Oracle Maximum
Availability Architecture

An Oracle White Paper
September 2011

Oracle Exadata Database Machine - Backup & Recovery Sizing: Tape Backups

Table of Contents

Introduction

Oracle Exadata Database Machine is a highly available and high performance system ideal for running mission critical Oracle databases. The databases are backed up to Exadata's fast recovery area (FRA) location which provides high bandwidth environment for faster backup and restore operations. However, keeping long term backups on local disks may not be feasible and creates capacity management issues. Secondly, these local copies do not protect the database in the event of a site or the Exadata Database Machine failure. In order to achieve high availability and business continuity, storing backup copies externally ensures the databases can be restored on to a different site or another Exadata Database Machine in the event of a failure.

Backup and recovery architecture to an external media requires additional hardware and software components. Each component has its limitation on how much data it can process and transfer at any given point in time. In order to suggest an optimal backup and recovery architecture, it is critical to select the right set of hardware and software components.

This document provides a methodology that can be used as a guideline to design a backup and recovery infrastructure for backing up Exadata Oracle databases to a tape media. The methodology is categorized into three stages.

1. **Gather Requirements**: The first stage is to gather application availability requirements. The relevant metrics are recovery time objective, recovery point objective, backup window, and retention period along with the size of the database to backup.

2. **Analysis**: Calculate the throughput, capacity, and future scalability needs which satisfy the availability requirements. Choose the right set of hardware and software components that are compatible and satisfy the overall requirements.

3. **Solution**: Architect a solution by putting together an end to end hardware and software infrastructure and demonstrate how it meets the requirements. The solution should also include relevant best practice recommendations.

Tape Backup Infrastructure Components

The following table illustrates the major components involved in a tape backup architecture.

TABLE 1: TAPE BACKUP INFRASTRUCTURE COMPONENTS

Component Name	Description
Exadata Database Nodes	Exadata physical servers that host the databases that needs to be backed up.
Network Switches	In this document, it is referred to the connectivity between Exadata and media servers.
Media Server(s)	It is the server that manages client's backup and restore to the attached tape media.
Administrative Server	It is a server that houses the catalog for the backup software.
FC Switches	Fibre Channel switches are used to connect FC cables between the Media Servers and the Tape Library or the tape drives
Tape Library	Storage device that contains tape drives, slots to hold tape catridges, along with other components such as bar code reader, robotic arm to move catridges etc.
Tape Drives	Devices that read and write data to the tape media.
Tape Media	Also called tape catridges, these are primarily used as offline backup storage device. The data is read and written sequentially.
Media Management Software	It is the backup application works with Recovery Manager for performing tape backups.

Requirements

1. Recovery Time Objective (RTO)

RTO indicates the downtime tolerance for an application. In the event of a failure, the data need to be restored, recovered and to be made operational within the defined RTO. For database backup & recovery sizing exercise, a worst case scenario of total time taken to restore the entire database and performing recovery is to be factored in the calculation. Data Warehouse (DWH)/ Decision Support System (DSS) databases typically have an RTO of 12 hrs or more. For databases used in Online Transaction Processing (OLTP) systems, the RTO is typically defined as minutes to few hours.

Depending on the RTO requirement, the backup location can be determined. It can be disk based, tape based, or a combination of disk and tape. For databases with RTO defined in hours can be met by the tape-based backup & recovery infrastructure outlined in the following sections. However, for databases with strict RTO requirement of minutes to seconds, other Oracle HA solutions such as Oracle Data Guard and Oracle Flashback are recommended.

2. Recovery Point Objective (RPO)

RPO indicates data loss tolerance for an application. RPOs are defined in terms of seconds, minutes or hours of data that the business can afford to lose in the event of a failure. Mission critical OLTP applications typically have a zero or near-zero data loss requirement. Depending on the nature of the downtime, under most circumstances, traditional backup and recovery methods can be used to meet the RPO requirement. However, for databases with lower RPO requirements such as zero data loss, deploying Oracle Data Guard is recommended in addition to the traditional backup and recovery architecture.

3. Backup Retention Period

Backup retention period defines how long the backup of an application is retained before the backup media is reused. It is typically defined in terms of days, weeks, months, and years. The retention period is also defined for the backup repository location. Backups stored on disks local to the production have shorter retention period. External disk or tape media typically have longer retention period. For example, backup retention period on Exadata FRA could be for 7 days, copies on an external disk could be for 6 months and copies on tape could be for several years.

4. Backup Window

Backup window is a service level agreement (SLA) which defines the amount of time allocated by the business/IT to complete the backup for a particular database, or a group of databases. As an example, there could be a requirement to backup a 500GB database within 1 hour. The backup infrastructure must be sized to meet the backup window SLA.

Backup infrastructure must be sized to perform backup and restore within the allotted time which are defined by backup window and RTO metrics. However, since the RTO includes restore and recovery processes, it is generally longer duration than the backup window.

Analysis

5. Network Connectivity

In order to meet the requirement, it is critical to choose the right connectivity options in the backup and recovery architecture. This section provides typical bandwidth numbers for different kinds of networks that one can expect to achieve in most configurations.

The network ports can be connected directly between the components or the ports can be bonded together to form a logical interface. The bonded ports can be configured for load balancing, fault tolerance, and better network utilization. If the bonding is done in active/passive configuration between two ports in a system, then the effective throughput between the ports is ½ of the total throughput.

The other factor that influences the overall network efficiency is the latency. Latency is the round trip of a data packet between two points. Increase in latency reduces the amount of data that can be transferred for a given time period.

5a. Exadata to Media Server connectivity

Exadata supports InfiniBand (IB), 1GigE, and 10GigE connectivity to external servers and NAS storage. Connectivity between Exadata and the Media Server can be done using one of the following network options:

TABLE 2: NETWORK CONNECTIVITY

Network Connectivity	Typical Network Bandwidth Per Port
InfiniBand QDR (40Gbps)	2GB - 2.5GB/sec
1GigE	100MB - 125MB/sec
10GigE	1GB – 1.25GB/sec

The network cables between the Exadata database nodes and the Media Servers can be directly connected (peer to peer) or over a network switch. A network switch is recommended wherever possible for easier network management and optimal performance. Depending on the backup and recovery requirements, either existing network infrastructure can be used (such as 1GigE) or a new InfiniBand/10GigE infrastructure needs to be deployed to meet the backup requirement.

5b. Media Server to the tape devices connectivity

Connectivity between the Media Servers and the tape devices are generally done over a Fibre Channel (FC) storage area network (SAN). The type of Host bus adapters (HBAs) installed on the Media Server determines the available throughput from the media server out to tape. A Media Server can be equipped with one or more HBAs to push the data to the tape drives. Additional tape drives may be effectively assigned to be managed by a Media Server until the throughput of the HBAs have been saturated. Table 3 lists typical bandwidth achieved per HBA port.

TABLE 3: FIBRE CHANNEL NETWORK

FC Connectivity	Typical Bandwidth Per Port
2 Gb/s	200MB/sec - 212.5MB/sec
4 Gb/s	400MB/sec - 425MB/sec
8 Gb/s	800MB/sec - 850MB/sec

In FC environments, usually one or more SAN switches are placed between the Media Servers and tape devices. The switch should have enough ports to accommodate immediate backup and recovery throughput needs, along with room for future for additional tape drives connectivity.

6. Backup Media and Administrative Servers

Media Server manages backup and restore of client's data to the attached media devices. Tape devices are connected to the Media Servers. Administrative Server houses the catalog for the backup software. Backup data flows from the Exadata database servers through the Media Servers out to tape devices.

Backup application software such as Oracle Secure Backup control messages flow between the Administrative Server, Database Servers and Media Servers.

It is good practice to allocate a minimum of two media servers for high availability in an active/active setup. The Media Server(s) and Administrative Server may utilize any platform / operating system supported by the backup software.

The network interface and HBA requirements varies for Administrative and Media Servers. An Administrative Server will need a 1GigE NIC and, if configured to control the library robot, also an HBA. In contrast, the Media Servers need a 1GigE NIC to communicate with the Administrative Server, HBA (s) for streaming data to tape devices and potentially an InfiniBand or 10GigE interface for connection to the Exadata. For best performance, InfiniBand connectivity between the media servers and Exadata is recommended.

7. Tape Drives and Media / Tape Libraries

7a. Tape Drives and Media

When selecting the most appropriate tape drive format, capacity and throughput are the key considerations affecting infrastructure sizing decisions. LTO is a popular drive type as is Oracle's StorageTek T10000C tape drive. Table 4 lists some of the popular tape drives/media and their typical capacity and speed characteristics. Compression speed varies depending on various factors. In this document, typical compression speed of 1.5:1 is used for sizing purposes.

TABLE 4: TAPE CAPACITY & SPEED

Tape Drive/ Media	Native Capacity	Compressed Capacity	Native Speed	Typical Compressed Speed (@ 1.5:1 ratio)
LTO4	800GB	1.6TB	120MB/sec	180MB/sec
LTO5	1.5TB	3TB	140MB/sec	210MB/sec
T10000C	5TB	10TB	240MB/sec	360MB/sec

The number of tape drives and tapes needed depends upon which tape drive is used. Choosing higher capacity tapes generally mean fewer tapes are needed to store the same information. Faster drives get the job done in less time so fewer may be needed to meet the RTO and backup window SLAs.

Beyond RTO / backup window sizing calculations, other use case considerations should be taken into account when determining how many tape drive are needed. For example, one may want additional tape drives to support requirements for:

a) Tape duplication

b) Performing concurrent backup and restore operations and meeting SLAs for both

7b. Tape Libraries

The tape library is chosen to accommodate tape drives in use and slot count requirements. It should also be able to scale to meet estimated growth over the next few years as defined by the business. Review the maximum number of drives, drive types and slots supported by the library which is an indicator of scalability for growing environments. For example, Oracle's StorageTek SL500 is a popular midrange tape library scaling from 24 to 863 TB of storage capacity. Oracle's StorageTek SL3000 provides throughput of up to 48.4 TB/hr and scales up to 56 drives and just under 6000 slots.

Oracle Secure Backup supports Oracle StorageTek libraries as well as most other popular virtual and physical tape libraries as listed on the OSB tape drive and library compatibility matrix (http://www.oracle.com/technetwork/database/secure-backup/learnmore/tape-devices-10-3-161824.pdf).

8. RMAN Configuration

RMAN is the central piece in the entire backup and recovery architecture for Exadata. Using RMAN is the only way the database can be backed up from an Exadata system. For tape backups, RMAN is used in conjunction with media management software such as Oracle Secure Backup to directly backup the database or backup the FRA to tape devices. Configuring right number of RMAN channels, tuning for performance, defining the right retention policies are critical to meet business SLA requirements.

RMAN has a number of features and tunable parameters for optimizing the backup and recovery operations. RMAN supports both full and incremental backups and can perform image copy backups or as backup sets. RMAN along with the media management layer can be configured to backup the database directly to the tape (or) can be configured to backup the backup components from the FRA to tape.

For non-image copy backups, RMAN performs implicit compression by ignoring null and empty blocks this saving a lot of storage space. Using either the default BASIC compression or utilizing Advanced Compression Option (ACO), users can choose additional compression options to even reduce the backup size. Enabling Block Change Tracking mechanism enhances the backup speed. Maximum Availability Architecture (MAA) best practices recommend configuring the same number of RMAN channels as the number of tape drives allocated for backing up the database. In other words, allocate one RMAN channel to one tape drive.

RMAN best practices (www.oracle.com/goto/rman) show various methods and tuning parameters to setup policies and improve the backup and recovery performance. Also refer to the performance tuning paper http://www.oracle.com/technetwork/database/focus-areas/availability/rman-perf-tuning-bp-452204.pdf.

Solution

9. Put It All Together - Complete Picture

With the information gathered regarding RTO, RPO, backup window and retention period, appropriate hardware and networking components are analyzed to determine the right solution which meets the requirement. While not explicitly addressed here, other considerations include IT budget limits/allocations, and estimated growth of the business in the next few years.

With multiple components involved in the architecture, care must be taken to avoid any major performance degradation influenced by any one component. For example, allocating fewer tape drives or configuring an incorrect or too few network links, can lead to potential performance bottlenecks.

In this section two scenarios are illustrated. The methodology outlined in this paper is used to arrive at the optimal backup and recovery configuration to meet the customer requirement. Since real numbers may vary at specific deployments, these scenarios should simply serve as sizing guidelines.

The first example involves providing a solution that can be satisfied with using InfiniBand and LTO5 compression. The second example requirement is met using multiple GigE links and native LTO5 speed.

Example Scenario 1: InfiniBand

Customer setup and requirement:

Exadata X2-2 Half rack system (4 Database nodes, 7 Storage Servers)
Database size to backup : 24TB
RTO : 4hrs, RPO : 30 minutes
Backup window of 4 hrs (6TB/hr)
Backup to tape using RMAN & OSB
All 4 database nodes are used for backup
InfiniBand connectivity between Database Servers and Media Servers

This requirement is met by using the methodology defined in this document. The lower range of bandwidth number is used for a conservative calculation. Table 5 illustrates the throughput values for the requirements gathered and the hardware components chosen to meet the requirements.

TABLE 5: SCENARIO 1 – BACKUP & RECOVERY SIZING USING INFINIBAND CONNECTIVITY

Step #	Description	Value
1	RTO	4 hrs
2	RPO	30 minutes
3	Backup Window & Required Backup Rate	4 hrs to backup 24TB @ 6TB/hr = **1.67GB/sec**
4a	Exadata to Media Servers over IB	1 x IB link per database node (total of 4 nodes) Total of 4 x IB links **= 8.4GB/sec** (@2.1GB/sec per QDR IB link) between Exadata Database Servers and Media Servers
4b	Network Switch	QDR Infiniband Switch
5a	Media Server	2 Media Servers (Active/Active) QDR HCA : 4 ports @ 2.1GB/sec each = 8.4GB/sec front-end 8Gbps HBA : 4 ports @ 800MB/sec each = **3.2GB/sec** back-end
5b	Administrative Server	1 Administrative Server 1GigE port used to connect to Media Servers and Exadata Database Servers
5c	Fibre Channel Switch	1 x 24 ports FC switch
6a	Tape Drives	12 x LTO5 Drives @ 210MB/sec each(Compressed – 1:1.5 rate) Total throughput : **2.52GB/sec** Each drive is assigned to one RMAN channel
6b	Tape Media	12 x LTO5 Tapes for full backup (@3TB per Tape)
6c	Tape Library	Oracle StorageTek SL3000
7	RMAN Channels	RMAN Channel per node = 3 Total RMAN channels = 3 x 4 = 12 One RMAN channel per tape drive

From the above table, the backup requirement of 1.67GB/sec is met by the architecture capable of performing @2.52GB/sec using InfiniBand connectivity, 2 Media Servers, and 12 LTO5 tape drives loaded in one Oracle's StorageTek SL3000 tape library. Figure 1 shows the connectivity diagram for this architecture.

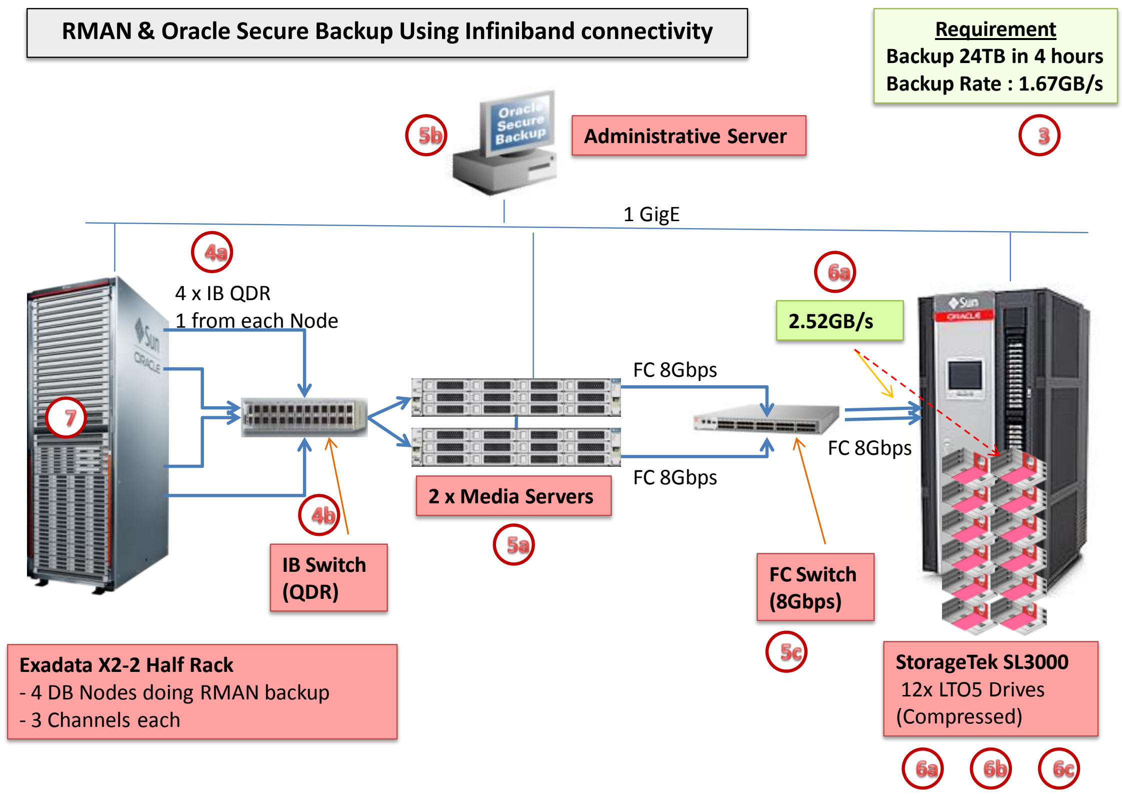

FIGURE 1 RMAN BACKUP USING INFINIBAND CONNECTIVITY AND ORACLE SECURE BACKUP

Exadata to Media Server Connectivity

Backup from Exadata is performed over InfiniBand QDR links to the Media Server via an InfiniBand switch. Each of the four database nodes is connected to the Media server over a single InfiniBand port. Each node is assigned 3 RMAN channels for a total of 12 RMAN channels.

Media Server and Tape Devices

Two Media Servers are configured for high availability in an active/active mode. Each server is configured with 2 IB QDR HCA and 4 x 8Gbps FC HBA ports. The ports are bonded for redundancy. The Media Servers are connected to the tape drives via a FC switch. Additional FC switches can be added for redundancy. Oracle's StorageTek SL3000 library with 12 active drives are assigned for the database backup. Compression is enabled at the tape drive level. To backup 24TB of data, 12 tape medias are needed per full backup. The backup software Oracle Secure Backup manages tape library, tape allocation, movement and vaulting.

Throughput Analysis

Between Exadata and Media Servers, based on 2.1GB/sec bandwidth per IB link, a potential rate of 8.4GB/sec can be achieved. The Media Server can write up to 3.2GB/sec using 4 ports at the rate of 800MB/sec per 8Gbps port. Media Servers write to 12 tape drives @ 2.52GB/sec with each tape drive performing @ 210MB./sec. Since the entire infrastructure performs at the rate of the slowest performance component, the overall data transfer is limited to 2.52GB/sec due to the number of tape drives (refer to line 6a in Table 5). In this configuration, that is acceptable since the required backup rate of 1.67GB/sec is achieved.

Meeting Requirements

The RTO of 4 hours is achieved by restoring 24TB of database @ 2.52GB/sec in 2.6 hrs followed by recovery. In most recovery conditions, only certain data files or tablespaces are restored and not the entire database. In those cases, the restore time is dramatically reduced.

Future Growth

If the customer wants throughput of more than 2.52GB/sec, then adding more drives will help until the 3.1GB/sec limitation on Media Server HBA ports is reached. To go beyond the 3.1GB/sec limitation, the number of HBA ports and the number of tape drives need to be increased.

Example Scenario 2: Gigabit Ethernet

Customer requirement:

Exadata X2-2 Half rack system (4 Database nodes, 7 Storage Servers)
Database size to backup : 8TB
RTO : 4 hrs, RPO : 30 minutes
Backup window of 4 hrs (2TB/hr)
Backup to tape using RMAN & OSB
Use 4 database nodes for backup
1GigE connectivity between Exadata Database Servers and Media Servers

This requirement is met by using the methodology defined in this document. The lower range of bandwidth numbers is used for a conservative calculation. Table 6 illustrates the throughput values for the requirements gathered and the hardware components chosen to meet the requirements.

TABLE 6 : SCENARIO 2 : BACKUP & RECOVERY SIZING USING MULTIPLE GIGE LINKS

Step #	Description	Value
1	RTO	4 hrs
2	RPO	30 minutes
3	Backup Window & Backup Rate	4 hrs to backup 6TB @ 2TB/hr = **560MB/sec**
4a	Exadata to Media Servers over 1GigE	2 x 1GigE ports per database node for 4 nodes Total for 8 x 1GigE ports = **800MB/sec** throughput between Exadata to Media Servers @ 100Mb/sec per port
4b	Network Switch	1GigE Network Switch
5a	Media Server	2 x Media Servers (HA) 1GigE : 8 x 1GigE ports @780MB/sec front-end 8Gbps HBA : 2 ports @ 800MB/sec each = **1.6GB/sec** back-end
5b	Administrative Server	1 Administrative Server 1GigE port to connect to Media Servers and Exadata Database Servers
5c	Fibre Channel Switch	1 x 24 Ports FC switch
6a	Tape Drives	8 x LTO5 Drives @ 140MB/sec (Native) Total tape drive throughput : **1.12GB/sec** Each drive is assigned to one RMAN channel
6b	Tape Media	8 x LTO5 Tapes per full backup (1.5TB per Tape)
6c	Tape Library	StorageTek SL500
7	RMAN Channels	RMAN Channel per node = 4 Total RMAN channels = 4x2 = 8 One RMAN channel per tape drive

From the above table, the backup requirement of 560MB/sec is met by the architecture capable of performing @1.12GB/sec using multiple GigE connectivity, 2 Media Servers, and 8 LTO5 tape drives loaded in StorageTek SL500 tape library. Figure 2 shows the connectivity diagram for this architecture.

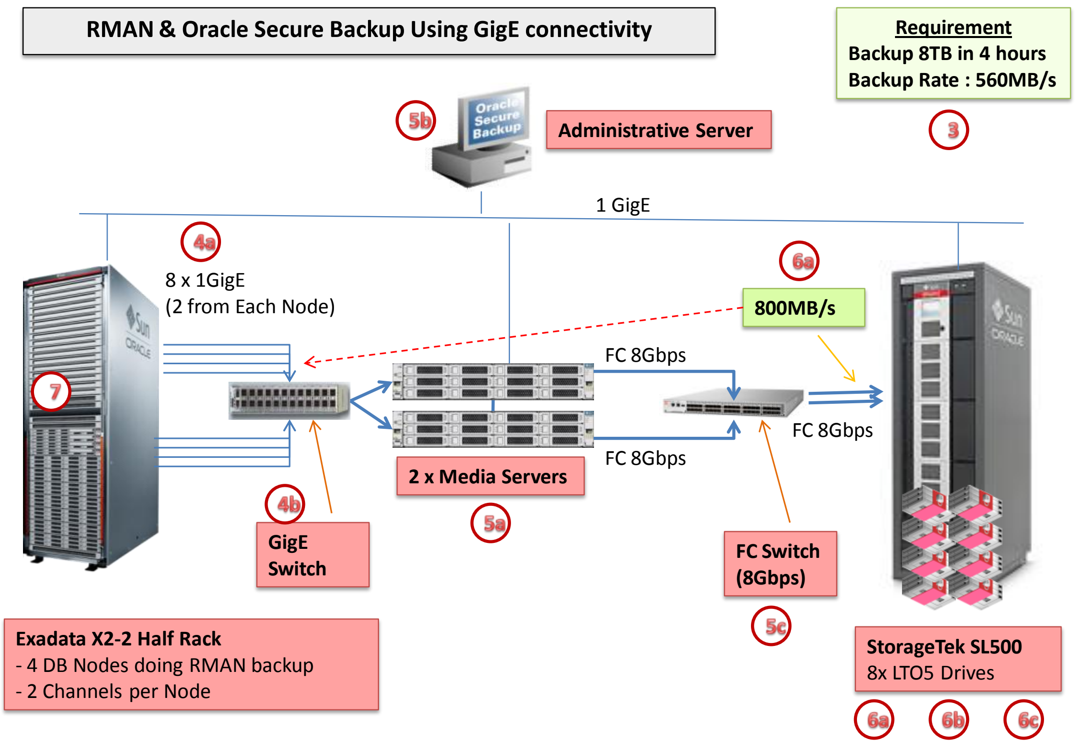

FIGURE 2 : RMAN/OSB BACKUP USING MULTIPLE 1GigE LINKS

Exadata to Media Server Connectivity

In this scenario, interconnect between Exadata and the Media Serves is established over 8x1GigE links via GigE switch. Each of the 4 database nodes has 4x1GigE ports, of which 2 GigE ports are connected to the media servers via a 1GigE switch.

Media Server and Tape Devices

Two Media Servers are configured with 8x 1GigE NIC ports and 2x8Gbps HBAs. The ports are bonded for redundancy. Oracle's StorageTek SL500 modular tape library is recommended for this configuration. 8 LTO5 drives are allocated for the backup operation for this database. No tape compression is enabled. Each LTO5 has the 1.5TB native capacity with a total of 12TB – which is sufficient to store a full database backup. Oracle Secure Backup software in conjunction with the tape library manages the tape allocation, movement and vaulting process.

Performance Analysis

In this architecture, data from Exadata is transferred at the rate of 800MB/sec to the media server over 8 x 1GigE links at the rate of ~100MB/sec. The media server can perform throughput of 1.6GB/sec using 2 x 8Gbps HBA ports at the rate of 800MB/sec per port. Each LTO5 tape drive can perform @ 140MB/sec of transfer with the total throughput of 1.12GB/sec.

Meeting Requirements

Since the whole infrastructure performs at the rate of the lowest performance component, the overall throughput is throttled to 800MB/sec due to the front-end limitation (refer to line 4a in Table 6). That is acceptable in this configuration as the required rate of 560MB/sec is achieved.

The RTO requirement of 4 hours is easily achieved by restoring 8TB database @ 800MB/sec in 2.8 hours followed by recovery.

Future Growth

If the throughput desired is more than 800MB/sec, then instead of 1GigE connectivity, 10GigE links can be deployed. That increases the throughput between Exadata and Media Servers. That may require changes to the number of HBA ports on the Media Servers, installing 10GigE ports on the Media Server and increasing the number of tape drives.

Conclusion

Properly sizing the backup & recovery infrastructure is critical to meet the RTO, RPO business requirements. In most cases, data backup is given more emphasis than the restore/recovery process. For a successful implementation, all the processes such as backup, restore, and recovery are to be considered and properly planned. It is also important to understand the limitation on all the hardware and software components. The examples in this document show the full database backup and restore process just to show the limits of the components. For typical customer deployments, it is recommended to do weekly full backups (e.g. on Sundays) followed by daily incremental backups. The time taken to perform incremental backups is usually lesser than full backups as it depends on the amount of data changed. Cost aspects for the solution are also taken into consideration when sizing the architecture.

Given all these parameters, this document can be used as a guideline for asking the right questions, choosing the right components, and architecting a right backup and recovery solution.

Reference

Oracle Maximum Availability Architecture Site

http://www.oracle.com/goto/maa

Backup and Recovery Performance and Best Practices for Exadata Database Machine - Oracle Database 11.2.0.2

http://www.oracle.com/technetwork/database/features/availability/maa-tech-wp-sundbm-backup-11202-183503.pdf

Recovery Manager Site

http://www.oracle.com/goto/rman

Recovery Manager Performance Tuning Best Practice

http://www.oracle.com/technetwork/database/focus-areas/availability/rman-perf-tuning-bp-452204.pdf

Oracle Secure Backup

http://www.oracle.com/goto/osb

StorageTek Tape Products

http://www.oracle.com/us/products/servers-storage/storage/tape-storage/index.html

Sun Server Networking Products

http://www.oracle.com/us/products/servers-storage/networking/index.html

Oracle is committed to developing practices and products that help protect the environment

Oracle Exadata Database Machine:
Backup & Recovery Sizing – Tape Backups
September 2011
Author: Sridhar Ranganathan
Contributing Authors: Ashish Ray, Donna Cooksey, Tim Chien

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

oracle.com

Hardware and Software, Engineered to Work Together