# Case Study: Oracle E-Business Suite with Data Guard Across a Wide Area Network

*"Data Guard 10g is a complete, reliable, DR solution for the mission critical 7TB database that supports Oracle's global business operations. High transaction rates, 7,000 concurrent users, and 1,000 miles separating the production and disaster recovery site – not a problem for Data Guard."*

*Raji Mani, Senior Manager*
*Database Services*

**Oracle Corporate Profile**

- **50,000 Employees**
- **$12Billion (US$) Annual Revenue**
- **Global operations are run on the Oracle E-Business Suite and a Global Single Instance of the Oracle Database**
- **www.oracle.com**

**Disaster Recovery Solution**

- **Oracle E-Business Suite**
- **Oracle Database 10*g***
- **Oracle Data Guard**
- **Oracle Real Application Clusters**
- **7TB Oracle Database**
- **7,000 concurrent users at peak**
- **Remote Disaster Recovery – Primary and Standby locations are separated by over 1,000 miles**

## OVERVIEW

Oracle Corporation saves $1 billion in operating costs each year as a result of its successful transition to an e-business. While the philosophy behind the effort is simple - do more with less, the implementation challenges were significant and are discussed in: "How We Saved a Billion Dollars" [1], by Larry Ellison.

A key enabler of this transformation is Oracle's own use of the Oracle E-Business suite and its ability to support worldwide business operations on a Global Single Instance (GSI), of the Oracle Database. The Oracle E-Business Suite and a single Oracle database have replaced more than 75 separate implementations of Oracle Applications running on hundreds of Oracle databases worldwide. Less hardware, reduced infrastructure cost, and lower maintenance expenses, all benefits that flow directly to Oracle's bottom line.

In the paper referenced above, Ellison also describes a fundamental principle of database synergy: A consolidated database contains more information, and can answer more questions, than the sum of the information in the smaller databases before combination. Consolidation of all of the company's mission critical data into a single database, however, conflicts with a fundamental principal of managing risk through diversification. Previously, a serious system failure would affect a single department or country subsidiary. In the new GSI world, without adequate data protection mechanisms in place, a serious outage would impact Oracle's business operations worldwide.

GSI cannot succeed without a highly available computing architecture that protects business operations against disruptions caused by any kind of system or data failure. For this reason, GSI follows Oracle best practices for deploying a Maximum Availability Architecture (MAA) [2]. MAA is specifically designed to reduce the cost and complexity of implementing a highly available computing environment.

In particular, Oracle Data Guard [3] is the MAA component that provides the Disaster Recovery solution for Oracle databases. It enables GSI to achieve its Service Level Agreements by protecting against events that would otherwise cause significant down time. Data Guard is a built-in feature of Oracle Database

Enterprise Edition is a disaster recovery solution that is database aware and integrated with other Oracle High Availability (HA) features.

## GSI SYSTEM ENVIRONMENT

GSI is a 7TB Oracle database supporting Oracle E-Business Suite applications. GSI is by definition a 24x7 business critical system, accessed by every Oracle subsidiary worldwide. Each Oracle line of business: development, marketing, sales, supply chain, manufacturing, customer service, accounting, and human resources, all utilize the same global database.

GSI runs on a 4 node Oracle Real Application Cluster (RAC). RAC enables multiple active Oracle instances, each running on a separate node, to share a common database. This provides High Availability by protecting against node failure as well as the obvious advantages of being able to dynamically add or subtract computing resources to the cluster as workload requires.

**System Configuration**

**Database Server**

- **Primary Database: 4-node Oracle Real Application Cluster**
  - **Oracle Database 10.1.0.3**
  - **(4) SUN 25Ks**
  - **Solaris 9**
  - **Each server configured with 36 dual-core 1.2GHz CPUs**
  - **144GB Memory**
  - **10TB EMC SAN storage**
- **Standby Database: Similarly configured 4-node Oracle Real Application Cluster.**

**Middle Tier**

- **(54) Dell Linux Servers**
- **2 X 3.06GHz CPUs/node**
- **6GB memory**
- **RedHat AS 2.1**
- **NetApps shared storage**

**Network**

- **Dual OC12**
- **1,000 miles between primary and standby sites**
- **35ms RTT**
- **TCP send/receive window size = 170k**

## GSI'S DISASTER RECOVERY REQUIREMENTS

GSI requires a Disaster Recovery strategy that prevents potential causes of downtime that are beyond the scope of RAC to address. Such events can include physical database corruptions caused by component failures (e.g. a rogue host bus adapter corrupting multiple data files), the failure of a storage subsystem shared by all nodes in the RAC cluster, or even the loss of all RAC nodes due to unforeseen external events.

The term "Disaster Recovery" conjures up visions of earthquakes, fires and 100-year floods. While these events are something that every Data Center should plan for, it is very likely that most Data Centers will never experience them. In contrast, it is another class of more mundane events; software, systems, or human failures, that more often produce unacceptable downtime for mission critical applications. Such events represent a greater challenge than a natural disaster because they occur much more frequently, and can be more difficult to predict. Oracle Corporation faces the same disaster recovery challenges that any other Oracle user would have with a similar configuration:

a. The primary database is a very active OLTP system supporting over 7,000 concurrent users, as well as running up to 225 concurrent batch jobs during peak periods. The DR solution must be able to keep up with the workload, keep primary and standby databases in synch, and generate no additional overhead on the primary server.

b. Peak redo rates of 8.2MB/sec

c. The WAN extends over 1,000 miles with round-trip network latency (RTT) of 35ms

An additional consideration is that the DBA team is always looking for ways to increase their efficiency. The DR solution needs to be self-managing. DBA and System Admin cycles cannot afford to be consumed by the overhead of keeping the primary and standby databases synchronized. The DR system must be resilient enough to handle peaks and valleys in application workload, and to recover from network outages and other failures without requiring time consuming manual intervention.

Additional Disaster Recovery requirements include:

**Workload**

- **Peak redo generation of 8.2MB/sec (500 MB/min)**
- **Sustained redo generation of 2.5MB/sec**
- **7,000 Concurrent Users at Peak Processing**
- **225 concurrent jobs at peak (675,000 jobs in the last week of a quarter)**

- The DR environment must isolate the standby database from any failure that could occur at the primary production site.
- Physical data corruptions must not be propagated to the standby database.
- Mechanisms must be provided to either prevent logical data corruption from being propagated, or provide for rapid repair without requiring extended downtime due to time-consuming point in time restore and recovery operations.
- Recovery Point Objective (RPO): A maximum of 5 minutes worth of transactions can be lost at failover time.
- Recovery Time Objective (RTO): Ideally, the maximum database recovery time following the decision to failover will not exceed a Service Level Agreement (SLA) of 15 minutes. However, this SLA can be extended to 1 hour if the extra time provides additional protection against physical corruption AND logical data corruptions, AND human error.
- Switchovers, used to test the DR solution and accommodate planned maintenance, must have zero data loss and meet the RTO objective. Furthermore, it must be possible to test the standby database by opening it in read-only mode while it is in the standby role, without any disruption to the primary database and without requiring the standby database to be rebuilt.
- The DR solution must not generate additional overhead on the primary server.
- The primary production database must be able to be recovered to its original state from a backup taken from the standby database.
- Management tasks must be automated.

## GSI DISASTER RECOVERY SOLUTION

Oracle elected to configure Data Guard using Redo Apply in Maximum Performance mode in order to meet the requirements described above.

**Data Guard Configuration**

- **Redo Apply (physical standby)**
- **Maximum Performance Mode**
- **LGWR ASYNC transport services**
- **50MB ASYNC in-memory network buffer (Data Guard 10g Release 1)**
- **8 ARCH processes enabled**
- **8 Online redo log groups with 1GB redo logs**
- **Standby Redo Logs**

**Standby System Configuration:** The standby database at the DR site is a 4-node RAC of SUN 25Ks each with 36 CPUs. The cluster is configured into 2 domains. The standby database domain is allocated 8 CPUs/node. The development & test domain is allocated 28 CPUs/node. The standby domain has sufficient processing capacity for the standby database to maintain synchronization with the primary database and can support production immediately following a failover, though at a much reduced level of service. The GSI primary and standby configuration is provided in Figure 1

GRID computing is the new computing architecture that effectively pools large numbers of servers and storage into a flexible, on-demand computing resource for all enterprise computing needs. GRID principles are evident in how GSI computing resources are utilized at the standby location. In the event that the

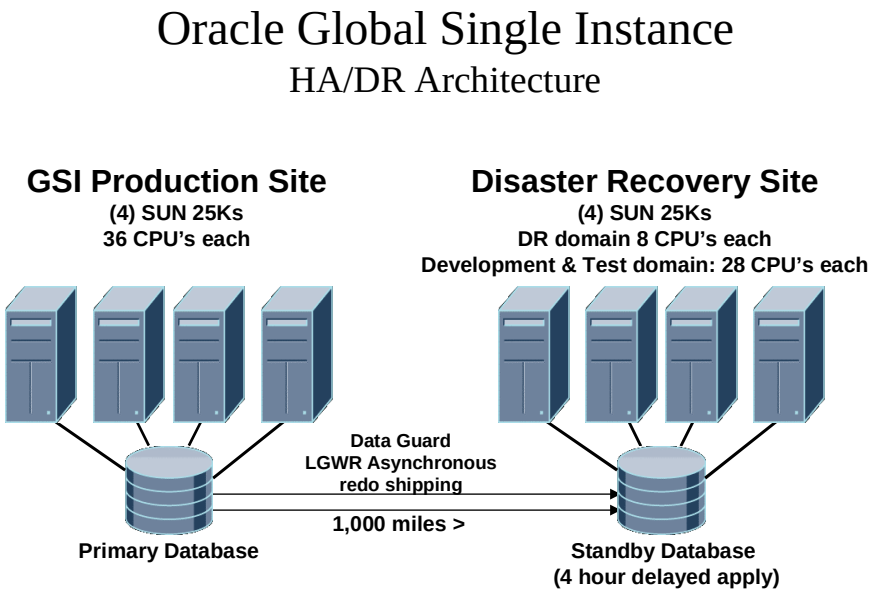


*Figure 1*

standby system needs to assume the role of primary database server, the systems admin staff reallocates CPUs from one domain to the next, until the capacity allocated to GSI production is back to the level that was present on the original primary. This process does not require any application downtime due to RAC's ability to dynamically add or remove nodes in a cluster without impacting the availability of other nodes. This strategy makes much more efficient use of computing resources, and significantly reduces the effective cost of maintaining the DR site.

**Data Guard Configuration Details:** Data Guard is configured using LGWR ASYNC redo transport service. In this asynchronous redo transport configuration, LGWR, while writing redo data to the online redo log, also places redo data in an in-memory network buffer on the primary server. A separate Data Guard process called the LGWR Network Server process (LNS) asynchronously reads from this network buffer and transmits the redo using Oracle Net Services and TCP/IP to the standby server. In asynchronous mode, LGWR does not wait for an acknowledgement that the redo has been received by the standby server.

The size of the network buffer used by Data Guard is configurable. The GSI configuration uses the maximum network buffer size for Data Guard 10g Release 1 of 50MB.

Asynchronous redo shipping achieves the important goal of no noticeable performance impact on the primary production system. It also limits potential data loss to the amount of data remaining in the network buffer on the primary server – an exposure that is well within GSI's recovery point objectives. The LGWR ASYNC architecture is provided in Figure 2, below.

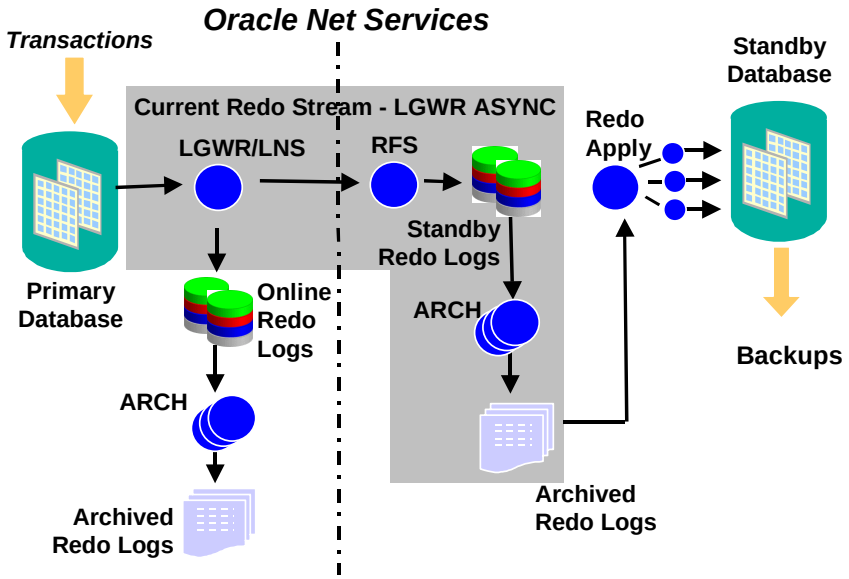


*Figure 2 – Oracle Data Guard LGWR ASYNC Redo Transport Architecture*

The parameters described above are all configured on the primary database by using Oracle Enterprise Manager, or by issuing the following initialization parameter:

```
LOG_ARCHIVE_DEST_2='SERVICE=[standbytns] LGWR
ASYNC=102400 NET_TIMEOUT=30'
```

Note: `[standbytns]` is the TNS entry pointing to the standby database and `ASYNC=102400` is the parameter that specifies the 50MB in-memory buffer in OS blocks.

**Standby Redo Logs:** A closer look at Figure 2 also shows that Oracle has followed best practices described more completely in the Maximum Availability Architecture (MAA) [2], by configuring Standby Redo Logs (SRLs) on the standby database. The RFS process on the standby database receives the redo data and writes it to a Standby Redo Log. Writing the redo data to disk as soon as it is received by the standby guarantees that data is protected should any event impact processing on either the primary or standby server. The application of redo data to the standby database is an asynchronous process relative to the highest priority task of insuring data is protected at the standby location. In this regard, the Standby Redo Logs are in effect a mirror of the primary database's online logs.

It is important to contrast Data Guard Redo Transport Services to traditional remote mirroring solutions used for disaster recovery. Data Guard only needs to transmit redo data to maintain a transactionally consistent copy of the primary database at the standby site. A remote mirroring solution must replicate every I/O to all database files, online logs, archive logs and the control file in order to maintain the standby database. This means that remote mirroring will send each database change at least three times to the remote site, increasing network I/O significantly.

The better network efficiency of Data Guard makes it possible to implement a DR site for high throughput applications across a Wide Area Network. The greater geographic separation between primary and standby sites provides a significant advantage of increased resiliency in the case of events that cause widespread outages – hurricanes, earthquakes, and power grid failures.

Finally, since Data Guard validates all redo before it is applied to the standby database it also makes it impossible for physical corruptions on the primary database to be propagated to the standby. Please reference Data Guard and Remote Mirroring Solutions [4] for more in-depth discussion.

**Completing the Apply Process:** At log switch time on the primary database an ARCH process archives the online log and another ARCH process on the standby server archives the standby redo log. The Data Guard Redo Apply process on the standby uses media recovery to apply the redo to the standby database, maintaining a physical replica of the primary database. A new Data Guard 10*g* feature discussed later in this paper, *Real Time Apply* , will shortly be implemented by GSI to eliminate the dependency between the primary log switch and standby apply. With Real Time Apply, redo data is applied to the standby database as quickly as it is received.

**Accommodating Network or Standby Outages:** There will be times when a the standby database is down or the network drops, and it is impossible to ship data from the primary to the standby location. Data Guard recognizes this condition, and rather than impact the LGWR's ability to continue processing on the primary server, it intentionally suspends shipping redo to the standby site. This

will causes the primary database to temporarily get ahead of the standby database but this is a situation Data Guard will resolve on its own, (as described in the following section).

**Automatic Resynchronization:** Data Guard proactively detects and resolves occurrences where the standby has not received all of the redo generated by the primary database. The resynchronization process is described in Figure 3. An ARCH process on the primary server continually pings the standby server to make sure it has received all of the redo generated by the primary database. If it discovers that the standby server has missing or incomplete logs, an ARCH

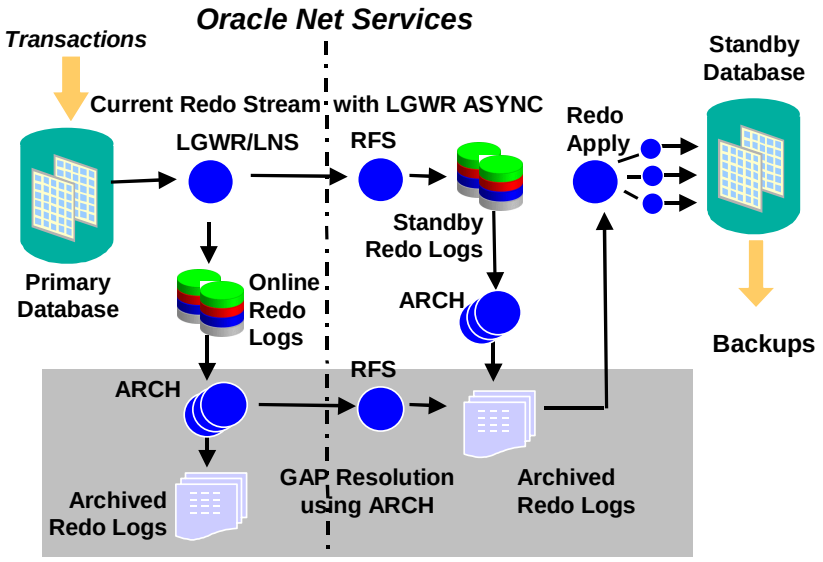


*Figure 3 – Automatic Resynchronization of Primary & Standby Databases*

process automatically resynchronizes the standby database by sending the required archive logs to the standby server. Meanwhile, at every log switch, Data Guard will attempt to reestablish a LGWR ASYNC connection with the standby server. When it succeeds, LGWR ASYNC resumes transmitting the current redo stream. This stops the primary from getting any further ahead of the standby databases, while ARCH completes the process of bringing the standby server completely up to date.

**Multiple ARCH Processes:** The high redo generation rate of the primary GSI database means that any interruption in network transmission can quickly result in the primary database getting ahead of the standby. Because the exposure to data loss is greatest during resynchronization, it is always desirable to complete the resynchronization process as quickly as possible. In Data Guard 10g Release 1, an

ARCH process ships one archive log at a time. Depending on the length of the network outage, it is possible for the primary database to get some number of log files ahead of the standby. Fortunately, up to 10 ARCH processes can be enabled in a Data Guard 10g Release 1 configuration.

One ARCH processes is used for local archiving, other ARCH processes are used to resynchronize the primary and standby database, shipping as many as 9 archive logs in parallel. In the GSI configuration, it was determined that enabling 8 ARCH processes is sufficient for this purpose. This parameter is:

```
LOG_ARCHIVE_MAX_PROCESSES=8
```

Note that in Data Guard 10g Release 2, the maximum number of ARCH processes that are configurable has been increased to 30. In addition, up to 5 ARCH processes can transmit a single archive log in parallel. These enhancements further accelerate gap resolution regardless if a single archive log has created a gap, or if a gap is the result of many archive logs generated during a prolonged standby outage.

**Data Guard and Wide Area Networks:** The flexibility of LGWR ASYNC and its ability to automatically resynchronize primary and standby databases is essential given two realities of wide area networks:

1. There will be times when the network connection between primary and standby site is lost
2. There will be periods of volatility in network bandwidth available to Data Guard combined with peaks in workload that can exceed available network bandwidth even under the best of circumstances.

Data Guard's flexible configuration options enables users to "dial in" the level of data protection that workload, systems, and network infrastructure will allow.

**Network Tuning:** Significant gain in network throughput can be achieved by tuning network parameters at the operating system and TCP/IP level. The Oracle system admin staff followed guidelines in the paper "Oracle Data Guard: Primary Site and Network Configuration Best Practices" [5]. An example of a parameter discussed in the paper is the TCP send/receive window size, set on both primary and standby systems. Oracle best practice guidelines provide understanding and a formula for calculating the optimum TCP send/receive window. The GSI setting is 170k.

**Protection Against Logical Corruptions:** A significant concern of the DBA staff is protecting the GSI database from logical corruptions and human error. Examples of this include running a batch job twice, or truncating a table, and then discovering that the wrong table was truncated by mistake. An Oracle Data Guard feature, "Delayed Apply" protects against such errors. The Redo Apply process can be delayed, allowing the DBA staff the time to discover logical corruptions before they can be applied to the standby database. Data Guard continues to ship the redo as it is generated, (so it is protected from primary server failure) but the apply is delayed on the standby database until the period of time specified in the parameter has passed. In GSI's case, there is a 4-hour delay specified using the following attribute of the `LOG_ARCHIVE_DEST` parameter:

```
LOG_ARCHIVE_DEST_2='SERVICE=[standbytns] LGWR
ASYNC=102400 NET_TIMEOUT=30 DELAY=240'
```

Note that the GSI instance is active enough that the 4-hour delay results in a backlog of redo on the standby server that will take longer to apply than the ideal Service Level Agreement (SLA) for Recovery Time (RTO) of less than 15-minutes. However, the delay makes possible an increased level of protection against logical data corruptions and human error, while still making it possible to complete recovery within GSI's Maximum RTO SLA of 1-hour.

The use of delayed apply has yielded dividends for Oracle IT. GSI experienced logical corruption caused by human error when a 160,000-row table was updated by mistake. Fortunately, the delayed apply on the GSI standby database made it possible to quickly resolve the problem as follows:

- Cancel recovery on standby and open read only.
- Stop the affected application on primary
- Export the required data from standby
- Recreate the table on primary; import data into the primary db after disabling triggers
- Restart the application on primary
- Restart recovery on the standby

The repair effort took less than 30 minutes compared to the estimated 10 hours or more that would have been needed with the conventional approach of restore from the previous nights backup, roll forward, etc.

In the future GSI will deploy another Oracle Database 10*g* feature discussed later in this paper, Flashback Technologies (Flasback Table and Flashback Database), [6] to provide the equivalent protection of the delayed apply. Using Flashback Database with Data Guard in place of a delayed apply enables administrators to quickly resolve user errors or logical corruptions by "rewinding" the database to a point in time before the problem occurred. Corruptions of more limited scope affecting a single table can be recovered even more quickly using Flashback Table. Flashback technologies eliminate the current compromise between data protection and recovery time since the standby database will always be up-to-date with the primary.

# IMPLEMENTING SITE FAILOVER

Figure 4 illustrates the various components used to implement a complete site failover. Oracle Data Guard is used to maintain the standby database and to perform database failover to the standby site. Unix rsync is used to synchronize oracle_homes between primary and standby sites. Network Appliance SnapMirror is used to synchronize file system resident application code. DNS push is used for network reconfiguration, redirecting users to the remote location (the new primary site) at failover time.

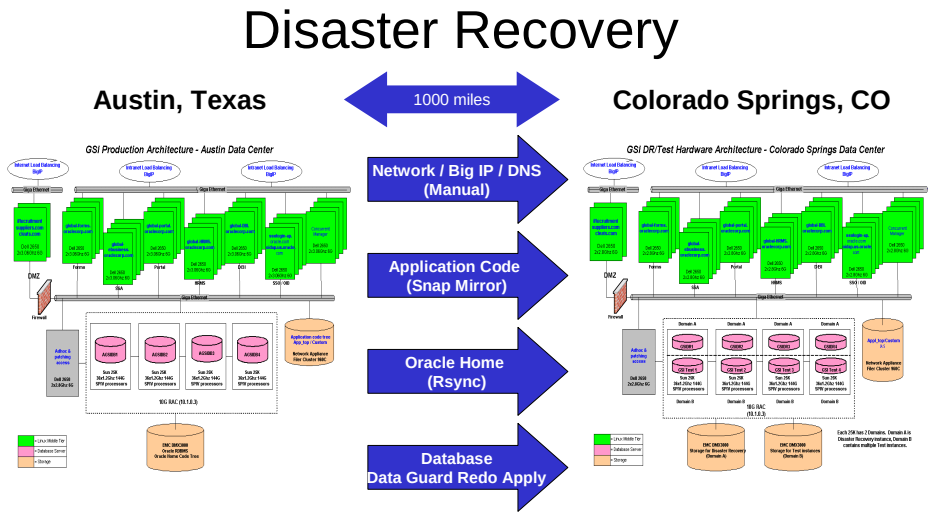


*Figure 4: Site Failover*

# ORACLE DATABASE 10g ENHANCEMENTS

At the time this paper was written, GSI was in the process of implementing the following new features of Data Guard 10g

**Real Time Apply:** Real Time Apply is the Data Guard 10*g* that eliminates the dependency on a log switch to initiate the process of applying redo data to the standby database. This enhancement, combined with the protection from user error and logical corruption provided by Flashback Database (described below), enables Data Guard to achieve an ideal result. Data Guard ships redo as fast as it is generated by the primary database and it applies redo to the standby database as soon as it is received.

**Flashback Technologies:**

Oracle Database 10*g* Flashback Table and Flashback Database will be used to protect against downtime due to logical corruptions and user error, eliminating the need for a 4-hour delayed apply. Eliminating the delay also eliminates the backlog of redo that needs to be applied at failover time and makes it possible for GSI to achieve its ideal recovery time SLA of less than 15 minutes.

**Flashback Table:** Provides the DBA the ability to recover a table or a set of tables to a specified point in time quickly, easily, and online. Flashback Table restores the tables while automatically maintaining its associated attributes such as - the current indexes, triggers and constraints, not requiring the DBA to find and restore application specific properties. Flashback Table alleviates the need to perform more complicated point in time recovery operations.

**Flashback Database:** Is a new strategy for doing point in time recovery for the entire database. It quickly rewinds an Oracle Database to a previous time to correct any problems caused by logical data corruption or user error. It is extremely fast, and easy to use, reducing recovery time from hours to minutes; no more need to restore from tape, no lengthy downtime, and no complicated recovery procedures.

A second advantage of Flashback Database is the ability to reinstate the original primary database after a failover, and bring it back into a Data Guard configuration without requiring a time consuming restore from a backup. As long as the original data files are intact, Flashback Database can be used to flash the original primary back to a point in time that preceded the failover. Data Guard can then automatically resynchronize it and makes it a standby to the new primary. Once synchronization is complete, a switchover can be used to reverse roles and restore the original primary to its role as the production database.

More information on Oracle Flashback Technologies is provided in the white paper: Flashback Technology, Recovering From Human Errors [6].

**LGWR ASYNC Redo Transport Enhancements:** As described above, the previous LGWR ASYNC architecture used an in-memory network buffer on the primary server. This buffer is configurable, but its maximum size in Data Guard 10*g* Release 1 is 50MB. In high workload environments on WANs with limited bandwidth and/or high latency this buffer may become full during peaks in workload. When this occurs, Data Guard's asynchronous shipping process cannot continue, and the Data Guard configuration automatically reverts to using the ARCH process rather than impact the performance or availability of the primary server. Data Guard automatically re-synchronizes the standby database with the primary, and will automatically re-establish a LGWR ASYNC connection at the next log switch. But during the period of resynchronization there is more exposure to data loss than if LGWR ASYNC had been able to continue uninterrupted.

The new LGWR ASYNC architecture in Data Guard 10g Release 2 eliminates the use of an in-memory buffer and significantly enhances the ability to handle peaks in redo shipping in WAN environments. LGWR performs its normal job of writing redo to the online redo log on the primary. As in previous releases, a completely separate Data Guard process (LNS) ships the data to the standby site, but instead of reading from an in-memory buffer of limited size, it now reads from the primary database online redo log as it transmits redo data asynchronously to the standby server. The new LGWR ASYNC architecture makes it possible to sustain asynchronous redo shipping through peaks in activity that would have filled the previous in-memory buffer. GSI will benefit from fewer interruptions in asynchronous redo shipping, thereby reducing exposure to data loss.

## CONCLUSION:

*"Data Guard Redo Apply is truly simple to manage. Once it is set up, there are only a few key areas to monitor. Data Guard automates the management of most tasks required to keep the primary and standby databases in sync.*

*Raji Mani*
*Senior Manager,*
*Database Services*
*Oracle Corporation*

Consolidation of applications and data into a single Oracle Global Instance creates tremendous competitive advantage by enabling better and faster decision making, while simultaneously reducing operating expenses.

However, without adequate data protection success may quickly turn into a business disaster if unforeseen events make GSI unavailable. All competitive advantage and cost savings rapidly evaporate should there be an extended outage. Data Guard is the fundamental building bock in the Maximum Availability Architecture that prevents this from happening. Data Guard:

**Protects Against Data Loss and Downtime:** An exact replica of the Global Single Instance is maintained at a remote location. The remote standby database can be transitioned to the primary role within GSI's Service Level Requirements.

**Reduces Cost:** Data Guard can be run on any server licensed for Oracle Enterprise Edition – there is no additional license or support cost. The same database licenses and servers are utilized by other applications and databases while in standby role, maximizing ROI for DR resources.

**Keeps it Simple:** Data Guard automates the process of synchronizing the primary and standby databases. It isolates the standby database from events that can make the primary production system unusable. It comes with a powerful, integrated, graphical user interface in Oracle Enterprise Manager that makes it easy to configure and manage.

**Proves it Works:** Data Guard can easily transition database roles from primary to standby, and back again. This makes it possible to test the reliability of the disaster recovery solution. It lets IT staff be 100% certain that the standby database is ready for to assume the production role.

# REFERENCES


1. How We Saved A Billion Dollars, by Larry Ellison – http://www.oracle.com/corporate/LECorpSingPgMag10_17.pdf
2. Oracle Maximum Availability Architecture - http://otn.oracle.com/deploy/availability/htdocs/maa.htm
3. Oracle Data Guard Overview - http://otn.oracle.com/deploy/availability/htdocs/DataGuardOverview.html
4. Data Guard and Remote Mirroring Solutions http://www.oracle.com/technology/deploy/availability/htdocs/DataGuardRemoteMirroring.html
5. Primary Site and Network Configuration Best Practices - http://www.oracle.com/technology/deploy/availability/pdf/MAA_DG_NetBestPrac.pdf
6. Flashback Technology, Recovering From Human Errors – http://www.oracle.com/technology/deploy/availability/pdf/TWP_HA_FlashbackOverview_10g_111503.pdf
7. Oracle Data Guard 10*g* Media Recovery Best Practices - http://www.oracle.com/technology/deploy/availability/pdf/MAA_WP_10gRecoveryBestPractices.pdf

**Oracle Global Single Instance - DATA GUARD PROFILE**
**November 2005**
**Authors: Joe Meeks, Ashish Ray, Lawrence To, Larry Carpenter**
**Contributing Authors: Raji Mani, Renzo Zagni, Jane Shen, Hemanta Parija, Tarun Mittal, Sean Mckeon, Venkatesh Bageppally**

**Oracle Corporation**
**World Headquarters**
**500 Oracle Parkway**
**Redwood Shores, CA 94065**
**U.S.A.**

**Worldwide Inquiries:**
**Phone: +1.650.506.7000**
**Fax: +1.650.506.7200**
**www.oracle.com**