Building a 24x7 Database

By Eyal Aronoff

QUEST SOFTWARE

Contents

QUEST SOFTWARE

Building a 24 X 7 Database

Eyal Aronoff, CTO of Oracle Database Products, Quest Software, Inc.

The Risk of Downtime

Over the last few years, corporations have invested billions of dollars to integrate the automations of core business systems into large Enterprise Resource Planning (ERP) applications. Increasing reliance on the availability of ERP environments and the advent of "around the world, around the clock" business transactions via e-commerce exposes organizations to a great risk. Losing access to the ERP system or the e-commerce application for an extended period of time may cause the entire business to collapse. This, however, is only one part of the picture. Most people in decision-making positions are totally dependent on the application to achieve their productivity goals. Successful enterprises reduce the risk of downtime while increasing the day-to-day application response time by combining multiple state-of-the-art technologies.

Your Definition of 24x7

When evaluating the need for high availability, you must first define what 24x7 means to your organization. Answer the following questions to determine your 24x7 needs:

1. Can you assign a dollar amount for application downtime?

Consider direct and indirect costs of downtime. Some applications may be internally or externally regulated with predefined penalties for periods of lack of service.

2. Do all application components/processes share the same level of importance?

Different parts of the application often bear different levels of mission-criticality. For your company, data entry may be available only between 8:00 a.m. and 5:00 p.m., or a particular batch job may need to be completed before 7:00 a.m. Such requirements may mean that although the criticality of the entire application does not warrant the substantial cost associated with higher availability, some investment should be made to minimize the potential of certain failures to occur during certain time frames. For example, with 2,400 data entry clerks working from 8:00 to 5:00, it is essential that the application does not experience more than two minutes of downtime during these hours. However, after 6:00 p.m., the criticality of this data entry application diminishes until the next morning.

3. Is the availability of the application more important than data consistency?

In financial applications, data integrity is paramount. Under no circumstances should a committed transaction be lost, even if it means more downtime for recovery. In many order entry systems, conversely, it is more important that the application remain available at all times even if a few orders are lost. This is especially true for e-commerce style applications. For many online shopping applications, it is essential to have the system accept new orders at all times. If the system goes down, another system must take over — even if it means that several transactions may be lost. The cost of downtime due to recovery is greater than the cost of lost transactions.

4. When can you do application and database upgrades?

Determine "quiet times" in which availability of the system is not critical. Is it long enough to accommodate an upgrade? How often is that quiet period of time — nightly? Every weekend? Monthly? On national holidays? A different solution may be required for systems that are maintained at night versus systems that do not have a reprieve. If your systems do not have quiet times, there should be some administrative time allocated.

Performance's Impact on Availability

Downtime is not the only concern when considering high availability— for many organizations, stringent processing needs force IT leaders to consider performance On one hand, online users require good response time from many short transactions. On the other hand, large batch jobs (e.g., reports and complex extracts) entail high throughput of a handful of very large transactions. These conflicting needs cause response time to fluctuate, decreasing the reliability and availability of the application. This is especially true with applications that provide services directly to end users and consumers, such as an e-commerce application.

Redundancy is the Key to Availability

The logical solution for increased availability is to maintain the data in more than one place. This enables high availability and one of the best techniques for improving application response time — separating batch reporting and extract processing from the OLTP processing.

The criteria for a comprehensive high availability and high performance solution include:

- Minimal impact on the availability and performance of the primary system

- Full copy of the primary database— primarily for reporting and extracts— that is accessible even when there is no “emergency
- The copy of the database should be an up-to-date image of the primary database
- Capacity to become the primary database (fail-over)in case of disaster
- Failover to the secondary database should be very fast, without data loss
- After the disaster, the solution should enable switching back to the primary system
- Ability to modify some aspects of the copy database to accommodate the different processing on it and the ability to reverse them when a failover occurs, e.g., construction of special indexes to support reporting needs
- The copy will not require its own database administration in addition to the administration of the primary system
- Redundancy in CPU as well as in the database
- Remote location of the secondary system

Range of Common Solutions

There is a wide range of solutions to the high availability problem. The most common methods are:

- **Local disk mirroring and/or RAID** – This solution provides protection against many disk-related failures, but the mirror is usually not breakable under normal circumstances. Once broken, the mirror becomes stale relative to the copy that is still operational. To resync (or re-silver), many disk mirror solutions perform a complete copy of the data from the operational copy to the stale copy. If the database is large, this process can take a very long time. Other disk mirroring techniques such as those provided by EMC and Veritas provide for a delta refresh, which is much faster. Local disk mirroring does not provide resolution for a local disaster. It also lacks protection against physical block corruption by Oracle or accidental loss of data due to a DBA error (such as dropping or truncating a production table).
- **Oracle standby database** – This solution provides some protection against catastrophe that makes the primary database unavailable, however, Oracle’s standby database has some shortcomings. The copy is only as current as the last log that was applied. Once the database is opened and modified, a complete image is required to reconstruct the standby database. Additionally, some administration is required for the standby database as the structure of the source database changes,

such as adding data files or auto-extending tablespaces. And, the standby database does not provide protection against certain types of database block corruption.

- **Local clustering** – Local clustering is a hardware solution that enables multiple computers to share a set of disks. Applications on these computers can freely migrate between the machines in the clusters using a technology known as "floating IP addresses." Unfortunately, the Oracle database relies on persistent memory structures, so when a switch happens, the database has to be brought down and restarted.

  This solution provides good protection against most common failures. However, since there is only one copy of the database, there should still be consideration for protection of the disks. Moreover, since there is only one copy of the database, any physical block corruption or accidental dropping of a database object will cause the application to fail. Finally, with a local cluster, there are no provisions for performance improvement by any load sharing.

- **Remote disk mirroring** –Two types of remote disk mirroring exist: synchronous and asynchronous. With asynchronous mirroring, the primary system does not wait for the data to be "committed" to the remote disk. For Oracle databases, however, asynchronous mirroring enables structural corruption in the mirrored database that would prevent a DBA from opening the remote database. For this reason, most remote mirroring implementations use the synchronous method, wherein the application waits for the data to be committed to both the local and the remote disk. To prevent slowing the primary system, however, this method requires a wide bandwidth between the source and destination. Most sites use the remote disk mirroring from EMC with one or more T3 lines.

- **Replication** – Replication provides a live remote database both to reduce the workload of the primary system and for fail-over when a disaster happens. However, Oracle replication is resource-intensive, and has a substantial impact to the primary system. Moreover, because of the many limitations imposed by Oracle replication, it cannot successfully replicate many of today's applications—particularly, large ERP sites.

  SharePlex® for Oracle from Quest Software overcomes these problems. SharePlex is a comprehensive and efficient solution that can support replication for most ERP sites. The live database on the remote site does require database administration, and application of patches to the application is not straightforward.

- **Local clustering with Oracle Parallel Server** – Oracle Parallel Server (OPS) offers another alternative for high availability systems. Using this facility, many instances of Oracle running on different hardware can access the same database on shared disks. This permits the hardware that would be allocated for a standby system to be actively used in production. Concurrent access to data by the different instances is managed by the Distributed Lock Manager (DLM), an application that

assures that an instance always accesses a consistent block from the database. The DLM causes the instance holding a dirty data block to flush it to disk, so the reading instance can access a clean copy. The difficulty in using OPS for highly available solutions is that the application needs to be designed so that transferring blocks between instances (pinging) is minimized. If not, application performance can be severely degraded. Also, with OPS, there is only one copy of the database that is not protected from disk failures, block corruption or human errors such as accidental table drops.

The Integrated High Availability and High Performance Solution

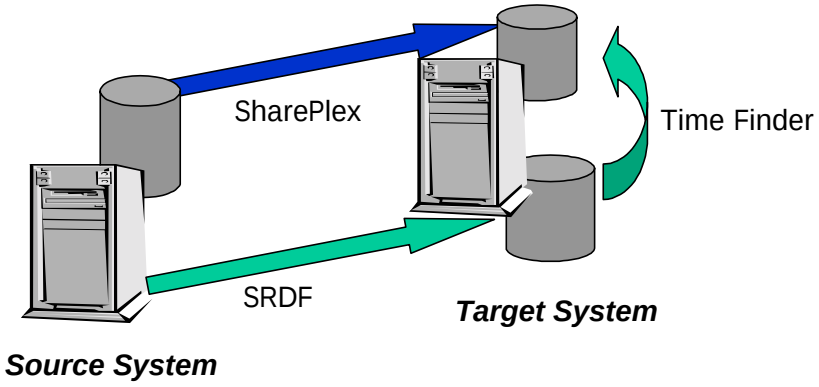
From the brief description of the high availability options shown above, it is clear that there is no one solution that can support all the requirements put forth above. Through a combination of hardware and software, EMC provides the disaster recovery components: **Symmetrix**™ disk drives on the production (source) and disaster recovery (target) systems with **SRDF™** running between these disks. A third mirror is created via EMC's **TimeFinder™** software, which is maintained as an accessible, up-to-date reporting instance by Quest Software's **SharePlex®**.

Quest Software's SharePlex minimizes the impact on the source system, the source instance and the network. This enables businesses to eliminate report processing overhead *without* replacing it with replication overhead. SharePlex includes:

SharePlex for Oracle – provides high-performance Oracle replication that maintains a continually updated, accessible target instance
SharePlex reconcile option – enables SharePlex to continue optimally after an EMC refresh

The Combined Solution

EMC's SRDF performs disk-level replication of the entire environment from the source system to the target system. In synchronous mode, no change is made to the source system that isn't also made to the target, guaranteeing that the target system's second mirror is an exact replica of the production server. If the production server fails, a secondary system is available with the same data, and the disk drives need only be mounted in order for business to continue despite the outage of the primary server. EMC's TimeFinder creates the Oracle database for the "third mirror," quickly generating the copy that SharePlex then maintains by replicating changes to selected tables and sequences from the source system to the third mirror.



In addition to greatly reducing the time required to create the initial image for the third mirror, TimeFinder refreshes the environment surrounding the Oracle tables and sequences periodically. TimeFinder facilitates the propagation of DDL changes, changes to stored procedures, and changes to applications, refreshing the image of the third mirror based on the image of the second mirror (which exactly reflects the primary system).

Benefits of the Proposed Solution

In the standard scenario in which the Reconcile Option is used, a customer has Symmetrix disk drives on the source and target systems, with EMC's SRDF between these disks. A third mirror is created via EMC's TimeFinder, which is maintained as an accessible, up-to-date reporting instance by Quest Software's SharePlex. This combination affords many benefits:

Easy initialization for 24x7 shops: SharePlex requires two matching copies of data with which to start replication. SRDF and TimeFinder, in conjunction with SharePlex, make this requirement easy to fulfill because SRDF and TimeFinder can create an initial replica instance. SharePlex's integration module for EMC will reconcile that instance with the information contained within SharePlex's queue files.

Reduced disaster recovery time: Since SharePlex maintains an available standby instance, if a disaster occurs on the source system, this standby instance can be used until the application can be restarted on the disaster recovery copy maintained by SRDF. This can reduce the application downtime significantly.

Provides more disaster recovery protection: SRDF, like every other mirroring solution, is prone to block corruption errors and to human errors such as accidentally dropping a production table. Since the replica maintained by SharePlex is a logical replica of the database, it provides a protection against both block corruption (known as Oracle Error 600) and accidental erroneous DDL.

Fast, easy migration without downtime: SharePlex can replicate between different versions of Oracle. When you need to upgrade an Oracle version, you can perform the upgrade on a secondary system. SharePlex keeps the data current on the upgraded database. Once the upgrade has been fully tested, you can use EMC's fast refresh capability to upgrade your production system to the new version. SharePlex minimizes downtime to the production system.

Flexible configurations with WAN support: SharePlex can replicate between instances on the same system, to instances on a local area network, or to remote instances through a wide area network. Additionally, through a cascading scenario, SharePlex can replicate from one system to several and from those onto further systems. With this type of configuration, a global company could replicate from New York directly to Boston, D.C., Atlanta, and London, and then have the London office replicate to Madrid, Paris, Rome, Brussels, and Berlin, limiting the traffic across the ocean while keeping all remote offices up to date.

About the Author

Eyal Aronoff, Chief Technology Officer of Oracle Database Products at Quest Software, Inc., is a recognized expert in Oracle database management. Eyal is a certified Oracle DBA who has performed benchmarking, tuning and capacity analysis for some of the world's largest Oracle databases for nearly 15 years. He is responsible for producing a multi-million dollar family of database tools that enhance Oracle database performance.

Eyal has written dozens of articles that address database management, application tuning, an better use of new Oracle features, and is published regularly in major Oracle magazines and newsletters. In addition, he is coauthor of *Advanced Oracle Tuning and Administration* published by Oracle Press, and *Advanced Oracle Tuning and Administration for Oracle8.* Eyal is a popular, sought-after presenter at Oracle user groups throughout the world.