

Oracle 10gR2 Databases on Hitachi High-performance NAS Platform, powered by BlueArc®

Best Practices Guide for Oracle on Linux

April 2008



Executive Summary

The Hitachi High-performance NAS Platform, powered by BlueArc®, delivers a high-performance, intelligent and scalable multiprotocol solution. This document describes best practices for using Oracle databases on Linux with NAS connectivity provided by the High-performance NAS Platform through the NFS protocol. The document discusses baseline configuration recommendations for both Oracle and High-performance NAS Platform heads.

This guide covers advanced Oracle and Linux topics. The reader is assumed to have an understanding of these technologies. All best practices outlined are a product of rigorous real world testing by leading BlueArc engineers and Oracle experts. Following the guidelines outlined within this document will help to ensure that deployments are optimized, and will therefore reduce the need for advanced troubleshooting and tuning down the road.

Contents

Hitachi High-performance NAS Platform Overview	1
RAID Storage	1
System Drives	1
Storage Pools	1
File Systems	2
Virtual Volumes	2
Oracle Tuning Settings for Hitachi High-performance NAS Platform	2
Oracle Design Considerations	3
General Oracle Configuration Guidelines	4
Database File Types	4
Choosing Mount Points	4
Choosing Mount Points for Oracle Software Files	4
Directory-specific Guidelines	5
Choosing Mount Points for Oracle Database and Recovery Files	6
Optimal File Placement	7
Backup and Recovery	7
Using Snapshots for Backup and Restore	7
File System Restore	11
Using Snapshots to Create Test or Development Systems	11
Performance Tuning	12
Hitachi Data Systems Recommendations	14
NFS Mount Options	14
Hitachi Data Systems Recommendations	15
Linux Options	15
Appendix	17
Configuring SSH Public Key Authentication with High-performance NAS Platform	17
Server Administration through Public Networks	17
File System Rollback	18

Oracle 10gR2 Databases on Hitachi High-performance NAS Platform, Powered by BlueArc®

Best Practices Guide for Oracle on Linux

Hitachi High-performance NAS Platform Overview

Hitachi High-performance NAS Platform, powered by BlueArc®, abstracts storage to allow for flexible allocation and growth strategies. In the following sections, the vital components of this strategy are described along with Hitachi Data Systems recommendations to leverage these capabilities to their fullest.

RAID Storage

The High-performance NAS Platform head relies on Hitachi Data Systems storage systems for disk capacity, scalability and performance. High-performance NAS Platform supports RAID-1+0, RAID-5 and RAID-6 for data protection.

For Fibre Channel disk, with the unique write-buffering effect on the High-performance NAS Platform, Hitachi Data Systems recommends RAID-5 providing the best combination of cost, performance and capacity utilization. For high performance Oracle OLTP environment, Hitachi Data Systems recommends RAID-10 RAID Configuration.

Also, the use of Fibre Channel or SAS disks is greatly preferred over the use of SATA disks for an Oracle OLTP environment. Hitachi Data Systems only recommends use of SATA disk in RAID-6 which provides dual parity to protect against double disk failures in Oracle Decision Support System (DSS).

System Drives

LUNs presented from the backend SAN are seen as system drives by the High-performance NAS Platform head. The server supports multiple paths to the backend storage for path failover and high availability. Hitachi Data Systems recommends having a one-to-one mapping between system drives and RAID groups.

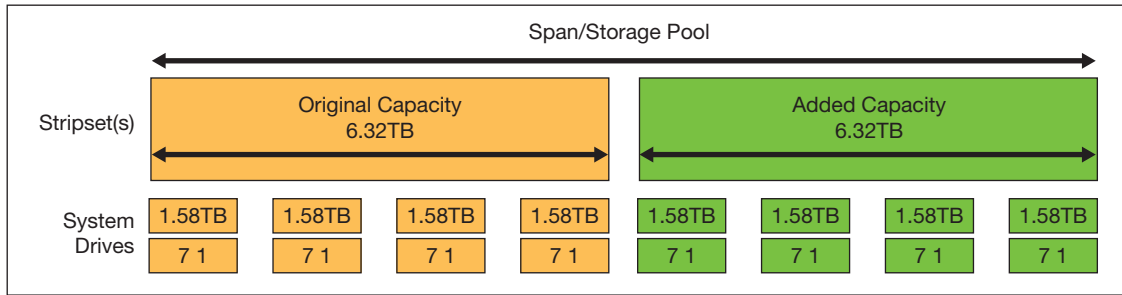
Storage Pools

Storage pools consist of one or more system drives. Data is striped evenly across all system drives in a storage pool. This allows the High-performance NAS Platform to leverage the aggregate throughput of many physical disks and provide optimal performance.

It is recommended that sufficient system drives with a sufficient number of physical disks capable of sustaining the required throughput of the databases are used in a storage pool.

A storage pool provides a logical container that can contain one or more file systems. Hitachi Data Systems supports dynamic expansion of storage pools and the file systems they contain. When expanding file system capacity, it may also be necessary to expand the storage pool itself. Hitachi Data Systems strongly recommends expanding the storage pools using the same number of system drives with the same characteristics. This will ensure consistent performance for all file systems residing within a storage pool (see Figure 1).

Figure 1. Adding Capacity to a Storage Pool



A storage pool provides a logical container that can contain one or more file systems.

File Systems

High-performance NAS Platform clusters support a maximum of 125 file systems. File systems are created within storage pools.

Each file system can be dynamically expanded and offers variable block sizes. High-performance NAS Platform provides two block size settings: 4KB and 32KB. In general, Hitachi Data Systems recommends using 4KB block sizes for Oracle databases.

Note: The block size of a file system cannot be changed after its creation. If the block size needs to be changed, the data should be migrated to a new file system with the desired block size.

Virtual Volumes

High-performance NAS Platform provides a feature called virtual volumes, which allows administrators to logically group directories within a base file system and apply independent properties such as quotas and replication to the group. The capacities of these virtual volumes can be increased or decreased dynamically.

Hitachi Data Systems recommends using virtual volumes to consolidate and ease management of many independent databases.

Note: Snapshots are created at the file system level, not at the virtual volume level.

Oracle Tuning Settings for Hitachi High-performance NAS Platform

There are several settings that can be tuned on the Hitachi High-performance NAS Platform server for Oracle. Hitachi Data Systems recommends the following:

Caution: Consult with a Hitachi Data Systems Professional Services representative to adjust these settings.

- If the High-performance NAS Platform head is only providing NFS or iSCSI connectivity:
 - Disable short name generation.
 - Set the file system security mode to UNIX Security Mode.
- Disable atime maintenance for Oracle file systems.
- Superflush:
 - Configure superflush for all system drives. This allows the High-performance NAS Platform to write an entire RAID stripe when applicable.

- Read ahead:
 - For small block, random workloads, disable read ahead.
 - For mixed workloads, leave the default settings for IO read ahead.
 - For workloads with a high degree of sequential read access, adjust read ahead values to pre-fetch more data.
- Storage pool configuration:
 - Hitachi Data Systems recommends dedicating each unique storage pool to individual nodes in a cluster. Using this method, no one storage pool is shared across nodes.

Oracle Design Considerations

When designing databases, there are many variables that must be taken into account and these variables need to be assessed based on your company's policies, objectives and service level requirements. Some important design variables include:

- Performance
- Backup
- Restore
- Replication

In general, limit the size of file systems (less than 1TB) to reduce time for maintenance operations, such as:

- Backups
- Restores

There is no need to split a database into multiple file systems. This simplifies management and backup. Historically, it was necessary to spread a database onto multiple mount points. The spreading onto multiple mount points was designed to enhance performance by utilizing as many physical spindles as possible. This is not needed with the High-performance NAS Platform head as it is already abstracting and combining the aggregate performance of many physical spindles. A down side to the historical spread was the inevitable space wastage (a few hundred megabytes would remain available on each mount point; wastage can become large when many mount points are present). With the High-performance NAS Platform head, because the all the data for a particular instance can reside on one file system, space utilization is optimized. Also, it was not uncommon to have two data files named the same but existing on more than one mount point (for example: /u01/oradata/TEST/system_01.dbf and /u02/oradata/TEST/system_01.dbf). Data file name duplication often causes problems when refreshing other database environments.

With NFS and High-performance NAS Platform technology, whenever possible, keep databases on one High-performance NAS Platform using one file system and one mount point. Doing so will significantly decrease administration effort and will not degrade performance. High-performance NAS Platform head can be clustered to provide highly available configurations that can be resilient to a host of failures whether on the network, the server or the storage. High-performance NAS Platform leverages RAID storage to ensure protection of data in the event of disk failure. Also, it is important to note that many databases from multiple host machines may be housed on a single NAS head or cluster. On occasion, a High-performance NAS Platform head may need maintenance performed. If a database is on multiple heads, maintenance requests may impact service level agreements (SLAs). If a database must span multiple heads, then carefully plan which portions of the database are on each head with infrequent maintenance in mind.

General Oracle Configuration Guidelines

This section provides excerpts from the Oracle installation guidelines for using a NAS storage device with Oracle software and database files. It includes Hitachi Data Systems best practices where applicable.

Use the following guideline to ensure that the performance of the Oracle software meets your requirements.

The performance of Oracle software and databases stored on NAS devices depends on the performance of the network connection between the Oracle server and the NAS device.

For this reason, Oracle recommends that you connect the server to the NAS device using a private dedicated network connection, which should be gigabit Ethernet or better.

Database File Types

Oracle Corporation endorses a standard known as OFA. Details for this OFA standard can be found by following this link, Optimal Flexible Architecture for non-RAC:

http://download-west.oracle.com/docs/html/B14399_01/app_ofa.htm#i633126

Each volume type is shown in Table 1 with an example. Note that the example is represented with the mount u01. This mount point name is for example purposes only. Any standard Oracle database mount point name can be utilized.

Table 1. Classification of Database Files

Volume Type	Example	High-performance NAS Platform
Power	<code>/u01/app/oracle/product/10.2.0/db_1</code>	High-performance NAS Platform head or local file system
Datafile Volume	<code>/u01/oradata/<DBNAME></code>	High-performance NAS Platform head
Redo Volume	<code>/u01/oradata/<DBNAME></code> <code>/u01/oradata/<DBNAME></code>	High-performance NAS Platform head (multiplex and “feel safe” to place on the same redundant and protected High-performance NAS Platform head)
Archive Volume	<code>/u01/oradata/<DBNAME>/archive</code>	High-performance NAS Platform head
Temporary Volume	<code>/u01/oradata/<DBNAME></code>	High-performance NAS Platform head; sparse files work well with High-performance NAS Platform technologies
Flash Volume	<code>/u01/oradata/<DBNAME>/flash</code>	High-performance NAS Platform head

Choosing Mount Points

This section provides guidelines on how to choose the mount points for the file systems that you want to use for the Oracle software and database files. The guidelines contained in this section comply with the OFA recommendations.

Choosing Mount Points for Oracle Software Files

Oracle software files are stored in three different directories:

- Oracle base directory
- Oracle inventory directory
- Oracle home directory

For the first installation of Oracle software on a system, the Oracle base directory, identified by the ORACLE_BASE environment variable, is normally the parent directory for both the Oracle inventory and home directories. For example, for a first installation, the Oracle base, Oracle inventory and Oracle home directories might have paths similar to those in Table 2.

Table 2. Oracle Home Directory and Path Examples

Directory	Path
Oracle Base (\$ORACLE_BASE)	/u01/app/oracle
Oracle Inventory	\$ORACLE_BASE/oraInventory
Oracle Home	\$ORACLE_BASE/product/10.2.0/db_1

For subsequent installations, you can choose to use either the same Oracle base directory or a different one, but every subsequent installation on a single host uses the original Oracle inventory directory. For example, if you use the /u02/app/oracle directory as the Oracle base directory for a new installation, the Oracle inventory directory continues to be /u01/app/oracle/oraInventory.

To enable effective maintenance of Oracle software on a particular system, Oracle recommends that the Oracle inventory be located onto a local file system as opposed to a NAS device that is shared with other hosts. A NAS device can be utilized to store the Oracle inventory; however, if the inventory is stored on a NAS, then ensure utilization of a different directory on each host (for example: /u01/app/oracle/<HostName>/oraInventory).

Directory-specific Guidelines

You can use any of the following directories as mount points for NFS file systems used to store Oracle software (Note that in the following examples, the paths shown are the defaults if the ORACLE_BASE environment variable is set before you start Oracle Universal Installer.):

- Oracle base directory or its parents (/u01/app/oracle)

If you use the Oracle base directory or one of its parents as a mount point, the default location for all Oracle software and database files will be on that file system. During the installation, you might consider changing the default location of the following directories:
- The Oracle inventory directory (oracle_base/oraInventory)

Specify a local file system or a host-specific directory on the NFS file system, for example:
oracle_base/hostname/oraInventory.
- The Oracle database file directory (oracle_base/oradata)

You might want to use a different file system for database files, for example, to enable you to specify different mount options or to distribute IO.
- The Oracle database recovery file directory (/u01/oradata/DBNAME/flash)

Oracle recommends that you use different file systems for database and recovery files. If you use this mount point, all Oracle installations that use this Oracle base directory will use the NFS file system. High-performance NAS Platform heads utilize snapshots to create backups. Snapshots taken on the same NAS head should be transferred to a tape device or another NAS head.
- The product directory (oracle_base/product)

By default, only software files will be located on the NFS file system. You can also use this mount point to install software from different releases, for example:
/u01/app/oracle/product/9.2.0
/u01/app/oracle/product/10.2.0/db_1

- The release directory (`oracle_base/product/10.2.0`)

By default, only software files will be located on the NFS file system. You can also use this mount point to install different products from the same release, for example:

```
/u01/app/oracle/product/10.2.0/db_1
```

```
/u01/app/oracle/product/10.2.0/companion_1
```

- The Oracle home directory (`oracle_base/product/10.2.0/db_1`)

By default, only software files will be located on the NFS file system. This is the most restrictive mount point. You can use it only to install a single release of one product:

```
/u01/app/oracle/product/10.2.0/db_1
```

When using a Hitachi High-performance NAS Platform to house Oracle binaries, options become available that were not previously available. One potentially convenient option that is available when an Oracle home is installed on a High-performance NAS Platform is the usage of a shared Oracle home.

It should be noted that Oracle Corporation only supports shared Oracle homes when running on a RAC system. A shared Oracle home is when a single set of binaries is used on more than one host machine.

Shared Oracle homes can be used to reduce installation time and decrease storage of redundant files. There are cons to utilizing a *shared* Oracle Home; these include:

- Patching of an Oracle home impacts many host machines.
- Running of scripts such as `root.sh` must be executed on multiple host machines.
- An accidental removal of a file in a shared Oracle home impacts multiple host machines.

While Hitachi Data Systems points out that a shared Oracle home can be used, Hitachi Data Systems does not contradict Oracle support practices.

Choosing Mount Points for Oracle Database and Recovery Files

To store Oracle database or recovery files on a NAS device, you can use different paths depending on whether you want to store files from only one database or from more than one database:

- Use the NFS file system for files from more than one database.

If you want to store the database files or recovery files from more than one database on the same NFS file systems, use paths or mount points similar to the following:

When Oracle Universal Installer prompts you for the data file and the recovery file directories, specify these paths. The Database Configuration Assistant and Enterprise Manager create subdirectories in these directories using the value you specify for the database name (`DB_NAME`) as the directory name. For example:

```
/u01/oradata/db_name1
```

```
/u01/oradata/db_name1/flash
```

- Use the NFS file system for files from only one database.

If you want to store the database files or recovery files for only one database in the NFS file system, you can create mount points similar to the following, in which `orcl` is the name that you want to use for the database:

```
/u01/oradata/orcl
```

```
/u01/oradata/orcl/flash
```

Specify the directory `/u01/oradata` when Oracle Universal Installer prompts you for the data file directory and specify the directory `/u01/oradata/db_name/flash` when the Oracle Universal Installer prompts you for the recovery file location. The `orcl` directory will be used automatically either by Database Configuration Assistant or by Enterprise Manager.

Optimal File Placement

After extensive IO benchmark testing there were no issues found in having the entire database on a single file system and single mount point. For this reason and for reasons of administration, it is recommended to create only one mount point per database. High-performance NAS Platform servers have redundancy built in such that a single mount point can be failed over in the event of network or hardware failure.

Backup and Recovery

It is strongly recommended that High-performance NAS Platform snapshots be used to execute backups of all database related files.

Oracle databases can be quickly and intelligently backed using a High-performance NAS Platform snapshot. Traditionally, Oracle databases are backed up using homemade scripting and a method called *hot backups*, or by using an Oracle-provided utility called RMAN (Recovery Manager). The down side to either of these two systems on their own is the amount of time and resources that it takes to execute a database backup. Typically database files are backed up one at a time or in small parallel groups. The operating system that hosts the database must expend valuable CPU and IO cycles to accomplish these backup tasks. When a database is large, the amount of time for backups often exceeds acceptable time windows. If a restore is ever needed, it is nearly impossible to complete the restore within popular service level agreements. Restores often take two to five days when accounting for all aspects of the restore.

The solution is a High-performance NAS Platform snapshot. High-performance NAS Platform can create a point-in-time snapshot of an entire file system quickly with no copying of data required. Holistic single file and file system restores are just as efficient. Snapshots are simple to set up and execute. All files that have been backed up with a snapshot can be restored at one time or individually. For maximum data protection, High-performance NAS Platform provides up to 1,024 snapshots per file system.

High-performance NAS Platform snapshots only use disk space when data is modified after the snapshot was created. There are no specific requirements to reserve disk space and earmark reserved disk space for snapshots. This means more capacity is available for database growth and capital investments are maximized. Hitachi Data Systems recommends planning ahead to ensure there is sufficient space on the file system to maintain the snapshots.

Using Snapshots for Backup and Restore

Each volume type referred to for Oracle files can have snapshots executed on different schedules.

Snapshots can be executed several ways. For easy integration with Oracle databases, Hitachi Data Systems recommends that SSH (Secure Socket Shell) be used. This allows for scripting while ensuring secure communication to the server. To prevent passwords from being used in scripts, High-performance NAS Platform supports public key authentication. See the “Appendix” for setup instructions.

Note: Another less secure option called SSC (Secure Support Component) is available. It also allows for remote execution of commands.

Snapshots taken can be registered with RMAN by way of the RMAN proxy command so that the database administrator has access to popular and fully functional Oracle database support tools.

Test and development systems that need to be refreshed from a production system can conveniently copy the snapshot files. This allows for much faster refresh procedures.

Snapshots and snap restores can be executed on a single High-performance NAS Platform, and they also can be replicated to remote High-performance NAS Platforms. This allows for simplified disaster recovery cases.

For Oracle database files, standard Oracle rules apply. This means that Oracle tablespaces must be put into backup mode prior to taking the snapshot and removed from backup mode after taking the snapshot.

To create a snapshot of the database:

1. Execute as the Oracle user.

```
$ su - oracle
```

The following is a sample script only, which places the database into hot backup mode:

```
$ ${ORACLE_HOME}/bin/sqlplus -s "/ as sysdba" <<EOF
set pagesize 0
set linesize 2000
set trimspool on
set feedback off

spool gen_beg_backup.sql
select 'alter tablespace '||tablespace_name||' begin backup;'
from dba_tablespaces
where contents <> 'TEMPORARY';
spool off
@@gen_beg_backup.sql
exit
EOF
```

2. Execute as the root user.

Note: root access is not required; other user IDs can be configured for this purpose.

```
$ su - root
$ ssh <IP-of-HNAS> snapshot mk --file-system <file-system> <snapshot-name>
For example:
ssh 192.168.18.79 snapshot mk --file-system ORA-HDS-SATA snap1
```

3. Execute as the Oracle user.

```
$ su - oracle
```

The following is a sample script only, which takes the database from backup mode to register the copy with RMAN (optional):

```

$ ${ORACLE_HOME}/bin/sqlplus -s "/as sysdba" <<EOF
set pagesize 0
set linesize 2000
set trimspool on
set feedback off
spool gen_end_backup.sql
select 'alter tablespace '||tablespace_name||' end backup;'
from dba_tablespaces
where contents <> 'TEMPORARY';
spool off
# Rem Assumes that the file system name is "orafs1" and the snapshot is called "snap1"
spool gen_catlog_files.rman
select
    'catalog datafilecopy '''||
    replace(reverse(substr(reverse(name),instr(reverse(name),'/'))),
        '/orafs1/', '/orafs1/.snapshot/snap1/') || -- Directory Housing the datafile
    reverse(substr(reverse(name),1,instr(reverse(name),'/'))-1)||''';' -- Datafile Name
from v\datafile;
spool off
exit
EOF

```

4. Optionally, run the following if you are using RMAN, and execute as the Oracle user. Register the snapshot with RMAN (optional).

```
$ rman target / cmdfile=gen_catlog_files.rman
```

Note that the **catalog archivelog '/path_to_arch_logs/log1.arc** command allows for efficient snapshots of archived log files. Once cataloged, RMAN uses these snapshot versions of the archive log files.

Recovering a Deleted File

This is recommended to create scripts to manage restores. In the following script example called *restore_file.sh*, the High-performance NAS Platform's IP address is 192.168.18.79:

```
#!/bin/bash
ssh 192.168.18.79 <<h1
```

The following selects the virtual server (EVS) through which the file system is exported and the recovery is performed:

```
evssel 1
snapshot recover-file --file-system $1 --confirm $2 $3
h1
```

Note that in the following sample call, the file system mount point is omitted as it is gathered from the file system parameter passed in.

```
$ ./restore_file.sh ORA-HDS-FS1 /.snapshot/snap1/app/oracle/oradata/oe104/users01.dbf
/app/oracle/oradata/oe104/users01.dbf
```

To restore a single file using SSH and public key authentication:

```
$ ssh <IP-of-HNAS> snapshot rollback-file --file-system
<file-system> [--confirm] <pathname> <snapshot-name>
```

To restore a single file using SSC:

```
$ ssc -u <user> -p <pass> <IP-of-HNAS> snapshot rollback-file
--file-system <file-system> [--confirm] <pathname> <snapshot-name>
```

To recover a deleted file using SSH and public key authentication:

```
$ ssh <IP-of-HNAS> snapshot recover-file --file-system <file-system> [--confirm]
<pathname-to-preserved-file> <pathname-for-recovered-file>
```

To recover a deleted file using SSC:

```
$ ssc -u <user> -p <pass> <IP-of-HNAS> snapshot recover-file
--file-system <file-system> [--confirm] <pathname-to-preserved-file> <pathname-for-
recovered-file>
```

Using RMAN to Recover the Database

After restoring a file or system, use RMAN to recover the database. The following is a sample as Oracle provides multiple methods to recover:

```
$ rman target /
RMAN> startup mount
RMAN> recover database;
RMAN> exit
```

High-performance NAS Platform provides a feature called FileSystem Rollback, which allows an entire file system to be restored from a snapshot to its previous state at the time the snapshot was created. To use this feature, a utility called accelerated data copy (ADC) is required. The utility is installed on the High-performance NAS Platform system management unit (SMU) by default. It also can be obtained from your Hitachi Data Systems representative, and can be run from Microsoft® Windows and UNIX clients on the network.

File System Restore

To execute a file system restore, run the following command:

```
$ ADC <adc-profile>
```

adc-profile is a user-created text file that identifies the file system, the IP address of the EVS and the snapshot from which to restore. See the “Appendix” for an example of this profile.

Note: This procedure is a full rollback of the file system or the directory specified and will remove any other data created that is not part of the snapshot.

Using Snapshots to Create Test or Development Systems

Combining the High-performance NAS Platform snapshot and replication features provides an easy and nondisruptive method to create a test or development copy of production Oracle databases.

To create a test or development copy:

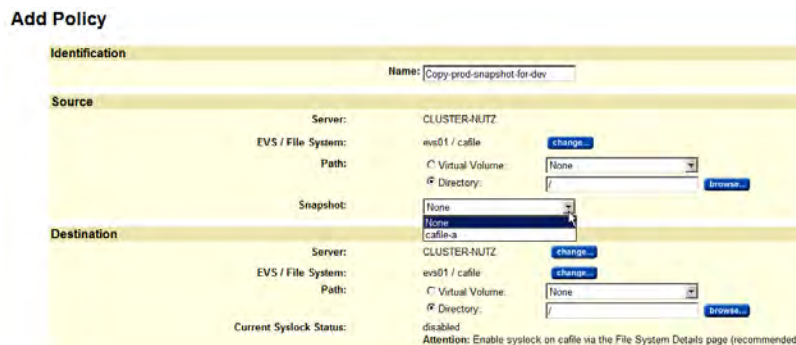
1. Determine which file system will house the test/development copy.

Hitachi Data Systems recommends either creating a new file system or using a file system that does not hold any production data. Also, if the test/development copy is expected to support a heavy test/development workload, then using a separate storage pool should be considered as the workload may impact production.

2. Create a replication policy. By default, any replication will create a snapshot before the replication job is executed. The administrator also has the option to replicate from an existing snapshot.

If the former is the case, then select the desired snapshot from the Source area (see Figure 2).

Figure 2: Creating a Replication Policy



Add Policy	
Identification	
Name:	Copy-prod-snapshot-for-dev
Source	
Server:	CLUSTER-NUTZ
EVS / File System:	evs01 / caffle
Path:	C: Virtual Volume: None Directory: /
Snapshot:	None caffle-a
Destination	
Server:	CLUSTER-NUTZ
EVS / File System:	evs01 / caffle
Path:	C: Virtual Volume: None Directory: /
Current Syslock Status:	disabled Attention: Enable syslock on caffle via the File System Details page (recommended).

Access the Add Policy screen to create, schedule and run policies and create the database.

3. Create a schedule policy.
4. Wait for the policy to be executed, or manually run the policy.
5. Use RMAN to create the database.

For example, using the duplicate database command, add the following to the init.ora file in the new database:

```
log_file_name_convert = ('/oracle/PROD/', '/oracle/DEV/')
db_file_name_convert = ('/oracle/PROD/', '/oracle/DEV/')
$ ssh oracle@dev_machine
# Startup the instance in mount mode
$ rman
RMAN> connect target sys/change_on_install@PROD
RMAN> connect auxiliary /
RMAN> run
{
  allocate auxiliary channel dev_restore device type disk;
  crosscheck copy of database;
  crosscheck backup;
  duplicate target database to DEV;
}
```

For more information on setting up replication, refer to the High-performance NAS Platform System Administration Guide.

Performance Tuning

NFS-based file systems do not use cached IO as a protective mechanism to prevent data corruption. This allows the Oracle database to have more memory available for cache buffering. An operating system utilizes about 5 to 10 percent of its total physical memory. Standard SAN-based file systems are typically configured to utilize upwards of 25 percent of a system's physical memory for buffering. This left about 65 percent of memory for applications. Oracle databases do a better job at buffering Oracle data than do operating system buffers. By utilizing a High-performance NAS Platform, the 25 percent of memory reserved for file system buffering is all but eliminated. This allows for applications to utilize upwards of 80 percent of physical memory.

On Linux 4, access High-performance NAS Platform NFS exports utilizing DirectIO. This is a highly efficient method for Oracle to execute IO calls. To ensure that applications are set up to take advantage of these performance benefits, set the Oracle initialization parameters shown in Table 3.

Table 3. Oracle Initialization Parameters for Performance Benefits

Parameter	Default	Description¹
DB_WRITER_PROCESSES	1	DB_WRITER_PROCESSES is useful for systems that modify data heavily. It specifies the initial number of database writer processes for an instance.
FILESYSTEMIO_OPTIONS	NA	FILESYSTEMIO_OPTIONS specifies IO operations for file system files.
DISK_ASYNCH_IO	TRUE	DISK_ASYNCH_IO controls whether IO to data files, control files and log files is asynchronous (that is, whether parallel server processes can overlap IO requests with CPU processing during table scans). If your platform supports asynchronous IO to disk, Oracle recommends that you leave this parameter set to its default value. However, if the asynchronous IO implementation is not stable, you can set this parameter to false to disable asynchronous IO. If your platform does not support asynchronous IO to disk, this parameter has no effect.
DB_CACHE_SIZE DB_CACHE_SIZE	0	DB_CACHE_SIZE specifies the size of the default buffer pool for buffers with the primary block size (the block size defined by the DB_BLOCK_SIZE initialization parameter). The value must be at least 4M * number of cpus * granule size (smaller values are automatically rounded up to this value). A user-specified value larger than this is rounded up to the nearest granule size. A value of zero is illegal because it is needed for the default memory pool of the primary block size, which is the block size for the system tablespace.
SGA_TARGET	0	SGA_TARGET specifies the total size of all SGA components. If SGA_TARGET is specified, then the following memory pools are automatically sized: <ul style="list-style-type: none"> • Buffer cache (DB_CACHE_SIZE) • Shared pool (SHARED_POOL_SIZE) • Large pool (LARGE_POOL_SIZE) • Java pool (JAVA_POOL_SIZE) • Streams pool (STREAMS_POOL_SIZE) <p>If these automatically tuned memory pools are set to nonzero values, then those values are used as minimum levels by Automatic Shared Memory Management. You would set minimum values if an application component needs a minimum amount of memory to function properly.</p> <p>The following pools are manually sized components and are not affected by Automatic Shared Memory Management:</p> <ul style="list-style-type: none"> • Log buffer • Other buffer caches, such as keep, recycle and other block sizes • Fixed SGA and other internal allocations <p>The memory allocated to these pools is deducted from the total available for SGA_TARGET when Automatic Shared Memory Management computes the values of the automatically tuned memory pools.</p>

¹ The Oracle parameter descriptions were taken from the Oracle Database Reference for 10g Release 2 (10.2) (part number B14237-03).

Hitachi Data Systems Recommendations

Hitachi Data Systems recommendations are shown in Table 4.

Table 4. Hitachi Data Systems Recommendations for Performance Benefits

Recommendation	Description
FILESYSTEMIO_OPTIONS =setall	Allows for asynchronous, DirectIO, or none depending on the specific file sets that are accessed.
DISK_ASYNC_IO=TRUE	Oracle can enlist the usage of asynchronous IO for database writer processes whenever the operating system and database allow for it. This IO method has the capability of speeding up Oracle systems by 10 percent or more.
DB_WRITER_PROCESSES=<CPU-Count>	Oracle databases can utilize many database writer processes. Each process is assigned to a different section of the database cache upon database startup. Utilizing multiple database writers alleviates latch-based contention on highly active databases. There is little additional overhead to using multiple processes. Multiple database writers and asynchronous IO can be used together to provide a robust and fast database.
Increase either DB_CACHE_SIZE and/or SGA_TARGET	To take advantage of the additional memory available to applications.

NFS Mount Options

Oracle recommends using certain NFS mount options for software and database files², as shown in Table 5.

Table 5. NFS Mount Options

Option	Requirement	Description
Hard	Mandatory	Generate a hard mount of the NFS file system. If the connection to the server fails or is temporarily lost, connection attempts are made until the NAS device responds.
Bg	Optional	Try to connect in the background if the connection fails.
Tcp	Optional	Use the TCP protocol rather than UDP. TCP is more reliable.
nfsvers=3	Optional	Use NFS version 3. Oracle recommends that you use NFS version 3 where available, unless the performance of version 2 is higher.
Suid	Optional	Allow clients to run executables with SUID enabled. This option is required for Oracle software mount points.
Rsize	Mandatory	The number of bytes used when reading from the NAS device. This value should be set to the maximum database block size supported by this platform. A value of 8,192 is often recommended for NFS version 2 and 32,768 is often recommended for NFS version 3.
Wsize	Mandatory	The number of bytes used when reading from the NAS device. This should be set to the maximum database block size supported by this platform. A value of 8,192 is recommended for NFS version 2 and 32,768 is often recommended for NFS version 3.
nointr (or intr)	Optional	Do not allow (or allow) keyboard interrupts to kill a process that is hung while waiting for a response on a hard-mounted file system.
actime=0 or noac	Mandatory	Disable attribute caching. This is needed by Oracle Universal Installer during software install. OUI will not install software on if this is not set.

² The NFS mount option table is from Oracle Database Installation Guide 10g Release 2 (10.2) (part number B15660-02).

Hitachi Data Systems Recommendations

Hitachi Data Systems recommends the following:

- Software mount points
defaults,nfsvers=3,rsize=32768,wsiz=32768,tcp,nointr,hard,noac,bg
- Database file mount points
defaults,nfsvers=3,rsize=32768,wsiz=32768,tcp,nointr,hard,noac,bg

Linux Options

A High-performance NAS Platform head supports all Linux operation systems that are supported by Oracle. It should be noted, however, that later releases of the Linux operating system include advancements in NFS settings. For this reason, Linux kernel 2.4.20 and later is suggested.

The following are a few recommended Linux versions:

- Red Hat 3
- Red Hat 4
- SUSE SLES 9

Oracle recommended kernel patches are suggested.

Ensure that the Linux system is set to allow and utilize *uncached* IO. Caching of IO for NFS file systems will impact performance and might cause data corruption. Caching of IO within the operating system is *expressly* unsupported.

On Linux, the following command should appear in the */etc/modules.conf* file:

```
options nfs nfs_uncached_io=1
```

To reiterate from a prior section in this document, the following options should be utilized for database file mount points:

defaults,nfsvers=3,rsize=32768,wsiz=32768,tcp,nointr,hard,noac,bg

The following are additional minimum recommended settings that allow for larger network packets:

```
$ cd /proc/sys/net/core
$ echo 1048576 > rmem_default
$ echo 1048576 > rmem_max
$ echo 262143 > wmem_default
$ echo 262143 > wmem_max
$ echo 0 > /proc/sys/net/ipv4/tcp_sack
$ echo 0 > /proc/sys/net/ipv4/tcp_timestamps
```

Also, to ensure that these setting persist after reboot, add the following lines to the */etc/sysctl.conf* file:

```
net.core.rmem_default = 1048576
net.core.rmem_max = 1048576
net.core.wmem_default = 262143
```

```
net.core.wmem_max = 262143
```

Notes:

- For these changes to be effective, the NFS file systems must be remounted.
- Ensure that the network interfaces are hard-set to utilize full duplex mode.
- Ensure that the network interface supports gigabit Ethernet or better.
- Jumbo frames are supported but not required to achieve maximum performance.

Appendix

Configuring SSH Public Key Authentication with High-performance NAS Platform

Scripts can perform tasks such as snapshots, virtual volume creation and quota management using SSH. Ordinarily, SSH requires a password: tasks can be performed unattended without passwords using SSH and public key authentication.

Prerequisites

Running scripts from systems on the public network requires an administrative IP address to be available on the gigabit interfaces.

A user needs to be created on the High-performance NAS Platform server. Use the **user add** `<username>` `<password>` `[role]` command.

Generating SSH Keys

Public key authentication uses a public/private key pair instead of passwords. The keys must first be generated on the client system. On UNIX systems, there should be a program called `ssh-keygen`, which is used to generate the needed keys.

There are three types of keys that can be generated: RSA, RSA2 and DSA. Discussion of the differences between these is beyond the scope of this article; however, RSA2 and DSA are preferred over RSA.

The keys generated are associated with the user logged in at that time. For example, if you are logged in as user `oracle` and use `ssh-keygen`, your SSH keys are stored in user `oracle`'s `.ssh` directory, located in the root of its home.

To generate a DSA SSH key pair:

```
$ ssh-keygen -t dsa
```

`ssh-keygen` prompts you to choose a location to save the keys. This is usually in `~/.ssh/`. Unless you experienced with this configuration, it is best to take the default.

Logging into High-performance NAS Platform

Now that you have your key, log in to High-performance NAS Platform. Once logged in, run:

```
$ ssh-register-public-key
```

This registers your public key with High-performance NAS Platform so that when you (or your script) attempt to log in next time, there are no prompts for passwords.

Server Administration through Public Networks

This section describes procedures to allow for CLI-based administration of a High-performance NAS Platform through the public-side gigabit Ethernet network. Also, it discusses precautions and advice on how to lock down a public-side administrative IP.

By default, a server can only be administered through the private management network. This physical port on the server is sometimes referred to as `mgmnt1`. Also by default, the front-side (or public-side) gigabit (or 10-gigabit) Ethernet ports are only used for file (or block) services. This default configuration ensures both a physical and logical separation of data access and server management networks.

While not recommended in most situations, it is possible to configure the server to enable administration through the gigabit Ethernet ports. To do so, it is necessary to install a public network IP address on the server's

administration service (admin EVS, also known as EVS 0), and associate the IP address with an aggregation group.

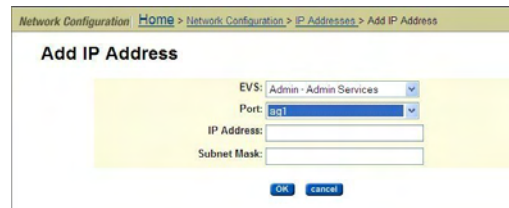
Enabling public-side administration can make the server more accessible to malicious attack. As such, when enabling public-side administration of the server, some extra steps should be taken to secure the public administrative IP address to allow access only to the desired services, and only by specific hosts. To secure a public administrative IP, use the *mscfg* CLI command on the High-performance NAS Platform.

Adding a Public-side Administrative IP Address

To add a public-side administrative IP address:

1. From the SMU home page, select **Network Configuration > IP Addresses**.
2. Click **add**.
3. From the EVS list, select **Admin - Admin Service**.
4. From the Port list, select a gigabit aggregation group: for example, **ag1** (see Figure 3).

Figure 3: Selecting the ag1 Gigabit Aggregation Group



Enabling public-side administration can make the server more accessible to malicious attack, so it is important to secure the public administrative IP address.

5. Enter an available IP address and subnet mask on your public network in the respective fields.

This is the IP address that will be accessible for server administration.

6. Click OK.

File System Rollback

This section describes how to roll back a file system or virtual volume using the FileSystem Rollback feature and the ADC utility.

For example, given the file system */fs1* and the virtual volume */vv1* residing on */fs1*, and snapshots A, B, C, D, the following depicts virtual volume and file system rollback scenarios:

To rollback virtual volume */vv1* to snapshot B:

1. Take a snapshot of */fs1*.

This is recommended but not required.

2. Remove shares and exports on virtual volume */vv1* or un-mount the exports on the clients.

This is not required; however, it is strongly recommended to prevent the client from accessing the file system while it is recovering.

3. Execute the following on the SMU or client that has the ADC binary:

```
$ adc vv1-rollback.txt
```

ADC is a binary, and the contents of the vv-rollback.txt profile are:

```
OPERATION copy

SOURCE -u ndmp -p ndmp 19.9.26.16 /fs1/vv1

DESTINATION -u ndmp -p ndmp 19.9.26.16 /fs1/vv1

ENVIRONMENT NDMP_BLUEARC_ROLLBACK B
```

To rollback file system /fs1, follow the same instructions and specify /fs1 in the parameter file.

ADC Profile

```
OPERATION <operation>

SOURCE -u <ndmp username> -p <ndmp user password> <source evs IP> <source filesystem>

DESTINATION -u <ndmp username> -p <ndmp user password> <destination evs IP> < destination
filesystem>

ENVIRONMENT <NDMP VARIABLE> [VARIABLE OPTIONS]
```

For additional information, see the *adc.readme* file that is part of the ADC distribution. This file contains detailed instructions for using the utility and defining profiles. There is also a copy available on the SMU located in */usr/local/bin/*.

Hitachi Data Systems Corporation

Corporate Headquarters 750 Central Expressway, Santa Clara, California 95050-2627 USA

Contact Information: + 1 408 970 1000 www.hds.com / info@hds.com!

Asia Pacific and Americas 750 Central Expressway, Santa Clara, California 95050-2627 USA!

Contact Information: + 1 408 970 1000 www.hds.com / info@hds.com !

Europe Headquarters Sefton Park, Stoke Poges, Buckinghamshire SL2 4HD United Kingdom

Contact Information: + 44 (0) 1753 618000 www.hds.com / info.uk@hds.com!

Neither BlueArc Corporation nor its affiliated companies (collectively, "BlueArc") makes any warranties about the information in this document. Under no circumstances shall BlueArc be liable for costs arising from the procurement of substitute products or services, lost profits, lost savings, loss of information or data, or from any other special, indirect, consequential, or incidental damages, that are the result of its products not being used in accordance with this document.

Hitachi is a registered trademark of Hitachi, Ltd., and/or its affiliates in the United States and other countries. Hitachi Data Systems is a registered trademark and service mark of Hitachi, Ltd., in the United States and other countries.

The following are trademarks licensed to BlueArc Corporation, registered in the USA and other countries: BlueArc, the BlueArc logo, and the BlueArc Storage System. Microsoft is a registered trademark of Microsoft Corporation.

Some portions of this document contain copyrighted material of the Oracle Corporation. Some portions of this document contain copyrighted material of BlueArc Corporation. Their usage is with the approval of the copyright holders.

All brand names and product names used in this document are trade names, service marks, trademarks or registered trademarks of their respective owners.

Notice: This document is for informational purposes only, and does not set forth any warranty, express or implied, concerning any equipment or service offered or to be offered by Hitachi Data Systems. This document describes some capabilities that are conditioned on a maintenance contract with Hitachi Data Systems being in effect, and that may be configuration-dependent, and features that may not be currently available. Contact your local Hitachi Data Systems sales office for information on feature and product availability.

Hitachi Data Systems sells and licenses its products subject to certain terms and conditions, including limited warranties. To see a copy of these terms and conditions prior to purchase or license, please go to <http://www.hds.com/corporate/legal/index.html> or call your local sales representative to obtain a printed copy. If you purchase or license the product, you are deemed to have accepted these terms and conditions.