



Technical report:
**Oracle 10g on HP-UX and IBM System
Storage N series**

Best-practice performance

• • • • • • • • •

Document NS3557-0

April 11, 2008



Table of contents

Abstract	3
Introduction and summary	3
Hardware and software environment	3
Server	3
Storage	4
HP-UX kernel parameters.....	4
HP-UX kernel-parameter recommendations.....	5
HP-UX NFS configuration	6
The <code>forcedirectio</code> mount option.....	6
Storage-system tuning	6
Storage provisioning details.....	6
Oracle tuning	7
HP-UX patches	9
Oracle performance over NFS	10
Conclusions	11
Trademarks and special notices	12



Abstract

This technical report provides best-practice tuning recommendations that can increase the performance of an Oracle 10g database on an HP-UX 11i v2 server in an IBM System Storage N series NFS environment. The focus of this paper is technical and the reader should be experienced with HP-UNIX system administration, Oracle database administration, network connectivity, and IBM N series storage administration.

Introduction and summary

This technical report provides performance-tuning recommendations for running Oracle databases on HP-UNIX servers and IBM® System Storage™ N series storage in a network file system (NFS) environment. A series of tests of an online transaction processing (OLTP) workload and an Oracle 10.2 database on an HP-UX server with an 11i v2 operating system (OS) were run to determine how various tuning settings affected NFS performance.

On the hardware used for this report, with the suggested tuning, NFS performance improved about 640 percent. This performance improvement resulted from tuning options that are made possible by the HP-UX 11i v2 server OS release. The improvements, however, are not out of the box; you must follow the simple tuning recommendations in this report to achieve them.

For NFS, the major factors in performance improvement involve the `forcedirectio` mount option, applying the NFS cumulative patch, and tuning kernel and Oracle parameters. Direct I/O is a new feature in HP-UX 11.23 for NFS. This report details the specific environment configurations, tuning recommendations, and performance results of the OLTP workload using NFS.

Hardware and software environment

Server

For this comparison, an HP Integrity RX2600 server was used. The configuration is shown in Table 1.

Component	Details
Operating system	HP-UX 11i v2 mission-critical operating-environment component
Version	B.11.23.0603 U ia64
System type	RX2600
Database server	Oracle 10.2.0.2.0 64-bit
Total physical RAM	6 GB
Processor	2 Intel® Itanium-2 CPUs 1.4 GHz
Storage network	2X 1 Gb Ethernet for NFS

Table 1) Server configuration.

The HP-UX host was configured with two HP A6825 gigabit Ethernet NICs, both dedicated to one IBM System Storage N series N5500 storage system. The link speed was set to 1 Gb/second, and jumbo frames were enabled with an MTU size of 9000.

Storage

The IBM N series storage-system configuration is described in Table 2.

Component	Details
Operating system	IBM System Storage N series with Data ONTAP® 7.2.1
Storage interconnect	2X 1 Gb Ethernet for NFS
Disks	8 DS14s of 72 GB, 10 KB RPM disks
Storage controller	Single IBM N series N5500
Disk shelf to storage-controller connection	2 backside FC-AL with 4 shelves per FC-AL loop
Storage switches	Server and storage system were direct connected with crossover cables

Table 2) Storage-system configuration.

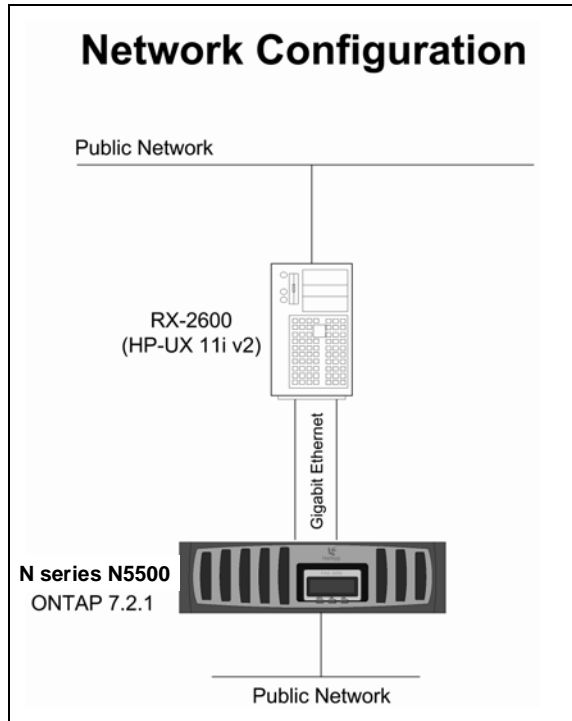


Figure 1) Schematic of HP-UX server and IBM N series storage for the NFS test environment.

IBM N series storage systems provide superior cost-effective data protection, availability, administration, and backup and recovery through IBM N series tools and features, which include the following:

- Fast reliable backups using IBM System Storage N series with Snapshot™ technology
- Disk redundancy through IBM System Storage N series with RAID-DP™ technology.

HP-UX kernel parameters

There are several ways to view/edit kernel parameters in HP-UX:

- Use the admin tool `sam`: -> Select Kernel Configuration -> Select Configurable Parameters.
- Use the `kmtune` or `kctune` commands.

The use of `sam` is recommended because of its ease of use. For example, it determines whether any parameter is "static" or "dynamic." In the case of static, `sam` will suggest a server reboot.

The `kmtune` command is used to query, set, or reset system parameters. `kmtune` displays the current values of all system parameters when used without any options or with the `-S` or `-l` option.



The `ktune` command is the administrative command for HP-UX kernel tunable parameters. It gives information about tunable parameters and their values and makes changes to tunable values.

HP-UX kernel-parameter recommendations

There are various kernel parameters that can improve the performance of NFS on HP-UX. The following kernel parameters were set or verified based on the minimum requirements for Oracle 10.2 installation and to improve NFS performance on HP-UX 11i v2 systems.

Kernel parameter	Value	Default value	Details
fs_async	0	0	Disable asynchronous file system writes.
dnlc_hash_locks	512	512	Size of the pool of locks used to control access to the directory name lookup cache (DNLC) data structures.
max_thread_proc	2257	256	Limits number of threads a single process can create. This protects the system from excessive use of system resources if a runaway process creates more threads than needed.
maxswapchunks	16384	16384	Determines maximum size of swap space on the system.
ncsize	36672	8976	DNLC—space needed for inodes.
nfile	65536	65536	Maximum number of files that can be open at any one time.
nflocks	4200	4096	Maximum number of file/record locks allowed on system at any time.
ninode	35648	4880	Maximum number of open inodes at any time.
Kernel parameter	Value	Default value	Details
nkthread	8416	8416	Maximum number of kernel threads that can be running on the system at any time.
nproc	4200	4200	Maximum number of processes that can be running on the system at any given time.
dbc_min_pct	5	5	Minimum percent of system memory used for dynamic buffer cache; inactive if nbuf and bufpages not equal to zero.
dbc_max_pct	25	50	Maximum percent of system memory used for dynamic buffer cache; inactive if nbuf and bufpages not equal to zero.
maxfiles	2048	2048	Specifies the “soft” limit for the number of files a process may open at any one time.
maxfiles_lim	4096	4096	Specifies the “hard” limit for the number of files a process may open at any one time.
ksi_alloc_max	33600	33600	Systemwide limit on the number of queued signals that can be allocated and in use.
executable_stack	0	0	Controls whether program stacks are executable by default.
maxuprc	3780	256	Limits the maximum number of processes per user.
vps_ceiling	64	16	Maximum (in kilobytes) of system-selectable page size.
msgmni	4200	512	Number of message queue identifiers.
msgseg	32767	8192	Number of segments available for messages.
msgtql	4200	1024	Number of message headers.
maxdsiz	1073741824	1073741824	Maximum size (in bytes) of data segment for any 32-bit user process.
maxdsiz_64bit	4294967296	4294967296	Maximum size (in bytes) of data segment for any 64-bit user process.
maxssiz	134217728	8388608	Maximum size (in bytes) of the stack for any 32-bit user process.
maxssiz_64bit	4294967296	268435v56	Maximum size (in bytes) of the stack for any 64-bit user process.
Kernel parameter	Value	Default value	Details
semmns	8400	4096	Maximum number of semaphores.
semmnu	4186	256	Number of semaphore undo structures.
shmmax	4294967296	1073741824	Maximum shared memory segment in bytes.
semmap	4098	4098	Size of the free-space resource map used for allocating new System V IPC semaphored in shared memory.
semvmx	32767	32767	Maximum value of any single System V IPC semaphore.
shmmni	512	400	Number of System V shared memory segment identifiers in system.
shmseg	120	300	Maximum number of System V shared memory segments per proc.

Table 3) HP-UX kernel-parameter settings.



HP-UX NFS configuration

The following mount options were used for the volumes in the HP-UX 11i v2 tests:

```
rw,bg,hard,nointr,forcedirectio,proto=tcp,vers=3,rsize=32768,wsiz=32768,timeo=600
```

The `forcedirectio` mount option

Database applications usually give better performance by allowing the database to manage its own cache. The `forcedirectio` mount option causes the NFS requests to bypass buffer cache on the server, thereby eliminating read-ahead and write-behind semantics, which helps dramatically with the database performance. Direct I/O (`forcedirectio`) saves memory and boosts performance of Oracle databases by allowing them to manage their own caches independently.

We ran tests with and without the `forcedirectio` mount option. The performance improvement by just enabling direct I/O was approximately 482 percent.

Storage-system tuning

No special tuning is recommended on the storage system. All the disks from the eight shelves of 72 GB, 10 KB RPM drives were placed in two aggregates. The aggregates were created using RAID-DP with 16 disks in a RAID group. One IBM System Storage N series with FlexVol™ root volume `vol0`, was created inside the aggregate `aggr0`, and two FlexVol volumes, `data1` and `log1`, were created inside the aggregate `aggr1`.

Storage provisioning details

```
aggr0 – 3 disks, RAID-DP, RAID size 16
        /vol/vol0      flexvol, root volume
aggr1 – 104 disks, RAID-DP, RAID size 16
        /vol/oradata1  flexvol, data1 1600GB data volume
        /vol/oralog1   flexvol, log1 200GB log volume
```

NFS mount	Oracle usage
/mnt/oradata1	Data files
/mnt/oradata2	Data files
/mnt/oralog1	Log and control files
/mnt/oralog2	Log and control files

Table 4) NFS mount and Oracle usage.



Oracle tuning

Below is a list of Oracle initialization parameters that were used for these OLTP tests. These parameters were used after extensive testing with our OLTP workload on our host configuration and were based on available system memory (6 GB), database size, and type of workload.

Various Oracle parameters, including `db_cache_size`, `db_16k_cache_size`, `log_buffer`, `shared_pool_size`, `db_writer_processes`, `sga_max_size`, `sga_target`, `pga_aggregate_target`, and `hpux_sched_noage` were tuned to improve performance.

```
compatible = 10.2.0.1.0
db_name = tpcc
control_files = (/mnt/oralog1/control_001, /mnt/oralog2/control_002)
parallel_max_servers = 10
db_files= 600
db_cache_size = 2640M
db_16k_cache_size= 480M
_log_parallelism = 4
Parallel_execution_message_size = 8192
dml_locks= 500
statistics_level = typical
log_buffer = 10048576
processes = 600
sessions = 600
transactions = 600
shared_pool_size = 250M
cursor_space_for_time = TRUE
db_block_size = 8192
undo_management = auto
undo_retention = 900
_undo_autotune=false
disk_asynch_io=false
db_writer_processes=8
db_file_multiblock_read_count=16
UNDO_TABLESPACE = undo_1
fast_start_mttr_target=2000
filesystemio_options=directio
pga_aggregate_target=210M
java_pool_size = 16777216
sga_max_size = 3800M
sga_target = 3800M
session_max_open_files = 300
transactions = 600
_kgl_large_heap_warning_threshold = 4194304
hpux_sched_noage = 178
```

The recommended value for the `db_writer_processes` parameter is to at least match the number of processor cores on the system. We tried testing with two, four, and eight `db_writer_processes` and got best performance with eight `db_writer_processes`.

The appropriate size of the SGA is limited by the available system memory and memory requirements of other applications running on the server. The total SGA size should be set to approximately 50 to 70 percent of system memory in case no other applications except an Oracle instance is running on the



server. Various parameters, including `db_cache_size`, `shared_pool_size`, `pga_aggregate_target`, and `log_buffer`, were set based on available system memory and database requirement. The total `db_cache_size` was set to 3120 MB, and `shared_pool_size` was set to 250 MB based on the total system memory of 6000 MB. For optimal performance, all these memory parameters should be tuned, and appropriate resources should be assigned based on total available system memory.

The following warning messages were reported multiple times in the alert log during initial testing:

“Memory Notification: Library Cache Object loaded into SGA
Heap size 2294K exceeds notification threshold (2048K)”

We found this problem documented in Oracle Metalink notes (Note: 330239.1). The problem occurs in Oracle 10g R2 and is the result of the KGL heap warning threshold default setting being too low. These warning messages appear in the alert log as a result of a new event messaging mechanism in Oracle 10g R2. The solution was to set the `_kgl_large_heap_warning_threshhold` parameter to a higher value. In Oracle 10.2.0.2, the threshold is increased to 50 MB after regression tests, so this should be a reasonable and recommended value.

We set the `hpux_sched_noage` parameter value to 178 as recommended by an Oracle Metalink note (Note: 217990.1). As per the Metalink note, this parameter should be suited to OLTP environments because OLTP environments can cause competition for critical resources. We conducted test runs with and without the `hpux_sched_noage` parameter and did not notice any increase or decrease in performance in our test environment. Application and server environments can be quite different. One should test and verify whether their environment does benefit from the `sched_noage` setting.



HP-UX patches

Figure 2 shows the list of HP-UX 11.23 patch bundle as well as patches required for NFS and Oracle 10g R2 that we installed prior to testing.

HP-UX patch bundle		
BUNDLE	B.11.23	Patch bundle
BUNDLE11i	B.11.23.0409.3	Required patch bundle for HP-UX 11i v2 (B.11.23), September 2004
FEATURE11i	B.11.23.0606.046a	Feature enablement patches for HP-UX 11i v2, June 2006
NFS patch		
PHNE_34550	1.0	NFS cumulative patch

Figure 2) Installed HP-UX 11.23 patch bundle and patches required for NFS and Oracle 10g R2.

We installed the NFS cumulative patch, PHNE_34500, during our testing. More recently, PHNE_35118, NFS cumulative patch has been released. The NFS cumulative patch helped with NFS read as well as write performance. Make sure the latest NFS cumulative patch is installed on the server. Even more current patches could have been released since the time of this report writing.

PHSS_31850	PHSS_31851
PHSS_31854	PHSS_31855
PHSS_33275	PHSS_33276
PHSS_32213	PHCO_32426
PHCO_31553	PHKL_33552
PHKL_32645	PHKL_32646
PHKL_32632	PHKL_33552
PHKL_32645	PHKL_32646
PHKL_31500	

Table 5) Patches required for Oracle 10g R2 installation.

The Oracle installation will not complete if any of the above HP-UX patches are missing.

It is a good practice to keep the HP-UX operating system up-to-date with available OS patches. There is lot of good information including documentation and latest patches on HP support websites.

Oracle performance over NFS

The test environment was designed to stress the database server software and hardware with the goal of utilizing all database-server CPU resources for each test run.

The database used for testing is best described as OLTP in nature with a physical size of approximately 350 GB. During the testing, we used a publicly available tool to generate an OLTP-type load consisting of a steady stream of small, random read and write operations (approximately 2:1 reads to writes) against the test database. This workload was designed to emulate the real-life activities of a wholesale supplier order-processing system in which inventory is spread across several regional warehouses.

The database utilized both primary and secondary keys for data access. In terms of measured database throughput, the metric of interest was defined as the number of orders processed per minute. Throughout this document, this measurement will be referred to as “order entry transactions per minute” or simply OETs. This is a measure of OETs per specified constant set of time (in this case one minute).

The benchmark ran in *server-only* mode — all users and the database engine ran on the RX2600 server. Server-only mode also implies that the user processes are running without any think times. This means that the users continually submit transactions without simulating any delay between transactions.

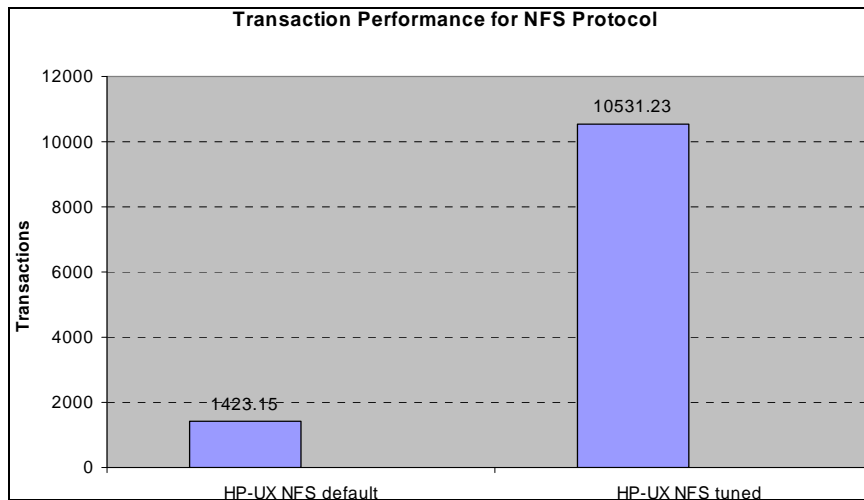


Figure 2) Order entry transactions for untuned and tuned NFS.

Figure 2 shows OETs for the NFS protocol. The chart contains results for both default (untuned) as well as the tuned configuration using the NFS protocol.

A series of preliminary tests were conducted to improve the performance of the base configuration by modifying a number of Oracle database and HP-UX kernel parameters from their default values. We then conducted a series of tuning iterations in which specific configuration parameters were modified and the resulting performance improvement measured. As a result of this tuning effort, we saw performance improve approximately 640 percent, going from 1423 OETs using default configuration parameters to 10531 OETs using our tuned configuration.

Performance improved about 158 percent by tuning Oracle parameters and HP-UX kernel parameters. Applying the latest patches improved performance around 482 percent by enabling direct I/O. NFS performance improved significantly by using direct I/O and tuning the kernel as well as Oracle parameters.



Conclusions

This paper demonstrates the direct I/O improvements on NFS for HP-UX 11i v2 environments.

Additionally, this paper outlines clear configuration guidelines for Oracle Database workloads on the NFS protocol that significantly improve performance and scaling over the default settings. Test results clearly support IBM System Storage N series as high-performing, low-cost, feature-rich storage for Oracle 10g running on HP-UX.

As with any environment, tuning a particular workload is an iterative and often complex process. The tuning guidelines in this paper should give good results compared to untuned HP-UX NFS performance. Individual results will vary depending upon type of the workload.



Trademarks and special notices

© International Business Machines 1994-2008. IBM, the IBM logo, System Storage, and other referenced IBM products and services are trademarks or registered trademarks of International Business Machines Corporation in the United States, other countries, or both. All rights reserved.

References in this document to IBM products or services do not imply that IBM intends to make them available in every country.

Data ONTAP, FlexVol, Network Appliance, the Network Appliance logo, RAID-DP and Snapshot are trademarks or registered trademarks of Network Appliance, Inc., in the U.S. and other countries.

Intel, Intel Inside (logos), MMX, and Pentium are trademarks of Intel Corporation in the United States, other countries, or both.

Other company, product, or service names may be trademarks or service marks of others.

Information is provided "AS IS" without warranty of any kind.

Information concerning non-IBM products was obtained from a supplier of these products, published announcement material, or other publicly available sources and does not constitute an endorsement of such products by IBM. Sources for non-IBM list prices and performance numbers are taken from publicly available information, including vendor announcements and vendor worldwide homepages. IBM has not tested these products and cannot confirm the accuracy of performance, capability, or any other claims related to non-IBM products. Questions on the capability of non-IBM products should be addressed to the supplier of those products.

Any references in this information to non-IBM Web sites are provided for convenience only and do not in any manner serve as an endorsement of those Web sites. The materials at those Web sites are not part of the materials for this IBM product and use of those Web sites is at your own risk.