Oracle Recovery Manager: Performance Testing at Sun Customer Benchmark Center

An Oracle White Paper
October 2001

Oracle Recovery Manager: Performance Testing at Sun Customer Benchmark Center

Oracle Recovery Manager: Performance Testing at Sun Customer Benchmark Center

EXECUTIVE OVERVIEW

In today's demanding market, downtime is not an option. Thus, developing a fast and reliable online backup solution is one of the most important aspects of achieving high availability. In addition, database recovery must be quick and efficient in case of a catastrophic failure. Moreover, a high level of availability requires automation of complex backup and recovery tasks. Oracle9*i* RMAN is the solution: it provides superior performance, maintains high availability of the database, and automates much of backup and recovery.

This paper will show that Oracle9*i* RMAN provides superior performance and that it can scale linearly to any amount of hardware. Specifically, this paper documents performance testing performed by Oracle that achieved the following results with Oracle9*i* RMAN:

1. Backup and restore speeds of 1 TB/hour.
2. Linear increase of speed as more tape drives are added.
3. Low CPU resource consumption during backup and restore operations, minimizing the impact on the production system.

INTRODUCTION

This paper represents results of Oracle9*i* RMAN performance testing conducted at Sun Customer Benchmark Center in Newark, CA in August 2001. The objective of the testing was to show that Oracle9*i* RMAN can achieve world-class backup and restore speed of 1 TB/hour. The results prove that Oracle9*i* RMAN is the right solution to achieve a high level of data availability: it provides fast and reliable backup and restore with minimum impact on production.

The chapter *Testing Environment* offers an overview of the hardware configuration and explains the setup of the software components used in the benchmark. The results of the tests are represented in the chapter *Results and Findings*. The *Conclusion* chapter summarizes the findings.

TESTING ENVIRONMENT

Testing was performed at Sun Customer Benchmark Center in Newark, CA. The following hardware and software components were used:

- Oracle9*i* Database Enterprise Edition
- One Sun Enterprise™ 10000 server
- Seven Sun A5200 disk arrays with 22 36GB disks each
- Two Sun StorEdge™ L700 tape libraries with 20 DLT7000 tape drives each
- VERITAS Volume Manager and VERITAS Database Edition™ for Oracle
- VERITAS NetBackup DataCenter 3.4 for Solaris

Hardware

The Oracle9*i* database was installed on a Sun Enterprise™ 10000 system connected to seven A5200 disk arrays. The tape subsystem consisted of two Sun StorEdge™ L700 tape libraries

The database server was a Sun Enterprise™ 10000 system with 64 SPARC 400 MHz processors on 16 independent system boards and 64 GB of memory (for specification, refer to [SME10K]). The disk storage system consisted of seven Sun StorEdge™ A5200 disk arrays (for specification, refer to [SMA5200]). Each disk array contained 22 Seagate 36 GB disks that were mirrored with VERITAS Volume Manager. The disk arrays were connected to the server by 14 FC-AL connections using a split-loop method, with 11 disk drives in each loop. The split-loop divided the original and mirror disk drives between two loops, to achieve better performance and availability. The tape subsystem consisted of two Sun StorEdge™ L700 tape libraries ([SML700]), each containing 20 DLT7000™ tape drives. All 40 tape drives were connected to the database server by 20 Fast/Wide SCSI-2 connections. Two tape drives were daisy-chained on each SCSI connection.

Figure 1 on the next page illustrates the hardware setup.

Database and Disk Layout

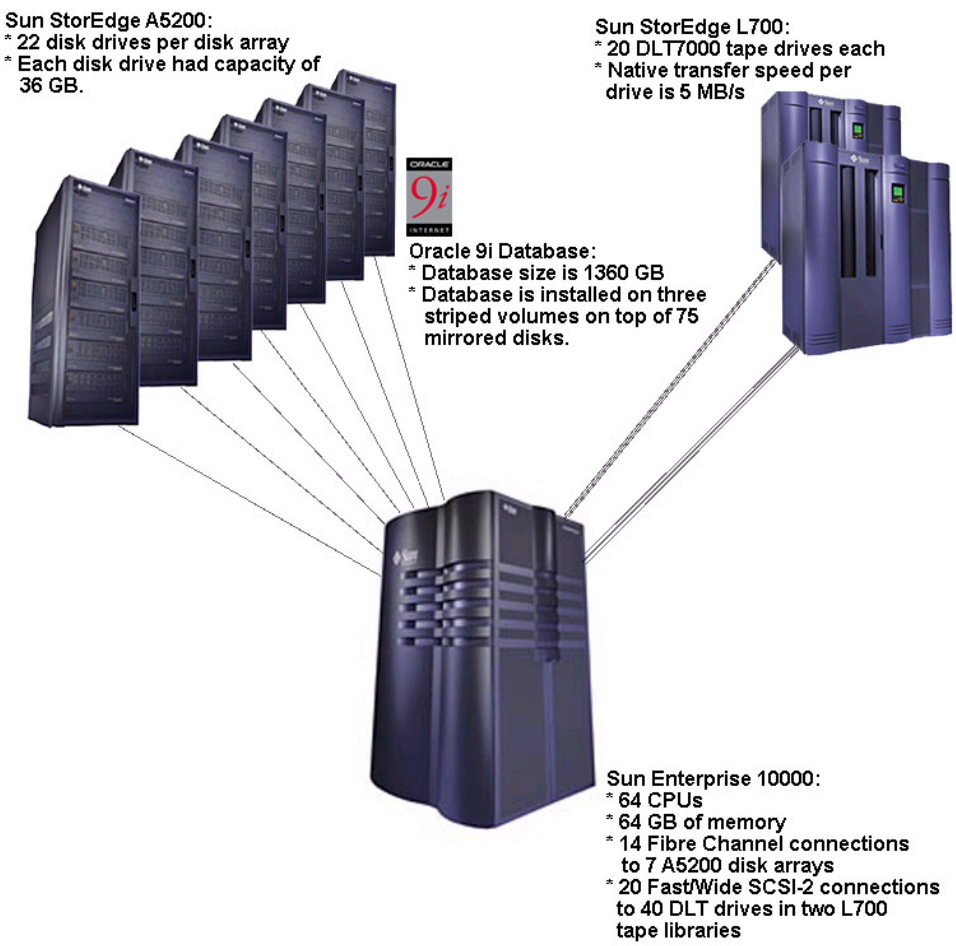
The Oracle database was Oracle9*i*, version 9.0.1.0. The database disk storage was configured as S.A.M.E.

We used Oracle9*i*, version 9.0.1.0, as the database server. The database size was 1360 GB, containing artificially generated data. The compression ratio of the data was 1.77:1 for the compression algorithm used by the DLT tape drives (Digital Level Ziv algorithm). The database files were located on three separate file systems. We used VERITAS File Systems™ with the VERITAS Quick I/O™ option. At the disk level, we mirrored 150 physical disks with VERITAS Volume Manager, resulting in 75 mirrored disks. The mirrored disks were then divided into three striped VERITAS volume groups. The volume groups were created as suggested in the S.A.M.E. paper ([JLSAME]), so that they were striped over all available disks with a stripe size of 1 MB.

Operating System and Backup Software

The operating system was Solaris8 with the Solaris8 Recommended Patch Cluster. The media management software was VERITAS NetBackup DataCenter 3.4 with the VERITAS NetBackup module for Oracle.

Figure 1: Hardware Configuration for RMAN Performance Testing



RESULTS AND FINDINGS

We tested RMAN backup and restore of the whole database, using varying numbers of tape drives, to demonstrate the scalability of RMAN. We also tested RMAN backup and restore commands with just 6 CPUs enabled to show that very fast backup and restore jobs can be done using a minimum of CPU resources.

During each run, we measured total average speed, average sustained speed, CPU utilization, disk and tape utilization, and other parameters. The **total average speed** is the backup/restore speed based on time differential between issuing a backup/restore command and the termination of the backup/restore. The **average sustained speed** is the rate of the actual data transfer. The graphs in this paper show only total average speed because the total average speed is the real-world measure of the actual backup and restore speed.

How Did We Tune?

Tuning backup and restore performance was easy and straightforward.

We tuned backup and restore performance according to the RMAN tuning paper (refer to [SDRMAN]). Therefore, we adhered to the following rules:

- The database files were striped over a large number of disks with a stripe size of 1 MB.
- The number of RMAN channels equaled the number of tape drives.
- We disabled RMAN multiplexing by setting the `MAXOPENFILES` parameter to `1`. The multiplexing of datafiles was unnecessary because the disk layout used the S.A.M.E. architecture.
- We used the `BLKSIZE` parameter to set the RMAN tape buffer size to 1 MB.
- We set the physical tape block size to 1 MB by entering `1048576` into the VERTIAS NetBackup configuration file `/usr/openv/netbackup/db/config/size_data_buffers.`
- We did not enable Oracle9*i* backup tape I/O slaves because the VERITAS NetBackup module for Oracle8*i* transfers data asynchronously.

The above changes indicate that you do not need to perform extensive tuning to achieve rapid RMAN backup and restore rates. RMAN tuning is simple and straightforward.

More details about hardware and software configuration can be found in the appendix at the end of this paper.

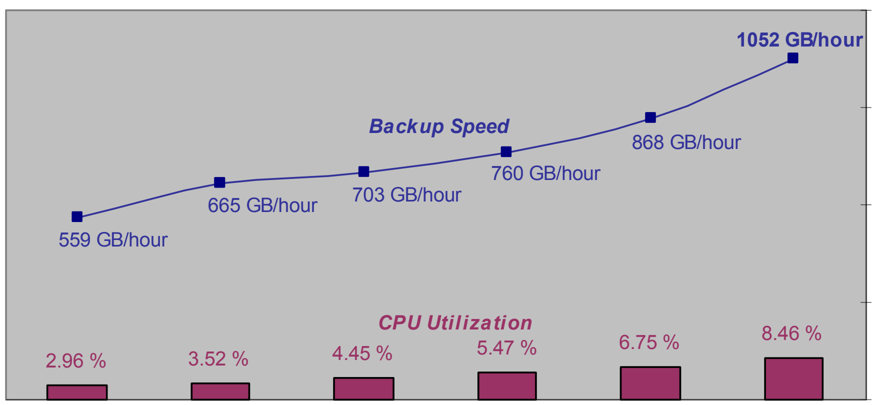
Results with All 64 CPUs enabled

The backup and restore speed on 40 tape drives exceeded 1 TB/hour.

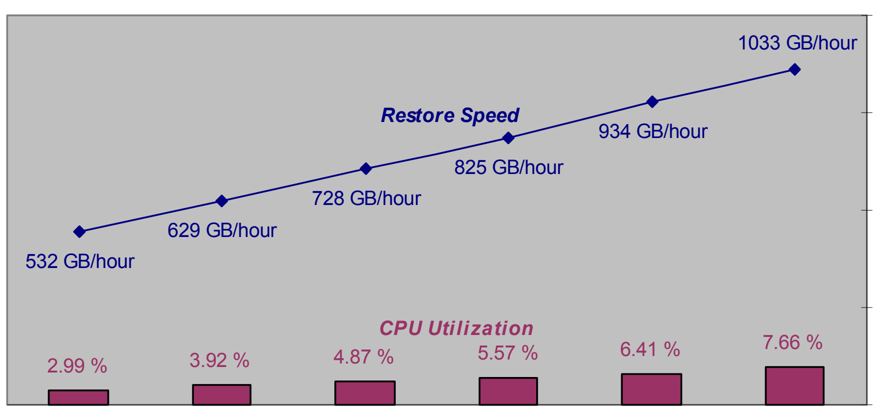
The tests with all 64 CPUs enabled showed that RMAN achieves a backup and restore speed of 1 TB/hour. The average sustained speed was even faster – 1.16 TB/hour with 40 tape drives. The tape drives were writing and reading data at their physical maximum speed of 8.3 MB/s. The tests showed that the CPU usage is very low – 8.6% for backups and 7.7 % for restores with 40 drives – which means that you can back up the database while it is online and actively used. Therefore, no backup window is required because RMAN does not consume excessive CPU cycles. Furthermore, RMAN online backups do not require the database to be placed in backup mode. The tests also showed that RMAN backup and restore speed increases linearly as you add tape drives. For example, a backup with 20 drives results in an average speed of 559 GB/hour with less than 3% CPU utilization.

We also measured the performance of RMAN validation, which is the process by which RMAN checks the integrity of all database blocks. The validation speed was 2 TB/hour with a CPU utilization of 10%.

The summary of the results is shown in the graphs below.



Parallelism	20	24	28	32	36	40
CPU Utilization	2.96 %	3.52 %	4.45 %	5.47 %	6.75 %	8.46 %
Speed	559 GB/hour	665 GB/hour	703 GB/hour	760 GB/hour	868 GB/hour	1052 GB/hour



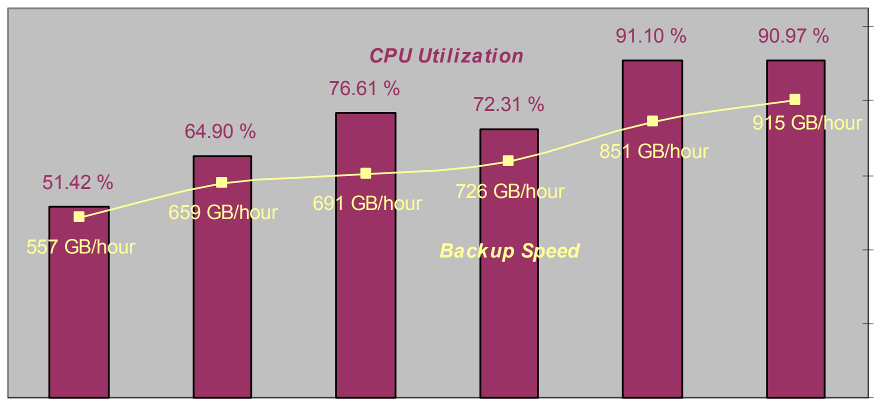
Parallelism	20	24	28	32	36	40
CPU Utilization	2.99 %	3.92 %	4.87 %	5.57 %	6.41 %	7.66 %
Speed	532 GB/hour	629 GB/hour	728 GB/hour	825 GB/hour	934 GB/hour	1033 GB/hour

The backup and restore speed on 40 tape drives approximated 1 TB/hour even with only six CPUs enabled.

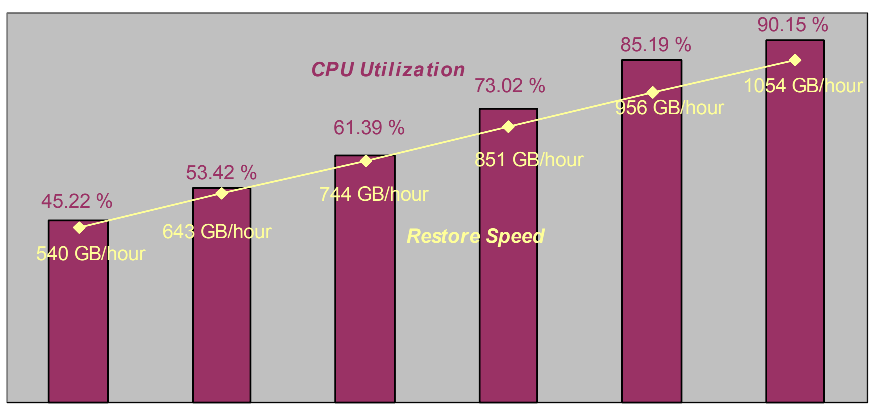
Results with 6 CPUs Enabled

We disabled 58 of the existing 64 CPUs to prove that RMAN can achieve extremely fast backup and restore using just a tiny fraction of the total processing power. With only 6 CPUs enabled, RMAN maintained a rate of 915 GB/hour with a total CPU utilization of about 91%. The sustained speed was 1037 GB/hour. The restore speed was the same as the backup speed. Thus, our tests showed that you can achieve high speed backup and restore using only a moderately powerful database (and backup) server. This also proves that the main requirement for fast backup/restore is a fast tape subsystem.

The summary of results is shown in the graphs below.



Parallelism	20	24	28	32	36	40
CPU Utilization	51.42 %	64.90 %	76.61 %	72.31 %	91.10 %	90.97 %
Speed	557 GB/hour	659 GB/hour	691 GB/hour	726 GB/hour	851 GB/hour	915 GB/hour



Parallelism	20	24	28	32	36	40
CPU Utilization	45.22 %	53.42 %	61.39 %	73.02 %	85.19 %	90.15 %
Speed	540 GB/hour	643 GB/hour	744 GB/hour	851 GB/hour	956 GB/hour	1054 GB/hour

Analysis of Results

The backup and restore bottleneck was the throughput of the DLT7000 drives

Our first priority was to determine the bottleneck in our system. In a correctly tuned backup/restore system, the tape drives should determine the total throughput. We found that this was indeed the case in this system. The results showed that the bottleneck was the throughput of the DLT7000 drives, which was slightly over 8 MB/s per drive. This speed is consistent with the technical specifications for DLT7000 drives: 5 MB/s is the native DLT7000 speed, and given that the data had a 1:1.77 compression ratio, the resulting speed is about 8 MB/s.

The average total throughput is less than the average sustained throughput because the latter excludes tape loading and position time as well as the startup of RMAN and NetBackup processes.

We also discovered that the average total throughput was 5-10% slower than the average sustained throughput. This difference occurred because the average total throughput included the time required for the Oracle processes to establish communication with NetBackup, for NetBackup to activate the 40 tape drives, and for tape loading and positioning. These initializations took approximately 6 minutes per run.

We also discovered that CPU utilization increases in a nonlinear fashion (2.8% for parallelism of 20, and 8.5% for 40 drives) in comparison with the backup speed. The nonlinear rate of increase in CPU utilization occurred because the drives were sharing SCSI connections, which created congestion that consumed CPU cycles. Therefore, we conducted additional tests with 8 to 16 tape drives and measured the CPU utilization when the drives do not share SCSI connections. The tests showed that if the drives do not share SCSI connections, then the CPU utilization grows linearly as we add more drives.

RMAN block validation uses about 40% of the total CPU cycles consumed by the backup.

We also examined the performance of Oracle9*i* RMAN data block validation. The tests showed that block validation consumes about 40% of the total CPU resources used by the backup.

Multiplexing of files installed on S.A.M.E does not improve performances.

It may seem that multiplexing files with VERITAS NetBackup or RMAN would improve performance – or, at least, do no harm. However, testing showed that datafile multiplexing (for example, with a level of 10) actually decreased performance. For example, a backup to 36 drives with a multiplexing level of 10 achieved 740 GB/hour – about 10% slower than a non-multiplexed backup. This is logical because the files are already striped on the disk volume level. In addition, we performed some tests with a non-S.A.M.E. disk layout. In that setup, each datafile was on a separate physical disk. These tests showed that, even when files are not striped, there is no need to multiplex them because a single disk is much faster than one tape drive.

CONCLUSION

In summary, our tests showed that RMAN:

- Achieves 1 TB/hour with minimal tuning efforts. This speed is achieved with less than 10% CPU utilization.
- Scales linearly as you add new tape drives.
- Achieves a speed of almost 1 TB/hour with only 6 CPUs enabled

These results prove that, in addition to its manageability benefits of automating database backup and recovery while maintaining high database availability, Oracle9*i* RMAN can efficiently utilize all of the available hardware in an enterprise-class backup system.

ACKNOWLEDGMENT

The testing environment was provided by Sun Customer Benchmark Center in Newark, CA. Veritas corp. provided volume and tape management software.

APPENDIX A: DETAILED DESCRIPTION OF TEST ENVIRONMENT

Hardware Configuration

The Sun Enterprise™ 10000 system used in the tests was configured with 64 SPARC 400 MHz processors on 16 independent system boards. Each processor can achieve up to 1.3 GB/s memory transfers and has 16 KB of non-blocking data cache. Each system board contained four memory banks of 1GB each, resulting in 64 GB of total memory. The system boards were linked with the Gigaplane-XB interconnect, which provides up to 12.8 GB/s of memory transfer bandwidth. The system had 4 SBuses per system board, resulting in 64 connections. Each SBus had 64-bit transfers with a maximum data throughput rate of 100 MB/s. Theoretically; the system can achieve 3.2 GB/s of I/Os.

The SBus connections were used as follows:

- The first and last connections were used for the `root` file system, which was mounted on D1000 disk arrays.
- One connection was used for the Ethernet network connection.
- 14 connections were used to connect seven A5200 disk arrays through an FC-AL split loop.
- 20 connections were used to connect 40 DLT7000 drives through an UltraSCSI Differential interface. The drives had only Fast Wide SCSI-2 interface, so UltraSCSI interface had maximum speed of 20 MB/s.

Oracle9*i* Database and RMAN Configuration

Our testing did not require special tuning of the OLTP parameters of the Oracle9*i* database. We used the default parameters created by the Oracle Database Configuration Assistant, and we made sure that some backup/restore related parameters were set correctly in the initialization parameter file:

```
#
# Ensure correct values for backup/restore parameters.  Note that
# these are the default values.
# - disable tape I/O slaves
# - enable asynchronous Disk I/O
backup_tape_io_slaves=false
disk_asynch_io=true
```

The Oracle9*i* RMAN configuration was straightforward and looked as follows:

```
CONFIGURE DEFAULT DEVICE TYPE TO SBT;
CONFIGURE DEVICE TYPE SBT PARALLELISM 40;
CONFIGURE CHANNEL DEVICE TYPE SBT MAXOPENFILES=1
          PARMS='BLKSIZE=1048576, ENV=(NB_ORA_CLASS=Oracle_backup)'
          FORMAT '%U_%t';
```

VERITAS NetBackup Configuration

We used the default VERITAS NetBackup Configuration except for increasing the VERITAS NetBackup tape block size to 1 MB in the file `/usr/openv/netbackup/db/config/size_data_buffers`. The default VERITAS NetBackup tape block size is only 64 KB, which did not give acceptable performance.

File System Configuration

All file systems with Oracle data were enabled with VERITAS Quick I/O™ and large file support:

```
mount -F vxfs -o largefiles,qio
                 /dev/vx/dsk/oracle/ora-volS /oracle/dataX
```

The file systems were also tuned as follows:

```
vxtunefs –o discovered_direct_iosz=1024k,
            initial_extent_size=16k,
            max_seqio_extent_size=1024k,
            max_direct_iosz=2048K,
            read_unit_io=1024K /oracle/dataX
```

Solaris8 Configuration

VERITAS NetBackup required that the parameter `TCP_TIME_WAIT_INTERVAL` of the TCP driver be set to 15000. We set this value with the Solaris command `ndd(1M)`.

The following changes to the `/etc/system` file were made to optimize the performance of the Veritas Volume Manager and File System:

```
**********************************************************************
*
* System kernel parameters.

* Soft Affinity - rechoose_interval
* This variable instructs the kernel on which cpu to select to run a
* thread if a choice needs to be made. So if a process hasn't run in
* rechoose_interval ticks it will be moved to another CPU. Otherwise
* it will continue to wait on the CPU it has been running on.
*
* A higher value of rechoose_interval "firms-up" the soft affinity.
* The down side is that you can end up with sluggish spreading out of
* processes on an application where a single processes forks a lots of
* children if this value is too high.
* A problem where threads migrate unnecessarily (Bugid 4103680) has
* been fixed in 2.7. The paramter should be set somewere between 3 and
* 300.
* DEFAULT: N/A
set rechoose_interval=30

* Enable kernel cage
* When set, this OS variable activates an OS mechanism that attempts
* to group OS memory that can't be relocated. This maximizes the
* number of CPU/memory boards that can be dynamically reconfigured.
* When unset, this variable allows permanent memory to spread
* throughout system memory, preventing the dynamically unconfiguration
```

```
* of most or all memory banks in the system. Activating this system
* variable can have a slight impact on application performance,
* depending on the application.
* DEFAULT: 0
set kernel_cage_enable=1

**********************************************************************
*
* VxVM System kernel parameters.

* vol_maxio
* IOs of a size larger than this are boken up in the Veritas VxVM
* layer. Physical IOs are broken up based on the disk capabilities are
* unaffected of the setting of the logical IO size.
* DEFAULT: 512 sectors
set vxio:vol_maxio=32768

* vol_maxkiocount
* This tunable controls the max amount of IOs that VxVM will
* do in parallel.
* DEFAULT 2048 IOs
set vxio:vol_maxkiocount=8192

* voliomem_kvmap_size
* This is a buffer for mapping IO memory. Must be larger than
* voliomem_max_memory.
* DEFAULT: 5 MB
set vxio:voliomem_kvmap_size=33554432

* vol_maxioctl
* The size of the largest ioctl that VxVM will handle.
* DEFAULT: 32 KB
set vxio:vol_maxioctl=131072

* vol_maxspecialio
* The size of the largest value handled by an ioctl. The ioctl itself
* may be small, but it can have requested a large IO operation.
* DEFAULT: 512 sectors (256KB)
set vxio:vol_maxspecialio=2048

* vol_default_iodelay
* Count in clock ticks that utilities will pause between issuing IOs,
* it they have been directed to throttle down speed but haven't been
* given a specific delay time. Utilities such as resyncronizing
* mirrors or rebuilding RAID-5 utilities will use this value.
* DEFAULT: 50 ticks
set vxio:vol_default_iodelay=5

************************************************************
* VxFS System kernel parameters
*

* lwp_default_stksize
* VxFS requires a stack size greater than the default 8K. The
* following values allow the kernel stack size for all threads to
* be increased to 16K.
* DEFAULT: 8 KB
set lwp_default_stksize=0x4000
```

APPENDIX B: REFERENCES

[SML700] Sun Microelectronics, Sun StorEdge L700 specifications, http://www.sun.com/storage/l700/

[SMA5200] Sun Microelectronics, Sun StorEdge™ A5200 specifications, http://www.sun.com/storage/a5200/

[SME10K] Sun Microelectronics, Sun Enterprise 10000 specifications, http://www.sun.com/servers/highend/10000/spec.html

[JLSAME] Juan Loaiza, Optimal Storage Configuration Made Easy, March 2000, http://technet.oracle.com/deploy/availability/pdf/oow2000_sane.pdf

[SDRMAN] Senad Dizdar, Tuning Oracle Recovery Manager, July 2001, http://otn.oracle.com/deploy/availability/pdf/rman_performance_wp.pdf

Oracle Recovery Manager: Performance Testing at Sun Customer Benchmark Center
October 2001
Author: Senad Dizdar
Contributing Authors: Steven Werthaimer, Lance Ashdown
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
www.oracle.com