

The Buffer Pool

On Recovery Service Levels

BY CRAIG S. MULLINS AND BRENDA HONEYCUTT

Special thanks to this issue's co-author and guest columnist, Brenda Honeycutt, of Software Engineering GmbH.

Business availability is more than just having a reliable hardware and database platform in place. Even the best high availability environment cannot safeguard itself from logical errors. Since most companies cannot afford downtime, it is important to plan ahead to ensure that critical enterprise data on which the company depends is always recoverable to ensure its availability.

Well-planned recovery procedures should be able to assure a complete recovery of enterprise-critical data within a predefined time window that provides for minimum disruption of the business. However, within complex environments, it is nearly impossible to perform recovery tests without disrupting the production system. Therefore, even the best-planned recovery scenarios can fail due to operational risks resulting from unforeseen and, therefore, immeasurable vulnerabilities.

With these issues in mind, we will examine the risk management of business availability, especially focusing on particularities of database health that we should look at differently in terms of service level agreements, as well as an approach to perform regular database audits that result in management level measurements of availability and recoverability. This article does not address data availability and recovery itself, rather the risk management process and the mechanized controls we can implement to assure high availability and recovery strategies that work when you need them.

WHY IS DATABASE HEALTH A MANAGEMENT ISSUE?

Because enterprises today rely so heavily on IT (and, indeed, database management systems) to conduct day-to-day business, it is not hyperbole to suggest that their very existence may be endangered by an IT failure. Therefore, from a business perspective, the sole responsibility for data loss or outages cannot be left totally to the IT department. The risk management team, which should also comprise high levels of management, needs to be involved. Under risk management, data availability and recovery are summarized with a number of other aspects like contingency planning and business continuity.

Enterprise data availability is so significant that a failure affects the whole or a major part of the business process, thus representing serious financial consequences. Added to that are possible

violations of laws or regulations. For example, Sarbanes-Oxley Section 404 requires that enterprises have enterprise-wide classification of data for security, risk, and business impact. To mitigate risk and assure ongoing business availability, the wise management team will require key performance indicators (KPIs) about data availability at all times and "at the touch of a button." Doing so provides the following advantages:

- Transparent and comprehensible data required to assess the risk situation;
- An early warning system that shows the need for action regarding data availability risk;
- An intelligent escalation routine when critical values are exceeded; and,
- Revision security by the continuous chronicling of the risk situation.

However, the complexity of today's IT systems makes this extremely challenging without a management tool to automate the process.

COMMON MISCONCEPTIONS ABOUT DATABASE HEALTH

There are numerous, lingering misconceptions about database health to which many organizations fall prey. These misconceptions occur because DBAs believe in, or apply concepts that are, objectively false. For example, many administrators will tell you that full recovery is possible and there are no problems because all backups are available and current. But, there isn't the slightest inkling of whether the recovery duration supports the business needs.

Another misguided notion is that recovering small objects, especially those with minimal updates, will always be fast. But this does not take into consideration situations where single updates to a page cover several archive logs. Recovering a huge object with high update rates, all of which are active, can be much faster.

Some organizations immediately migrate current backups to tape in order to free up disk space. Well, this approach will free up some DASD, but the time to recall that tape will need to be added to the recovery time. So what is more important: disk space or time-to-recover?

Others probably are shaking their heads thinking something like “virtual tape storage is as good as DASD.” Not necessarily; do not overestimate the speed of VTS. And parallel recovery using VTS is not possible if the archive logs are stored there too.

What about active logs? Some organizations mistakenly believe that single active logging is fine if mirroring is used. But what happens if you experience a technical error with the channel, the device or the device firmware? The data will be mirrored as inconsistently as the source. You should always do dual logging, without exception.

Perhaps the most significant misconception of all is equating a healthy database system with one that performs well with regard to SLAs. This view ignores a very critical aspect of database health, recovery health. For example, no one should consider the following DB2 system health:

1. Improperly set system parameters (DSNZPARMs);
2. Backups are missing, or are too old; and,
3. Recovery is possible, but the recovery duration won't support the business needs.

If you cannot recover your databases after a problem then it won't matter how fast you can access them, will it? Anybody can deliver fast access to the wrong information. Database health is a combination of keeping the information in databases accurate, secure and accessible. If any one of these three legs of the administrative stool is cut off, the stool – and, therefore, the database – topples over. And recoverability is a component of both accuracy and accessibility.

Just like for performance management, we need to understand the availability needs of our data in terms of the business. In the event of a failure, how rapidly must we be able to recover from that failure? Keep in mind that the failure could be either physical, such as a failed disk drive, or logical, such as applying the wrong input to a process which corrupts the database.

Only after we know the impact to the business can we develop an appropriate backup and recovery plan. We need service level agreements (SLAs) for recovery just like we have SLAs for performance. The recovery SLA needs to be from an application perspective, such as “Time to restore application availability after a failure for application X cannot exceed 2 hours (or 10 minutes or ...).”

To create effective SLAs you will need to be able to answer the question: “What is the cost of not having this data available?” When we know the expectations of the business we can work to create a backup and recovery plan that matches the requirements. There are multiple techniques and methods for backing up and recovering databases. Some techniques, while more costly, can enhance availability by recovering data more rapidly.

It is imperative that the DBA team creates an appropriate recovery strategy for each database object. This requires mapping database objects to applications so we can adopt the proper strategy in accordance with the application recovery SLA. Some database objects will participate in multiple applications, and their recovery strategy will therefore be more complex.

RISK MANAGEMENT FOR ENTERPRISE DATA

Many of the biggest and most critical applications in the world rely on DB2 for their persistent data requirements. As the main data storage for many large enterprises, DB2 data is a requirement for business continuity and assurance of minimal financial loss. Put another way, for many enterprises, a DB2 outage is equivalent to a business outage.

Given this large responsibility DB2 must bear, a well-thought-out recovery plan is essential. The risk management process for a DB2 recovery must be a systematic application of management policies and practices, combined with the specific IT tasks necessary to ensure that the chosen recovery procedures are ready to run, as fast as possible, and take full advantage of technology investments.

The risk analysis for such recovery procedures must include the systematic usage of available information about the pertinent databases. Such analysis must include the determination of specific key performance indicators (KPIs) for planning a recovery scenario, as well as reliable estimates on how long a recovery will actually require in order to measure the magnitude of the recovery consequences. Proactive measures demand that all conditions that play a role in the recovery must be known and evaluated in advance. For example, you will need to factor in such issues as missing backups, too infrequent backups, hardware limitations, and so on.

The risk evaluation must be able to systematically compare the results of the above recovery risk analysis with acceptable levels of recovery criteria. This evaluation should provide a transparent risk indicator to management, as well as detailed reports and optimization hints necessary for the IT staff to implement preemptive countermeasures.

In addition to recoverability, these processes must also take into account critical internal settings required to ensure that enterprise data availability is as high as possible at all times. Are you aware of the settings in your environment of recovery-related DSNZPARMs such as LOGAPSTG and CHKREQ? And are you aware of the impact on recoverability of improper settings?

WHAT IS THE PROBLEM?

The complete risk management process should be an ongoing, proactive, daily process, which only automation can provide. Many times, such comprehensive risk analysis processes are done as periodic audits provided by outside consultants, at very high costs. But such audits are merely a snapshot of a given moment. Changes to the environment can invalidate the results within hours. Whether done by external or internal staff, and regardless of the cost and frequency, the comprehensive collection of the relevant recovery information is a time consuming process, and some vulnerabilities remain obscured in every event, for example:

- Actual recovery times are always variable based on database size, CPU and I/O capacity;
- Backup and recovery strategies; and,
- Key performance indicators pertaining to each and every DB2 system.

All of the pertinent information, including DB2 parameters and internals, need to be collected, analyzed and interpreted into a transparent and comprehensible management report.

A PICTURE OF HEALTH

A useful DB2 recovery health check process should be able to automatically audit all components of availability and recoverability for all DB2 subsystems enterprise-wide. The process should be set up to present the results combined into a single overall availability indicator; in other words, a picture of the health of your DB2 environment in terms of its recoverability. Consider, for example, the management-level report shown in Figure 1. The chart shows,

at a glance, throughout the day where we exceed, meet, or fail against the recovery service levels we have set. The management team can readily look at this report and rapidly determine whether pre-emptive measures are required.

Producing such a report requires sophisticated processes to transform complex risk factors of the infrastructure into measurable, calculable and transparent KPIs. There are solutions on the market today that can provide such an automated early warning system for management, and technicians, to ensure the recoverability of your DB2 systems and databases. And in this day-and-age of data breaches, regulations, and round-the-clock availability requirements, it is unforgivable to have a prolonged outage due to preventable recoverability problems.

Are you sure your DB2 systems are healthy and recoverable?

ABOUT THE AUTHORS

Craig S. Mullins is a data management strategist at NEON Enterprise Software, Inc. He is the author of two books: *DB2 Developer's Guide* and *Database Administration: The Complete Guide to Practices & Procedures*. You can contact him via his Web site www.craigsmullins.com.

Brenda Honeycutt manages the technical publications department and does DB2 product consulting for Software Engineering GmbH, a DB2 solutions provider specializing in DB2 tools and consulting for over 20 years. You can contact her at b.honeycutt@seg.de.

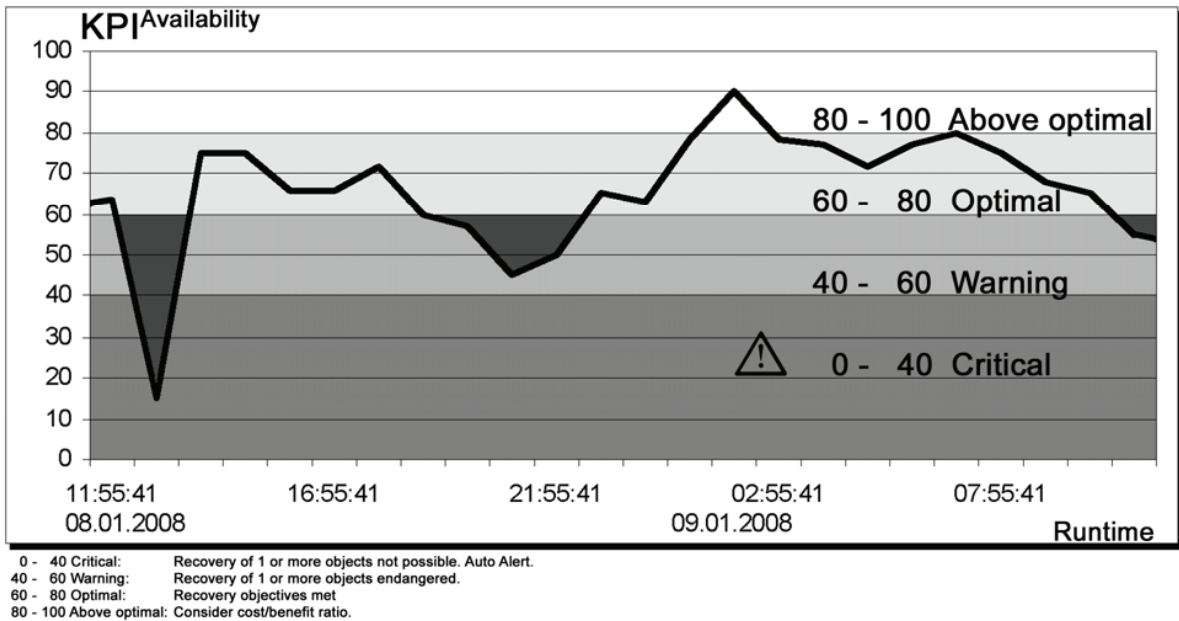


Figure 1. Management Level Data Availability Report.