

The Availability Corner

What Reliability Do We Really Need?

January/February 2005

Dr. Bill Highleyman
Dr. Bruce Holenstein
Paul J. Holenstein

System reliability doesn't come for free. As we strive for higher reliability, we often have to add redundant hardware. In the NonStop world, this means additional processors, mirrored disks, and redundant ServerNet fabrics, for a start. For extreme reliability, we even buy a whole new backup system.

Performance can be compromised as well. Though we don't usually think about it, disk mirroring, checkpointing, server classes, and error-correcting communication protocols all impact performance. We can have high reliability, high performance, and low cost. Pick any two.

Therefore, it is important not to demand reliability to the point of overkill. Every application has its own reliability requirements often expressed as the amount of application downtime it can suffer and the amount of data loss it can sustain in the event of an outage.. Systems that protect life and property require extreme reliability. These include 911 systems, hospital patient monitoring systems, and nuclear facility control systems. Then there are those systems that have large financial penalties in the event of a failure, such as stock exchanges and many financial systems. These must always be up during business hours but often can come down in off hours (though the 24x7 world is quickly changing this). Failures of online stores and ATM systems can drive customers away. Batch systems can probably stand some downtime. In some cases, bad publicity can be more expensive than out-of-pocket costs (what if eBay or Google should go down?).

Each company knows its business and should know the cost of system failures (if they don't, they should quickly initiate the steps to understand system failure ramifications – a topic which we will address in a later column). They each make their own decisions as to the reliability that they require and structure their systems accordingly. Unfortunately, in many cases, up-front costs appear to overshadow costs of potential downtime; and systems are often under-configured for reliability.

The Standish Group maintains an impressive on-going study of the total cost of ownership (TCO) of a wide variety of systems. This study includes the cost of downtime. This study shows clearly that NonStop systems have a competitive advantage when the cost of downtime is considered. But the recent Worldwide Survey of HP Users (run by the HP User Group communities) showed that NonStop does not fair well when it comes to TCO. Either the message is not getting out, or more likely corporate managers are still looking at their quarterly bottom line rather than the long term consequences of failures.

Therefore, it is with great interest that we look at what users really feel they need in terms of reliability as opposed to what they really have. A recent survey sponsored by the Business Continuity SIG, run on the Online Advocacy web site, revealed the following among NonStop users:

- It would take more than four hours – and sometimes days - to resume main business functions following a critical outage for half the respondents.
- Half of the respondents felt if they were down for more than four hours, it would make the newspapers. 10% thought that this would happen if they were down for more than ten minutes.
- Downtime costs ranged from \$1,000 per hour to \$14 million per day.
- One-third of all respondents had experienced a critical service outage in the past ten years.
- A third of the respondents felt that their company was not well-prepared for a disaster.

Supplementing these results were several Instapolls also run on the Online Advocacy web site. They showed the following:

- Almost 50% of respondents could not lose any data as a result of a failure. Another 25% could lose no more than 30 seconds of data.
- 25% of respondents had to recover in less than 1 minute. Two thirds of the respondents had to recover in less than 30 minutes.
- 30% of respondents could be down no more than 5 minutes per year. 50% of respondents could be down no more than one hour per year (this is four 9s availability, which is the NonStop norm).
- The cost of downtime for 20% of respondents exceeded \$100,000 per hour. It was over \$1,000 per hour for half the respondents.

Reliability requirements vary widely. Achieving what you need is what this series is about. Join us next time as we discuss the various measurements of availability and reliability to understand where we are and where we want to go.”