

Tape Backup Performance and
Best Practices for Exadata Storage
and the HP Oracle Database
Machine

*Oracle Maximum Availability Architecture White Paper
July 2009*

Maximum Availability Architecture

Oracle Best Practices For High Availability

ORACLE

Introduction	1
Background	2
Key Performance Observations	5
Tape Backup Best Practices	5
Test Environment and Methodology	7
Backup and Restore Architectures.....	8
Architecture for Back Up to Tape or Virtual Tape Libraries	
Using InfiniBand	8
Architecture for Back Up to Tape Using Gigabit Ethernet	12
Configuration Best Practices	14
RMAN Tape-Based Configurations	14
Backups from a Subset of Database Instances	15
System Configuration Changes	17
Configuring IB Backups.....	17
Configuring GigE Backups	22
Conclusion	24
Appendix A: The Impact of Tape Initialization on Effective	
Backup Rates	25

Introduction

The HP Oracle Exadata Storage Server is a storage product that is highly optimized for use with Oracle Database and speeds up I/O and SQL processing for data warehousing applications. You can build a data warehouse either by using Exadata Storage Servers and providing database servers and a storage network, or by deploying an *HP Oracle Database Machine*. The HP Oracle Database Machine is a fully integrated platform, including all of the components required to quickly and easily deploy a data warehouse in the enterprise.

The Oracle Database has very sophisticated and scalable backup technologies. These technologies work especially well combined with HP Oracle Exadata Storage Servers. On an HP Oracle Database Machine Full Rack we were able to attain backup to tape rates of 11.2 TB/hour for full backups, and over 100 TB/hour for incremental backups using only two tape media servers. Higher performance is possible using more media servers. The technologies that help to attain these rates include:

- Recovery Manager (RMAN) automatically parallelizes backup operations across all database nodes and Exadata storage cells. This allows all the disks, all the network connections, and all the CPUs in the system to contribute to performing backup operations.
- The InfiniBand network provides an extremely high performance network for transferring backup data from storage servers to database servers and then to tape media servers. Not only are the transfer rates high, the CPU utilization of InfiniBand network transfers is very low.

- The block change tracking feature of the Oracle Database allows incremental backups to run very quickly and efficiently. With block change tracking, only the areas of the database that have been modified need to be read from disk.
- Exadata storage has very highly optimized disk I/O capabilities. Each Exadata Storage Server can achieve a disk transfer rate of over 1000 MB/sec with just 12 disk drives.
- Exadata has offload capabilities that further speed up incremental backups. When changed blocks and unchanged blocks are near each other on disk, The Exadata offload capability combines with block change tracking to perform very efficient large I/Os at the storage level, while returning only individual changed blocks to the database level.

Overall, backup runs much faster in an Exadata environment than other database environments. In all our tests, the performance bottleneck was the tape media servers. The HP Oracle Database Machine ran at less than 25% of its maximum data throughput, leaving plenty of bandwidth for concurrent user workloads.

This paper provides:

- Architecture recommendations for performing backups to tape devices with Exadata Storage Server systems and the HP Oracle Database Machine
- Guidelines and configuration best practices to optimize tape backup rates

Background

The goal of this white paper is to provide best practices for scaling tape backups performed with RMAN against Exadata Storage Servers. This paper is not intended to be a benchmark for comparing different products for tape backups. To simplify testing and

improve repeatability, we did not measure the effects of a user workload running concurrently with the backup and restore operations. In the future, the MAA team plans to revise this white paper to include disk backup and restore best practices, and best practices around managing concurrent workloads during backup with the I/O Resource Manager.

You must use RMAN to perform backup operations with Exadata. Any tape backup product that integrates with RMAN is automatically supported for use with Exadata Storage Server systems. Tape backup rates are affected by different media management products, media servers, interconnects, and the tape devices used. (A media server is defined as a server attached to tape devices.) Nevertheless, the architectures recommended in this white paper—as well as the achieved throughput rates and best practices—provide guidelines that are applicable to the various solutions on the market. Oracle Secure Backup, a centralized tape backup management solution, was used in the testing. A separate and more detailed white paper about best practices for RMAN and Oracle Secure Backup with Exadata will be made available in the future.

Much of the testing and analysis documented in this paper was conducted on a HP Oracle Database Machine Full Rack. The results and recommendations are also applicable to Exadata warehouses that are custom-built with database servers and InfiniBand fabric provided by the customer.

This white paper contains the following sections:

- [Key Observations](#)—Provides key rates
- [Tape Backup Best Practices](#)—Describes best practices and recommendations
- [Test Environment and Methodology](#)—Describes the test environment
- [Backup and Restore Architectures](#)—Describes the high-level architectures using InfiniBand or Gigabit connections

- [Configuration Best Practices](#)—Recommends configuration settings to optimize performance
- [System Configuration Changes](#)—Describes steps to configure InfiniBand or Gigabit based tape backup configurations

The sections in this white paper do not need to be read in sequence. Readers who are interested only in configuring tape backups with Exadata can jump to the last three sections.

Key Performance Observations

The HP Oracle Database Machine with Exadata storage provides the performance and technology to quickly back up very large databases. The read rate per Exadata storage server (also referred to as an Exadata cell) is up to 1 GB/sec for SAS-based Exadata cells, and up to 750 MB/sec for SATA-based Exadata cells. The throughput scales linearly as Exadata cells are added to the configuration because when you add more Exadata cells, it not only increases the storage capacity, but it also increases I/O bandwidth. Thus, a configuration using two SAS-based cells doubles the raw capacity and increases the I/O bandwidth to as much as 2 GB/sec total. Additionally, the work of filtering out the modified data blocks during incremental backups is offloaded from the database server to the Exadata storage servers. This reduces CPU usage on the database instances and results in faster backups.

An HP Oracle Database Machine attains very high data transfer rates to a tape backup infrastructure over the InfiniBand network. With an HP Oracle Database Machine Full Rack using SAS drives and 2 media servers, an effective incremental backup rate of approximately 104 TB/hour was achieved by properly configuring the number of RMAN channels, tape drives, and database instances. A full backup rate of 11.2 TB/hour was measured on the same system. The rate can be scaled further by adding more media servers.

Note: The effective backup rate or effective incremental backup rate is defined as the $(\text{Total Database Size})/(\text{Backup Time})$. An incremental backup only backs up the changes from last cumulative or full backup. Backup time starts at the moment when data started streaming to tape and completes when the backup operation completes. The backup time includes ramp up, ramp down, and tape startup processing that is common to all tape and media servers. Excluding the initialization times, backup transfer rates approach theoretical network bandwidth. For full backups, the backup rate is constrained by the number of media servers and network bandwidth to the media servers. With incremental backups, the time to complete the backup reduces dramatically; hence the effective backup rate can increase by the same factor. Incremental backups ran so quickly on our 10 TB test database that tape startup processing times became a large percentage of the total backup time. Therefore, incremental backup rates will be even higher than the rates we report in this paper for larger databases where tape startup processing is a smaller percentage of the total time.

Tape Backup Best Practices

Oracle recommends the following best practices for performing backups to tape in a data warehouse environment:

- **Perform level 0 and differential backups**

- Create a level 0 (full) backup once a week
- Create a differential incremental backup daily

The effective backup rate varies depending upon the change differential since the last incremental backup. For example, using one media server, the effective incremental backup rate was 47 TB/hour (13.3 GB/sec) with a 10% change differential and was 24 TB/hour (6.8 GB/sec) with a 20% change differential.

- **Use Infiniband to connect from the HP Oracle Database Machine to the media servers for best performance**

With the available InfiniBand ports in an HP Oracle Database Machine, media servers can be directly connected to the InfiniBand fabric by adding an InfiniBand double data rate host channel adapter (DDR HCA) to the media server. For high availability, connect the HCA to two HP Oracle Database Machine Infiniband switches to eliminate the switch as a single point of failure. This provides transparent failover if connectivity is lost to one of the ports.

There are seven bonded InfiniBand ports available in an HP Oracle Database Machine Full Rack and nine bonded InfiniBand ports in an HP Oracle Database Machine Half Rack. This is more than enough connectivity to achieve extremely high backup rates.

- **Add more media servers to achieve higher aggregate backup rates**

Backup rates are limited by the throughput of the media server and the connection between the HP Oracle Database Machine and the media server. The practical data throughput rate of a single InfiniBand DDR HCA is about 1.6 GB/sec. You can achieve higher aggregate rates for your backup infrastructure by adding more media servers, and consequently by adding more Infiniband DDR HCAs and connections.

- **Configure one RMAN channel per tape drive**

A single tape drive, such as the HP LTO3 or LTO4, can deliver from 160 to 240 MB/sec of read/write bandwidth per drive, for compressed data. Backup performance scales when you add more tape drives and RMAN channels, assuming available throughput on the media server. Configure one RMAN channel per tape drive. A single RMAN channel in an HP Oracle Database Machine can stream data at a throughput rate of 400 MB/sec or greater. Therefore, the performance limit on a single channel is likely to be the tape drive or media server, not the HP Oracle Database Machine or RMAN.

- **Tune tape initialization**

As with any tape backup software and tape devices, RMAN must initialize each channel with the media server. Initializing the RMAN channels may take from seconds to minutes depending on the number of RMAN channels since each channel must be initialized serially. For example, we observed a 4-minute initialization time with 24 channels when using Oracle Secure Backup. In fact, because the backup rates are so high on the HP Oracle Database

Machine, for smaller backup sets, we observed that the data being backed up by the first channels completed even before the last channels were allocated. The initialization times become negligible when the backup set is greater than a terabyte in size.

Follow your vendor's recommendations for tuning tape initialization.

See Also: [Appendix A](#) for examples showing the impact of tape initialization to the effective backup rates.

Test Environment and Methodology

The test environment consists of the following:

- **Database Tier**— HP Oracle Database Machine Full Rack consisting of 8 x HP DL360 G5 database servers and 14 x HP/Oracle DL180 G5 Exadata Cells. Database size was 10 TB before mirroring.
- **Media Server Tier**—2 x HP DL360 G5 servers with one Infiniband DDR HCA per server.
- **Software**— Oracle Database 11g Release 1 (11.1.0.7 plus the recommended database patches, as described in My Oracle Support (formerly Oracle*Metalink*) [Note 791275.1](#).
- **Oracle Secure Backup Release 10.2.0.3** was used as the media manager for the tape backup performance tests. Instead of using physical tapes, virtual tapes were configured to backup to `/dev/null`. This was done because the goal of this paper is to evaluate database and media server performance and scalability rather than tape drive performance.

Backup and Restore Architectures

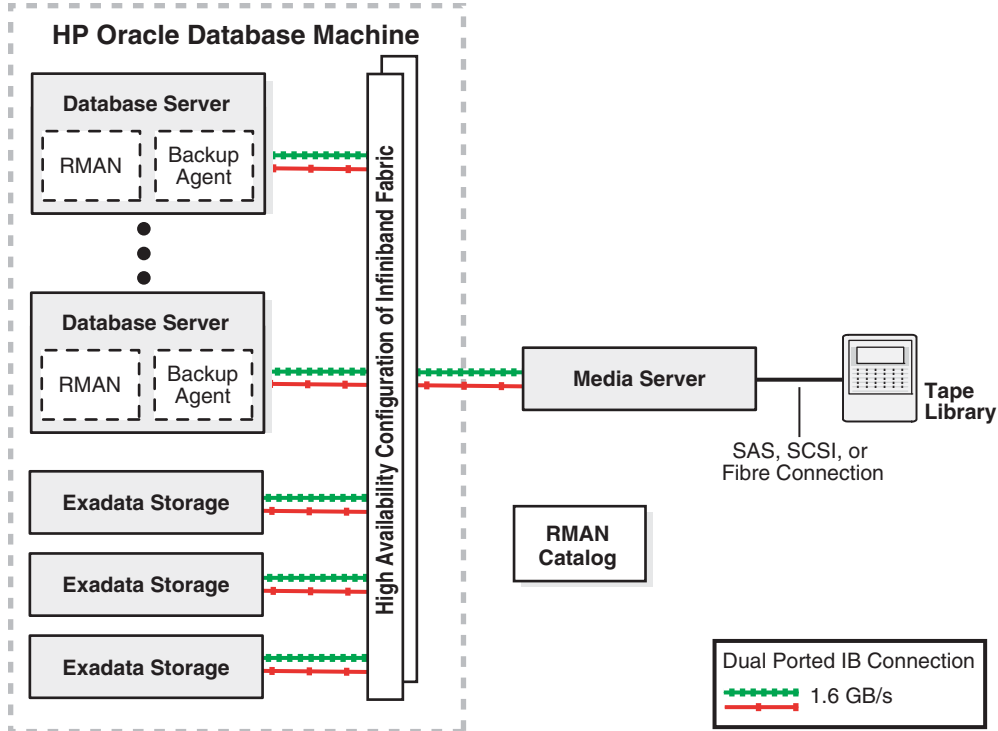
This section recommends architectures for backing up Exadata Storage Server systems. It describes the following topics:

- [Recommended architecture for backing up to tape or virtual tape libraries using InfiniBand connection](#)
- [Secondary architecture for backing up to tape using Gigabit connection](#)

Architecture for Back Up to Tape or Virtual Tape Libraries Using InfiniBand

The diagram in Figure 1 represents a single media server configuration for customers that require a data transfer rate of up to 5.6 TB/hour (1.6 GB/sec). The effective backup rate depends on the type of RMAN backup. For a full backup, the backup rate is 5.6 TB/hour (1.6 GB/sec). For a differential or cumulative incremental backup where only 10% of the database has changed, the effective incremental backup rate was measured to be 47 TB/hour (13 GB/sec) and where 20% of the database has changed the effective incremental backup rate was 24 TB/hour (6.8 GB/sec).

Figure 1: Small-to-Medium Size Backup Configuration to a Single Media Server



In Figure 1, the media server is connected directly by two Infiniband links to the existing highly available InfiniBand fabric architecture of the HP Oracle Database Machine full rack, requiring the media server to have a dual ported Infiniband DDR HCA. The highly available Infiniband fabric architecture configures a minimum of two switches for high availability purposes. Adding more hardware components to an HP Oracle Database Machine (such as Fibre Channel cards) to attach tape drives locally, is not supported.

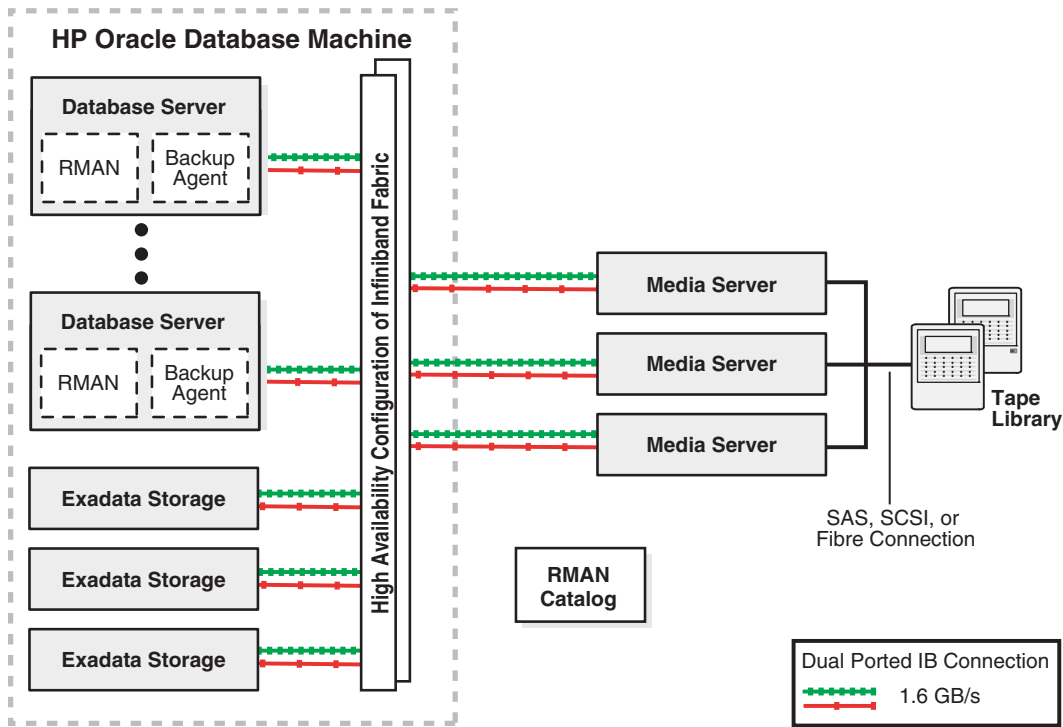
Oracle recommends using an external RMAN recovery catalog repository. See the [Oracle Database Backup and Recovery User's Guide](#) for more information about the RMAN repository.

Figure 2 represents a tape backup solution with two or more media servers providing an actual data transfer rate of up to 5.6 TB/hour (1.6 GB/sec) per media server. To connect media servers directly to the existing InfiniBand fabric, you can use:

- Nine available ports for HP Oracle Database Machine Half Racks
- Seven available ports for HP Oracle Database Machine Full Racks

Each media server requires an Infiniband DDR HCA or the recommended dual-ported Infiniband DDR HCA. Infiniband provides excellent connectivity between Exadata and the media server. Infiniband delivers both very high bandwidth and very low CPU utilization, even at Gigabyte data rates. The network protocol used for backups over Infiniband is the standard TCP/IP protocol, so it is transparent to the backup software on the database servers and the media servers. The backup software operates identically to the way it operates over a Gigabit Ethernet network.

Figure 2: Larger Backup Configuration with Multiple Media Servers

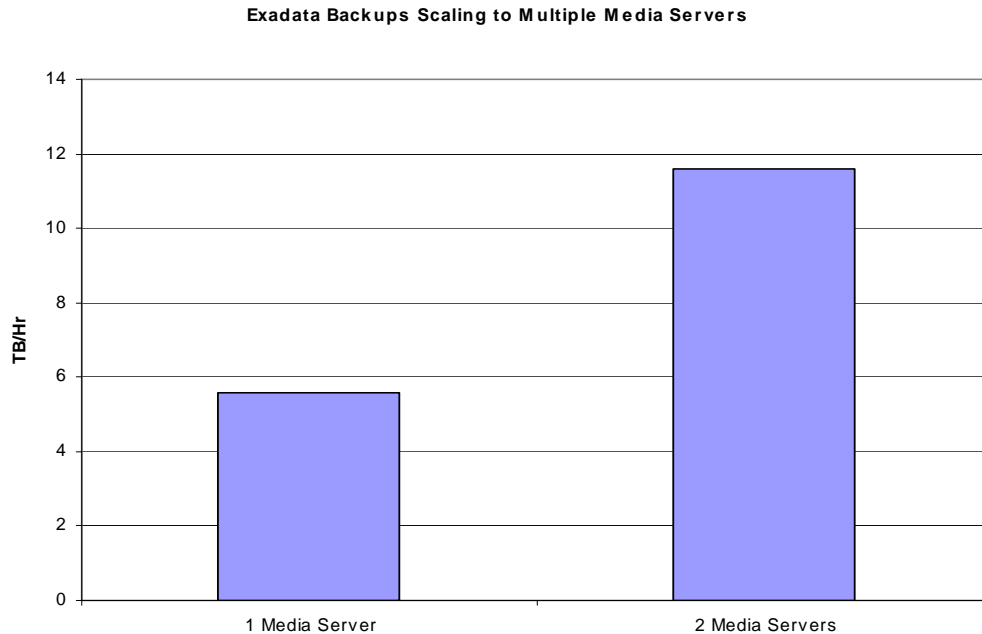


The three Media Server configuration in Figure 2 has an effective maximum data transfer rate of 4.8 GB/sec, or 16.8 TB/hour.

Allow a sufficient number of tape drives so that a media server can achieve its maximum backup and restore rates. For example, if a tape drive backup rate is 240 MB/sec of compressed data¹, you need at least 7 tape drives to achieve the maximum data transfer rate of the media server of approximately 1.6 GB/sec, limited by the bandwidth of a single HCA.

Figure 3 shows the scaling ability of media servers. In both cases represented in the graph, the same full database backup was completed but the number of media servers was doubled. The scaling factor achieved was approximately 100%, with the backup rate going from 5.6 TB/hour to 11.2 TB/hour.

¹ An LTO4 Tape Drive is capable of writing approx 240 MB/s compressed data to tape, while an LTO3 Tape Drive is capable of writing approx 160 MB/s of compressed data to tape.

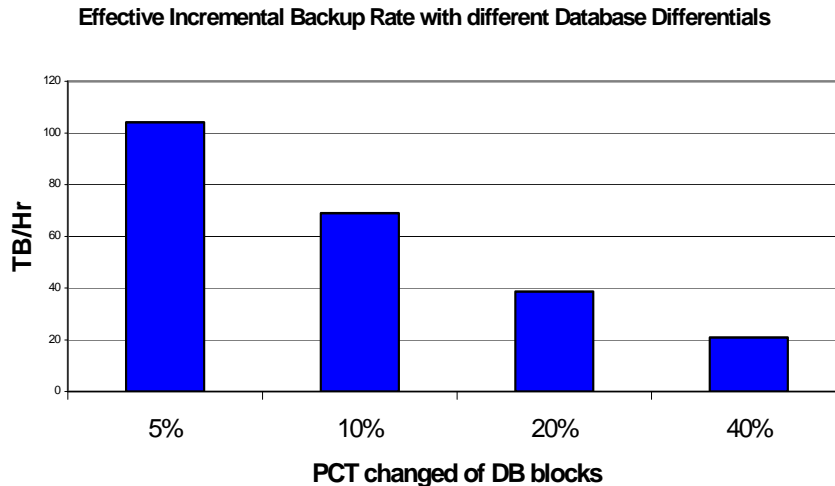
Figure 3: Exadata Backups Scaling to Multiple Media Servers

Oracle further optimizes incremental backups by:

- Using incremental backups, which:
 - Reduced system utilization in all cases
 - Reduced backup times in cases where the percent of database blocks changed between incremental backups was less than 40%
- Using a [block change tracking file](#) to identify changed blocks for incremental backups. By reading the small bitmap file to determine which blocks changed, RMAN avoids having to scan every block in the data file that it is backing up. With Exadata, more data block inspection is offloaded from the database server.

Figure 4 shows the effect of the database differential backup on the backup rate.

Figure 4: Database Differential Impact on Effective Backup Rate



Architecture for Back Up to Tape Using Gigabit Ethernet

Customers have asked how to use their existing Gigabit Ethernet infrastructure to back up the HP Oracle Database Machine. This architecture in Figure 5 is configured similarly to the Infiniband architecture. As you might expect, the maximum network bandwidth between the HP Oracle Database Machine and the external media servers can become a bottleneck. Also, Gigabit Ethernet consumes much more CPU to transfer data than using Infiniband.

The maximum transfer rate over Gigabit Ethernet (GigE) is 120 MB/sec per Ethernet link and with a dual-ported Gigabit Ethernet connection, as described Figure 5, the maximum transfer rate is 240 MB/sec. A full database backup rate of 0.4 TB/hour (120 MB/sec) per Ethernet link is achievable. An effective incremental backup rate where 5% of the database has changed would be 8.2 TB/hour (2.3 GB/sec) per Ethernet link.

To increase the speed of the backup, use either multiple Ethernet ports per media server or multiple media servers. Exadata backup rates can scale linearly as you increase from a GigE single link to a GigE bonded link (total of 2 GigE links) and then to 2 GigE bonded links (for a total of 4 GigE links). As shown in Figure 6, the backup rates increased from 0.4 TB/hr to 0.8 TB/hr, and then to 1.5 TB/hr respectively. There are up to 16 Gigabit Ethernet links in each HP Oracle Database Machine. Therefore, the maximum theoretical full backup rate using Gigabit Ethernet is 1920 MB/sec or 6.6 TB/hour. The effective maximum backup rate will be lower than this because running backups at this rate leaves no network bandwidth for the application.

Note that the maximum potential rate of backup using the 16 Gigabit Ethernet links is approximately the same as a single Infiniband link. Therefore, Infiniband is generally preferred for large databases or to perform very fast backups and restores. Nevertheless, a full backup rate of at least a Terabyte per hour is achievable using Gigabit Ethernet, and this is sufficient for many databases, especially those that are single-digit terabytes or low double-digit terabytes in size. A staggered full backup schedule—where a full backup of different tablespaces is performed on different nights of the week—can be implemented to spread the load and time of a full backup.

Figure 5: Exadata Backups to Tape Using Gigabit Ethernet Connection

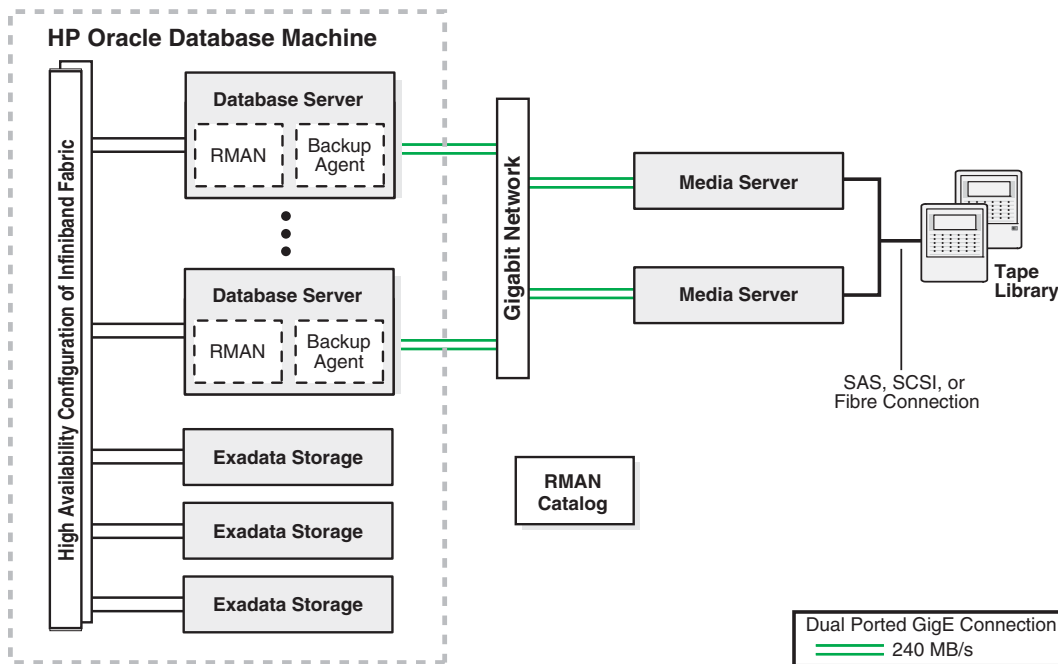
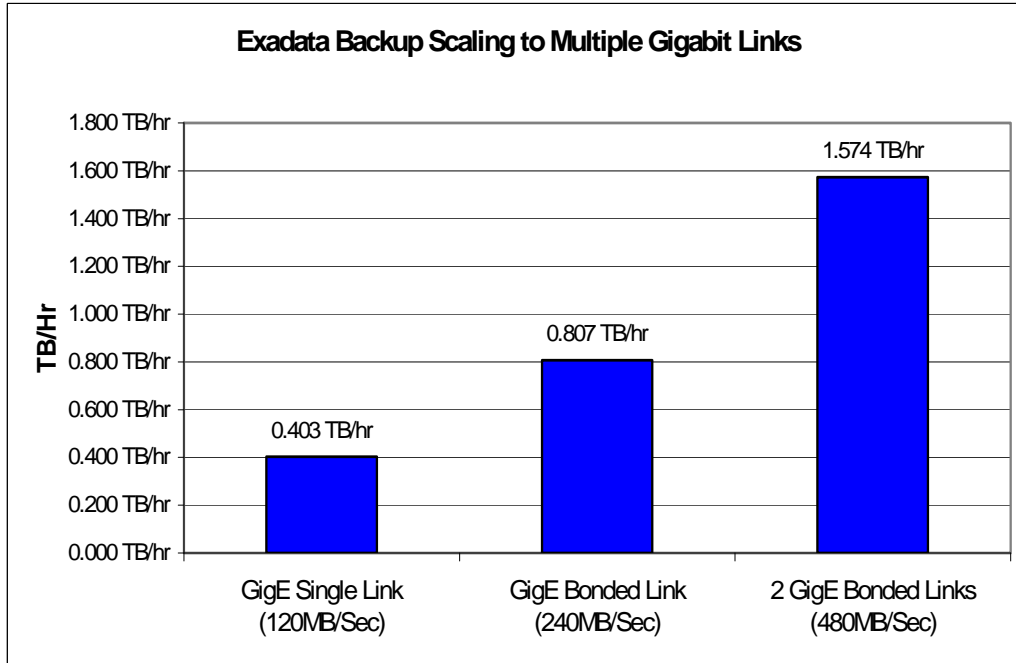


Figure 6 shows Exadata Backup Scaling across multiple Gigabit Links bonded together. In our testing, the two GigE Bonded Links was achieved over multiple Media Servers because the test environment could not support four GigE Links on a single system, but a similar throughput should be achievable if four GigE Links in a single host are bonded together.

Figure 6: Exadata Backup Scaling to Multiple Gigabit Links



Configuration Best Practices

The following sections document the best practices for backing up to tape from an HP Oracle Database Machine and expand on the information in the [“Key Observations”](#) section.

RMAN Tape-Based Configurations

Use the following best practices for tape-based backups:

- Configure one RMAN channel per tape drive for tape backups.
- Scale backup rates by adding tape drives.

Typical tape drive backup rates are between 100 MB/sec and 240 MB/sec, depending on the drive type and compression options. Note that tape drive compression becomes less effective when backing up tables that are compressed at the database level.

- Allocate channels equally across all database nodes to distribute system utilization. RMAN works with SQL*Net service load balancing to distribute channels load evenly among all the instances offering the service at the time of the backup. See the [Example](#) that follows this list.
- Use RMAN incremental backups to reduce backup time and backup space:

- If the daily incremental change to the database is less than 20% of the database size, then enable Block Change Tracking on the database. You may still benefit by using Block Change Tracking with higher percentages (> 20%) but testing is recommended to ensure that backup times are reduced.
- For database recovery using these backups, RMAN first restores the most recent full backup. Recovery then applies a cumulative incremental backup. Then, Redo Apply proceeds automatically.

Example

1. Configure an Oracle Service to run against all nodes in the cluster. The service is used by the RMAN Backup command. RMAN automatically spreads the backup load evenly among all the instances offering the service. For example:

```
$ srvctl add service -d <db_unique_name> \  
-s <service_name> \  
-r <list of preferred instances>  
$ srvctl start service -d <db_unique_name> \  
-s <service_name>
```

2. Connect to RMAN using the service name:

```
$ rman target sys/<sys_password>@<service_name>
```

3. Configure the default device type to be SBT (for tape only):

```
rman> CONFIGURE DEFAULT DEVICE TYPE TO SBT;
```

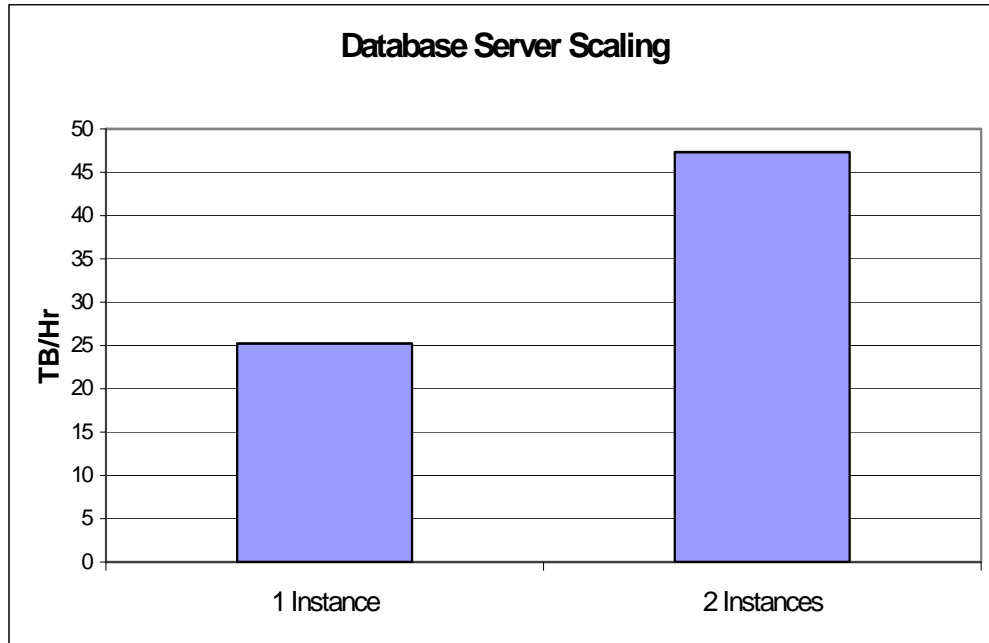
4. Configure number of channels to the number of tape drives. The following example assumes eight tape drives:

```
rman> CONFIGURE DEVICE TYPE SBT PARALLELISM 8;
```

Backups from a Subset of Database Instances

Instead of the simplicity of using all instances for backup and restore operations, some customers select a subset of available database instances on which to perform backup and restore operations. The primary reason to use only a subset of database instances is to lower the cost of the backup software in cases where the cost is based on the number of CPUs or on node access. Oracle Secure Backup, which is licensed per tape drive with no charge for client machines, does not have this issue.

Figure 7 illustrates how effectively incremental backup rates can scale by adding instances and RMAN channels. The graph shows a 10% change differential in database server scaling during an incremental backup.

Figure 7: Backup Rates When Adding Instances and RMAN Channels

To configure RMAN to run on a limited number of nodes, specify the instances when you define the service to be used by the RMAN Backup command.

For example:

```
$ srvctl add service -d <db_unique_name> \  
-s <service_name> \  
-r <list of preferred instances>  
$ srvctl start service -d <db_unique_name> \  
-s <service_name>
```

System Configuration Changes

The following sections document the best practices for database server and media server system configurations:

- [Configuring IB Backups](#)
 - [Database Server IB Configurations](#)
 - [Media Server IB Configurations](#)
- [Configuring GigE Backups](#)
 - [Ethernet Network System Configuration](#)
 - [Gigabit Ethernet Switch Configuration](#)
 - [Database Server Gigabit Configuration](#)
 - [Media Server Gigabit Configuration](#)

Configuring IB Backups

Database Server IB Configurations

No changes are required to the Database Servers of the HP Oracle Database Machine running Exadata 11g Release 1 Patch 1 (11.1.3.2.0) and later. In a custom configuration running Oracle Enterprise Linux v5.1 (or later) or RedHat Enterprise Linux v5.1 (or later), make the following configuration changes to the database servers, if not already present. If your database server is running a different operating system, contact your vendor for the appropriate InfiniBand configuration.

InfiniBand IPoIB Connected Mode

TCP/IP over InfiniBand (IPoIB) is a requirement for backup and recovery in an Exadata environment. The IP protocol is started by default in “DataGram” mode. This option is for generic compatibility between InfiniBand solutions. In a Linux environment, you should configure the database servers for “Connected” mode for faster performance. For example:

1. Edit the `/etc/ofed/openib.conf` file and search for `SET_IPOIB_CM` and change to “yes”:

```
# Enable IPoIB Connected Mode
SET_IPOIB_CM=yes
```

2. Verify that Connected Mode is enabled on the system, as follows:

```
# cat /sys/class/net/ib0/mode
connected
# cat /sys/class/net/ib1/mode
connected
```

If the returned value is not “connected” then Connected Mode has not been enabled on the system. Verify the configuration file and reboot the system.

Configure MTU Size on IB Interfaces

To speed up data transmission, increase the MTU size to 65520 (64K–Packet Overhead).

1. Edit the `/etc/sysconfig/network-scripts/ifcfg-ib*` and the `/etc/sysconfig/network-scripts/ifcfg-bond0` files to add an entry for “MTU=65520”. For example:

```
MTU=65520
```

2. Verify that the MTU size is 65520:

```
# ifconfig ib0 | grep MTU
UP BROADCAST RUNNING SLAVE MULTICAST MTU:65520 Metric:1
# ifconfig ib1 | grep MTU
UP BROADCAST RUNNING SLAVE MULTICAST MTU:65520 Metric:1
# ifconfig bond0 | grep MTU
UP BROADCAST RUNNING MASTER MULTICAST MTU:65520 Metric:1
```

If the output does not show the MTU size as 65520, then verify the configuration files and reboot the system.

Media Server IB Configurations

The following recommendations are only applicable for media servers running Oracle Enterprise Linux v5.1 (or later) or RedHat Enterprise Linux version 5.1 (or later). If your specific media server is running a different operating system, contact your vendor for the appropriate InfiniBand configuration.

Infiniband Network System Configuration

Within the existing highly available InfiniBand fabric architecture of the HP Oracle Database Machine there are available Infiniband ports for the media servers to connect to. In an HP Oracle Database Machine Full Rack, there are seven highly available Infiniband ports available, and in an HP Oracle Database Machine Half Rack, there are nine highly available Infiniband ports available.

Connect the two redundant Infiniband cables from each media server to free ports on two different Infiniband switches. It does not matter which switches are used, provided the cables are connected to different switches.

Bonding the InfiniBand Interfaces

Take the following steps to configure bonding of `ib0` and `ib1` into a bonded device `bond1`:

1. Modify the `/etc/modprobe.conf` file to add the following two lines to the bottom of the file. This adds another bonding alias and options.

```
alias bond1 bonding
options bonding max_bonds=2
```

The file will appear similar to the following example. This file assumes bonding was previously established on `bond0`:

```
alias eth0 tg3
alias scsi_hostadapter cciss
alias scsi_hostadapter1 ata_piix
alias scsi_hostadapter2 usb-storage
alias ib0 ib_ipoib
alias ib1 ib_ipoib
alias bond0 bonding
alias bond1 bonding
options bonding max_bonds=2
```

2. Create the `/etc/sysconfig/network-scripts/ifcfg-bond1` file as follows. Ensure that you include the two comment lines at the top of the file.

```
DEVICE=bond1
USERCTL=no
BOOTPROTO=none
ONBOOT=yes
IPADDR=<IP Address for bond1>
NETMASK=<Netmask>
NETWORK=<Network calculated using ipcalc-n ip_address netmask>
GATEWAY=<Gateway IP address>
BONDING_OPTS="mode=active-backup miimon=100 downdelay=5000
              updelay=5000"
IPV6INIT=no
```

3. Make copies of the current `ib0` and `ib1` configuration files. Ensure the copied files do not start with `ifcfg-ib0`. Prefix the file name with `backup-` or a similar word, and do not add a suffix such as `-backup`. For example:

```
cd /etc/sysconfig/network-scripts/  
cp ifcfg-ib0 backup-ifcfg-ib0  
cp ifcfg-ib1 backup-ifcfg-ib1
```

4. Modify the current `ib0` and `ib1` configuration files so they are configured to act as slaves to the `bond1` interface. The files should appear as follows:

* File `ifcfg-ib0`:

```
DEVICE=ib0  
USERCTL=no  
ONBOOT=yes  
MASTER=bond1  
SLAVE=yes  
HOTPLUG=no  
BOOTPROTO=none  
MTU=65520
```

* File `ifcfg-ib1`:

```
DEVICE=ib1  
USERCTL=no  
ONBOOT=yes  
MASTER=bond1  
SLAVE=yes  
HOTPLUG=no  
BOOTPROTO=none  
MTU=65520
```

5. Restart the system.
6. Log in as the `root` user after the system restarts to verify that NIC bonding is running correctly.

```
# cat /proc/net/bonding/bond1
```

```
Ethernet Channel Bonding Driver: v3.2.1 (October 15, 2007)  
Bonding Mode: fault-tolerance (active-backup)  
Primary Slave: None
```

```
Currently Active Slave: ib0
MII Status: up
MII Polling Interval (ms): 100
Up Delay (ms): 5000
Down Delay (ms): 5000

Slave Interface: ib0
MII Status: up
Link Failure Count: 1
Permanent HW addr: 80:00:00:48:fe:80

Slave Interface: ib1
MII Status: up
Link Failure Count: 1
Permanent HW addr: 80:00:00:49:fe:80
```

OpenFabrics Enterprise Distribution

You must use an OpenFabrics Enterprise Distribution (OFED) version compatible with the version found in the HP Oracle Database Machine in the media server. You can download the OFED from Oracle Support.

InfiniBand IPoIB Connected Mode

Configure the media servers for “Connected” mode. For example:

1. Edit the `/etc/ofed/openib.conf` file, search for the `SET_IPOIB_CM` parameter, and set it to “yes”:

```
# Enable IPoIB Connected Mode
SET_IPOIB_CM=yes
```

2. Verify that Connected Mode is enabled on the system:

```
# cat /sys/class/net/ib0/mode
connected
# cat /sys/class/net/ib1/mode
connected
```

If the returned value is not “Connected” then Connected Mode has not been enabled on the system. Verify the configuration file and reboot the system.

Configure MTU Size on IB Interfaces

To speed up data transmission, increase the MTU size to 65520 (64K–Packet Overhead).

1. Edit the `/etc/sysconfig/network-scripts/ifcfg-ib*` and the `/etc/sysconfig/network-scripts/ifcfg-bond0` files to add an entry for “MTU=65520”. For example:

```
MTU=65520
```

2. Verify that the MTU size is 65520:

```
# ifconfig ib0 | grep MTU
    UP BROADCAST RUNNING SLAVE MULTICAST MTU:65520 Metric:1
# ifconfig ib1 | grep MTU
    UP BROADCAST RUNNING SLAVE MULTICAST MTU:65520 Metric:1
# ifconfig bond0 | grep MTU
    UP BROADCAST RUNNING MASTER MULTICAST MTU:65520 Metric:1
```

If the output does not show the MTU size as 65520, then verify the configuration files and reboot the system.

Configuring GigE Backups

Ethernet Network System Configuration

When connecting the media servers to the HP Oracle Database Machine via Ethernet, connect the `eth1` interfaces from each Database Server directly into the data center network. For high availability, the two network interfaces on the Database Servers as well as the two network interfaces on the media server may be bonded together. In this configuration, configure the `eth1` interface as the preferred or primary interface and configure `eth0` as the redundant interface.

If throughput is a concern, then connect both `eth0` and `eth1` interfaces from each Database Server directly into the data center’s redundant network. The two interfaces can then be bonded together in a redundant and aggregated way to provide increased throughput and redundancy.

Gigabit Ethernet Switch Configuration

For optimal throughput and availability, configure hardware Link Aggregation in the gigabit switch. Link Aggregation Control Protocol² (LACP) is defined as part of IEEE 802.1AX-2008. Other software enabled bonding options are available within the operating system of the Database Servers and media server, which may also be used.

² See http://en.wikipedia.org/wiki/Link_aggregation.

If LACP is to be used, ensure that LACP is supported and configured on the Ethernet switch for `Src XOR Dst TCP/UDP Port`. See your vendor's Gigabit switch documentation for information about configuring source and destination port load balancing.

On a Cisco 4948 switch, use the following commands to implement `Src XOR Dst TCP/UDP Port`:

```
swi-2 (config)#port-channel load-balance src-dst-port
swi-2#wr mem
swi-2#sh etherchannel load-balance
EtherChannel Load-Balancing Operational State (src-dst-port):
Non-IP: Source XOR Destination MAC address
  IPv4: Source XOR Destination TCP/UDP (layer-4) port number
  IPv6: Source XOR Destination IP address
swi-2#
```

Additionally, if LACP is to be used, when configuring the `ifcfg-bond1` file, change the `BONDING_OPTS` setting to `mode=4`.

Database Server Gigabit Ethernet Configurations

No specific changes need to be made to the Database Servers, but to obtain higher backup rates, create a Dual Ports Gigabit Ethernet Configuration. See the *Oracle Exadata Storage Server Software User's Guide* about Bonding of `eth0` and `eth1` on Database Server Nodes (`database nodes`) in HP Oracle Database Machine.

If LACP is to be used, when configuring the `ifcfg-bond1` file, remember to change the `BONDING_OPTS` parameter setting to be `mode=4`.

Media Server Gigabit Ethernet Configurations

The following recommendations are only applicable for media servers running Oracle Enterprise Linux v5.1 (or later) or RedHat Enterprise Linux v5.1 (or later). If your specific media server is running a different operating system, contact your vendor for the appropriate Gigabit configuration.

As with the Database Server Gigabit Ethernet Configuration, no specific changes must be made to the media servers, but to obtain higher backup rates, create a Multiple Ported Gigabit Ethernet Configuration. The steps to configure bonding on the media server are the same as on the Database Servers. See the *Oracle Exadata Storage Server Software User's Guide* for a detailed procedure.

Conclusion

With the Exadata Storage Server, Oracle provides an architecture that allows customers with large databases to scale their tape backup to any desired performance level. The number and connectivity of media servers, and the number and speed of tape drives will define the performance limit of backup, not the Database Machine. With two media servers, effective full backup rates from 11.2 TB/hour and effective incremental backup rate of over 104 TB/hour were achieved.

Appendix A: The Impact of Tape Initialization on Effective Backup Rates

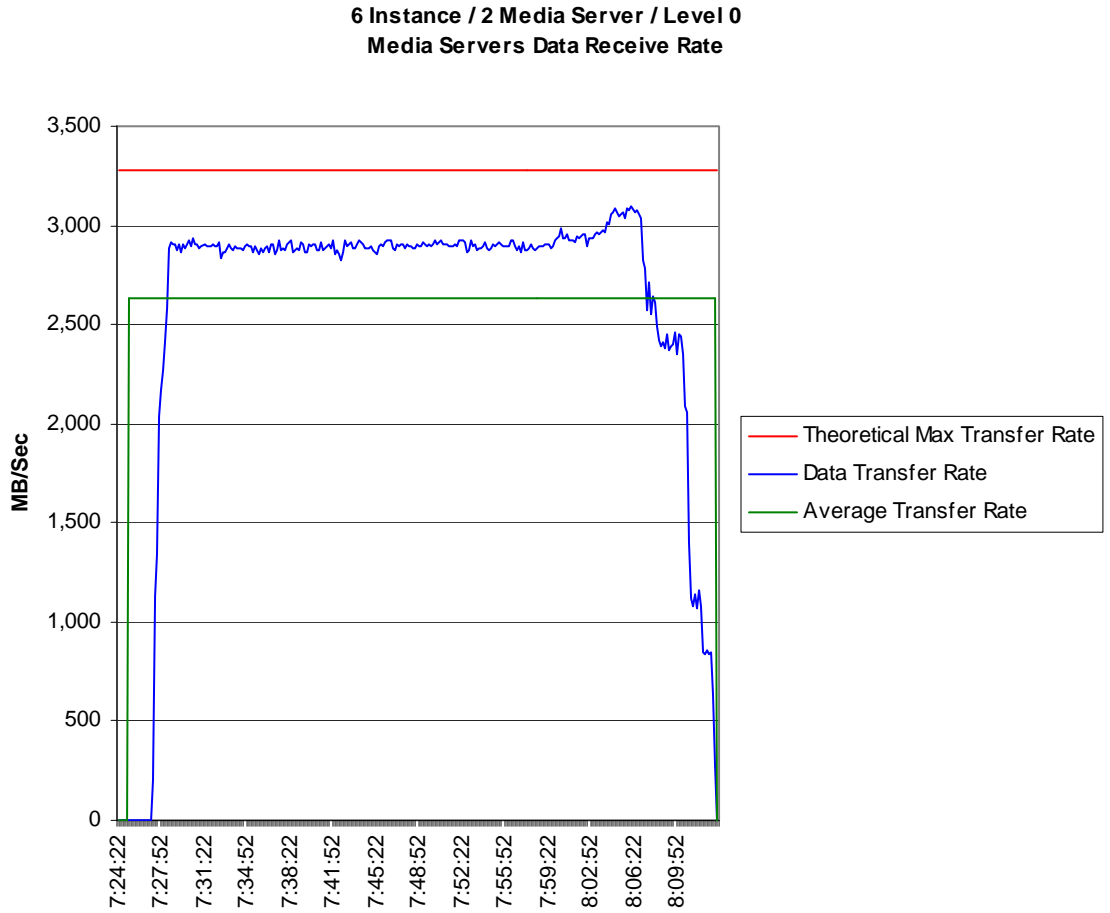
When starting a backup operation to media servers, the *tape initialization* occurs while RMAN establishes its relationship with the media server software, mounts the tape drives, and positions the tape for write I/Os. This process is serial so one RMAN channel needs to be initialized completely before moving to the next RMAN channel. The tape initialization varies with third-party tape backup software and tape devices, and can take seconds to minutes per RMAN channel or tape drive.

For a large backup that runs for one hour or more, the tape initialization cost is insignificant compared to overall backup time. Peak steady state is reached relatively quickly and we can scale effective backup rate with additional media servers. For smaller backups that use many RMAN channels, peak steady state may never be reached because RMAN channels are being started, loaded and eventually closed at different intervals.

Most customers are satisfied by either low backup duration for small backups or high effective backup rate for large backups.

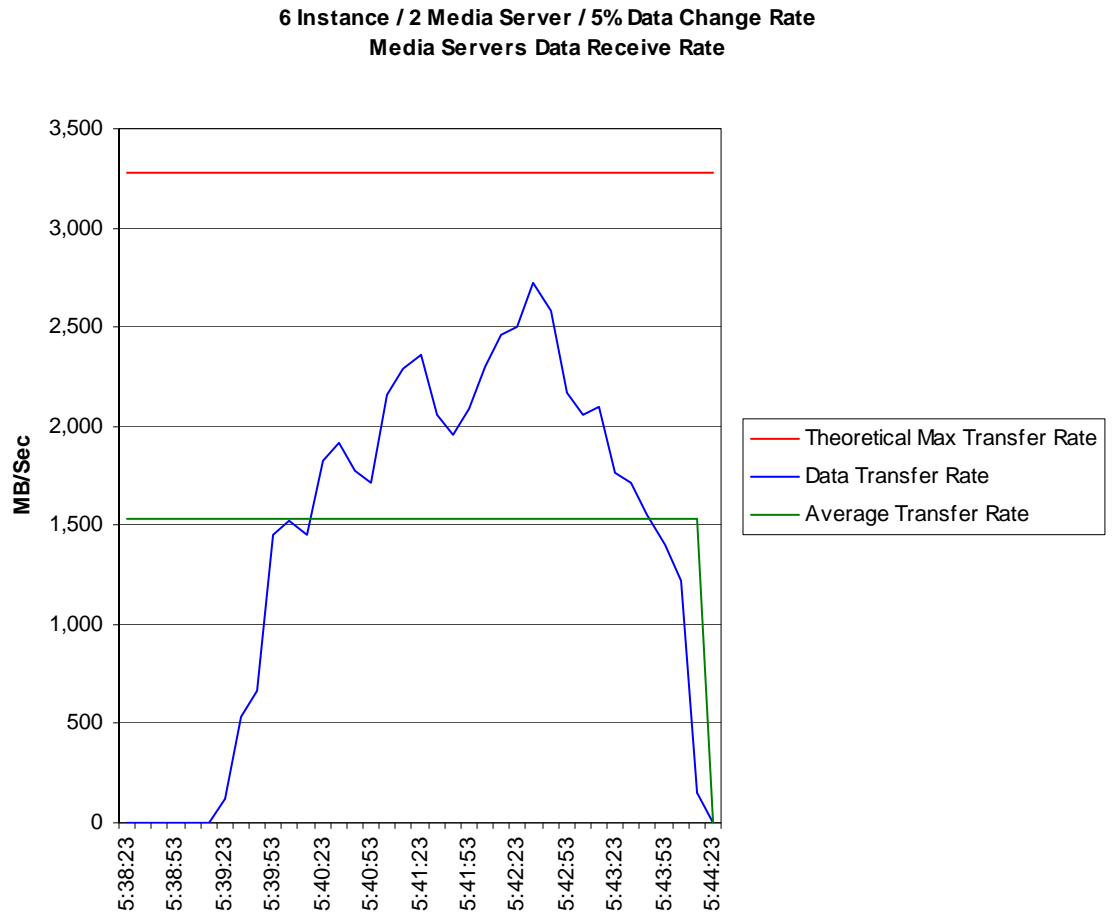
Figure 8 shows that the effective backup rate of 2.6 GB/sec is close to the theoretical maximum transfer rate of 3.2 GB/sec. If the actual backup size was greater, we observe that we get closer to theoretical maximum transfer rate. This full database backup completed in approximately 50 minutes.

Figure 8: Full Database Backup



When looking at the total backup duration in Figure 8, the initialization phase of the backup time is small compared to the total run time of 50 minutes. Steady state is reached very quickly. Compare this to backup example in Figure 9 for which the overall backup time is approximately 5 minutes. Due to the short backup duration, steady state was never reached because of the initialization costs of different RMAN channels and the extremely fast backup rates of each RMAN channel. In comparison, the effective backup rate looks poor in comparison to the maximum theoretical transfer rate.

Figure 9: Effective Incremental Backup with Only 5% Data Change





Tape Backup Performance and Best Practices
for Exadata Storage and the HP Oracle
Database Machine

July 2009

Author: Andrew Babb

Contributing Authors: Doug Utzig, Lawrence To,
Michael Nowak, Viv Schupmann

Reviewers: Herman Baer, Tim Chien, Donna
Cooksey, Juan Loaiza, George Lumpkin, Ron
Weiss, Steve Wertheimer

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

Worldwide Inquiries:
Phone: +1.650.506.7000
Fax: +1.650.506.7200



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

Copyright © 2009, Oracle and/or its affiliates. All rights reserved. This document is provided for information purposes only and the contents hereof are subject to change without notice. This document is not warranted to be error-free, nor subject to any other warranties or conditions, whether expressed orally or implied in law, including implied warranties and conditions of merchantability or fitness for a particular purpose. We specifically disclaim any liability with respect to this document and no contractual obligations are formed either directly or indirectly by this document. This document may not be reproduced or transmitted in any form or by any means, electronic or mechanical, for any purpose, without our prior written permission.

Oracle is a registered trademark of Oracle Corporation and/or its affiliates. Other names may be trademarks of their respective owners.