DB2 for AIX

By Gene Kligerman and Ming Wu

This article describes the capabilities of IBM DATABASE 2™ for AIX (DB2 for AIX) in uniprocessor and SMP environments. It highlights the enhancements incorporated into Version 2.1, provides an introduction to the DB2 process architecture, and discusses performance issues.

DB2® is an integrated family of Relational Database Management Systems (RDBMSs) for a variety of platforms, including MVS™, VM, VSE, OS/400®, OS/2®, AIX, Sun® Solaris®, and HP/UX. IBM plans to support additional platforms in the future.

Introduction

IBM brings the dependability, integrity, and proven design of DB2 to the UNIX® environment. DB2 for AIX delivers reliable performance, high availability, outstanding recoverability features, and seamless access to DB2 on the host through the DB2 Distributed Database Connection.

DB2 for AIX has many features to help customers fine-tune application performance while ensuring data integrity: a cost-based optimizer, stored procedures, row blocking, compound SQL, multiple levels of concurrency, and a granular locking scheme.

Advanced recovery features protect data, even between backups. High availability is achieved through online backup. DB2 for AIX delivers full transaction support, ensuring that transactions are completed normally before changing any data. Row-level locking and declarative referential integrity contribute to even safer data handling.

DB2 for AIX supports an SQL that is compatible with DB2 for MVS. Since it also provides an interface that is similar to DB2 on other platforms, DB2 applications and skills are easily transferred to a client-server environment. DB2 for AIX also supports a wide variety of applications running on DOS, Windows™, OS/2, AIX, HP/UX, and Solaris clients. New applications can be developed on these clients using DB2 software developers' kits, which have a variety of languages and access methods.

Both DB2 Version 1 and Version 2 support AIX Versions 3.2.4, 3.2.5, and 4.1.1 (in uniprocessor and SMP environments). Because Symmetric Multiprocessor (SMP) hardware supported by AIX became available after DB2 Version 1 and before Version 2, the implementation of SMP support differs slightly between the two DB2 versions.

DB2 Version 1 Support of AIX 4.1

DB2 Version 1 was architected as a multiprocessing application, with each asynchronous unit of activity implemented as an AIX process.

DB2 creates AIX processes called *database agents* that each manage one database connection. The DB2 Database Administrator (DBA) can specify a maximum number of database agents (the total number of users that can connect to the database) and the maximum number of concurrent agents (the number of agents that can actively execute SQL statements simultaneously). If more connections are attempted than the maximum value allows, the additional agents must wait. This enables a DBA to control the use of system resources.

Since DB2 was implemented as a multiprocessing application from the beginning, it was not necessary to redesign the product to support the new AIX SMP environment.

When DB2 Version 1 was first released in October 1993, it supported AIX 3.2.4, AIX 3.2.5, and uniprocessor IBM RISC System/6000s available at the time. Providing support for the new SMP machines has two implications.


Ming Wu


Gene Kligerman

- DB2 must support AIX 4.1.1 in both uniprocessor and SMP environments.
- DB2 must use the new AIX 4.1.1 operating system services to enable it to run safely in an SMP environment.

Supporting AIX 4.1.1 was easy because DB2 uses only standard AIX interfaces that are highly compatible with AIX 3.2.5. IBM verified DB2's support for AIX 4.1 in a uniprocessor environment by successfully running a full regression test.

The main advantage of IBM's SMP architecture is the direct portability of most software developed for uniprocessors. Few changes were required for DB2 Version 1 to support the SMP environment. Care was taken to ensure that SMP-enabling of DB2/6000™ did not affect the performance of the uniprocessor DB2 installations. The multiprocessing architecture of DB2 Version 1 provided a ready-made basis for exploiting SMP.

Two changes were necessary for Version 1 of DB2 to support SMP hardware under AIX 4.1:

- The code was modified to use the new latching services that were introduced in AIX 4.1 to serialize access to shared DB2 data structures.
- The DB2 installation process was modified to ensure that the code installed was appropriate for the level of operating system (AIX Version 3.2.4, 3.2.5, or 4.1.1) and hardware (uniprocessor or SMP).

DB2 for AIX Version 2.1

DB2 for AIX Version 2, with improved features and performance enhancements, is accompanied by a rich suite of tools for database administration and performance tuning. At the time this article is being written (January 1995), DB2 for AIX Version 2.1 is in beta testing with customers worldwide. DB2 for AIX Version 2 comes as a single binary that supports all levels of AIX—from 3.2.4 to 4.1.1 and beyond—ensuring that all users can benefit equally from the DB2 for AIX Version 2 performance and functional enhancements.

Highlights of Version 2 are described in the following sections.

Innovative Multimedia and Object-Oriented Applications

Extensions to the relational capability of DB2 enable multimedia objects to be manipulated and object-oriented programming techniques to be used to increase programmer productivity and data integrity. These extensions include the following.

- User-Defined Functions (UDFs), User-Defined Triggers (UDTs), check constraints, and Large Objects (LOBs)
- The ability to write recursive SQL queries so that a single SQL statement can replace complex application logic in applications such as bills-of-material processing and path expression queries

Flexible Management of Large Databases

Managing very large databases (limited only by the size of physical storage) is easier when a database can be partitioned into separately managed parts called *tablespaces*. DB2 Version 2 has enhancements for large database administration and improved database availability, including the following:

- Ability to place tables, indexes, and LOBs into different tablespaces, and flexibility in assigning storage devices to tablespaces, enabling administrators to maximize database availability and performance
- Support for online or offline backup and recovery of all or part of the database using tablespaces
- Support for backup and recovery utilities using several devices in parallel, reducing the time for these activities
- A high-speed load utility to perform bulk loading of data into tables

A key advantage of IBM's SMP architecture is the direct portability of most software developed for uniprocessors.

Distributed Applications and Data Replication

With DB2 for AIX Version 2.1 and accompanying products, the application developer and database administrator have great flexibility in distributing data and applications across various workstation and host platforms. Enhancements in this area include the following:

- Support for Distributed Relational Database Architecture™ Application Server (DRDA™ AS) to provide access to DB2 databases on AIX and OS/2 from MVS, VM, and OS/400 applications
- Support for data replication products such as Data Propagator™ Non-Relational, Data Propagator Relational, and DataRefresher™, enabling extensive data replication facilities from DB2 for MVS, DB2/400™, IMS, and VSAM to AIX and OS/2 DB2 databases
- Distributed Unit of Work (DUOW), also known as two-phase commit support

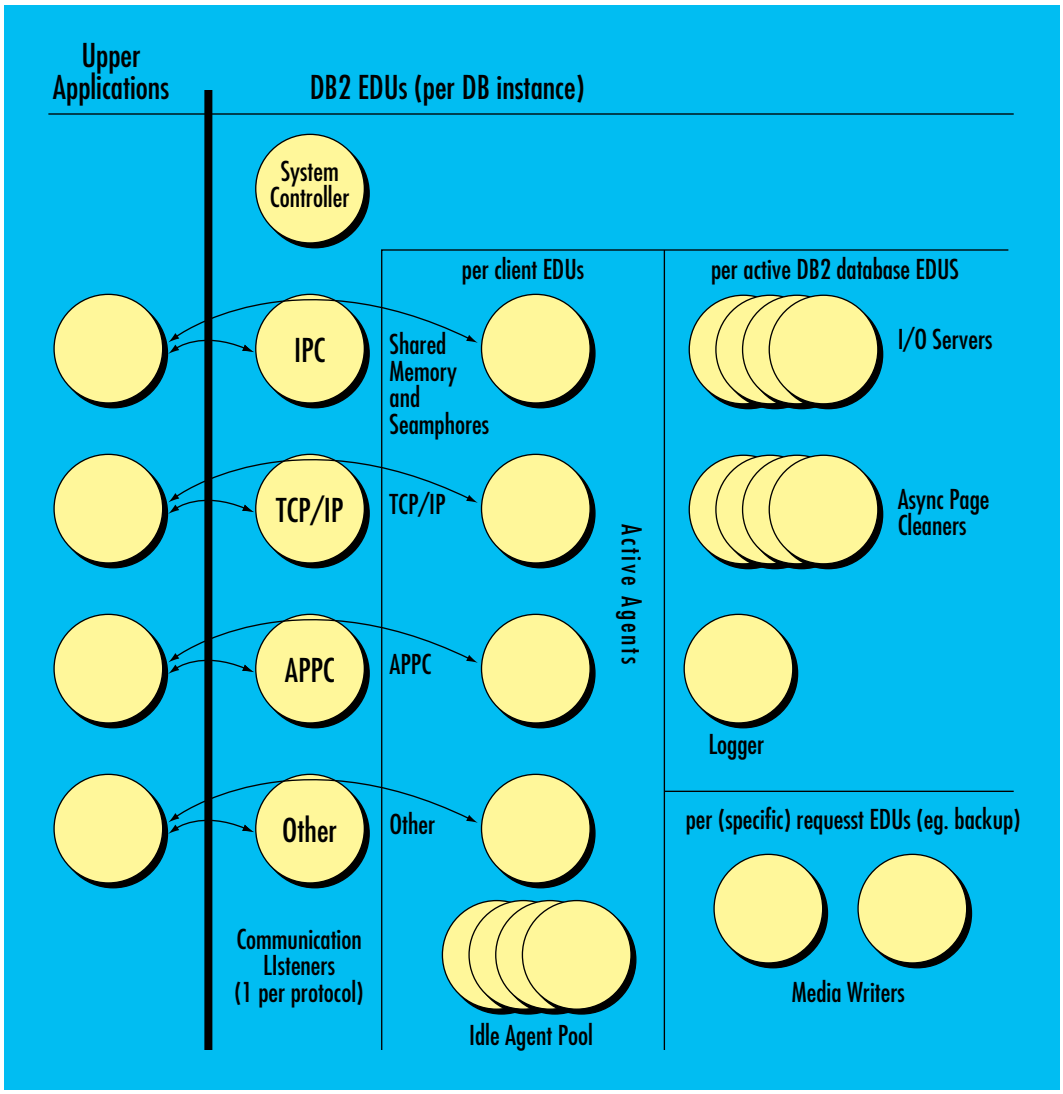

Figure 1. The DB2 process model architecture

Performance

The new SQL optimizer architecture uses sophisticated techniques to transform complex queries into their most efficient form and to accurately determine the best access path to data. Support for I/O prefetch and parallel I/O further improves the performance of Online Transaction Processing (OLTP) applications, decision-support applications, and database utilities.

DB2 Version 2 Architecture

The architecture of DB2 Version 2 incorporates Engine Dispatchable Units (EDUs), the fundamental dispatching unit for all asynchronous activities performed by the DB2 database engine.

EDUs are a logical concept. Physically, they are implemented as either processes or threads depending on the platform. In the OS/2 environment, EDUs are implemented as threads within a single process. In the AIX environment, EDUs are implemented as AIX processes. Although AIX 4.1 provides thread support, the need to support DB2 under AIX 3.2 (which does not have thread support) dictates the use of processes in DB2 for AIX Version 2. Threads will be used in the future on thread-efficient platforms.

The DB2 engine creates different types of EDUs for specific tasks.

- **System controller**: The master EDU controls the starting and stopping of DB2 and owns the shared resources.
- **Agents**: An agent EDU is assigned to every connected client application. A pool of idle agents can be defined. When a new database connection is made, one of these idle agents is dispatched to handle the connection. Using idle agents reduces the connect time of applications to the database.
- **Asynchronous page cleaners**: EDUs off-load buffer page writing operations from the agent EDU executing the SQL query. This eliminates the need for the agent EDU to wait synchronously until modified pages are written from the buffer to disk to make room for query data. Using asynchronous page cleaners is particularly advantageous for user environments in which the arrival of database requests is irregular. The page cleaners take advantage of lulls in database activity to flush buffer pages to disk in preparation for future requests.
- **I/O servers**: These EDUs perform parallel I/O and big-block reads, which can dramatically improve performance in decision-support environments. For example, if DB2 is performing a table scan, I/O server EDUs will be used to prefetch multiple pages in a single request (sequential prefetch). I/O server EDUs are also invoked by the DB2 index manager to prefetch data based on an index sequence (index prefetch). The big-block read capability (reading several disk pages using a single I/O request) can reduce the CPU overhead and improve query response time.
- **Media writers**: The capability to perform backup and recovery using multiple devices in parallel is built into the DB2 engine. DB2 dispatches media writer EDUs to perform the backup and recovery operations in parallel, dramatically reducing the elapsed time to perform these functions. The high-speed load utility also uses media writer EDUs to store data directly onto disk, bypassing all the processing normally associated with SQL `INSERT` statements.
- **Communication listeners**: One communication listener EDU is assigned to each type of communications protocol supported by DB2, such as Interprocess Communications (IPC) for local clients, and Advanced Program-to-Program Communications (APPC), TCP/IP, and others for remote clients.
- **Database logger**: The logger EDU writes a log of the changes to the database data performed by agents for user applications. There is one logger EDU for each DB2 database controlled by the DB2 instance.

DB2 has no concept of checkpoints. Unlike other open systems databases, DB2 does not have performance degradations at periodic intervals to flush buffer pages to disk and to commit the changes to the log. All committed transactions are guaranteed to be recoverable, preserving the Atomicity, Consistency, Isolation, and Durability (ACID) transaction properties of the DB2 database.

Figure 1 illustrates the DB2 process model architecture. The circles represent EDUs (processes under AIX). A DB2 instance is a single system that controls one or more DB2 databases. The System Controller is responsible for creating all DB2 EDUs, including communication listeners, media writers, and idle agents.

EDUs are the fundamental dispatching unit for all asynchronous activities performed by the DB2 database engine.

When a user application attempts to connect to a DB2 database, an appropriate communication listener will detect the conversation attempt, and assign one of the idle agents (or create a new agent) to own the connection and perform database work.

The database is inactive (no runtime resources assigned to it) until at least one user is connected to it. The System Controller activates a database at the time of the first connect to it by creating the EDUs associated with a database and assigning resources to it.

Each DB2 database has a logger EDU and a configurable number of I/O servers and asynchronous page cleaners. All database EDUs and allocated memory are released when the last application disconnects from the database.

Media writer EDUs are created when a specific task needs to be done, such as backup or recovery, which differs from other DB2 EDUs.

DB2 Performance Results

In January 1995, IBM announced the completion of industry-standard TPC-C benchmark testing of DB2 for AIX Version 2.1.

A RISC System/6000 POWERserver™ R24, which has a single POWER2 processor and DB2 for AIX Version 2.1 achieved 1,470.06 transactions per minute (tpmC) at $896.22/tpmC on the TPC-C benchmark. The transaction rate of 1,470 tpmC

tops the previous best-reported uniprocessor result by 31%, which is on IBM's RISC System/6000 POWERserver 59H with Sybase SQL Server Version 10.0.1

On the TPC-C benchmark, DB2's new result is the highest of any uniprocessor using either open-system or proprietary databases, and tops many of the multiprocessor systems.

The IBM solution of DB2 and POWERserver R24 achieves 22% better tpmC than an eight-processor Sun SPARCserver® 1000E (1,204 tpmC at $890/tpmC) using Sybase® SQL Server 10, 13% better tpmC than a four-processor AT&T® 3555/4 (1,296 tpmC at $1,040/tpmC) using INFORMIX® On-Line 5.03, and 5% better tpmC than a two-processor HP™ Model H70 (1,403 tpmC at $758/tpmC) using SQL Server 10.0.1.

DB2 Version 2 features can increase the performance of transactional and decision-support workloads in both SMP and uniprocessor environments.

These TPC-C results show the excellent transaction processing performance of DB2 for AIX Version 2.1. The test demonstrates sustained rapid response time under a heavy workload from 1,280 users accessing a large database configured with 124 gigabytes of disk. DB2 demonstrates that it can take on the challenge of managing data in very demanding environments.

DB2 for AIX Version 2.1 has been designed to exploit the capabilities of both SMP and uniprocessor architectures to deliver excellent performance and sustained scalability as more SMP processors are added to the system. (**Note**: Measurements of DB2 performance under AIX SMP show outstanding performance; however, they were not completed at press time. Therefore, official numbers were not available for this article. For additional information, contact one of the authors of this article or your IBM marketing rep.)

The AIX Journaled File System (JFS) was used to manage DB2 data on disk during the TPC-C benchmark. The excellent performance of AIX JFS in the heavy workload TPC-C scenario was instrumental in achieving the outstanding TPC-C performance results above.

For ultimate performance and flexibility, DB2 for AIX Version 2.1 users can choose between three mechanisms for managing data on disk: AIX JFS, logical volumes managed by the AIX Logical Volume Manager (LVM), or physical devices managed directly by the DB2 database.

SQL Optimizer Technology

DB2 Version 2 features can increase the performance of transactional and decision-support workloads in both SMP and uniprocessor environments. DB2 Version 2 benefits do not stop there. For a database to perform well on complex decision-support workloads, it is not enough to merely work harder by parallelizing operations, off-loading processing to ancillary processors, and relying on faster CPUs and disks. Working smarter is also essential, and this is exactly what the DB2 SQL optimizer technology is designed to do.

All of IBM's relational database offerings on all platforms have always had a cost-based SQL optimizer. The cost-based optimizer eliminates the need for tailoring the query to the operational environment. Without a cost-based optimizer, poor response times and significant unnecessary work could result.

The increasing importance of decision-support applications is resulting in more complex database queries. These queries may have been written by end users, generated by automatic tools, or produced by the many point-and-click application interfaces popular in DOS, Windows, OS/2, and UNIX environments.

The DB2 optimizer incorporates a sophisticated query rewrite phase that automatically transforms a complex query into a different query that is analyzed to generate the best possible access plan to the data. As a result, the end user does not need to be concerned with coding the optimal SQL query.

The Version 2 optimizer also looks at many alternatives in its search for the best query execution plan. In addition, it engages more sophisticated techniques of modeling the cost of different ways to fetch data from disk. All these techniques can result in an order-of-magnitude performance improvement for complex queries as compared to existing SQL optimizer technology.

The power of choosing the best access plan is not free—it takes time. To allow users the flexibility of balancing the time spent in generating the access plan versus the execution time, users can specify the class of optimization techniques that the DB2 optimizer should apply during access plan generation. Users can also specify the optimization class for the whole application or for an individual SQL statement.

Tools for Managing Performance

To assist database administrators and developers in obtaining the best possible performance from the database under a variety of operating conditions, IBM has developed two new tools: DB2 Visual Explain and DB2 Performance Monitor.

DB2 Visual Explain

An easy-to-use, graphical tool, DB2 Visual Explain provides details about the access plan

chosen by the DBMS optimizer to access data. DB2 Visual Explain supports Version 2 of DB2 for AIX with these capabilities:

- **Graphical presentation of Explain output.** Explain output, which reveals the SQL access plan chosen by the optimizer, is usually textual or tabular in nature. For most SQL statements, this output is difficult to interpret.

  DB2 Visual Explain presents this output in an easy-to-understand graphical format. Relationships between database objects such as tables and indices are instantly clear, as are various operations—such as table scans and sorts—that the optimizer has chosen for accessing the data. With this information, users can tune the structure of their databases for maximum efficiency and performance.

- **Detailed optimizer information.** DB2 Visual Explain provides optimizer information to inform the user of the cost of executing the SQL statements. This information includes I/O and CPU cost estimates for each operation, bind-time, and current table statistics.

  An administrator or developer can easily identify the most expensive operation for a given SQL statement and focus on tuning that operation. For example, tuning might consist of creating an index that would avoid the expense of a full table scan.

- **What-if modeling capability for SQL statements**. With DB2 Visual Explain, users can model the effect of various changes in the database environment on SQL statements. For example, a user can determine the estimated time needed to execute a query in a production environment with a million rows—all without the need to add data to a test environment of 100 rows.

Figure 2 shows a graphical representation of an SQL query that is optimized by DB2 to produce an access plan. The rectangles represent database tables, while the ovals represent database operations such as table scans, index scans, joins, or sorts. The graphical presentation of the access plan makes it easy to identify possible SQL problems. For example, creating an index would eliminate the requirement to do a full table scan. Users can click on any access plan object (in this case, SCAN 42) and view the access plan details associated with the operation.

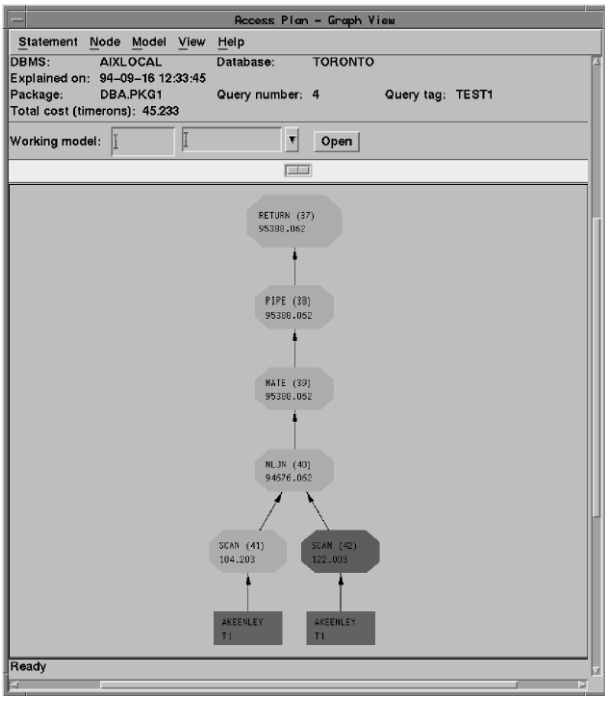

Figure 2. DB2 Visual Explain

DB2 Performance Monitor for AIX

The Performance Monitor provides comprehensive performance data collection, viewing, reporting, analysis, and alerting capabilities for Version 2 of DB2 for AIX.

- **Comprehensive, flexible data collection.** DB2 Performance Monitor for AIX supports more than 200 performance attributes including buffer pool, lock and deadlock, sorting, communication, agent, and logging information. Data is shown for database managers, databases, tablespaces, tables, connections, transactions, and statements.

  Support for two types of performance data, snapshot and event, also provides flexibility. Snapshot data measures performance characteristics at times, while event data summarizes performance attributes for a certain duration (for example, from the time a statement begins executing to the time it completes).

  For a statement event, for example, the DB2 Performance Monitor can monitor the start and stop time, total CPU time, total number and CPU time of sorts, rows read and written, and the SQLCA for the statement. For dynamic statements, the SQL statement text is also captured.

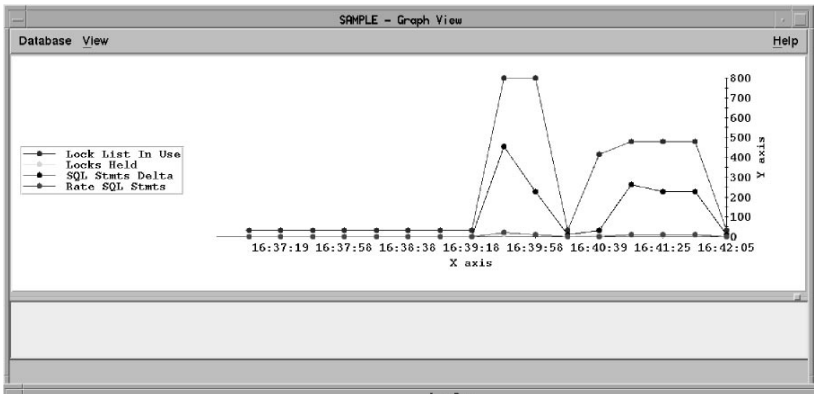

Figure 3. Performance Monitor

- **Easy-to-use, intuitive viewing and reporting.** Collected data can be viewed in real time and recorded for later playback and analysis. Recorded data can also be loaded into an SQL performance database for further querying and reporting.

  When viewing data, users have the option of graphical or textual presentation. Graphical User Interface (GUI) views help users interpret complex data, while textual output provides information for reporting and archival purposes.

- **Powerful data analysis capabilities.** With DB2 Performance Monitor for AIX, users can customize measurements with spreadsheet-like formulas. For example, rather than monitoring an absolute measurement, a user can monitor a ratio calculated from two related measurements. A variety of formulas are supported, including functions such as average, minimum, and maximum.

  DB2 Performance Monitor for AIX has a starter set of commonly used measurements so users can perform useful monitoring immediately. Data can be analyzed to filter out only those records applicable to certain database objects or to a certain period. For example, users tracing a performance problem can quickly focus on event records applicable to a given database.

  The data produced by the DB2 Performance Monitor can be stored in a DB2 database, providing an opportunity for additional analysis. For example, event records showing I/O cost for every statement in a given application can be sorted in descending order, quickly identifying the most expensive statements.

- **Robust alerting capabilities.** For any performance measurement, users can define exception conditions by specifying a threshold value. When that value is reached, the user can specify any or all of the following actions:

  -Notification through a window or audible alarm
  -Logging of a record in a log file
  -Execution of a command or program
  -Notification to a management product such as NetView®

With a comprehensive, flexible, yet easy-to-use array of monitoring options, administrators and developers have a powerful tool to aid in performance tuning and problem determination.

Figure 3 shows, in graph form, the level of database activity over the last five minutes using selected indicators of database performance.

Summary

DB2/6000 has the following benefits in both uniprocessor and SMP environments:

- **Solid architecture**. DB2 can support many users in client-server OLTP and decision-support environments without compromising the efficient use of system resources.
- **Excellent performance**. DB2 for AIX can handle sustained rapid response time under a heavy workload from hundreds of users accessing a large database.
- **Good scalability**. As the number of processors increase, DB2 for AIX takes advantage of them, resulting in greater throughput.

Gene Kligerman, IBM Canada, 1150 Eglinton Avenue East, Toronto, Ontario, Canada, M3C 1H7. Internet: genie@vnet.ibm.com. Mr. Kligerman is currently a planner for the DB2 for AIX Version 2.1 server. He has a BSc and an MSc in Computer Science from the University of Toronto.

Ming Wu, IBM Canada, 1150 Eglinton Avenue East, Toronto, Ontario, Canada, M3C 1H7. Internet: mingwu@torolab2.vnet.ibm.com. Ms. Wu is a member of the workstation DB2 performance group. She has an MASc in Electrical Engineering and a BASc in Engineering Science from the University of Toronto.