

OnLine XPS is the third in an ongoing series of compatible database products based on Informix's Dynamic Scalable Architecture (DSA). OnLine XPS extends DSA to loosely coupled or shared-nothing computing architectures including clusters of symmetric multiprocessor (SMP) and massively parallel processor (MPP) systems, such as SP systems.

The Informix Extended Parallel Option provides:

Parallel Sort: Sorting is a fundamental activity for such common database operations as building indexes, sort-merge joins, and ORDER BY in SQL queries. Improvements in sort speed are leveraged into improvements in many applications.

Parallel Scan: In most applications today, reports, joins, and index builds all require long scan operations when working with large databases. Parallel scan reduces the scan time dramatically by taking

advantage of table partitioning, which enables scanning of multiple nodes and disks in parallel.

Parallel Insert: When inserting a large number of records into the database using the INSERT INTO statement, or when selecting and inserting data into a temporary table using the SELECT...INTO TEMP statement, parallel insert can significantly speed up the time to complete the transaction by inserting the record in parallel across multiple nodes.

Parallel Delete: Parallel delete can speed up the time required to delete a large block of records. Also leveraging Extended Parallel Option's data partitioning capability, parallel delete utilizes multiple nodes to delete records from disks in parallel.

Parallel Join: Informix Dynamic Server pluralizes a join by using advanced algorithms that permit multiple nodes to scan data and then join the selections in parallel.

Parallel Aggregation: Query operations that incorporate a GROUP BY command and aggregate functions, such as SUM, AVERAGE, MIN, or MAX, can be executed on separate nodes in parallel, thus, completing the database task much faster.

9.3.1.4 Sybase MPP for the IBM RS/6000 SP

Sybase MPP provides powerful, high-performance parallel database processing with the scalability, flexibility, and manageability to meet the needs of an enterprise data store. Sybase MPP leverages SMP, SIMP cluster, and MPP platforms. It deploys multiple SQL Servers that work in unison to process queries, transactions, inserts, updates, and deletes in parallel, as well as parallel load, create index, backup, and recovery. Whatever your current bottleneck, Sybase MPP can break through it to deliver outstanding performance on massive databases.

The powerful technology of Sybase MPP is made easy to manage through another component of the product, the Sybase MPP Manager. This graphical environment provides an easy way for database administrators to perform many server management tasks including monitoring and recovering the system, backing up and restoring the system in parallel, managing logs, and initiating an extensive library of stored procedures.

Sybase MPP pluralizes all operations for significantly faster reads, inserts, updates, and deletes, as well as load, create index, backup, and recover. Sybase architecture for excellent multi-user support provides easy-to-use management and administration services. It offers graphical utilities for easy access to management functions. Sybase MPP system schema, through

Sybase stored procedures, uses a global repository to catalog data placement and data partitioning.

9.3.1.5 Why use the SP for VLDB using Parallel DBMS?

For large data warehouse applications, the SP offers a broad range of scalability. Linear scalability with DSS has been demonstrated many times with different Relational Data Base Management Systems (RDBMS). An example is shown in Figure 83 on page 299. For identical data in the database, response times to requests are inversely proportional to the number of nodes.

Scalability means that you get constant response times as the database grows. You just need to increase the number of nodes. For example, the response time for N nodes and D data in the base will remain the same for 2N nodes and 2D data.

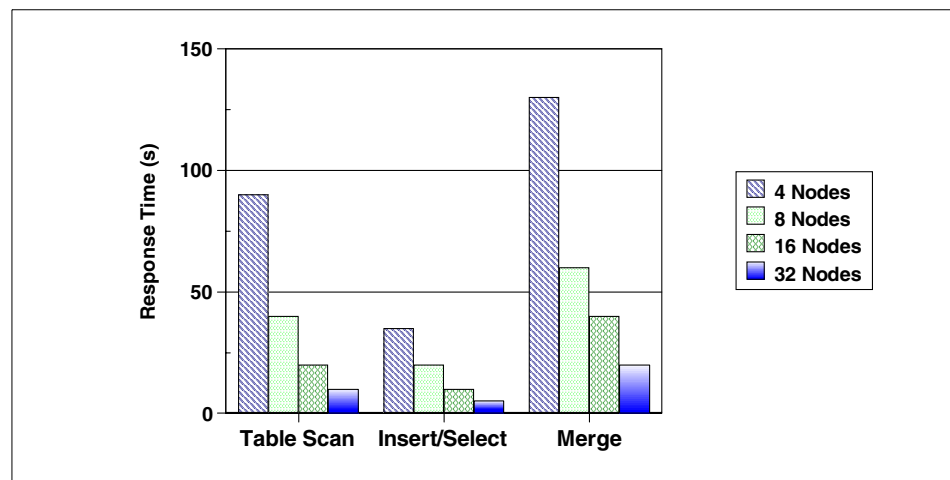


Figure 83. Scalability with DSS using DB2 PE

Scalability is not the only reason for choosing an SP. Other reasons are:

- The SP allows linear scalability with more than 90 percent efficiency.
- DSS is both CPU- and I/O-intensive, and the SP architecture supports both.
- The High Performance Switch is effectively utilized.
- This enables fast movement of many rows of table data between nodes.
- Relational Database Management Systems can essentially support all operations in parallel.
- This is needed for very large database (VLDB) and/or complex queries.
- High volume parallel batch processing is possible.

- The SP helps with loading, updates, printing, valuation, backups, data cleansing, and so on.
- Interoperability with mainframes allows bulk data transfer over channels as well as interoperability with legacy systems and networks.

9.3.2 DB connectivity tool for data integration

9.3.2.1 Distributed Relational Database Architecture (DRDA)

DRDA support defines protocols for communication between an application program and a remote relational database. DRDA support provides distributed relational database management in both IBM and non-IBM environments. SQL is the standard IBM database language. It provides the necessary consistency to enable distributed data processing across like and unlike operating environments. Within DRDA support, SQL allows users to define, retrieve, and manipulate data across environments that support a DRDA implementation.

9.3.2.2 EDA/SQL

Information Builders' EDA/SQL offers a broad range of interoperability. With complete transparency, users can access and join data located in more than 70 database structures (both legacy and relational) on more than 35 operating platforms. EDA also provides access to many file structures resident in major vertical applications, such as SAP R/3, J.D. Edwards, PeopleSoft, Dun & Bradstreet Financials, Baan, and Hogan Financials. Many of the world's leading database and network vendors, including Oracle, Informix, Microsoft, IBM, Novell, and Netscape, have chosen to partner with Information Builders to extend data access and integration capabilities for data warehousing, decision support, electronic commerce, and other business information systems.

9.3.2.3 SAG CLI

The X/Open company and the SQL Access Group (SAG) jointly developed a standard specification for a callable SQL interface referred to as X/Open CLI or SAG CLI. The goal of this interface is to increase the portability of applications by enabling them to become independent of any one database vendor's programming interface. Most of the X/Open Call Level Interface specification has been accepted as part of the ISO Call Level Interface International Standard. Microsoft developed a callable SQL interface called Open Database Connectivity (ODBC) for MS Windows based on a preliminary draft of X/Open CLI. ODBC has expanded X/Open CLI and provides extended functions supporting additional capability. ODBC is not limited to Microsoft operating systems; other implementations are available on various platforms.

9.3.2.4 Open Database Connectivity (ODBC)

ODBC is an industry standard that allows applications that are written using ODBC, such as the Net.Commerce server, to connect to multiple types of databases that support ODBC.

In an ODBC environment, the ODBC Driver Manager provides a linkage between the ODBC application and the underlying databases. The user decides which database the ODBC applications are to access. When an ODBC application sends a request to the ODBC Driver Manager to access a database, the driver manager dynamically loads the appropriate ODBC driver to connect to the requested database. The driver also provides a set of standard application programming interfaces (APIs) to perform database functions that connect to the database, perform dynamic SQL functions, commit or roll back database transactions, and so forth. Each database that supports ODBC has its own ODBC drivers.

9.3.2.5 Java Database Connectivity (JDBC)

JDBC is a standard SQL database access interface providing uniform access to a wide range of relational databases. JDBC is a Java API for executing SQL statements. As a point of interest, JDBC is a trademarked name and is not an acronym; nevertheless, JDBC is often thought of as standing for Java Database Connectivity. It consists of a set of classes and interfaces written in the Java programming language. JDBC provides a standard API for tool/database developers and makes it possible to write database applications using a pure Java API.

Using JDBC, it is easy to send SQL statements to virtually any relational database. In other words, with the JDBC API, it isn't necessary to write one program to access a Sybase database, another program to access an Oracle database, another program to access an Informix database, and so on. One can write a single program using the JDBC API, and the program will be able to send SQL statements to the appropriate database. And, with an application written in the Java programming language, one also doesn't have to worry about writing different applications to run on different platforms. The combination of Java and JDBC lets a programmer write it once and run it anywhere.

9.3.2.6 Common Gateway Interface (CGI)

CGI is a standard that is supported by almost all Web servers. It defines how information is exchanged between a Web server and an external program (CGI program). The CGI specification dictates how CGI programs get their input and how they produce any output. CGI programs process data that is received from browser clients.

Gateway programs, or scripts, are executable programs that can be run by themselves, but it is not advisable. These programs have been created by external programs in order to allow them to run under various (possibly very different) information servers interchangeably.

Gateways conforming to this specification can be written in any language that produces an executable file. Some of the popular languages are:

- C/C++
- PERL
- TCL
- The Bourne Shell
- The C Shell

Figure 84 shows the architectural evolution for the Web Server Logic programming environment.

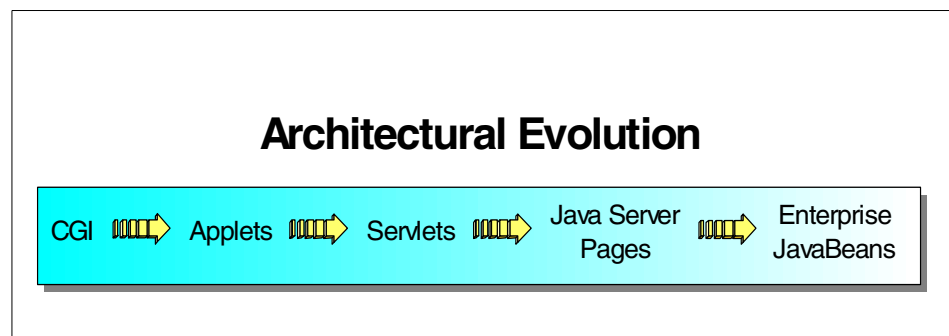


Figure 84. Web Server Logic programming environment evolution

9.3.2.7 Java Servlet

Servlets are Java programs that use additional packages, and associated classes and methods, in the JavaSoft Java Servlet Application Programming Interface (API). Similar to the way applets run on a browser and extend the browser's capabilities, servlets run on a Java-enabled Web server and extend the server's capabilities. Servlets extend the capabilities of the server by creating a framework for providing request/response services over the Web. When a Web browser sends a request to the server, the server can send the request information to a Servlet and have the Servlet construct the response that the server sends back to the Web browser. A Servlet can be loaded automatically when the Web server is started, or it can be loaded the first time a Web browser requests its services. After loading, a Servlet continues to run, waiting for additional Web browser requests. A Servlet can be loaded

automatically at server startup, when a Web browser first requests the Servlet, or when the Servlet is reloaded.

9.3.2.8 Enterprise Java Beans (EJB)

The Enterprise Java Beans Specification provides a detailed description of the services needed to support enterprise beans. It separates the enterprise bean's business logic from the intricacies of persistency, transactions, security, and other middleware-related services. EJB clients interact with enterprise beans through the interfaces defined by the EJB container and the services provided by the EJB server. The EJB server environment and containers are discussed in the sections that follow. The EJB server manages containers and provides the services required by the EJB specification. Mandatory services are as follows:

- Java Naming and Directory Interface (JNDI)
- Transaction service compatible with the Object Transaction Service
- Security

The EJB server can also optionally provide access to a data store through JDBC.

9.3.3 Sizing and capacity planning tools

There are several third-party tools for sizing and capacity planning available. For example, you might want to look into BEZ's tools for Oracle on UDB DBMS performance and sizing.

9.3.3.1 BEZ

BEZ Systems, Inc. provides innovative, practical, and top quality solutions to support customers' needs in performance, capacity management, and database performance optimization for data warehouses based on MPP and SMP systems.

BEZ specializes in performance and capacity management software used to design, build, manage, and grow data warehouses and distributed data marts. BEZ's strength is in its expertise with relational technology, massively parallel processing systems, and performance management.

BEZPlus Software

BEZ provides comprehensive performance and capacity management solutions for data warehouses and data marts. At the center of their strategy is the BEZPlus software products.

BEZPlus is a set of tools used to assist with performance and capacity management for Oracle, Teradata, UDB, and MS SQL Server environments.

INVESTIGATOR is a performance management tool that extracts measurement data, provides daily reporting, performs workload characterization, generates graphs, and creates a baseline model. An analyst can identify the existing performance bottlenecks, critical workloads, users, and SQL that use an excessive amount of system resources. INVESTIGATOR narrows down the scope of tuning efforts to both reduce the time and increase the effectiveness of performance management efforts.

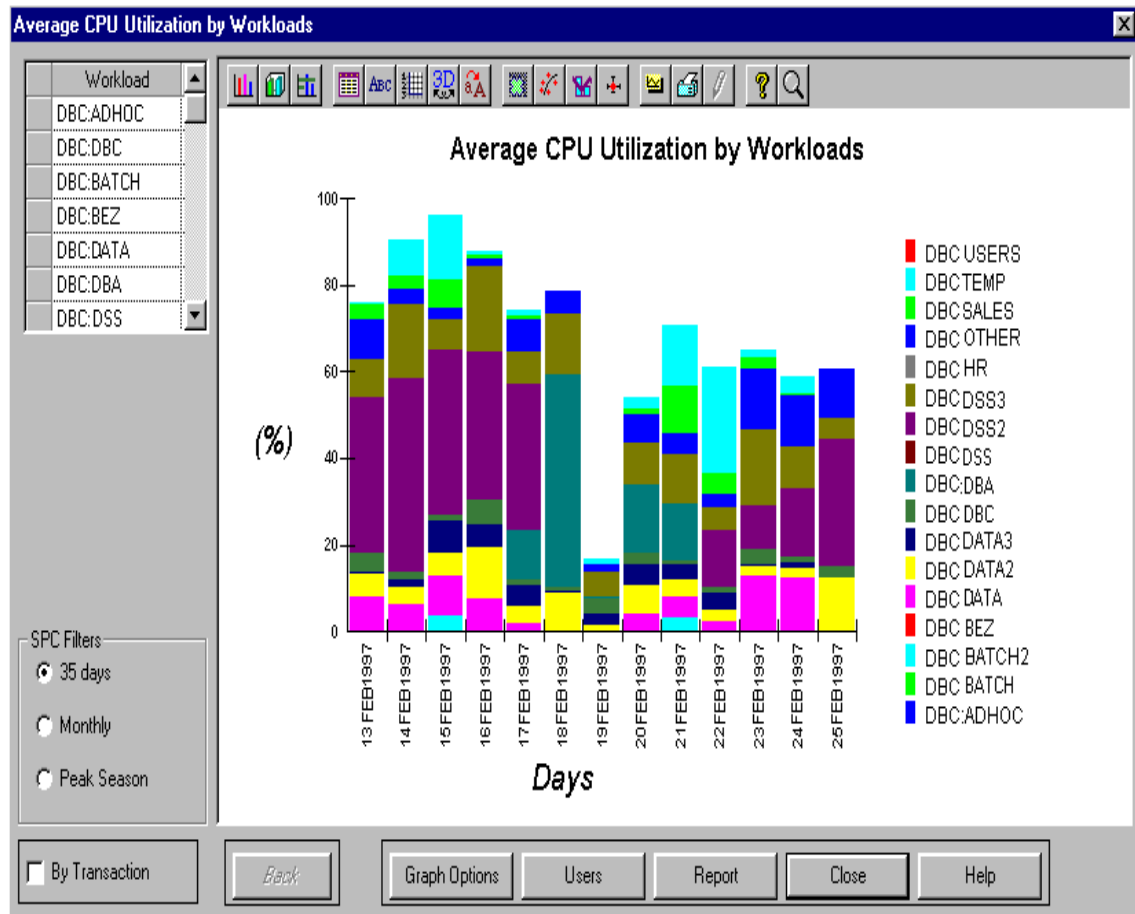


Figure 85. Average CPU utilization by workloads of BEZ

SerView DBA is designed to optimize database administration functions. This product conducts a detailed analysis of SQL and tables accessed by different users. SerView DBA helps to decide what should be tuned in order to make tuning efforts more effective. It provides in-depth table use analysis, helps to

create the right indexes, and reduces the time and effort required to tune the complex environment.

STRATEGIST is a performance prediction tool used to plan a DB environment for the future. The STRATEGIST tool can be used to get the best possible performance environment today. STRATEGIST models, evaluates, and predicts performance, therefore enabling you to:

- Answer a variety of *what-if* questions about the impact of workload and database growth vs. different hardware configurations
- Evaluate database and application design alternatives
- Analyze the impact of changes in scheduling
- Justify hardware upgrades and data warehouse solutions

STRATEGIST's automatic model calibration and performance prediction simplify daily performance and capacity planning functions. The ability to predict expected performance allows you to develop and implement measures optimizing the utilization of available resources, tune critical databases and applications and change scheduling procedures to increase the effective capacity of the RS/6000 SP system, therefore, reducing the frequency of hardware upgrades. Capacity planning recommendations often include an evaluation of the impact of new technology or justification for reengineering the application or DB integration.

With STRATEGIST, you can easily evaluate different alternatives in order to find the best solution for satisfying your service level objectives. Each evaluation of a model allows you to organize multi-step experiments with the results presented in graphs and tables.

APPRAISER predicts performance for new applications. APPRAISER can also be used to estimate the impact of database and application tuning on existing applications. It uses DDL, DML, and expected arrival rates as input to estimate the service time and resource utilization by new applications while the applications are still in the design stage. Output from APPRAISER is used by STRATEGIST to predict the impact new applications will have on the existing workloads and to estimate the impact of tuning on the performance of other workloads.

CorpView DBA contains a performance data mart with current and historical data characterizing 24*7 performance enterprise wide. Statistical process control (SPC) is used to identify unusual increases in activity or resource consumption. Data mining technology allows the analyst to drill down and identify critical workloads, users, SQL, tables, and columns. CorpView

includes chargeback reports for workloads and users and automatically generates performance management reports for distribution through the intranet.

9.3.3.2 BMC BEST/1

Among the third party tools available, you have the option to look into BMC's BEST/1 tools for capacity planning.

BMC Software provides management solutions that ensure the availability, performance, and recovery of business-critical applications. They call this application service assurance, and it means that the applications customers rely on most stay up and running around the clock.

The BEST/1 products from BMC Software use sophisticated graphical analysis, modeling, and reporting to provide complete and accurate insight into enterprise-wide application performance. BEST/1 automated analysis helps business-critical systems and applications operate at peak performance, thus, increasing production and reducing the rate of infrastructure upgrades.

BEST/1 addresses performance management requirements across UNIX, Windows NT, OS/390, Parallel Sysplex, VM, and AS/400 environments. Specialized performance modules with specific application intelligence provide response time modeling to predict when applications' response times will *hit the wall*, therefore, providing ample time to manage changes and upgrades of application and resource performance bottleneck identification. For testing of effects of redistributed workloads, such as adding desktop clients before making changes, BEST/1 is the perfect complement to BMC Software's PATROL solution to provide a strong arsenal for assuring business availability.

Product Components

BEST/1 Performance Console:

Used with the BEST/1 for Distributed Systems for Unix and Windows NT product. It manages the automated functions and provides the user interface for the analysis, modeling, and reporting capabilities for resolving critical application performance problems.

BEST/1 for Distributed Systems

Gives IT professionals the ability to understand how business applications perform today and predicts how they will perform in the future. It can help maximize system resources, minimize response times, and stay ahead of service-level demands, all within budget constraints. BEST/1 for Distributed

Systems is the first integrated performance solution in the marketplace for combined Unix and Windows NT environments.

Workload CPU Utilization Workload [APPLICATION] on 10/29

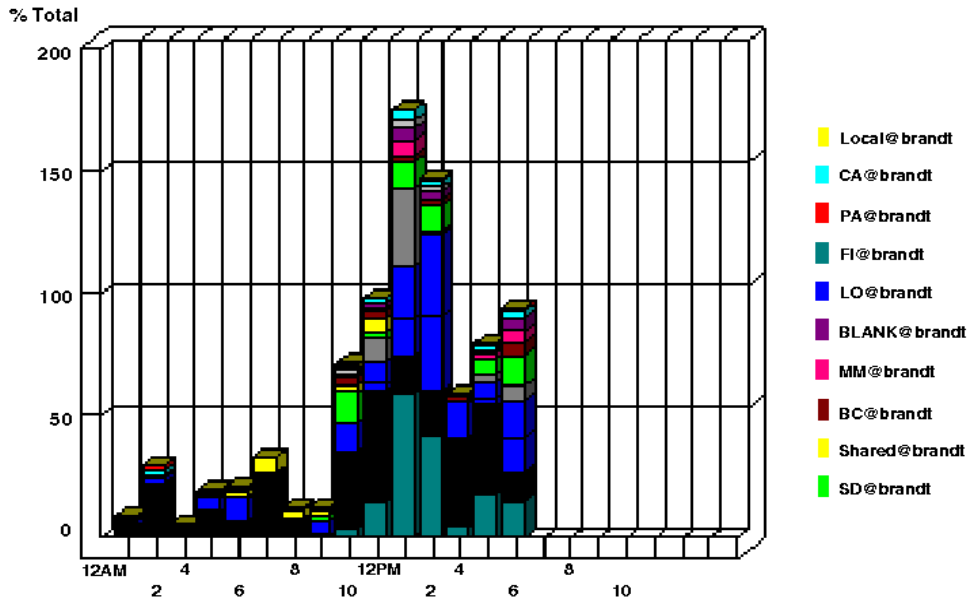


Figure 86. Workload CPU utilization

BEST/1+ for Tivoli Enterprise:

Provides high-level integration between Tivoli Enterprise Software and BEST/1 for Distributed Systems product. It provides customized management and control capabilities that enhance system performance and reliability. BEST/1 event output is merged with other event sources (applications, databases, networks, and systems), and the results are presented centrally on the Tivoli Enterprise Console.

BEST/1 Performance Modules:

The BEST/1 Performance Module product is a support option to the BEST/1 for Distributed Systems product. Available for Informix, Oracle SAP R/3, and Sybase, these performance modules provide current analysis, performance history visualization, analysis, and what-if modeling of system resource

usage, users, and applications accessing the RDBMS or ERP. It is available across multiple UNIX variants and also for Windows NT for Informix.

9.3.4 Lotus Domino Enterprise data integration

Lotus provides services for connecting Domino to a variety of enterprise systems, such as relational database management, transaction processing, resource planning applications, and unstructured data. The services consist of Domino Enterprise Connection Services (DECS), which provide a real-time forms-based interface to enterprise data, and Lotus Enterprise Integrator (LEI), which provides scheduled and event-driven high-speed data transfer capabilities between Domino and enterprise systems. DECS is based on Domino's open architecture for enterprise integration and the associated products. The services also include the programming interfaces used in enterprise integration including the Lotus Connector Lotus Script Extension, the Lotus Connector Java classes, and the Lotus Connector Toolkit, as well as the product-specific Lotus Script Extensions (SAP R/3, MQSeries, DB2, and ODBC). Additional integration methods are Java Database Connectivity, CORBA, ActiveX Data Object, NotesSQL, and Servlets.

Lotus Enterprise Integrator is the successor to NotesPump.

NotesPump is a general purpose data transfer engine. It runs as a server on either Notes server or client. It allows bi-directional data transfer among all data sources including Notes, DB2, Oracle, Sybase, Informix, and others. It consists of two components: NotesPump Server and NotesPump Administration databases.

Figure 87 on page 309 shows the Domino NotesPump Server configuration.

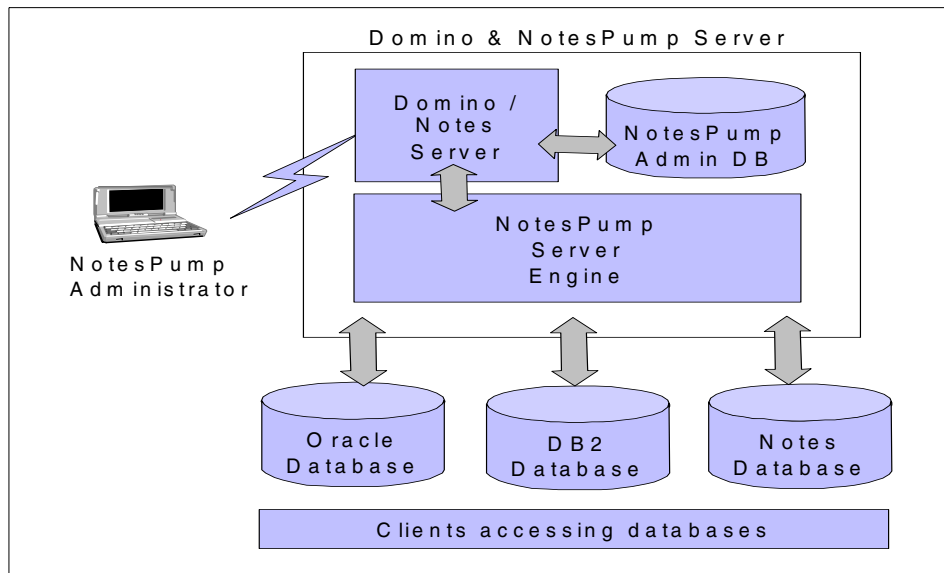


Figure 87. Domino NotesPump server

9.3.5 Storage solution for data integration

Storage Area Network (SAN) is a new storage solution for distributed data environments. SAN is an evolutionary solution from Network Attached Storage (NAS). SAN is an emerging storage solution for data integration.

SAN is a centrally managed, high-speed storage network consisting of multi-vendor storage systems, storage management software, application servers, and network hardware that allow companies to exploit the value of their business information. Connectivity, management, exploitation, and services make up the foundation of SAN deployment. SANs combine connectivity hardware, such as Fibre Channel hubs, switches, and gateways, with software management that accommodates both IBM and non-IBM products and software solutions exploiting the storage, access, flow, and protection of information seamlessly, any time, anywhere.

Leveraging its extensive IT planning, design, and implementation experience, IBM provides support, services, and education required to support end-to-end SAN solutions. IBM laid the groundwork for storage area networks in 1990 with the announcement of ESCON, a serial, Fibre Optic, point-to-point switched network connecting tape, disk, and printing devices to MVS hosts. ESCON has matured to become truly heterogeneous, supporting either native ESCON attachment or ESCON attachment via converters or

gateways to multiple systems including UNIX servers from IBM, Hewlett-Packard, Sun Microsystems, DEC, and Sequent as well as Windows NT platforms. Today, IBM is aggressively working with associations and standards organizations, including the Storage Networking Industry Association (SNIA) and the Fibre Channel Association (FCA), to develop SAN standards.

The SAN Data Gateway: Utilizes Ultra SCSI channel and Fibre channel bandwidth for attachment of the IBM Versatile Storage Server, the Magstar 3590 Tape Subsystem, the Magstar 3494 Tape Library, the Magstar 3590 Silo Compatible Tape Subsystem environments, the Magstar MP 3570 Tape Subsystem, or the Magstar MP 3575 Tape Library Data server.

IBM Fibre Channel RAID Storage Server: The IBM Fibre Channel RAID Storage Server meets the requirements of high-performance systems running large database applications, such as ERP and data warehousing. With RAID levels of 0, 1, 3, 5, and 0+1, the IBM Fibre Channel RAID Storage Server provides industry-standard Fibre Channel attachment for small clusters of UNIX-based servers from IBM, Sun Microsystems, Hewlett-Packard, and Intel-based servers running NT.

IBM Fibre Channel Storage Hub: The IBM Fibre Channel Storage Hub, utilizing technology from VIXEL (an IBM Partner) provides flexible connectivity options for configuring multiple Fibre Channel host and storage server attachments. The seven-port hub provides a cost-effective, single-point Fibre Channel solution that supports up to 100 MB per second data transmission speeds between system servers and storage servers.

SAN Management Software Solutions: The IBM StorWatch Fibre Channel RAID specialist, a network-based integrated storage management tool, allows storage administrators configure, monitor, dynamically change, and manage multiple Fibre Channel RAID Storage Servers from a single Windows 95 or NT Workstation.

For more information, visit: <http://www.ibm.com/storage>

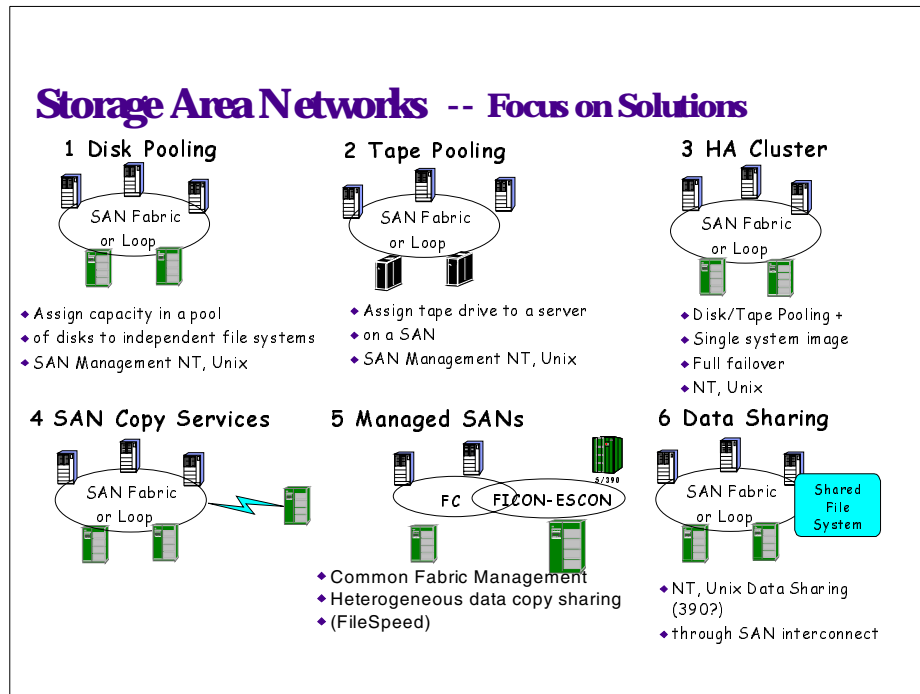


Figure 88. Storage Area Network (SAN) for data integration

9.4 Server consolidation for business application integration

Business environments are changing rapidly. Because of this, customers have implemented business applications such as MIS, ERP, DW, Web Serving, and so on. These applications were often implemented separately. Emerging business environments, however, require business integration. For example, ERP is one of sources of data for DW. In order to meet the rapidly changing business requirements, DW will require tight integration with ERP and legacy systems. ERP and DW can not be separate applications. These systems should be cautiously tied together. It is also required that ERP should be integrated with CRM, SCM, and e-business. This structure is called Extended ERP.

And Enterprise Application Integration (EAI) will be the evolutionary solution for Business Integration. More Server Consolidation opportunities will be generated from business integration solutions.

9.4.1 Extended ERP

Companies all over the world are redesigning and reengineering their business processes to reflect the changes their markets are undergoing. Many companies recognize that to remain competitive into the 21st century they must begin integrating processes across disparate divisions and departments. This requires a new generation of IT solutions that allow your organization to work more efficiently, respond to customers more quickly, and bring new products to market more rapidly. Businesses can get this by implementing technology based on Enterprise Resource Planning (ERP).

ERP has been implemented in a wide range of manufacturing, process, distribution, financial, and retail businesses. ERPs are used to identify and plan the enterprise-wide resources needed to service customers. ERP covers just about every application area of your business including financial functions, such as general ledger, accounts payable, accounts receivable, and cost accounting along with fixed-asset management; production planning, purchasing and inventory control, human resources, payroll, and customer order-management and billing.

The key is integration

Integration is the defining characteristic of ERP solutions. As information is added or changed within one business application, other related functions and applications are immediately updated to reflect the new input. This real-time integration of divergent applications helps streamline the entire business process and helps reduce the traditional time delays incurred by manual methods of information recording and communication among non-integrated islands of automation. Integration allows your business to operate more efficiently, from order entry to production scheduling to final delivery, which reduces overhead and daily operating costs.

Extensions to ERP applications

Organizations expect to integrate their enterprise resource planning (ERP) systems with their new e-commerce initiatives. They want to make their information available to users across extended enterprises. The ERP vendors are working to offer viable, fully-integrated e-business solutions.

SAP announced mySAP.com. Oracle announced its e-business solution, and Baan announced its e-enterprise suite. PeopleSoft aligned itself with Commerce One to offer a promising solution called PeopleSoft Procurement Community. All of these future offerings sound enticing, but customized integration is still needed to ensure a seamless flow of data between the various applications.

It is clear that as organizations focus on e-commerce, a major requirement will be the integration of all these transaction-based and analytical systems. As e-commerce will also require tight integration with ERP and legacy systems for most companies, these systems should be carefully tied together. It is also required that ERP should be integrated with CRM, SCM, and e-business. ERP and Customer Relationship Management (CRM) vendors are rushing to introduce *Webified* versions of their software.

The Internet is a gold mine with new; *unlimited* opportunity. The Web is changing the way organizations do business. The e-business solution offers Web-based business-to-business (B2B) and business-to-consumer (B2C) solutions. Additionally, many Java development-tool vendors are touting rapid development solutions for organizations to build their own. These solutions allow organizations to integrate the existing ERP solutions with e-commerce.

Figure 89 shows the Extended ERP solution.

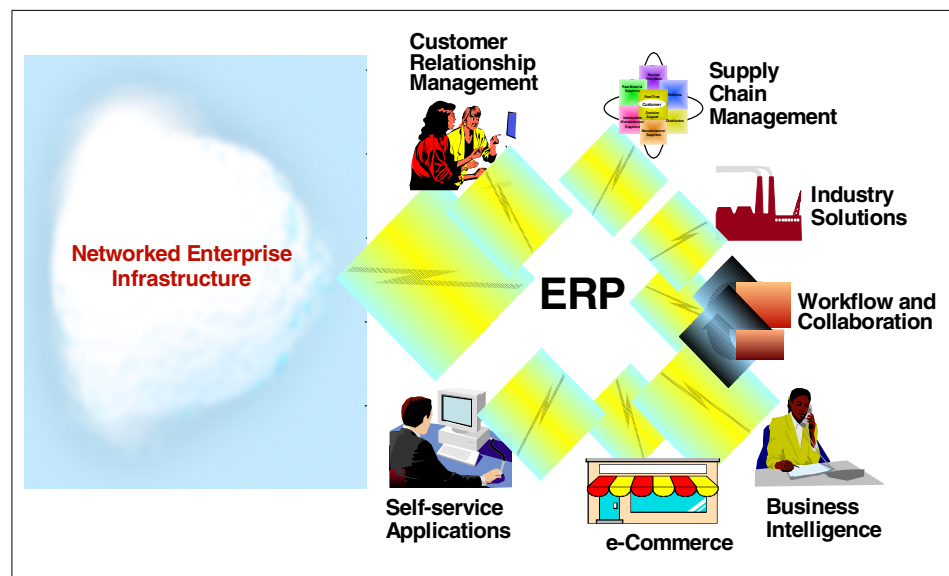


Figure 89. Extended ERP solutions

IBMs ERP strategy combines its high-performance RS/6000 business computers, support for leading ERP software packages, and unparalleled customer services. One elements of IBMs ERP solution is translation of your business needs into IT requirements. IBM recently announced Commerce Integrator, one of IBMs MQSeries, that ties e-commerce and ERP together. IBM MQSeries is a successful solution for Extended ERP and EAI.

9.4.2 Enterprise Application Integration (EAI)

EAI is a new paradigm to integrate existing applications, new point solutions, or a combination of both. Effective application integration is vital to an organization's ability to respond to changing market demands, seize new market opportunities, improve customer service, and achieve business growth potential. New business opportunities inevitably require the flexibility in business processes. EAI provides such flexibility.

Enterprise application integration tools (EAI) can be broken down into four categories:

- **Middleware:** Uses encapsulation to move data.
- **Connectors:** Specific interfaces for specific data structures.
- **Message brokers:** Hub approach that takes messages and forwards them to appropriate databases.
- **Process automation:** Converts data from one format to another by using source codes from both applications.

Prior to integrating e-business applications, such as customer and employee self-service, procurement, and order-entry to front-office and back-office applications, an organization must ensure that its legacy applications are fully integrated. A successful e-business relies on the complete transfer of data across the enterprise. Tools can be used to ensure this integration.

The EAI tool kit extends the vital application integration capabilities of message queuing to reach across full business processes so that they can be adapted flexibly to exploit any new business opportunities.

The benefits of EAI can be realized both within and beyond the enterprise:

- By making it easy to integrate application and data enterprise-wide, and by providing faster access to information, it can shorten time to market, therefore, improving customer service and reducing overall costs.
- By opening up the information in IT systems to suppliers and customers, it can help leverage the value chain to improve quality and accelerate responsiveness to change.
- By providing relief from the burden of modifying applications every time they are integrated, that is, connected in new ways. Transformation and routing of data is performed outside the application without the need for scarce programming and communications skills; so, application testing and assurance costs are reduced.

The organization can decide to put together an integrated system using any of the EAI categories mentioned above.

9.4.3 EAI solutions - MQ integrator and BEA eLink

9.4.3.1 MQSeries Integrator - IBMs EAI solution

IBM MQSeries Integrator is powerful message brokering software that ensures business-critical applications and processes can understand each other. Based on MQSeries' messaging and queuing capabilities, the MQSeries Integrator is a real-time, intelligent, rules-based message routing and dynamic message content transformation and formatting system that allows you to integrate all types of applications and systems into robust, flexible, and scalable information networks. MQSeries Integrator is EuroReady and year 2000 ready. MQSeries, IBMs industry-leading, messaging-oriented middleware, enables diverse applications to communicate securely and reliably, with enterprise-level performance, over a wide range of platforms.

The value of an MQSeries Integrator solution is that it easily shares, and can act on, knowledge. In other words, it provides the core of enterprise intelligence. Knowledge of the applications enables transformation of message formats. Knowledge of business rules and information requirements enables intelligent routing of information to where it is needed. And knowledge of packaged application documents enables a quick start to integrating these applications with the rest of the enterprise. Collectively, capabilities like these in the hub are usually known by the term message broker. MQSeries Integrator has all the capabilities to be a full message broker.

The following are provided by MQSeries Integrator Features:

- Heterogeneous platform support
- Layered design
- Database repository for rules and format definitions
- Assured delivery
- Transactional integrity maintained
- Message transformation and translation
- Intelligent routing

Figure 90 on page 316 shows the MQSeries Message Routing Efficiency.

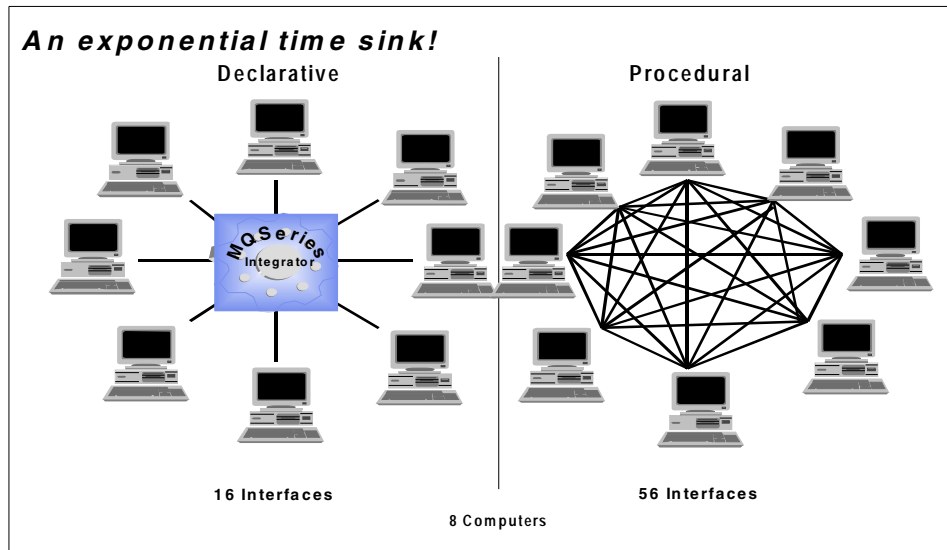


Figure 90. Efficient Message Routing of MQSeries

The following are the different points of view of the MQSeries Integrator Features:

Transformation: Transformation is important because of the way that applications work. Most enterprises have applications that have developed over the years, on different systems, using different programming languages and different methods of communication. Standard message queuing technology can bridge differences like these, but each message queue has to be explicitly told about the characteristics of each message destination.

Intelligent routing: Intelligent routing encapsulates business knowledge of how information should be distributed between message-sending and receiving applications throughout the enterprise. This knowledge is stored in the hub as a set of rules that are applied to each message that passes through the hub.

Application Templates: Many enterprises make use of packaged applications, like PeopleSoft GL and SAP's R/3. Application templates encapsulate knowledge of all the data that is carried in the forms, therefore, making it easy for other (non-packaged) applications to gain access to the information that is in the forms. Application templates ensure that information flow can encompass every type of application in the enterprise, both packaged and custom-made.

9.4.3.2 BEA EAI Solution

BEA was founded in January, 1995. Its growth has been fueled by the following acquisitions:

- Tuxedo from Novell (February, 1996)
- ObjectBroker, MessageQ from DEC (February, 1997)
- Top End from NCR (March, 1998)
- WebLogic, Inc. (September, 1998)

Figure 91 shows BEA EAI Solution Framework.

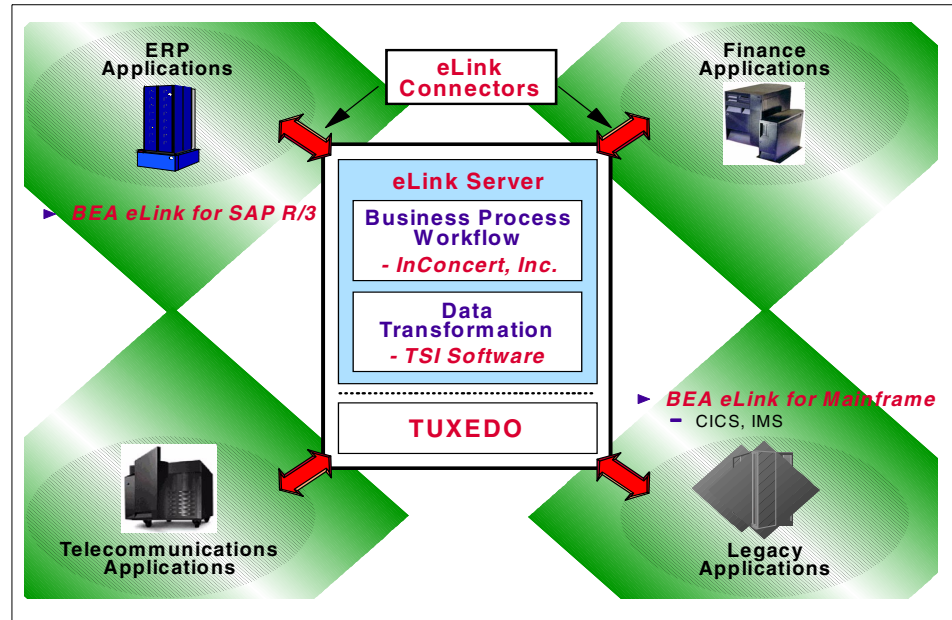


Figure 91. BEA EAI solution framework

New division created BEA WebXpress from WebLogic acquisition. BEA provides an eLink Enterprise Application Integration (EAI) set of offerings.

BEA also provides the following WebLogic Web application server offerings:

- BEA WebLogic family of Web application servers
- BEA WebLogic Express (formerly Tengah/JDBC)
- BEA WebLogic Server (formerly Tengah)
- BEA WebLogic Enterprise (Tengah + M3)

BEA has been merging Top End into TUXEDO. BEA increases its focus on application integration needs and solutions.

Figure 92 shows the IBM-BEA product comparison.

| IBM | | BEA |
|---|----------------------------------|--|
| WebSphere Enterprise <small>(formerly TXSeries)</small> | TP Monitor | TUXEDO |
| | CORBA/OTM | WebLogic Enterprise <small>(formerly M3)</small> |
| MQSeries | Message Queuing | TUXEDO |
| | Message Broker | Enterprise Application Integration (eLink) <small>NEW</small> |
| | Business Process Workflow | Enterprise Application Integration (eLink) |
| WebSphere ▶ <i>Standard</i> ▶ <i>Advanced</i> ▶ <i>Enterprise</i> | Web Application Server | WebLogic ▶ <i>Express</i> ▶ <i>WebLogic Server</i> ▶ <i>Enterprise</i> |

Figure 92. IBM - BEA product comparison

9.4.4 Server consolidation for Web application integration

SP system has been used for Web environments as a consolidation server. SP system is a good solution for Web Application Server. Application Service Provider (ASP) is a new IT trend. More Server Consolidation opportunities will be generated from Web Application Integration.

9.4.4.1 Typical Web server consolidation using SP

There are three main application areas for which the SP-based Web server acts as a good solution:

- Internet service provider
- Content hosting
- Mega Web site management

Internet Service Provider (ISP)

An Internet Service Provider connects end users to the Internet. The services offered with a subscription include access to news, chat room and bulletin board services, some authority checking and service, a home page, the ability to send and receive mail, and Web access via a browser, such as Netscape's Navigator.

Figure 93 shows the ISP Internet Access and Web Content Hosting Services. RS/6000 SP with Interactive Session Support/SecureWay Network Dispatcher (ISS/ND) is an ideal solution for ISP services because of the ease with which these various services can be split across RS/6000 SP nodes, the asymmetry of configuration that is possible, and the high availability solutions provided with HACMP.

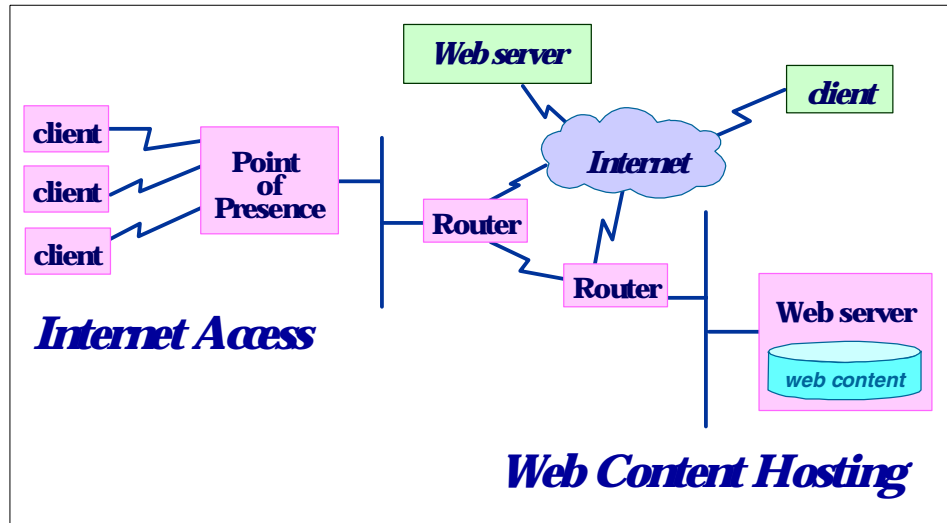


Figure 93. ISP Internet access and Web content hosting services

Content Hosting

Customers in either the Internet or intranet, who have a lot of content to be delivered, can also be well-served with a RS/6000 SP platform. These customers fall into two categories:

- Those that house and serve up their own content.
- Those that provide the content hosting service to smaller companies with the resulting aggregate bandwidth being quite large.

In the first case, an SP-based solution makes sense if there are a large number of accesses to content requiring the scalable performance of the RS/6000 SP for acceptable Web response time. However, it can also be a good fit for a solution where the quantity of data is large, and response time is not a critical factor. In this case, a single node could serve as the Web server, but you might employ several back-end database servers that each have access to the desired data, which can be retrieved by any of these servers. The client request can come into the ISS/ND node of RS/6000 SP and be routed to the back-end server nodes responsible for that IP Web-site. Then,

the Web server that is listening for that IP address can handle the Web request.

Mega Web Site Management

The RS/6000 SP is a perfect solution for extremely busy Mega Web sites. Because of its excellent scalability, RS/6000 SP with ISS/ND provides a winning combination for problems that are large in scope. One such example is socks servers. Almost all companies that allow their employees access to the Internet from their intranet do so with the use of a socks server. The client's browsers point to the socks server, which then takes the client's Web requests and sends them out on the Internet. For a large corporation, a large number of socks servers would be required to implement this service.

A solution to this socks server problem is to house all of them in an RS/6000 SP and allow one of the nodes on the RS/6000 SP to run the ISS/ND software. All the corporation browsers would point to this node, which would then route the requests to the heaviest-loaded socks server. This solution to the socks server problem insures that the load is adequately balanced across the servers, and it also reduces the support costs compared to a regular cluster of workstations.

Figure 94 on page 321 shows the ISS/ND network balancing on RS/6000 systems.

Likewise, if an Internet Service Provider wants to provide excellent response time to customers, they may elect to place some proxy servers between the clients and the Internet. Of course, in such a solution, you do not want your proxy server to become a bottleneck. A natural response is to replicate the proxy servers and then balance the client requests across each of the proxy servers. The advantages described in the socks server example also apply to the scalable proxy server solution: Balanced work across the servers and lower support costs.

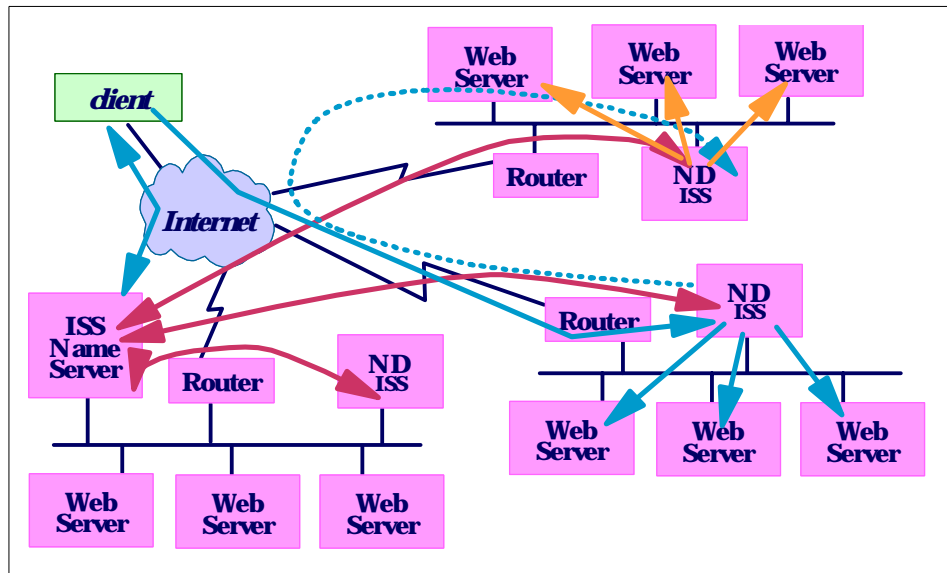


Figure 94. Network balancing of ISS and SecureWay Network Dispatcher

9.4.4.2 Web Application Server

Initial Internet scenarios focused on Web publishing, world-wide access to essentially-static information, and Internet access service. There was not much distinction between clients, and clients could get read-only, non-transactional services. Now, Internet scenarios focus on Web Application areas that can provide the distinct, transactional services with business logic.

Whether it is business integration, Web self-service or e-commerce, the IBM WebSphere Application Server provides the software and tools for building and deploying Web-based applications. With Standard, Advanced, and Enterprise editions, Application Server supports all your e-business needs ranging from simple Web transaction processing to enterprise-wide Web applications. An open, extensible solution providing the highest levels of performance, security, availability, and scalability, WebSphere Application Server leverages your existing IT investment to create new opportunities.

The WebSphere Engine solution allows for the creation, test, and deployment of transactional e-business applications that can scale to meet a business' needs demand. This solution helps customers deploy and manage Web-based applications ranging from simple publishing and self-service interactions to powerful e-business solutions. It is best suited for customers needing a deployment platform for applications they have built (or contracted to be built) that contain significant new business logic. Software developers

are quick to realize the advantages of building applications using open standards, such as Enterprise Java Beans (EJB), Java Servlets, and Java Server Pages (JSPs). You will find the WebSphere Engine is a superb deployment platform for those new applications not just in terms of ease of management of those Web-applications, but also in the base function/performance of the platform including reliability and scalability. Examples of customers who would want to avail themselves of this technology would be companies that currently use a Web-site to take an order and use batch and/or manual techniques to bridge the Web-request to the legacy process/systems to process the order. Business logic could be developed using JSPs, Servlets, and EJB to bridge these two environments when deployed with the WebSphere engine.

Figure 95 shows the WebSphere Application Server Architecture.

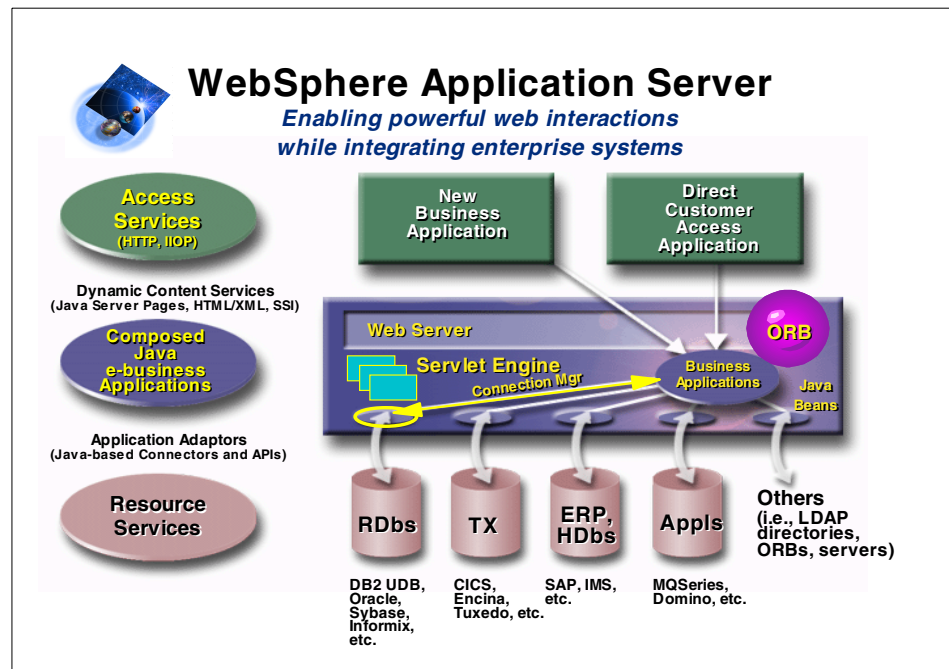


Figure 95. WebSphere application server architecture

9.4.4.3 Application Service Provider (ASP)

Internet service is rapidly evolving to differentiated Web applications. Business Integration ties customers, business partners, and employees together using the Web. It needs transactions, persistence, and security

under program control. It drives the need for a Web Application Server to provide these services and necessary infrastructures.

ASP provides more advanced services. Whereas Internet Service Provider (ISP) vendors provided the services to connect end users to the Internet, ASP vendors provide one-to-many application services and delivers a packaged software product over the internet, primarily to business customers, which is a trend to move complex applications from enterprise to service providers.

What business problems exist?

- Customer relationship
- Supply chain management
- Electronic commerce
- Knowledge management

What Client types?

- Any browser
- Windows only
- Thin devices

What Data source types?

- ERP systems
- Relational databases
- Transaction-based data sources

ASP vendors bundle the value added services, such as installation, training, integration, and support. ASP service is more than hosting static Web sites. The service covers the key business management applications: Sales, order entry, distribution, customer services, and so on.

ASP also delivers application outsourcing services. These kinds of application outsourcing services imply the advantage and the benefit of server consolidation.

Figure 96 on page 324 shows the ASP service environment.

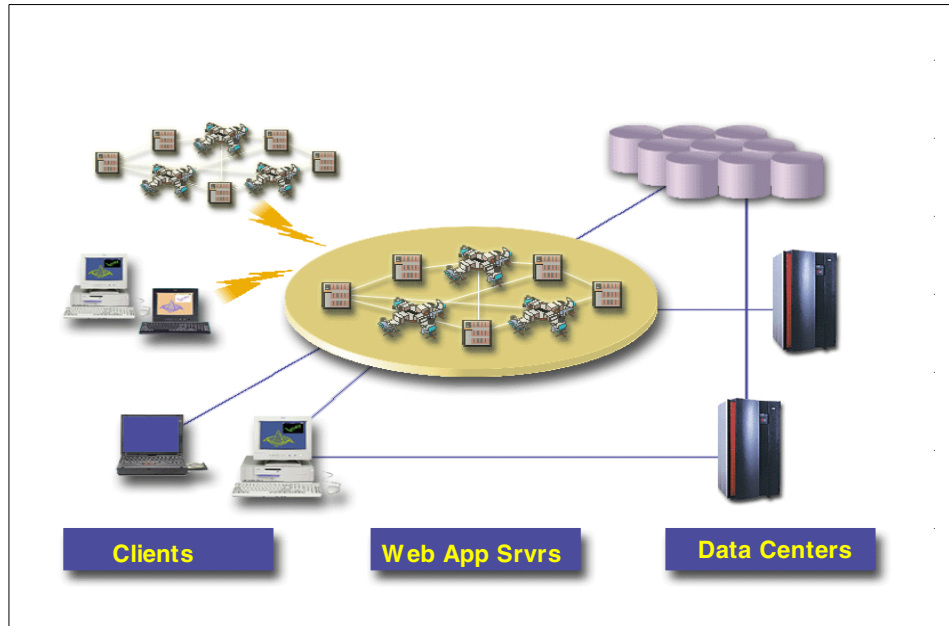


Figure 96. Web application service architecture

9.4.5 Application development and automated testing tools

Compuware is a provider of application development tools. UNIFACE is Compuware's development environment for business-critical application.

UNIFACE is the environment that allows organizations to assemble large-scale applications from best-of-breed components constructed using different tools and programming languages. Developers can not only use components constructed using UNIFACE's powerful construction tools, but they can also use other tools and languages, such as Visual Basic, Java, C++, and COBOL.

UNIFACE features a highly productive development environment that allows teams of developers to work in parallel on complex applications. Developers assemble components built in UNIFACE and other leading construction tools including Visual Basic, Java, and legacy languages. When new business needs arise, individual components can be replaced without needing to redevelop and retest the whole application. Because you can never predict what the future holds, your business-critical applications must be technology independent. UNIFACE applications run on all major hardware platforms, operating systems, network configurations, databases, and component

models. And because UNIFACE supports all major enterprise architectures including host-based, two-tier and multi-tier client/server and the Internet, application integration and redeployment are greatly simplified.

For the Automated testing process, Compuware provides File-AID/CS and QACenter.

File-AID/CS enables developers to extract production data to re-create errors and test conditions in a client/server environment. With File-AID/CS, developers can quickly and accurately identify and correct data-related problems without being an expert in the specific database environment they're working in. The spreadsheet-like editor of File-AID/CS provides a clear view into the data because it uses table definitions or COBOL record layouts as templates. Designed to abide by currently established security mechanisms, File-AID/CS can be used with confidence.

QACenter automated technology for understanding and resolving production problems extends to the Web and client/server environments. QACenter testing tools can help you achieve consistent, dependable performance of applications in the Windows, Windows 95, Windows NT, DOS, and UNIX environments. QACenter also supports packaged applications and applications developed with 4GLs and middleware.

9.4.6 Outsourcing vs. server consolidation

Service providers are outsourcing many data centers already. The outsourcing service is more than hosting static Web sites. It is a outsourcing of key business management applications and it is trend to move complex applications from enterprise to the service providers. This means that outsourcing is becoming acceptable. Internet transport is also becoming acceptable thanks to the emerging security solutions.

Web hosting is also a form of outsourcing services. One provider of application outsourcing services in ASP. Application outsourcing is a service wherein the responsibility of the deployment, management, and enhancement of a packaged customized software application is handed off contractually to an external service provider.

Companies can utilize application outsourcing services in the following ways:

- Payroll
- Accounting
- Human Resources
- Sales Force Automation
- ERP

- Electronic Commerce
- Decision Support System
- Web Billing

Outsourcing services help businesses become more competitive by allowing them to concentrate on their business strategy. IT Outsourcing Services provides management and support for IT operations. Outsourcing contracts range from high-end server (host) operations and midrange services to distributed and desktop information technology services. It includes system monitoring, system operations, system management controls, technical support and help desks, and can be tailored to your needs with optional services.

Outsourcing service provides the Capacity Services to help relieve companies from the task of managing computing resources, thus, allowing them to focus on their core business. Capacity Services furnishes the flexibility needed to expand your data processing capabilities on a seasonal or ongoing basis, adopt new technologies, migrate key applications to new platforms, and quickly develop, test, and deploy new applications.

Outsourcing offer customers complete, integrated user services encompassing advice, planning, procurement, preparation and delivery of hardware and software, training, support and maintenance, security and asset management, and administration of their distributed computing environment.

This is driven by the growth and proliferation of decentralized and distributed servers throughout a customer's enterprise that can no longer be managed cost effectively or securely. Nor can these servers meet the high availability demands of emerging business processes, such as e-business, business intelligence, ERP, customer relationship management solutions, supply chain management solutions, and the like.

Enterprises can benefit from outsourcing using Server Consolidation. Outsourcing services imply the advantage and the benefit of Server Consolidation.

Server Consolidation is an enabling, background solution for the emerging key business applications and IT trends.

Appendix A. Information sources

Information on software and product solutions mentioned in this redbook were gained from Web sites and also from promotional materials provided by each of the companies mentioned in this redbook.

A.1 Independent software vendors

The following list is a corporate profile on each of the Independent Software Vendors (ISVs) that we detailed along with the products that were mentioned in this redbook:

- **ADIC** www.adic.com

Advanced Digital Information Corporation (ADIC) is a market leading, drive-independent manufacturer of innovative data storage products and specialized storage management software that integrates into a wide range of rapidly evolving network computing environments.

Founded in 1983 to market a mini-computer data cartridge backup drive, ADIC has since grown to become the leading technology-independent supplier of automated tape libraries for PC/LAN and UNIX-based networked systems. ADIC has approximately 600 team members, and the current annual run rate is over \$200 million. Revenue for the last fiscal quarter (ending April 30, 1999) was \$54 million.

The following are the components of ADIC:

- AMASS for Unix

- **BEZ Software** www.bez.com

BEZ Systems, Inc. provides innovative, practical, and top quality solutions to support customers' needs in performance and capacity management and database performance optimization for data warehouses based on MPP and SMP systems.

BEZ specializes in performance and capacity management software used to design, build, manage, and grow data warehouses and distributed data marts. BEZ's strength is its expertise with relational technology, massively parallel processing systems, and performance management.

BEZ Systems was founded in 1984.

The following are the components of BEZ Software:

- BEZPlus software
 - Investigator

- SerView DBA
- CorpView DBA
- Strategist
- Appraiser
- **BMC Software** www.bmc.com

BMC Software is a market leading provider of management solutions that ensure the availability, performance, and recovery of business-critical applications. They call this *application service assurance*, and it means that the applications their customers rely on most stay up and running around the clock.

For more than 18 years, the largest and most successful companies have relied upon BMC Software. BMC Software is among the world's largest independent software vendors, a Forbes 500 company, and a member of the S&P 500, with revenues of \$1.3 billion in fiscal year 1999. The company is headquartered in Houston, Texas, with over 100 offices worldwide.

The following are the components of BMC Software:

- BEST/1 for Distributed Systems
- BEST/1 Agents
- BEST/1 Analyze
- BEST/1 Collectors
- BEST/1 Performance Console
- BEST/1 Performance Modules
- BEST/1 Predict
- BEST/1 Visualizer
- COMMAND/POST
- CONTROL-M for Distributed Systems
- PATROL
- SQL BackTrack
- **Candle Corporation** www.candle.com

The Candle Corporation is a worldwide software and services company providing comprehensive solutions to the business community. Recognizing that, in most industries, their businesses are their applications, Candle's offerings in software and services, expertise, and support help customers get their applications to work and work together.

With a 23-year history, Candle is committed to providing the products and services needed for successful application connectivity and integration.

The following are the components of Candle Corporation:

- Command Center for Distributed Systems

- **Computer Associates** www.cai.com

Computer Associates International, Inc., is a world leader in mission-critical business computing and provides software, support, and integration services in more than 100 countries around the world.

CA solutions traditionally make up the software that runs businesses, and they are now a market leading developer of client/server solutions.

CA has more than 14,000 employees and had a revenue of \$5.3 billion in fiscal year 1999.

The following are the components of Computer Associates:

- ARCserve/IT

- **Compuware** www.compuware.com

Compuware has been in business for 26 years, and they are one of the largest independent software vendors in the world, earning over \$1.6 billion in revenue in the fiscal year 1999.

They develop, market, and support seven families of software products including more than 110 offerings, and their productivity solutions help 14,000 of the largest corporations more efficiently maintain and enhance their most critical business applications.

They employ more than 15,000 employees, and their professional services staff of over 10,500 develops 28 staff-years of commercial application software every day. They have more than 100 offices located in 45 countries.

The following are the components of Compuware:

- EcoTOOLS
- EcoSYSTEMS
- UNIFACE
- QACenter
- File-AID

- **Legato** www.legato.com

Legato Systems, Inc., is a market leader in the enterprise storage management software market. Legato's products enable *Information*

Continuance, a seamless approach to the movement, management, and protection of data throughout an enterprise.

Founded in 1989, Legato's storage management software products have become the recognized industry standard with the largest installed base representing over 65,000 customers.

The following are the components of Legato:

- SmartMedia
- NetWorker
- **VERITAS Software** www.veritas.com

As the leading provider of enterprise-class application storage management software, VERITAS Software ensures the continuous availability of business-critical information by delivering integrated, cross-platform storage management software solutions.

Founded in 1982, the California based company has grown to over 2,300 employees residing in 17 countries worldwide. Revenues exceeded \$210 million in the fiscal year 1998 and, for the second consecutive year, the company ranked as one of Forbes' 200 Best Small Companies in America with an overall ranking of 12.

The following are the components of VERITAS Software:

- NetBackup

A.2 IBM Softwares

The following list is a corporate profile on the IBM software vendors detailed and products that were mentioned in this redbook:

- **IBM** www.ibm.com
 - IBM Netview
 - IBM LoadLeveler
 - IBM Shared Disk
 - IBM Parallel System Support Program (PSSP)
 - IBM Job Scheduler
 - IBM Infoprint Manager for AIX
 - IBM NetTAPE
 - IBM SecureWay Network Dispatcher
 - HACMP

- HAGEO
- Sysback/6000
- **Tivoli** www.tivoli.com

Tivoli Systems Inc., an IBM company, provides the industry's leading open, highly scalable, and cross-platform management solutions that span networks, systems, applications, and business-to-business commerce. Tivoli is a global company dedicated to providing products, services, and programs that enable companies of any size to manage their networked PCs and distributed systems from a single location.

Tivoli is the sixth largest software company in the world. Their products are used by 96 percent of Fortune 500 companies. Headquartered in Austin, Texas, Tivoli distributes its products world-wide through a network of global sales offices, systems integrators, resellers, and IBM sales channels. They have a workforce of over 4,000 employees.

The following is the list of the components from the Tivoli Enterprise software described in this book.

- Tivoli Software Distribution
- Tivoli Inventory
- Tivoli User Administration
- Tivoli Global Sign-On
- Tivoli Security Management
- Tivoli Global Enterprise Manager (GEM)
- Tivoli Enterprise Console (TEC)
- Tivoli NetView for UNIX
- Tivoli Decision Support
- Tivoli Workload Scheduler
- Tivoli Distributed Monitoring
- Tivoli Output Manager
- Tivoli Storage Manager
- Tivoli Service Desk
 - Tivoli Change Management
 - Tivoli Asset Management

Appendix B. Special notices

This publication is intended to help consultants and IT managers, their technical team, and RS/6000 sales teams in IBM who need to identify requirements and opportunities and to plan for server consolidation on RS/6000 platforms. The information in this publication is not intended as the specification of any programming interfaces that are provided by described products. See the PUBLICATIONS section of the IBM Programming Announcement for each described product for more information about what publications are considered to be product documentation.

Illustrations reflect the capabilities and interfaces of the current release of AIX 4.3.3. Changes may be incorporated in future updates.

References in this publication to IBM products, programs or services do not imply that IBM intends to make these available in all countries in which IBM operates. Any reference to an IBM product, program, or service is not intended to state or imply that only IBM's product, program, or service may be used. Any functionally equivalent program that does not infringe any of IBM's intellectual property rights may be used instead of the IBM product, program or service.

Information in this book was developed in conjunction with use of the equipment specified, and is limited in application to those specific hardware and software products and levels.

IBM may have patents or pending patent applications covering subject matter in this document. The furnishing of this document does not give you any license to these patents. You can send license inquiries, in writing, to the IBM Director of Licensing, IBM Corporation, North Castle Drive, Armonk, NY 10504-1785.

Licensees of this program who wish to have information about it for the purpose of enabling: (i) the exchange of information between independently created programs and other programs (including this one) and (ii) the mutual use of the information which has been exchanged, should contact IBM Corporation, Dept. 600A, Mail Drop 1329, Somers, NY 10589 USA.

Such information may be available, subject to appropriate terms and conditions, including in some cases, payment of a fee.

The information contained in this document has not been submitted to any formal IBM test and is distributed AS IS. The information about non-IBM ("vendor") products in this manual has been supplied by the vendor and IBM

assumes no responsibility for its accuracy or completeness. The use of this information or the implementation of any of these techniques is a customer responsibility and depends on the customer's ability to evaluate and integrate them into the customer's operational environment. While each item may have been reviewed by IBM for accuracy in a specific situation, there is no guarantee that the same or similar results will be obtained elsewhere. Customers attempting to adapt these techniques to their own environments do so at their own risk.

Any pointers in this publication to external Web sites are provided for convenience only and do not in any manner serve as an endorsement of these Web sites.

Any performance data contained in this document was determined in a controlled environment, and therefore, the results that may be obtained in other operating environments may vary significantly. Users of this document should verify the applicable data for their specific environment.

This document contains examples of data and reports used in daily business operations. To illustrate them as completely as possible, the examples contain the names of individuals, companies, brands, and products. All of these names are fictitious and any similarity to the names and addresses used by an actual business enterprise is entirely coincidental.

Reference to PTF numbers that have not been released through the normal distribution process does not imply general availability. The purpose of including these reference numbers is to alert IBM customers to specific information relative to the implementation of the PTF when it becomes available to each customer according to the normal IBM PTF distribution process.

The following terms are trademarks of the International Business Machines Corporation in the United States and/or other countries:

| | |
|-------------------------|------------------------|
| ADSTAR | AIX |
| AIX/6000 | AIXwindows |
| AS/400 | CICS |
| DB2 | DB2 Universal Database |
| ESCON | HACMP/6000 |
| IBM | IBM Global Network |
| Infoprint | IMS |
| LoadLeveler | Micro Channel |
| MVS | OS/2 |
| OS/390 | POWERparallel |
| PowerPC Architecture | POWER2 Architecture |
| Print Services Facility | RISC System/6000 |

| | |
|--------------------------------|-----------|
| RS/6000 | S/390 |
| Scalable POWERparallel Systems | SecureWay |
| SP | SP1 |
| SP2 | WebSphere |

The following product names are trademarks of Tivoli Systems, Inc.:

| | |
|-----------------|-------------------|
| NetView | Tivoli |
| Tivoli ADSM | Tivoli Enterprise |
| Tivoli NetView | Tivoli Ready |
| Tivoli Reporter | TME |

The following product names are trademarks of Lotus Development Corporation:

| | |
|-------------|-------|
| Domino | Lotus |
| Lotus Notes | Notes |

The following terms are trademarks of other companies:

Tivoli, Manage. Anything. Anywhere., The Power To Manage., Anything. Anywhere., TME, NetView, Cross-Site, Tivoli Ready, Tivoli Certified, Planet Tivoli, and Tivoli Enterprise are trademarks or registered trademarks of Tivoli Systems Inc., an IBM company, in the United States, other countries, or both. In Denmark, Tivoli is a trademark licensed from Kjøbenhavns Sommer - Tivoli A/S.

C-bus is a trademark of Corollary, Inc. in the United States and/or other countries.

Java and all Java-based trademarks and logos are trademarks or registered trademarks of Sun Microsystems, Inc. in the United States and/or other countries.

Microsoft, Windows, Windows NT, and the Windows logo are trademarks of Microsoft Corporation in the United States and/or other countries.

PC Direct is a trademark of Ziff Communications Company in the United States and/or other countries and is used by IBM Corporation under license.

ActionMedia, LANDesk, MMX, Pentium and ProShare are trademarks of Intel Corporation in the United States and/or other countries.

UNIX is a registered trademark in the United States and other countries licensed exclusively through The Open Group.

SET and the SET logo are trademarks owned by SET Secure Electronic Transaction LLC.

Other company, product, and service names may be trademarks or service marks of others.

Appendix C. Related publications

The publications listed in this section are considered particularly suitable for a more detailed discussion of the topics covered in this redbook.

C.1 IBM Redbooks publications

For information on ordering these publications see “How to get IBM Redbooks” on page 341.

- *Disaster Recovery with HAGEO: An Installer's Companion*, SG24-2018
- *Consolidating UNIX Systems onto OS/390*, SG24-2090
- *Backup, Recovery, and Availability on RS/6000 SP*, SG24-4695
- *Implementing High Availability on RISC/6000 SP*, SG24-4742
- *Understanding IBM RS/6000 Performance and Sizing*, SG24-4810
- *Selecting a Server - The Value of S/390*, SG24-4812
- *AIX and Windows NT Solutions for Interoperability*, SG24-5102
- *Inside the RS/6000 SP*, SG24-5145
- *SP Perspectives: A New View of Your SP System*, SG24-5180
- *AS/400 Consolidation Strategies and Implementation*, SG24-5186
- *Monitoring and Managing IBM SSA Disk Subsystems*, SG24-5251
- *HACMP Enhanced Scalability Handbook*, SG24-5328
- *PSSP 3.1 Announcement*, SG24-5332
- *IBM Storage Solutions for Server Consolidation*, SG24-5355
- *The RS/6000 SP Inside Out*, SG24-5374
- *Implementing the Enterprise Storage Server in Your Environment*, SG24-5420
- *S/390 Server Consolidation - A Guide for IT Managers*, SG24-5600
- *Elements of Security: AIX 4.1*, GG24-4433

C.2 IBM Redbooks collections

Redbooks are also available on the following CD-ROMs. Click the CD-ROMs button at <http://www.redbooks.ibm.com/> for information about all the CD-ROMs offered, updates and formats.

| CD-ROM Title | Collection Kit Number |
|--|-----------------------|
| System/390 Redbooks Collection | SK2T-2177 |
| Networking and Systems Management Redbooks Collection | SK2T-6022 |
| Transaction Processing and Data Management Redbooks Collection | SK2T-8038 |
| Lotus Redbooks Collection | SK2T-8039 |
| Tivoli Redbooks Collection | SK2T-8044 |
| AS/400 Redbooks Collection | SK2T-2849 |
| Netfinity Hardware and Software Redbooks Collection | SK2T-8046 |
| RS/6000 Redbooks Collection (BkMgr) | SK2T-8040 |
| RS/6000 Redbooks Collection (PDF Format) | SK2T-8043 |
| Application Development Redbooks Collection | SK2T-8037 |
| IBM Enterprise Storage and Systems Management Solutions | SK3T-3694 |

C.3 Other resources

These publications are also relevant as further information sources:

- *AIX Version 4.3 Network Installation Management Guide and Reference*, SC23-2627
- *HACMP Version 4.3 AIX Planning Guide*, SC23-4277
- *HACMP Version 4.3 AIX: Installation Guide*, SC23-4278
- *PSSP Administration Guide*, SA22-7348
- *Managing Shared Disks*, SA22-7349
- *IBM AIX Parallel I/O File System: Installation, Administration, and Use*, SH34-6065
- *IBM Systems Journal*, SJ34-2, G321-0120
- *IBM Redpiece: PSSP 3.1 Survival Guide*, SG24-5344
- *IBM RS/6000 Performance in Focus*, SG24-5511 (available at a later date)

C.4 Referenced Web sites

These Web sites are also relevant as further information sources:

- http://www.rs6000.ibm.com/hardware/enterprise/s80_specs.html
- <http://www.rs6000.ibm.com/estimator>
- <http://www.rs6000.ibm.com/hascripts>
- <http://www.lotus.com>
- <http://www.ibm.com/storage>
- <http://www.ibm.com/servers/clusters>

How to get IBM Redbooks

This section explains how both customers and IBM employees can find out about IBM Redbooks, redpieces, and CD-ROMs. A form for ordering books and CD-ROMs by fax or e-mail is also provided.

- **Redbooks Web Site** <http://www.redbooks.ibm.com/>

Search for, view, download, or order hardcopy/CD-ROM Redbooks from the Redbooks Web site. Also read redpieces and download additional materials (code samples or diskette/CD-ROM images) from this Redbooks site.

Redpieces are Redbooks in progress; not all Redbooks become redpieces and sometimes just a few chapters will be published this way. The intent is to get the information out much quicker than the formal publishing process allows.

- **E-mail Orders**

Send orders by e-mail including information from the IBM Redbooks fax order form to:

| | e-mail address |
|-----------------------|---|
| In United States | usib6fpl@ibmmail.com |
| Outside North America | Contact information is in the "How to Order" section at this site: http://www.elink.ibm.com/pbl/pbl |

- **Telephone Orders**

| | |
|---------------------------|--|
| United States (toll free) | 1-800-879-2755 |
| Canada (toll free) | 1-800-IBM-4YOU |
| Outside North America | Country coordinator phone number is in the "How to Order" section at this site: http://www.elink.ibm.com/pbl/pbl |

- **Fax Orders**

| | |
|---------------------------|--|
| United States (toll free) | 1-800-445-9269 |
| Canada | 1-403-267-4455 |
| Outside North America | Fax phone number is in the "How to Order" section at this site: http://www.elink.ibm.com/pbl/pbl |

This information was current at the time of publication, but is continually subject to change. The latest information may be found at the Redbooks Web site.

IBM Intranet for Employees

IBM employees may register for information on workshops, residencies, and Redbooks by accessing the IBM Intranet Web site at <http://w3.itso.ibm.com/> and clicking the ITSO Mailing List button. Look in the Materials repository for workshops, presentations, papers, and Web pages developed and written by the ITSO technical professionals; click the Additional Materials button. Employees may access MyNews at <http://w3.ibm.com/> for redbook, residency, and workshop announcements.

Glossary

ADSM. Adstar Distributed Storage Management (now called TSM).

AIX. Advanced Interactive Executive.

AMASS-OMM. AMASS Off-line Media Manager.

API. Application Programming Interfaces.

ATM. Asynchronous Transfer Mode.

CPU. Central Processing Unit.

CRC. Cyclic Redundancy Check.

CSPOC. Cluster Single Point of Control.

DARE. Dynamic Automatic Reconfiguration Event.

DCE. Distributed Computing Environment.

DFS. Distributed File System.

DNS. Domain Name server.

DRM. Disaster Recovery Manager.

ECC. Error Checking and Correction.

EM. Event Management.

EMAPI. Event Manager Application Programming Interface.

ERP. Enterprise Resource Planning.

ESS. Enterprise Storage Server.

FCAL. Fibre Channel Arbitrated Loop.

FDDI. Fibre Distributed Data Interchange.

FFDC. First Failure Data Capture.

FTP. File Transfer Protocol.

GDPS. Geographically Dispersed Parallel Sysplex.

GEM. Tivoli Global Enterprise Manager.

GODM. Global Object Data Manager.

GPFS. General Parallel File System.

GUI. Graphical User Interface.

HACMP. High Availability Cluster Multi Processing.

HACMP/ES. HACMP Enhanced Scalability.

HACWS. High Availability Control WorkStation.

HAGEO. High Availability Geographic Cluster.

HAI. High Availability Infrastructure.

HSD. Hashed Shared Disk.

HSM. Hierarchical Storage Manager.

HTTP. Hypertext Transfer Protocol.

IBM. International Business Machines.

IP. Internet Protocol.

IPAT. IP Address Takeover.

ISS. Interactive Session Support.

ISV. Independent Software Vendor.

ITSO. International Technical Support Organization.

JFS. Journaled File System.

KM. Knowledge Module.

LDAP. Lightweight Directory Access Protocol.

LSF. Log Structured File.

LVM. Logical Volume Manager.

NFS. Network File System.

NIM. Network Installation Manager.

NIS. Network Information System.

NNTP. Native Internet News Protocol.

NSM. Network Storage Manager.

OLTP. Online Transaction Processing.

PFA. Predictive Failure Analysis.

PIOFS. Parallel I/O File System.

PMAN. PSSP Problem Management.

POP3. Post Office Protocol Version 3.

PPRC. Peer to Peer Remote Copy.

PSSP. Parallel System Support Programs.

PTPE. Performance Toolbox Parallel Extensions.

PTX. Performance Toolbox.

RAS. Reliability, Availability, and Serviceability.

RDBMS. Relational Database Management System.

ROLTP. Relative Online Transaction Processing.

RS/6000. RISC System/6000.

RSCT. RS/6000 Cluster Technology.

RVA. RAMAC Virtual Array.

RVSD. Recoverable/Virtual Shared Disk.

SDR. System Data Repository.

SLIP. Serial Line Internet Protocol.

SMIT. System Management Interface Tool.

SMTP. Simple Mail Transfer Protocol.

SNMP. Simple Network Management Protocol.

SOCC. Serial Optical Channel Converter.

SP. Scalable PowerParallel.

SSA. Serial Storage Architecture.

SSL. Secure Socket Layer.

T0. Time 0.

TCP. Transmit Control Protocol.

TDM. Tivoli Distributed Monitoring.

TEC. Tivoli Enterprise Console.

TSM. Tivoli Storage Management (previously called ADSM).

UDF. Universal Data Format.

UDP. User Datagram Protocol.

VFS. Virtual File System.

VMM. Virtual Memory Manager.

VSD. Virtual Shared Disk.

VSS. Virtual Storage Server.

VTS. Virtual Tape Server.

WLM. Workload Management.

XRC. Extended Remote Copy.

Index

Symbols

/etc/inittab file 196
/etc/qconfig file 52

Numerics

2-tier and 3-tier concepts 275
2-tier concept solution 277
32-node support 245
3-tier concept solution 277
64-bit support 9

A

accounting 54
ACLs 62
ActiveX Data Object 308
Adaptive Server 274
Adding a Class (WLM) 198
ADIC 168, 178, 327
Administration (WLM) 194
Adobe Acrobat Reader 9
ADSM 167, 168, 169
Advanced ClusterProven 255, 257
Advisor 223
AFPTM 160
AFS 229
AIX 9
ALIGN 76
AMASS Offline Media Manager 168, 178, 327
API 132, 278
Application and Database Protection 236
Application Errors 180
Application Integration 36
Application pathname (WLM) 187
Application Planning 253
Appraiser 328
Architecture 45
ARCserveIT 168, 177, 329
Arcus 171
ASP 318, 322
ATM 13, 240
ATM Support 246
Automation Solutions 146, 148
Automounter 123
azizo 130

B

B2B 313
B2C 313
Baan 152, 153
Backup Sites 184
Batch Processing 151
BEST/1 69, 133
BEST/1 Agents 328
BEST/1 Analyze 328
BEST/1 Collectors 328
BEST/1 for Distributed Systems 328
BEST/1 Performance Console 328
BEST/1 Performance Modules 134, 328
BEST/1 Predict 328
BEST/1 Visualizer 328
BEZ 327
BEZPlus 69, 327
BMC 117, 124, 133, 134, 148, 150, 151, 152, 158, 328
boot/install server 57
BOP 66
Bourne Shell 302
BSA 76
Business integration 6
Business Intelligence 69
Business Recovery Service 184

C

C Shell 302
C++ 324
CA 153, 168
CA-Ingress 10
Candle 148, 328
Candle Command Center 240
cascading access 297
CBR 219, 231
CDE 11
Central manager server (LoadLeveler) 216
Centralization 32
Centralized Backup and Recovery 166
Centralized Backup and Recovery Solutions 168
CGI 301
Change Management 164
Change Management Solutions 164
Checkpoint Firewall 69
chuser 121
CICS 278

Class Assignment Rules (WLM) 189
Class Limits (WLM) 192
Class Shares (WLM) 191
Classes (WLM) 187
client/server computing model 3
Cluster (LoadLeveler) 215
Cluster Disks 252
Cluster Networks 251
Cluster Nodes 249
Cluster Single Point of Control 184
ClusterProven 255
COBOL 324
Command Center 148, 329
COMMAND/POST 150, 151, 153, 158, 328
COMMAND/POST Enterprise Server 150
COMMAND/POST Explorer 151
COMMAND/POST PhonePoint 151
Commercept 69
Computer Associates 168, 177, 329
Compuware 136, 164, 329
Concurrent Access 246
concurrent access 297
Configuration Management 162
Configuration Management Solutions 163
Consolidated Software Errors and Recovery 180
Content Hosting 318
Control Workstation 57, 64
CONTROL-M 69, 152
CONTROL-M for Distributed Systems 328
CORBA 308
CorpView DBA 69, 328
CPU Options 250
CRC 183
CRM 313
CSPOC 184, 234
CSS 64
Customer example 264
Cyclic Redundancy Check 183

D

DARE 243
Data Encryption 172
Data integration 35
Data Marts 282
Data mining 282
Data Partitioning 292
Data Warehouse 281
Database instance 292

Datasafe 171
datawarehouse systems 69
DB2 291
DB2 Database Server 272
DB2 Universal Database 273
DBMS 271
DCE 9, 11, 278
DDR 280
Decision Support Systems 69
DECS 308
Dedicated Temporary Tablespaces 296
Default class (WLM) 187
DES 64
Disaster Recovery 170, 183
Disk Drive Predictive Failure Analysis 181
Disk Mirroring/Logical Volume Mirroring 247
Disk Space Management 139
Disk Space Management Solutions 139
Dispatcher 220
Distributed computing 3
Distributed Lock Manager 294
Distributed system 275
DLM 294, 295
DRDA 300
DSA 297
dsh 58
DSS 281
DTP 279
Dynamic Reconfiguration 184, 234

E

EAI 311, 314
e-business solutions 272
ECC 181
EcoSYSTEMS 137, 329
EcoTOOLS 69, 136, 164, 329
EDA/SQL 300
EM 154
email 157
EMAPI 154, 155
enccp 279
Encina 277, 278
Enterprise Java Beans 303, 322
Enterprise Storage Server 258
EOP 66
ERP 6
Error Checking and Correction 181
error log 52, 180

ESCON 13, 260
Event Management 153, 154, 243
Event Management Solutions 153, 158
Event Manager Application Programming 155
Event Manager Application Programming Interface 154
Executing server (LoadLeveler) 216
Executor 221
Extended ERP 312

F

F50 10
FCS 251
FDDI 13, 240
FFDC 183
File-AID 329
First Failure Data Capture 183
Frame 47

G

Gartner Cost of Ownership 77
GDPS 261
GEM 147, 331
General Parallel File System 139, 141
GODM 246
GPFS 139, 141
Group Services 243
Groups (WLM) 187

H

H50 12
HACMP 13, 179, 182, 184, 293, 330
HACMP/ES 243
HACWS 241
HAGEO 184, 240, 331
Hardware Recovery 181
Hashed Shared Disk 139, 140
Heartbeat rate 247
heterogeneous 115, 118, 158, 171, 172, 176
High Availability 13
High Availability 142, 233
High Availability Cluster Multi-Processing 182, 184, 293
High Availability Geographic Cluster 184
High Availability Options at the Operating System Level 247
High Performance Switch 144, 295, 299

Hot Standby 236
HP OpenView 151, 153, 158
HSD 139, 140
HTTP 8

I

IBM DB2 10
IBM LoadLeveler 330
IBM NetTape 330
IBM Netview 330
IBM Parallel System Support Program 330
IBM Shared Disk 139, 330
ICOM 76, 77
Improved snapshot facility 246
Independent Software Vendors 154, 327
Info/Management 153
Infoprint Manager 160
Infoprint Manager for AIX 330
Informix 10, 172, 297
Informix Online Dynamic Server 273
Informix Online Extended Parallel Server 273
Informix Universal Server 273
Initiating Workload Manager 196
Instance Owner 292
Intelligent Miner IM 282
Interactive Service Support 183
Internet Service Providers 318
Investigator 327
IP Address Protection 239
IP Sec 9
IPAT 239
ISP 229, 318
ISS 183, 219
ISS/ND 319
ISV 154, 181, 327

J

Java Development Kit 228
Java Server Pages 322
Java Servlet 302
JD Edwards' OneWorld 283
JDBC 301
JDK 9
JFS 141
JIT 9
JNDI 303
Job (LoadLeveler) 215
Job classes (LoadLeveler) 217

Job Command File (LoadLeveler) 215
Job Manager 55
Job Scheduler 330
Job Scheduler 152
Job Step (LoadLeveler) 215
Journaled File System 141

K

Kerberos 62
KM 158
Knowledge Modules 158

L

LDAP 9, 125
LEDs 47
Legato 168, 174, 176, 329
LEI 308
LoadLeveler 131, 133, 154, 215
Localized Automation 150
Logical Volume Manager 139
Lotus Domino 287
Lotus Domino Enterprise Data Integration 308
Lotus Domino Go Webserver 9
Lotus Notes 286
LPP 133
LRU 294
LSF 259
lsuser 121
LU 6.2 279
LVM 139, 233

M

Mega Web-site management 318
metadata 282
MIB 230
Micro Channel 45
Microsoft Exchange 172
Microsoft SQL Server 172
Migration Service 85
MIMD 45
mkuser 121
Monitor Workload Management 205
MPE 152
MPP 297
MQSeries 69
MQSeries Integrator 315
mutual access 297

Mutual takeover 236

N

NAS 309
NetBackup 167, 168, 171, 330
Netscape FastTrack 9
Netscape Navigator 9
NetTAPE 168, 169
NetView 146
Network Failure Protection 239
NetWorker 168, 174, 176, 330
NFS 57, 116, 117, 141
NIM 56, 116
NIS 56, 120, 122, 123, 124
Node 47
Node Considerations 250
Nodegroups 292
NotesPump 308
NotesSQL 308
Novell NetWare 172
Novell Network Services 9
NSM 259

O

ObjectBroker 317
ODBC 301
ODM 56
OLAP 282
OLTTP 71
Operating System Install and Updates 116
OPQ 294
OPS 294
Oracle 10, 152, 153, 172
Oracle 8i 273
Oracle Applications 283
Oracle Express 273
OS/390 158
Outsourcing 184, 325

P

Packaging 246
pager 150, 151, 158
Parallel Cost Model for the Optimizer 296
Parallel Create Index 295
Parallel Data Load 295
Parallel database 291
parallel file system 142

Parallel I/O File System 139, 143
parallel mode server 294
parallel query 294, 295
Partitioned Views 296
Partitioning Key 292
Partitioning Map 292
PATROL 150, 151, 153, 158, 328
Patrol 69
PCI 45
PeopleSoft 152, 153
PeopleSoft Applications 283
Performance Management 127
Performance Management Solutions 128
Performance Measurements 127
Performance Requirements 253
Performance Toolbox 127, 129, 132
Performance Toolbox Parallel Extensions 127, 131
PERL 302
PFA 181
Physical Consolidation 33
Physical Server Consolidation 34
PIOFS 139, 143
PMAN 156
Positioning 259
POWER3 45
PowerPC 45
PPC 279
PPRC 259
Predictive Failure Analysis 181
Print Management 51
Print Services Management 159
Print Services Management Solutions 159
Printer/Plotter Device 51
Progress 10
Project Definition Workshop 85
ProjectWatch 85
Property files (WLM) 194
PSSP 48, 120, 121, 122, 130, 132, 133, 142, 154, 179, 181, 330
PSSP Problem Management 155, 156, 343
PTPE 131
PTX 132

Q

QACenter 69, 329
QALoad 69
QoS 226
Queue 52

Queue Device 52

R

RAID 234, 247
RAS 181
RDBMS 163, 299
Recoverable/Virtual Shared Disk 139, 140
Red Brick 273
Regaining Flexibility 21
Remedy 153
Remote Agents 150
Resource Group Options 253
Resource Monitors 154
Resource Planning 252
rlogin 122
rmuser 121
Rotating Standby 236
RQS 279
RSCT 243
RVA 259
RVSD 139, 140

S

S80 7
SAG CLI 300
SAN 309
SAP R/3 13, 152, 153, 172, 283
schedtune command 194
Scheduling server (LoadLeveler) 215
Scorpion 76
SCSI disk 234
SDA 228
SDR 55, 155
SDR Independency 246
Secure Sockets Layer (SSL) 125
SecureWay Network Dispatcher 330
SecureWay Network Dispatcher 218
Security 31
Serial (RS232) 252
Server Consolidation Savings Estimator 77
SerView DBA 69, 328
SFS 279
Shared pseudo-class (WLM) 188
Single Point of Control 20, 50
SISD 45
SLIP 240
SmartMedia 168, 176, 330
SMIT 50, 116

SmoothStart 85
SMP 45, 182
SNA 148, 279
SNMP 129, 131, 148, 151, 158, 230
SOCC 240
Software Installation Management 115
Software Installation Management Solutions 115
Software Installation Products 117
Software Recovery 179
SP 45
SP Access Control 122
SP File Collections 120, 122
SP Switch 59, 154, 155
spacs_cntrl 122
spchuser 121
spluser 121
spmuser 121
sprmuser 121
SQL BackTrack 69, 328
SSA 11, 234
Standard UNIX Performance Tools 128
Star Queries 296
Storage Area Networks 263
Storage Consolidation 260
Strategist 328
svmon command 207
Switch adapter 45
Sybase 10, 172, 298
Sybase Adaptive Server 273
Sybase IQ 273
Sybase MPP 273
Sysback/6000 168, 331
Sysctl 61
syslog 180
SYSPRIO 217
System class (WLM) 188
System Data Repository 155
System Errors 179
System logging 52
System Management 48
System Monitor 63
System Recovery 166
Systems Management Transformation Projects 85
SystemWatch 85

T

Tape Management 167
TAR 171, 172

Target-mode SCSI 252
Target-mode SSA 252
TCL 302
TCO 22
TDM 156
TEC 156, 157, 331
Tiers (WLM) 189
Tivoli 153, 181, 331
Tivoli Asset Management 163, 331
Tivoli Change Management 164, 331
Tivoli Decision Support 148, 331
Tivoli Destiny/Output Manager 161, 331
Tivoli Distributed Monitoring 156, 331
Tivoli Enterprise Console 148, 156, 157, 331
Tivoli Framework 157, 163
Tivoli Global Enterprise Manager 147, 331
Tivoli Global Sign-On 125, 331
Tivoli Inventory 163, 331
Tivoli Maestro 152
Tivoli Management Framework 117, 126
Tivoli NetView 147, 331
Tivoli Security Management 331
Tivoli Service Desk 331
Tivoli Software Distribution 117, 163, 331
Tivoli Storage Management 168, 169, 331
Tivoli TME 151, 158
Tivoli User Administration 124, 126, 331
Tivoli User Administration OnePassword 125
Tivoli Workload Scheduler 331
TM-XA 279
Topology DARE 245
Topology Services 243
TP Monitors 277
Types of Consolidation 29

U

Unclassified pseudo-class (WLM) 188
UNIFACE 69
Uniface 329
User Management 118
User Management Solutions 119
Users (WLM) 187
UUCP 137

V

Vantive 153
VERITAS 168, 171, 330
VFS 141, 144

Virtual File System 141, 144
Virtual Printer 52
Virtual Shared Disk 131, 139, 154, 294
VLDB 275, 290, 299
VMM 185
vmtune command 194
VSD 55, 131, 139, 294
VSS 259
VTS 259

W

Web 136, 161
Web Cache Manager 259
Web Server Logic 302
WebSphere Application Server 287
WebSphere Performance Pack 229
Windows NT 152, 172
WLM 185
wlmcntrl command 195
wlmstat command 205
Workload Management 185
WTE 228

X

X/Open 279
xmperf 130
xmservd 131
XPS 297
XRC 259

Y

yppasswd 123
ypupdated 123

IBM Redbooks evaluation

Server Consolidation on RS/6000
SG24-5507-00

Your feedback is very important to help us maintain the quality of IBM Redbooks. **Please complete this questionnaire and return it using one of the following methods:**

- Use the online evaluation form found at <http://www.redbooks.ibm.com/>
- Fax this form to: USA International Access Code + 1 914 432 8264
- Send your comments in an Internet note to redbook@us.ibm.com

Which of the following best describes you?

Customer **Business Partner** **Solution Developer** **IBM employee**
 None of the above

Please rate your overall satisfaction with this book using the scale:
(1 = very good, 2 = good, 3 = average, 4 = poor, 5 = very poor)

Overall Satisfaction _____

Please answer the following questions:

Was this redbook published in time for your needs? Yes___ No___

If no, please explain:

What other Redbooks would you like to see published?

Comments/Suggestions: (THANK YOU FOR YOUR FEEDBACK!)

SG24-5507-00
Printed in the U.S.A.

Server Consolidation on RS/6000

SG24-5507-00

