THE RIGHT CHOICE FOR DISASTER RECOVERY: DATA GUARD, STRETCH CLUSTERS OR REMOTE MIRRORING

Ashish Ray, Oracle Corporation

INTRODUCTION

In today's market conditions, business continuity and disaster recovery (DR) are a top priority for senior management of most global enterprises. A global enterprise has to deal with economic fluctuations, rapid changes in market trends and competitive pressures on a 24x7 basis, and must be able to swiftly and efficiently deal with unforeseen business interruptions.

Various solutions are available today to protect business-critical data, and enable enterprises to quickly restore their business operations in the event of outages or disasters. This paper discusses three such technologies – *Oracle Data Guard* in a Maximum Availability Architecture (MAA) configuration (i.e. Data Guard combined with Oracle Real Application Clusters (RAC)), *Stretch Clusters* based on RAC, and *Remote Mirroring*. It describes their capabilities in terms of data protection, data availability and data recovery, and makes best practice recommendations regarding their applicability to various business situations.

IMPACT OF DISASTERS

In order to assess any DR solution, what must be determined is how effective this solution is, in protecting the enterprise from the common outages that affect its IT infrastructure. This is especially important since an enterprise today operates in an extremely complex and a highly networked, global economy, and is more susceptible to interruptions than in the past. With tightly-knit supply chains and just-in-time inventory models, an outage that causes business downtime adversely impacts not only the particular enterprise where it occurred, but also its suppliers, partners, customers, and in some cases the global market. No wonder for certain industries the cost of business downtime can be as much as millions of dollars in an hour.

CAUSES OF DOWNTIME

What are the causes of such business downtime? While significant disasters such as hurricanes or earthquakes capture news headlines, majority of business downtime is caused by much more mundane activities such as server failures, network glitches, human errors, system maintenance, etc. For any organization evaluating a DR solution, it is important to understand the nature of downtime occurrences that are typical in its IT infrastructure, so that it can implement a solution, which, among other capabilities, offers the best protection from the factors that cause this downtime.

Downtime that affect a business can be categorized into:

(I) Unplanned downtime
(II) Planned downtime

Unplanned downtime is caused by unexpected failures or outages of one or more components of the IT system. Examples of unplanned outages are:

(A) Hardware failure
- Server hardware (e.g. CPU, memory, fan, NIC)
- Network components (e.g. cable, router, switch, hub)
- Storage components (e.g. HBA, RAID, disk-array)

(B) Software failure
- OS, firmware, device drivers
- Database/Middleware/Applications

(C) Human errors (unintentional/malicious)
- Logical data corruption (e.g. wrong batch job)
- Physical data corruption (e.g. installing unsupported components)

(D) Site disasters
- Natural (e.g. earthquake, hurricane, flood)
- Unnatural (e.g. power outage, fire, terrorism, accidents)

It may be noted that some of these factors are inter-related. For example, there may be physical data corruptions because of failure in storage components; software components may fail because of human errors, etc.

Planned downtime may occur when one or more components of the IT system undergo scheduled maintenance. Examples of planned outages are:

(A) Hardware upgrade
- Server/Network/Storage

(B) Software upgrade/patches
- OS, firmware, device drivers
- Database/Middleware/Applications

(C) Other maintenance activities (scheduled/on-demand)
- Backup/restore
- Logical data maintenance (e.g. schema change, data transformation)
- General data center/site maintenance

Enterprises should do a categorization of their downtime profile as well as an analysis of their business processes to better understand the critical factors that may cause, or continue to cause significant interruptions to their businesses[1].

In the next section, an overview of Remote Mirroring technologies, Stretch Clusters based on RAC, and Data Guard in an MAA configuration is provided. Following this, a general framework to evaluate DR solutions is discussed, and thereupon these solutions are comparatively assessed using this framework. In certain cases, best practice recommendations are provided for configuring them in the most optimal manner to meet various business continuity requirements.

[1] In the DR industry, such an analysis is referred to as Business Impact Analysis (BIA).

OVERVIEW – REMOTE MIRRORING SOLUTIONS

There are primarily two types of remote mirroring solutions that differ based on where the mirroring-related processing takes place:

- Host-based
- Hardware-based.

HOST-BASED MIRRORING

There are two kinds of *Host-based Remote Mirroring* solutions, which are: (a) Volume Mirroring, and (b) Host-based Replication. For both of these categories, the mirroring-related processing occurs in the database servers, also known as the primary servers.

For *Volume Mirroring*, every time there is a write activity in the primary server, the Logical Volume Manager (LVM) at the time of writing to the local volumes, also mirrors the writes, synchronously, to a set of remote volumes. The LVM makes the local and remote disks appear as single set of disks independent of location. Because of the synchronous nature of the writes, and protocol limitations of the underlying Fibre Channel transport, such remote volumes can only be located across small distances, typically up to a few kilometers.

It may be noted that starting with Oracle Database 10*g*, the Automatic Storage Management (ASM) feature provides the capability to do host-based mirroring of all block-level Oracle files such as datafiles, log files and controlfile. The mirroring is done at the extent level, allowing the mixing of primary and mirrored extents in each disk.

For *Host-based Replication*, specialized file system drivers or volume manager components in the primary server intercept local writes, package them in logical messages, and synchronously or asynchronously send them over IP to remote (or secondary) hosts. Such solutions need to maintain specialized logs to keep track of write-ordering. The data volumes on the secondary server cannot be used (even for read-only access) while replication is in progress. To use the data on the secondary host while the primary is active, a volume-level snapshot may be used at the secondary host. Since IP is the underlying transport, and asynchronous replication is supported, such configurations may be deployed over wide distances.

In host-based mirroring, the primary and secondary storage devices do not need to have the same physical characteristics, nor do they have to be provided by the same vendor.

The following diagram gives an example of a host-based mirroring solution.

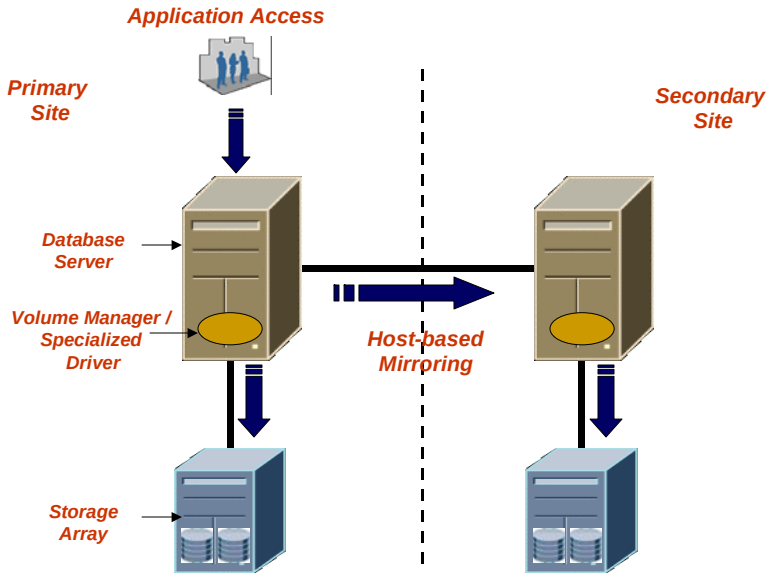

Fig. 1: Host-based Remote Mirroring

HARDWARE-BASED REMOTE MIRRORING

In *Hardware-based Remote Mirroring,* also known as *Storage Array-based Mirroring*, storage array controllers at the primary site mirror changed disk I/O blocks to a similar storage array at the secondary site. These changes are sent using protocols such as ESCON, FICON and Fibre Channel, although in some recent versions iSCSI and IP-based transport are also supported. The mirroring over the appropriate communication links is controlled by specialized link adapters loaded with appropriate microcode. As I/Os occur at the primary server, data is written to the cache of the source array, and placed in a queue. The link adapter takes the first entry of the queue and moves it across the link to the mirrored array.

The processing related to this mirroring is limited to the storage array. Both synchronous and asynchronous writes are supported. However, since Fibre Channel can be extended only over limited distances (approximately a few kms), specialized channel extenders have to be deployed if the target site is located beyond this distance. Storage Array-based mirroring typically requires homogeneous storage subsystems on both sites and a dedicated network between the sites. Similar to Host-based Remote Mirroring, the target storage arrays cannot be accessed by any application while the mirroring is going on.

Hardware-based Remote Mirroring is a more popular and established solution compared to Host-based Remote Mirroring, and in the rest of this paper, unless specifically called out, if the term "Remote Mirroring" is used, it applies to Hardware-based Remote Mirroring.

The following diagram shows example of a *Hardware-based Remote Mirroring* configuration.

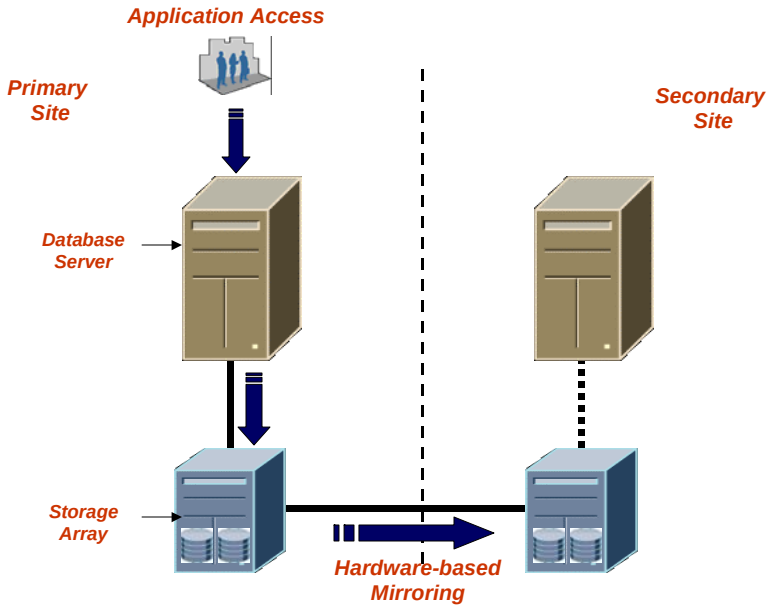

Fig. 2: Hardware-based Remote Mirroring

It may be noted that recently another kind of remote mirroring technology has surfaced – known as *SAN Switch-based Remote Mirroring,* which is similar to Storage Array-based Mirroring. This is achieved by embedding specialized application microcode in the SAN switch to enable replication to remote switches. Since this technology is yet to gain mainstream adoption, it will not be discussed in this paper.

OVERVIEW – STRETCH CLUSTER

In a conventional RAC implementation [1], multiple RAC nodes are located in the same data center and access a shared data storage, offering excellent high availability and scalability benefits to applications, without requiring any application-level code change. While this configuration offers good protection against events such as node failures, it does not protect against events that could destroy or damage the data center. Stretch Cluster configurations[2] solve this problem by allowing these nodes to be located at different sites, which can be located tens of kilometers apart, while still maintaining a unified cluster configuration over a single database. The storage subsystems, or the disk arrays, are also duplicated across the sites, and kept in synch by mirroring facilities provided by the underlying cluster-aware volume manager. The advantage of such a Stretch Cluster configuration is that it allows all nodes in this configuration to be used to serve production application workload, while offering protection from localized disasters such as data center fires.

It may be noted that if instead of volume mirroring, hardware-based mirroring is used in a Stretch Cluster configuration, the servers at the DR site cannot access their local volumes at the DR site (since mirrored volumes cannot be accessed while the hardware-based mirroring is in progress), and hence must access the source volumes across the network. This produces additional network and disk latencies for applications accessing the DR servers, and hence hardware-based mirroring is not recommended in Stretch Cluster configurations.

[2] Some terms that are also used to refer to Stretch Clusters are *Extended Clusters*, *Distance Clusters*, and *Geo-Clusters*.

The two major types of network traffic between the sites in a RAC-based Stretch Cluster configuration are the *Cache Fusion* [2] traffic and volume manager mirroring traffic for data I/O. For optimal cluster operation and performance, a Stretch Cluster configuration requires reliable, redundant, low-latency network between the two sites (typically Gigabit Ethernet for the Cache Fusion traffic and DWDM/Dark Fiber for volume mirroring). The following diagram shows a typical 4-node RAC Stretch Cluster configuration.

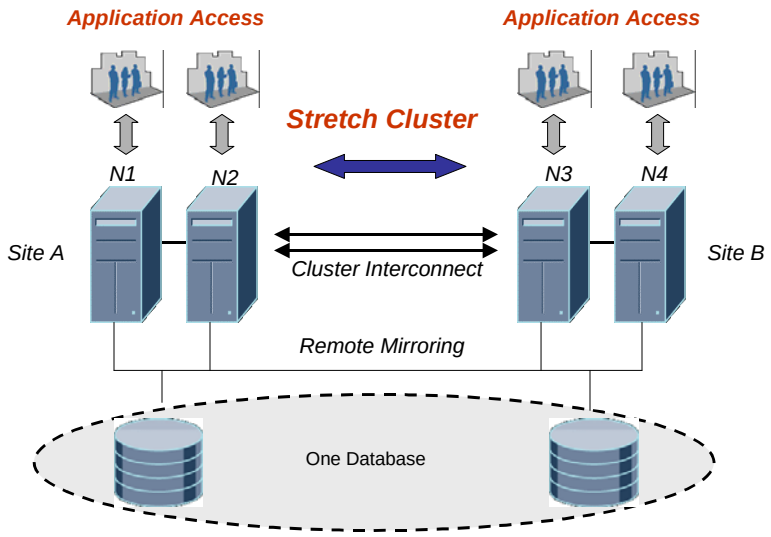

Fig. 3: A 4-node RAC Stretch Cluster Configuration

OVERVIEW – ORACLE DATA GUARD

Oracle Data Guard [3] is a built-in feature of the Oracle Database Enterprise Edition. It is a software solution that creates, maintains, and manages one or more standby databases to protect enterprise data from failures, disasters, errors, and corruptions. It maintains the consistency between the primary and standby databases by synchronously or asynchronously transmitting transactional redo data from the primary to the standby databases. These standby databases can be located on the same local area network (LAN) as that of the primary database, or they may be located several thousands of miles away, connected over a wide area network (WAN). If the production database becomes unavailable because of a planned or an unplanned outage, Data Guard can switch any standby database to the production role, thus minimizing the downtime associated with the outage, and enabling zero data loss.

The following diagram shows an example of a Data Guard configuration.

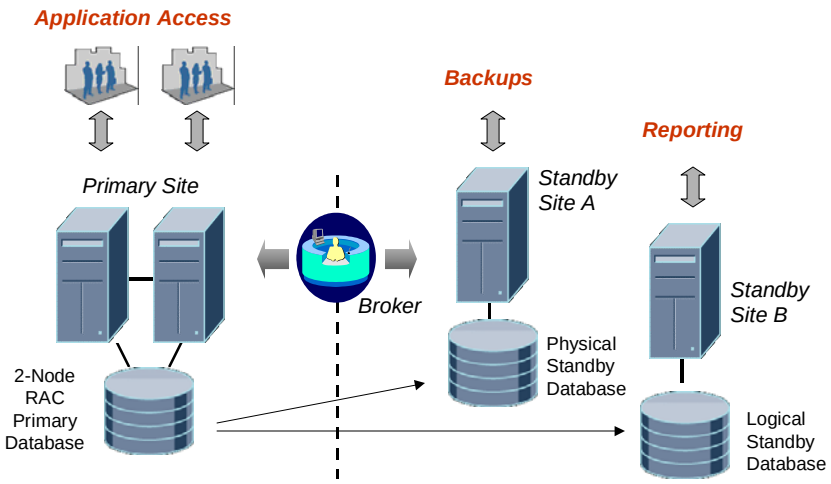

Fig. 4: Oracle Data Guard Configuration with a 2-node RAC primary database

In a Data Guard configuration, a primary database can support up to nine standby databases, which can be either a physical standby database, or a logical standby database. A physical standby database is kept synchronized with the primary database by applying redo data on a block-for-block basis, while the logical standby database is kept synchronized using SQL statements corresponding to the redo data. A physical standby database may be used to offload backup activities from the primary database to it, while a logical standby database can be used for reporting; however, production application workload can only be served by the primary database.

Data Guard and RAC are integrated with each other, and the primary or the physical/logical standby databases may be RAC databases. Such a combined RAC and Data Guard configuration offers an excellent HA and DR architecture, and is the foundation of *Oracle Maximum Availability Architecture* (MAA), which is Oracle's HA best practices blueprint [4].

Data Guard offers a distributed management framework, called the Data Guard Broker, which automates the creation, maintenance and monitoring of Data Guard configurations through Oracle Enterprise Manager. Also, since Data Guard is based on transmitting and applying Oracle redo blocks, the underlying storage arrays in a Data Guard configuration can be different.

Data Guard being a built-in feature of the Oracle database, it protects data that is resident in the Oracle database. Filesystem data, or data that is resident in non-Oracle databases, is not protected by Data Guard.

FRAMEWORK TO EVALUATE DR SOLUTIONS

In the remaining sections of this paper, Remote Mirroring, Stretch Cluster and Data Guard will be comparatively assessed on their effectiveness as disaster recovery solutions. To do that, what is needed is a set of criteria to evaluate disaster recovery solutions, such that the solution chosen by the enterprise is the best one available to maintain business continuity and protect business critical data in the event of one or more of the planned/unplanned outages listed previously.

Oracle's High Availability (HA) team recommends that any evaluation of a DR solution be made using the following criteria:

1. **Data protection** – This indicates to what extent business critical data is protected by the solution, in the event of planned/unplanned outages. A related metric is known as the Recovery Point Objective (RPO), which measures the maximum amount of data an enterprise may afford to lose in the event of an outage, before it causes material impact to its operations.

2. **Data availability** – This measures to what extent the data is available for use by end-users at the production and DR servers, *while* it is being protected. This measures the effective utilization of the DR resources implemented by the enterprise.

3. **Data recoverability** – If an outage was to occur at the production data center, this indicates how soon the data could be recovered at the DR data center, and be available for end-user access. This is often referred to as the Recovery Time Objective (RTO), which measures the maximum duration the application may be unavailable for end-user access, before it causes material impact to the business operations.

4. **Manageability and Operability** – Management and administration issues may significantly increase the total cost of ownership of DR solutions. Management issues, along with functional capabilities, should be carefully evaluated before making the final selection. Evaluation metrics in this category include how simple/complex it is to manage the DR configuration on an ongoing basis, whether the implementation requires additional personnel, training, and systems integration, whether the solution provides optimal degrees of automation with built-in alerting mechanisms to warn administrators of unexpected system changes, whether the functionality offered by the solution is flexible enough to adapt to changing business requirements, etc.

5. **Total Cost of Ownership (TCO)** – For today's high-end IT systems, the initial licensing/acquisition costs are only a fraction of the total costs required to own and maintain the systems. This is especially true for DR solutions that must provide adequate protection from the outages typical in the enterprise. For example, a DR solution may offer low licensing costs, yet – if it is unable to protect the enterprise from critical outages that may in turn cause the enterprise significant costs of downtime, the total costs for owning and maintaining such a system may run into millions of dollars. Because of this, along with management costs, downtime costs must also be considered while doing any TCO analysis of a DR solution.

The following sections provide a comparative analysis of Data Guard in an MAA configuration, Stretch Clusters based on RAC, and Remote Mirroring, using this DR Solution Evaluation Framework.

EVALUATING DR – DATA PROTECTION CHARACTERISTICS

A fundamental data protection requirement for any DR solution is whether it is able to provide bullet-proof isolation of faults, outages and data corruptions.

DATA GUARD IN AN MAA CONFIGURATION

- **Protection from Data Corruptions** – Most outages are not caused by natural disasters such as earthquakes or hurricanes, but by more mundane activities such as human errors, logical and physical data corruptions. Since the databases in a Data Guard configuration are loosely coupled, and the Redo Apply and SQL Apply technologies validate the redo before applying it on the standby database, Data Guard offers excellent fault-isolation and protection from data corruptions.

- **Protection from Widespread Disasters** – Data Guard, which is based on standard TCP/IP protocols, has no distance limitation. It also supports asynchronous redo transport to minimize any impact on the production throughput. The standby databases may be located thousands of miles away from the production database, enabling enterprises to leverage remote data centers. This enables Data Guard to provide highly effective protection from widespread disasters – e.g. regional earthquakes, or regional power outages such as the US East Coast power outage in 2003, or regional hurricanes that affected Florida in 2004.

- **Protection from Network Outages** – With so many components (e.g. switches, routers, bridges, etc.) present in a network, outages and/or glitches in a network are not uncommon, and the probability of such outages increases with the network size. Data Guard offers various configuration options through which administrators may choose the right protection mode (e.g. the Maximum Availability mode or Maximum Performance mode) that allows the primary database to continue its operations without any impact even if the network is disconnected between the primary and standby servers. When the network is restored, Data Guard automatically resynchronizes the standby database with the primary.

- **Protection from Server Failures** – When Data Guard is used in an MAA configuration, i.e. Data Guard combined with RAC, if one or more servers in the primary cluster fail, the surviving nodes can quickly take over the workload of the failed nodes, with minimal impact on the application. Besides, using the Fast-Start Failover feature of Data Guard in Oracle Database 10*g*, following an outage at the primary site, Data Guard can automatically fail over to a chosen standby very quickly, without requiring any manual intervention.

- **Minimizing Downtime during Database Upgrades** – A significant amount of downtime is caused by maintenance activities in the data center. This includes the upgrade of databases from one version to the next higher version. Data Guard in Oracle Database 10*g* supports database upgrades for major release and patchset upgrades (from Oracle Database 10*g* onwards) in a rolling fashion – with near zero downtime, by using Data Guard SQL Apply. The steps involve upgrading the logical standby database to the next release, running in a mixed mode to test and validate the upgrade, doing a role reversal by switching over to the upgraded database, and then finally upgrading the old primary database. While running in a mixed mode for testing purpose, the upgrade can be aborted and the software downgraded, without data loss.

  Besides database rolling upgrades, the switchover feature in a Data Guard configuration allows it to be used also for other maintenance activities such as hardware upgrades, application upgrades, etc.

- **Protecting Non-database Data** – Data Guard offers protection for only data resident in Oracle databases.

STRETCH CLUSTERS

- **Protection from Data Corruptions** – In a Stretch Cluster configuration, any corruption at one volume will be propagated to the remote volume by the underlying remote mirroring technology, thus rendering the entire configuration unusable.

- **Protection from Widespread Disasters** – In a Stretch Cluster configuration, all I/O in any volume will have to be synchronously propagated to the remote volume to keep the two data volumes in synch. Because of the synchronous nature of the transport, it is impractical to deploy a Stretch Cluster configuration over distances beyond a few kilometers. Because of this, such a configuration does not provide an effective solution for regional disasters such as destructive hurricanes. Stretch Cluster configurations however offer excellent protection from localized disasters, such as data center fires.

- **Protection from Network Outages** – For a RAC Stretch Cluster configuration, since the cluster-interconnect, which enables Cache Fusion communication and cluster management services, has to span the inter-site distance, this network has to be extremely reliable with very low latencies (recommended to be less than 2 ms). For high latencies of the network, or any network glitches, the Cache Fusion communication, along with other communication regarding management of the cluster, is impacted, which in turn adversely impacts the entire cluster. To minimize the scope of this impact, it is recommended to keep the Cache Fusion communication on dedicated redundant network channels.

- **Protection from Server Failures** – A RAC-based Stretch Cluster configuration offers excellent protection from server failures. If one or more servers in a Stretch Cluster configuration fail, the surviving nodes perform recovery on behalf of the failed nodes, the workload gets re-distributed among these surviving nodes and processing continues, with no application downtime. Of course, the servers have to be provisioned in advance with excess capacity such that they can take on the additional load without degrading application performance.

- **Minimizing Downtime during Database Upgrades** – Since a Stretch Cluster configuration is based upon a single database, rolling database upgrades, or rolling application upgrades, is not possible in such a configuration. However, starting with Oracle9*i*, RAC offers the capability of applying one-off database software patches in a rolling manner [5]. Thus, this is a benefit of Stretch Cluster as well as an MAA configuration.

- **Protecting Non-database Data** – RAC Stretch Clusters offer protection for only data resident in Oracle databases.

REMOTE MIRRORING

- **Protection from Data Corruptions** – Remote Mirroring configurations are unable to protect against data corruptions – any corruption at the primary storage array will be transmitted to the remote storage array.

- **Protection from Widespread Disasters** – Storage Array-based Remote Mirroring solutions also have similar distance restrictions, because of the limitations of the underlying protocol, typically Fibre Channel, and hence they may not provide adequate protection from regional disasters. To alleviate this problem, protocol converters (converting Fibre Channel to IP) may be used, however that increases the overall cost of the configuration. It is also possible to do asynchronous mirroring with Remote Mirroring solutions, however it should be noted that some asynchronous mirroring solutions may not preserve database write-ordering and will corrupt the target database. Oracle customers interested in implementing a Remote Mirroring solution must use a solution that is validated as part of Oracle Storage Compatibility Program (OSCP) [6] to ensure that their remotely protected database is always consistent. OSCP-validated synchronous remote mirroring solutions provide excellent protection from widespread disasters.

- **Protection from Network Outages** – Remote Mirroring solutions offer good protection from network outages, especially if they are validated in OSCP to operate in an asynchronous mode.

- **Protection from Server Failures** – Remote Mirroring solutions offer no intrinsic capability such as RAC or Data Guard to offer protection from server failures. Such a solution either has to be separately integrated with RAC, or, after a server failure, a manual failover process to the remote host has to be executed.

- **Minimizing Downtime during Database Upgrades** – Since the target volume in a Remote Mirroring configuration matches block-for-block the source volume, rolling database upgrades, or rolling application upgrades, is not possible in a Remote Mirroring configuration.

- **Protecting Non-database Data** – Remote Mirroring solutions offer excellent protection for filesystem data, or data that is resident in databases other than Oracle, since such solutions are database agnostic.

SUMMARY

To summarize, data protection is the most critical requirement of any DR solution. Based on the discussions above, the following data protection effectiveness characteristics can be profiled for each of MAA, Stretch Clusters and Remote Mirroring for the following set of key outages that can affect the data center:

Sample Outage Profile	Effectiveness in Protecting From the Outages		
	MAA	**Stretch Cluster**	**Remote Mirroring**
Data corruptions	Excellent	Poor	Poor
Disasters (Local/Regional)	Excellent	Fair	Excellent
Network outage	Excellent	Good	Excellent
Server failure	Excellent	Excellent	Fair
Database upgrade	Excellent	Fair	Poor
Application upgrade	Fair	Poor	Poor

Mapping and aggregating these effectiveness-values to percentage values of outages protected by each solution, the above table can be visually represented by the following diagram.

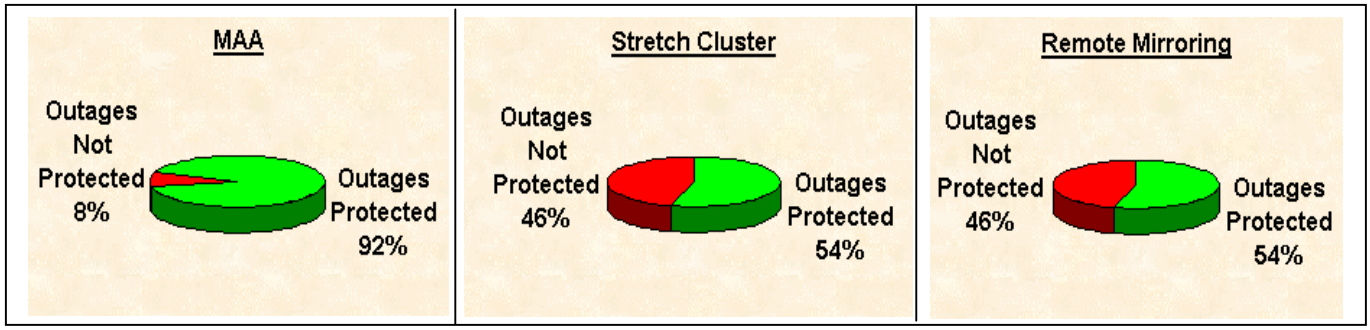

Fig. 5: Effectiveness of each DR Solution in Protecting from Outages

This points out the general effectiveness of an MAA configuration (i.e. RAC and Data Guard combined) in protecting enterprises from a wide variety of outages and maintaining business continuity. However, each business case is unique regarding high availability requirements, and hence such an evaluation should be made only after a clear understanding of the outage risks and downtime impacts for the particular enterprise.

EVALUATING DR – DATA AVAILABILITY CHARACTERISTICS

The Data Availability metric has two aspects:

- To what extent the data/systems are available for use at the DR site, and
- Whether there is any impact on the primary applications while the data is protected.

Since any HA and/or DR configuration involves deploying redundant systems, administrators desire that all the systems employed in this configuration are utilized effectively, so that the maximum value can be extracted out of the DR investment. A Stretch Cluster configuration offers best system utilization in that regard. On the other hand, Data Guard offers variouse options to reduce any impact on the production application throughput.

DATA GUARD IN AN MAA CONFIGURATION

- **Systems Utilization** – Data Guard offers limited availability of its standby databases for application processing. A physical standby database can be opened up for read-only access, but redo cannot be applied at that time. However, using RMAN [7], a physical standby database can be used to offload backups from the primary database, while redo is being applied at the same time. A logical standby database can be used for simultaneous reporting while SQL Apply is going on, but any updates of protected data can only occur in the primary database.

- **Impact on Application Throughput** – Data Guard can use asynchronous redo transport over long distances spanning thousands of miles, with minimal impact on the application throughput, and with less network bandwidth requirements compared to solutions based on Remote Mirroring.

STRETCH CLUSTERS

- **Systems Utilization** – Stretch Cluster configurations offer excellent systems utilization. In such configurations, all the RAC nodes are fully active, and can support production applications. Besides, with RAC load-balancing, the workload can be distributed among the nodes, thereby improving application throughput and systems utilization.

- **Impact on Application Throughput** – In a Stretch Cluster configuration, which relies on synchronous mirroring – just because of the synchronous nature of the mirroring, there is a higher probability of impact on the throughput of database applications, with increased network latency and increased distances. Because of this reason, a dedicated network with extremely low latencies and spanning smaller distances is advisable for such configurations. Otherwise, latencies/glitches in the interconnect will adversely impact Cache Fusion communication, which in turn will directly impact the scalability and performance of production applications.

REMOTE MIRRORING

- **Systems Utilization** – Remote Mirroring solutions offer very poor systems utilization since the target volumes cannot be directly accessed by any application (e.g. reporting, read-only operations, or backups) while mirroring is in progress. Such solutions typically suggest using the snapshot feature to create a version of the data on a tertiary volume for subsequent access, thereby requiring investment in additional storage devices. Note that administrators must ensure that database consistency is always preserved while taking these snapshots.

- **Impact on Application Throughput** – Remote Mirroring configurations that can be operated in an asynchronous mode will have minimal impact on the throughput of database applications. Hardware-based Remote Mirroring solutions have the additional advantage in that they can offload the processing to the storage arrays and save on server processing cycles.

SUMMARY

To summarize, data availability allows enterprises to extract value out of their DR investments. Based on the discussions above, the following data availability effectiveness characteristics can be profiled for each of MAA, Stretch Clusters and Remote Mirroring:

Data Availability Characteristics	Effectiveness in Providing Data Availability		
	MAA	**Stretch Cluster**	**Remote Mirroring**
Systems utilization	Fair	Excellent	Poor
No impact on application throughput	Excellent	Good	Excellent

EVALUATING DR – DATA RECOVERABILITY CHARACTERISTICS

Following a disaster, services must be restored as soon as possible. This is the moment when downtime costs are mounting, time demands are high, and resources are limited. The DR systems must work reliably and flawlessly at this time. The enterprise cannot afford to go through another disaster.

DATA GUARD IN AN MAA CONFIGURATION

- **Recovery Following an Outage** – Data Guard offers robust capabilities in this regard. Integrated role management support through the switchover and failover features allows the standby database to quickly be the new primary database and continue serving the data needs of the enterprise. All the data that has been applied on the standby database has already been validated as part of Redo Apply or SQL Apply, so applications can immediately start working on data that is assured to be consistent and clean.

  In Oracle9*i*, the failover process is manually initiated, and there may be accumulated redo data that has to be applied on the standby database before the failover process is complete. In Oracle Database 10*g* Release 1, using Real Time Apply, all redo is immediately applied to the standby database, instead of waiting for a log switch, and thus it speeds up the failover/switchover time. Additionally, using the Fast-Start Failover feature in Oracle Database 10*g* Release 2, following an outage at the primary site, Data Guard can automatically fail over to a chosen standby in a matter of seconds, without requiring any manual intervention. The combination of Real Time Apply and Fast-Start Failover means Data Guard can resume business operations very quickly after a disaster.

- **Fast Automatic Reinstatement of Old Primary** – Another very critical aspect of recovery following an outage is how quickly the old primary database can be reinstated back in service, to ensure DR protection against a possible second disaster. Data Guard in Oracle Database 10g Release 2, following a failover and after the restarting of the old primary database, allows it to be automatically reinstated as a new standby database in the configuration, synchronized with the current primary database, and resume DR protection for the entire configuration. This eliminates the need to recreate the database from a backup, or by copying data files across a network.

STRETCH CLUSTERS

- **Recovery Following an Outage** – A Stretch Cluster configuration, if it can survive the disaster, also offers excellent recovery capabilities. If a subset of the nodes fails, applications connected to those nodes can immediately start accessing the surviving nodes and continue their operations. One of the surviving instances automatically performs instance recovery for any failed nodes or instances, rolling forward all transactions recorded in the redo logs of the failed instances since the last checkpoint. Transactions touching disjointed data sets will continue to be processed while a small brownout (reduced by optimal setting of the FAST_START_MTTR_TARGET parameter) may be incurred for transactions that are dependent on the data that needs to be recovered. After the roll-forward phase is complete, all transaction processing can recommence, while all uncommitted transactions are rolled back in the background.

- **Fast Automatic Reinstatement of Old Primary** – For Stretch Cluster configurations, the data volumes at the old primary site have to be completely resynchronized with the data volumes at the new primary site before the servers at the old primary site can start accessing their local data volumes. This may be a manual operation.

REMOTE MIRRORING

- **Recovery Following an Outage** – In a Remote Mirroring configuration, following a disaster, the database has to be mounted on the target volumes and restarted. This is a manual operation. Besides, since the data was never validated during the mirroring (unless a split mirror was done and the data validated on that split image), the database restart may fail if the data was somehow corrupted during the mirroring process. This of course is the worst possible time for such a "second disaster" to occur. Because of this, for remote mirroring solutions, administrators must periodically validate the mirrored data to avoid such situations.

- **Fast Automatic Reinstatement of Old Primary** – Following a disaster at the primary site and a subsequent failover, the old primary data volumes have to be resynchronized with changes at the new primary data volumes. Depending on the remote mirroring solution, this may be a manual operation. Also, if these changes are not tracked, these volumes have to be recreated by copying data files across the network, or from a backup of the new primary database. For very large databases, this can be a very time consuming operation, network intensive, and it keeps the new primary database vulnerable to a second disaster for the entire duration.

SUMMARY

To summarize, data recoverability measures how fast enterprises can resume business services after a disaster. Based on the discussions above, the following data recoverability effectiveness characteristics can be profiled for each of MAA, Stretch Clusters and Remote Mirroring:

Data Recoverability Characteristics	Effectiveness in Providing Data Recovery		
	MAA	**Stretch Cluster**	**Remote Mirroring**
Fast integrated automatic recovery after an outage	Excellent	Excellent	Poor
Fast automatic reinstatement of old primary	Excellent	Good	Good

EVALUATING DR – MANAGEABILITY & OPERABILITY CHARACTERISTICS

Since DR solutions involve redundant system components across multiple sites, they must be easy to install, configure and manage on a continuing basis. They should also provide the necessary configuration options such that they can easily meet changing business requirements.

DATA GUARD IN AN MAA CONFIGURATION

- **Manageability** – Setting up an MAA configuration requires separate configuration of Data Guard and RAC. However, once the configuration is set up, the automation built within MAA keeps the standby databases in synch with the primary by having the multiple RAC primary instances send their redo to a designated standby instance (called the *apply instance*) and automatically merging and applying these redo threads to the standby database. An MAA configuration can handle network disconnects gracefully based on protection mode settings, and handle outages/disasters at the primary site without requiring little or no manual intervention. An MAA configuration does not require integration with any third party software or hardware.

- **Operability** – Data Guard, with its full suite of disaster recovery features – e.g. Redo Apply (physical standby databases), SQL Apply (logical standby databases), multiple protection modes, push-button automated switchover/failover capabilities, automatic gap detection and resolution, GUI-driven management and monitoring framework, cascaded redo log destinations, etc., is a very comprehensive and effective solution optimized for data protection and disaster recovery.

  Data Guard is implemented on top of pure commodity hardware. It only requires a standard TCP/IP-based network link between the two computers. No specialized hardware is required. It also allows the storage on the standby server to be laid out in a different fashion from the primary. For example, customers can put the files on different disks, volumes, file systems, etc.

STRETCH CLUSTERS

- **Manageability** – Stretch Clusters use RAC, a built-in feature of the Oracle database; however Stretch Cluster configurations require integration with a volume mirroring software. Also, setting up a Stretch Cluster configuration has some unique design issues. The network between the inter-site nodes has to be extremely reliable with redundancies built on dedicated channels for each type of communication (e.g. Cache Fusion and volume mirroring), and must have extremely low latencies to avoid any impact on the overall cluster operation and OLTP application throughput. Administrators also have to be careful when the applications and/or the usage profile change, for potential impact on the new applications.

  Special considerations also need to be given to placement of voting disks. For example, in case of network connectivity problems between the two sites, building a proper quorum mechanism to avoid split-brain situations[3] is important consideration in this configuration. Firstly, the mirroring solution needs to be cluster-aware. Secondly, if the quorum device is located in one of the two sites, and that site is struck with a disaster, the entire cluster is brought down and must be manually reinstated. For maximum availability, the quorum device ought to be placed at a third location, which increases the operational complexity as well as the cost.

- **Operability** – There are some platform-related issues in a Stretch Cluster configuration. To support host-based mirroring, it needs a Logical Volume Manager (LVM), that has to be cluster-aware, and hence, integrated with the clusterware used in the configuration. If Oracle Clusterware is used, there are some restrictions. For example, in Oracle9*i*, if Oracle Clusterware is used, using an LVM is not an option since there is no such integration between an LVM and Oracle Clusterware in Oracle9*i*. In that case, the only mirroring option in a Stretch Cluster configuration is hardware mirroring, which has limited availability characteristics, as explained previously. In Oracle Database 10*g*, if Oracle Clusterware is used, the only LVM option is ASM.

  If non-Oracle Clusterware is used, there are no such restrictions, since there are cluster-aware LVMs available that are integrated with non-Oracle Clusterware, and hence such Stretch Cluster configurations can use host-based mirroring.

[3] A split-brain condition occurs when all communication between clustered nodes is lost and the cluster becomes partitioned into sub-clusters, each believing that it is the only partition. This may cause conflict on the shared data, leading to data corruption.

REMOTE MIRRORING

- **Ease of Management** – A Remote Mirroring solution is an extra cost purchase that requires separate integration, acquiring new set of product knowledge, and hiring extra resources to administer this solution.

- **Operability** – In a Data Guard configuration, only the redo data need to be sent to the remote site. However, if a remote mirroring solution is used, typically the database files, the online logs, the archive logs and the control file must be mirrored. This means that remote mirroring will send each change at least three times to the remote site. In an internal analysis of Oracle's corporate e-mail systems, as shown in the following graph, it was demonstrated that 7 times more data was transmitted over the network and 27 times more I/O operations were performed using a remote mirroring solution, compared to using Data Guard.

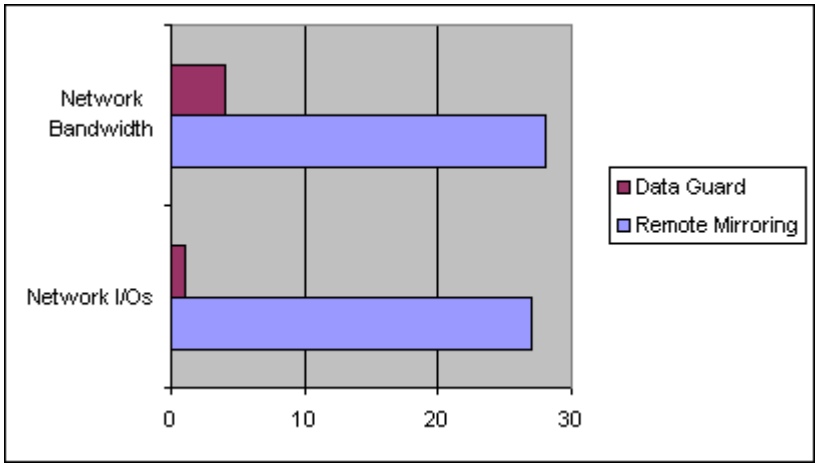

Fig. 6: Data Guard vs. Remote Mirroring – Network Resource Consumption

Hardware Remote Mirroring solutions are also restrictive in the sense that typically these remote mirroring solutions can be used with only the identically configured storage systems from the same vendor that manufactures these remote mirroring solutions.

SUMMARY

To summarize, manageability and operability measure the ease-of-use with which the DR solution can be administered and managed. Based on the discussions above, the following manageability and operability effectiveness characteristics can be profiled for each of MAA, Stretch Clusters and Remote Mirroring:

Manageability & Operability Characteristics	Effectiveness in Manageability & Operability		
	MAA	**Stretch Cluster**	**Remote Mirroring**
Ease of configuration, management and operation	Good	Fair	Good
Flexibility in operability	Excellent	Good	Good
Network resource utilization	Excellent	Fair	Fair

EVALUATING DR – TOTAL COST OF OWNERSHIP

The final metric to evaluate a DR solution is Total Cost of ownership (TCO) of that solution. TCO analysis is especially important for a DR solution, since, with the redundancies built in this configuration, total costs to maintain this solution over its useful life may easily exceed the initial estimate and cause budget overruns. For an effective TCO analysis of an HA or a DR solution, the following components must be considered:

(I) Acquisition & Service costs
(II) Install & Configuration costs
(III) Management & Administration costs
(IV) Downtime costs

Acquisition and Service costs are what the enterprise paid to acquire the solution – typically, it is the upfront technology licensing costs and the service contract expenses (renewable every year). Technology licensing includes servers, OS, database, the particular DR solution, storage arrays, network, etc.

Installation and configuration costs are incurred when the enterprise spends time to implement the solution. This may involve sending employees for training, hiring consultants to do the initial implementation, having one or more employees work on this project for some time, etc.

Management and administration costs are related to expenses incurred to maintain the DR configuration over its life. Once configured and set up, if the DR solution needs little manual intervention to successfully operate, then its management costs will be minimal. In contrast, if the configuration needs to be reinitialized for any kind of reconfiguration, that adds to the management costs. If it needs one or more dedicated administrators to maintain it, that increases its management costs as well. Similarly, if every change of configuration parameters requires applications/databases to be restarted – that also increases the operational costs.

Finally, downtime costs are related to the unplanned and planned downtime elements that the solution could not protect the enterprise from. Depending on the nature of the business of the enterprise, and depending on the outage, these downtime costs could be very high for an enterprise. For example, if a Remote Mirroring solution is implemented without any clustering support, then server failures could cause some downtime in that configuration since it cannot automatically fail over to another server as done in a cluster configuration. Similarly, if a Stretch Cluster configuration is implemented, and data corruption occurs, that would add up the downtime costs for that configuration, since a corruption will adversely affect the entire database.

For a TCO analysis of a disaster recovery solution in the enterprise, focus has to be given especially to Management and Downtime costs, because those tend to be the most significant factors in the overall cost equation.

ACQUISITION & SERVICE COSTS

Considering the TCO component (I), i.e. *Acquisition & Service costs*, a Stretch Cluster configuration seems to have the least hardware costs, since it does not need to have servers in the "standby" mode as is the case with the other two configurations. For example, a 4-node Stretch Cluster configuration is functionally equivalent to a 4-node primary database + 2-node standby database in an MAA configuration. However, the Stretch Cluster configuration has two extra expensive components compared to an MAA configuration – a cluster-aware volume manager that can do mirroring to remote disks, and specialized hi-speed, low-latency and redundant network for the cluster and I/O traffic.

For a Remote Mirroring configuration, there are expenses involved for the solution itself, and in addition – the standby servers in the target site are basically idle since the target data volumes cannot be accessed while the mirroring is in progress. Compared to this, the standby servers in a Data Guard configuration can be used for reporting, backups, testing the DR configuration, etc. Another unique issue related to Remote Mirroring solutions is that they may be licensed on capacity pricing – i.e. costs increase with the amount of the data replicated, which significantly adds up the costs to replicate large databases for high-growth enterprises.

Finally, since Remote Mirroring solutions (as well as the volume manager mirroring for Stretch Clusters) send more network traffic compared to Data Guard, they have more bandwidth requirement and hence more network-related expenses than an MAA configuration.

CONFIGURATION & MANAGEMENT COSTS

Considering the TCO components (II) and (III), i.e. *Install & Configuration costs*, and *Management & Administration costs*, a Remote Mirroring configuration tends to have higher costs compared to an MAA configuration, since it involves extra integration, may require sending personnel for specialized training or hiring extra resources instead of leveraging existing database administration skills, and requires manual intervention for operations such as splitting the mirror, performing a failover, etc.

DOWNTIME COSTS

This brings us to the final TCO component (IV), i.e. Downtime costs, which is the most critical TCO component of a DR or HA solution. For example, assume a hypothetical enterprise, which, without any disaster protection solution, is estimated to suffer 40 hrs of downtime in a year, with an average downtime cost of $50,000 per hour, leading to total downtime costs of $2M per year. As noted previously, if an MAA configuration offers 92% protection from outages, compared to 54% protection in Stretch Clusters and Remote Mirroring, the enterprise stands to save much more on downtime costs with an MAA configuration. Clearly, an MAA configuration has the lowest Downtime Costs component of the TCO, and the difference in absolute dollar amounts is even more significant when considered over the entire life of such a configuration.

SUMMARY

To summarize, a careful TCO analysis of a DR solution should be done that should take into account how the solution helps minimize downtime costs for the enterprise. Based on the discussions above, the following TCO components can be profiled for each of MAA, Stretch Clusters and Remote Mirroring:

TCO Component	TCO Characteristics		
	MAA	**Stretch Cluster**	**Remote Mirroring**
Low Acquisition and Service costs	Excellent	Excellent	Fair
Low Install & Configuration costs	Excellent	Excellent	Good
Low Management & Administration costs	Excellent	Good	Fair
Low Downtime costs	Excellent	Fair	Fair

SUMMARY

High availability and disaster recovery are universal requirements for any global enterprise today. This paper analyzed the DR and HA capabilities of three solutions available for Oracle customers – Data Guard in an MAA configuration, RAC-based Stretch Clusters, and Remote Mirroring. The following table summarizes this analysis.

DR/HA Capabilities	Data Guard in MAA	Stretch Cluster	Remote Mirroring
Data Protection			
Protection from data corruptions	Excellent	Poor	Poor
Protection from local disasters	Excellent	Excellent	Excellent
Protection from regional disasters	Excellent	Poor	Excellent
Protection from network outages	Excellent	Good	Excellent
Protection from server failures	Excellent	Excellent	Fair
Reduction in downtime for database upgrades	Excellent	Fair	Poor
Protection of non-database data	Poor	Poor	Excellent
Data Availability			
Systems utilization	Fair	Excellent	Poor
Reduction in impact on application throughput	Excellent	Good	Excellent
Data Recovery			
Fast integrated automatic recovery after an outage	Excellent	Excellent	Poor
Fast automatic reinstatement of old primary	Excellent	Good	Good
Manageability & Operability			
Ease of configuration, management and operation	Good	Fair	Good
Flexibility in operability	Excellent	Good	Good
Network resource utilization	Excellent	Fair	Fair
Total Cost of Ownership			
Low Acquisition and Service costs	Excellent	Excellent	Fair
Low Install & Configuration costs	Excellent	Excellent	Good
Low Management & Administration costs	Excellent	Good	Fair
Low Downtime costs	Excellent	Fair	Fair

As can be seen from the above table, each of these solutions has some unique benefits and considerations. In general, for the best protection from a wide variety of unplanned and planned outages, least management costs and lowest total cost of ownership, Data Guard in an MAA configuration is recommended. RAC-based Stretch Cluster configurations allow very good utilization of all servers in that configuration, however it does not provide adequate protection from a critical set of outages. Remote Mirroring is an effective solution for protecting file system data, or data that is resident in databases other than Oracle, however for data resident in Oracle databases, it is an unnecessary extra-cost purchase, without any effective utilization of the remote servers.

Enterprises interested in these solutions must first perform a thorough analysis of the high availability requirements of their critical business applications, the outage risks in their data centers, and the impact of downtime from these outages. Using such analyses, enterprises should perform an assessment of these solutions using the evaluation framework described in this document, to select and implement the solution that best meets their business continuity requirements.

REFERENCES

1. Oracle Real Application Clusters (RAC): http://www.oracle.com/technology/products/database/clustering/index.html
2. Oracle RAC 10g Technical White Paper: http://www.oracle.com/technology/products/database/clustering/pdf/TWP_RAC_Overview_10gR1_112503.pdf
3. Oracle Data Guard: http://www.oracle.com/technology/deploy/availability/htdocs/DataGuardOverview.html
4. Oracle Maximum Availability Architecture (MAA): http://www.oracle.com/technology/deploy/availability/htdocs/maa.htm
5. Database Rolling Patch Updates with Real Application Clusters: http://www.oracle.com/technology/deploy/availability/pdf/Rolling_Patch_Update_Data_Sheet.pdf
6. Oracle Storage Compatibility Program (OSCP): http://www.oracle.com/technology/deploy/availability/htdocs/oscp.html
7. Oracle Recovery Manager (RMAN): http://www.oracle.com/technology/deploy/availability/htdocs/rman_overview.htm