RN

Decision Framework, DF-06-1449
K. Fitzpatrick

Research Note
29 October 1998

Platform Availability Data: Can You Spare a Minute?

Data regarding platform availability is key to understanding relevant costs of downtime and recovery for business systems.

Core Topics
IT Measures and Metrics: Measurement of IT; Measuring IT Effectiveness

Choosing the right technology to improve availability and attendant levels of spending requires an understanding of the direct and indirect costs of business system downtime. While much effort can be spent by an organization performing a formal, rigorous business impact assessment (BIA), grouping business systems into rough downtime cost categories and comparing these with platform downtime or availability data can be a relatively simple, efficient indicator of which business systems provide the greatest availability — or are at the greatest risk, and therefore increased cost, of downtime. Understanding acceptable technology spending limits for each business system greatly assists in narrowing choices of recovery alternatives.

Currently, the GartnerMeasurement database, as depicted in Figure 1 shows a wide range of availability among the most-used platforms. These numbers are from 240 observations covering 190 firms.

Figure 1
Availability Ranges for Top IT Platforms

Platform	Outages/Server/ Year	Availability (24x 365 basis)
S/390 (sysplexed)	10 minutes	99.998%
Tandem	1.7 hours	99.98%
AS/400	5.2 hours	99.94%
S/390 (nonsysplexed)	8.9 hours	99.90%
VAX	18.9 hours	99.78%
Unix (all)	23.6 hours	99.73%
NT	224.5 hours	97.44%

Source: GartnerMeasurement

GartnerGroup

Behind the Data
With the exception of S/390, the data reported for the various platforms has not been grouped into clustered/nonclustered segments. Less than 10 percent the more than 150 installations studied have implemented this technology. The VAX platform is an exception to this in that over 50 percent of these installations have clustered servers accessing shared data.

An important factor affecting the numbers is that close to 70 percent of servers in the study are more than two years old. As such, many of these servers do not have the remote monitoring/problem management capabilities being incorporated in the latest generation.

One reason clustering has not been more prevalent among Unix and NT servers, (technology capabilities aside), becomes apparent when the duration of an individual outage is reviewed. This review shows that the recycle times of a failed application or server in these environments are much shorter compared to traditional legacy environments. Given this ,it is human nature to forget the little bumps when is comes time to acquire additional capacities, make platforming decisions, or set strategic direction.

The outages occurring on both the Unix and NT platforms ,while more frequent, tend to be of shorter duration. Likening this to an automobile, while other platforms will occasionally get into major accidents, Unix and NT have more-minor fender benders. However, these minor fender benders, happening as frequently as is seen in the numbers, can also keep an organization from getting to Point A to Point B.

An aspect missing from the numbers is the amount of scheduled downtime that is incurred after a device is made operational from an unplanned outage. Preliminary information suggests that this points out an even greater disparity between platforms. This is due primarily to the lack of regularly tested recovery plans. Future work in this area is being undertaken to clarify this important aspect of overall availability.

What Do the Data Mean?
This chart only compares availability times as indicated by data gathered in GartnerMeasurement benchmarking studies. This is not a comparison chart for indicating which platform may be better or worse than others. The data is presented only to indicate relative availabilities. No attempt is made to compare the abilities of each platform type to run specific applications.

Why Availability Is Important: Decisions are being made today solely on acquisition price, marketing claims, and the desire for the "latest and greatest" technologies. Surprisingly — even frighteningly — little attention is being paid to understanding the impact of new or altered technologies on business systems and processes, let alone legacy systems. Increased complexity leads to increased downtime.

While much attention has been paid to the total cost of operation (TCO) of user desktop appliances, little regard is given to the availability aspect of the back-end devices that are truly running the business. Only a handful of user organizations to date have taken even an initial stab at the cost to the organization of unavailability. System availability is key to success in the emerging global, electronic marketplaces. The future is 24x7 for most firms.

How to Measure Availability: *Account for required hours of operation.* Are users accessing the system around the clock? Does the system only have to be accessible during core business hours — five days a week, eight to 10 hours a day? *Start with the basics.* How many times has the server required a restart? How long did it take? *Strive to take an end-user perspective.* Move beyond a platform view to a business system view. How long does it take for the end user to access the system on average? What applications are required on a daily basis, and how available are the relevant servers?

Recovery Strategies Currently Reaping Improvements in Availability: *Electronic vaulting:* the electronic transfer of daily backups to an alternate facility, thus reducing hours of transport time in the event of a disaster. *Electronic journaling:* the electronic transfers of database or file system changes that have occurred since the last backup, either continuously or at some predefined interval. This is typically accomplished through the use of dual writes to dedicated server/storage systems. *Hot Standby Systems:* shadowing or mirroring of processes occurring from one server to a duplicate server that can be switched to in the event of a failure within the primary server.

Bottom Line: The integration of technology with business functions has placed increased pressure on the IS organization to provide high availability and continuous 24x7 access to business applications and data. Any downtime at all can result in significant, direct losses to revenue, profit, opportunity and productivity, as well as indirect losses such as misinformed decision making and degraded image and reputation. As use of the Internet and intranets increases, enterprises will not be able to survive without near 100 percent availability levels.