DBMS High Availability: Issues and Strategies for Maximizing Reliability and Minimizing Downtime

Tom Clark
Informix Software, Inc.
tclark@informix.com

As information increasingly becomes the lifeblood of business, the databases that store business critical information become vital to business operation, with downtime for the DBMS creating critical productivity and opportunity impacts on the company's business. With the move to deploy both new and old applications on the web, this dependency on DBMS availability escalates as the increased access to information causes more databases to become mission critical. The heightened availability requirements have forced open systems databases to provide the types of HA once reserved for mainframe environments. The nature of databases complicates HA because it must be achieved in a way that maintains the consistency of the data in the database.

This presentation focuses on the issues involved in developing and deploying highly available databases, strategies and solutions for maximizing availability, and specific technologies that contribute to making databases highly available. We divide events affecting DBMS availability into five categories:

- **Software Failure**
  As hardware components become increasingly reliable, most system failures are caused by software. This implies that the ability to isolate and recover from software failures is a critical part of DBMS availability.
- **Data Corruption**
  Data corruption is a less common form of failure but can be disastrous. Data can be corrupted due to bugs in the DBMS software, an application gone awry, or a failure in the disk system. It is important to find corruption as soon as possible to prevent it from spreading to other systems, and recovery from data corruption must cause as little system downtime as possible. Single points of failure such as shared disk systems must be evaluated on the basis of the cost of downtime versus the cost of replicating data.
- **Hardware Failure**
  In spite of the increasing reliability of hardware components, hardware failure in power supplies, memories, etc., must be planned for and recovered from. Catastrophic hardware failures caused by events such as fires or natural disasters require a disaster recovery solution. The disaster recovery solution must be able to be brought online quickly, with little or no loss of inflight data.
- **System Administration**
  The broad areas of general system maintenance (indexing, backups, etc.), evolution of the system (schema changes, data loads, etc.), and software upgrades are a common cause of DBMS downtime. These operations must be required infrequently by the DBMS and cause as little system downtime as possible.
- **Performance**
  In an e-business application, success often means a significant increase in the number of users and the amount of data managed by the DBMS. A lack of scaling or other performance problems appears to users as a lack of availability. The DBMS system must be able to scale to the increased workload, provide tools for finding performance problems, and provide utilities to tune the system without requiring downtime. The DBMS architecture is a critical factor in the scalability and performance characteristics of the system.

Some of the strategies and technologies discussed for maximizing availability include fault isolation and failure independence, backup/restore, disks systems such as RAID, replication based availability techniques, online administration features, cluster managers, and DBMS architectures that lend themselves to increased availability.