inomass.com
Inomass
Data Security-Availability Consultants

inomass.com
Inomass
Data Security-Availability Consultants

Inomass GmbH
Moosbachstrasse 60b
8910 Affoltern a/A Switzerland

Where Information and Business Continuity goes on...

White Paper about Lan-Less-Backup and caching influences with Oracle and SUN Solaris

TABLE OF CONTENTS

1 IST-ANALYSIS

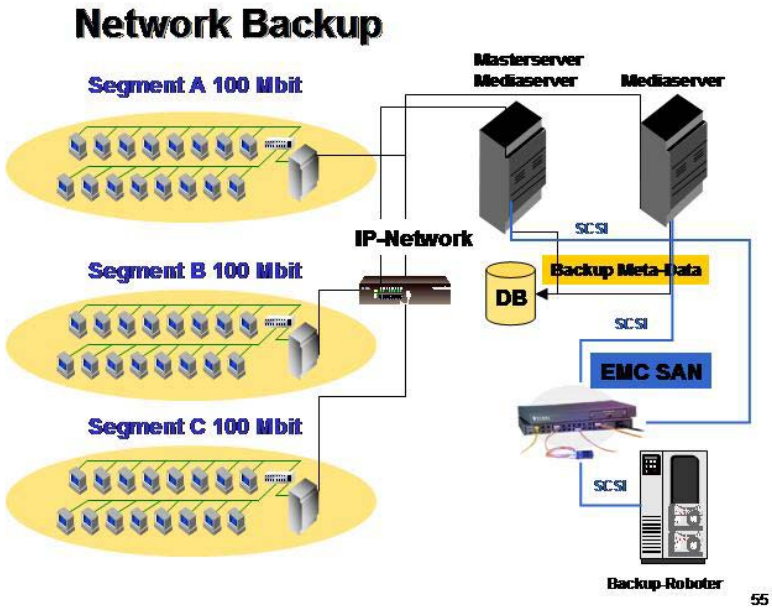
1.1 Backup Software

Backup environment based on Veritas Netbackup Datacenter on Windows NT 4.0. Approximately 100-200 Windows and UNIX servers are saved trough network to this central backup-server and one NT-media server. As network and backup speed is not sufficient, sometimes backup-jobs are failing. Slow network-speed has at consequence that tape devices kept to long busy even if only small amount of Data are written.

Veritas tries to prevent this problem with bundling multiple jobs (Multiplexing) to one (multiple) device, queuing but can't handle this number of clients.

The Verias Job queues get especially on weekend to large as than full-backups were done and not all jobs are finished until Monday morning. This is a problem of slow network-performance and not a problem of not enough drives.

1.1.1 Schematic set up in place



The central master server contains the volume Meta database, file indexes, robotic control, global device database host, job –configurations, schedules etc. If master server fails, backup/restores are not possible anymore. If master DB is not available no meta-data can be written/read and even with additional media-server no backup/restore will be possible. It is for this reason that a master-server should be only master server and not media-server as well, and be clustered to provide higher availability.

1.1.2 Master Server (Backup) -Server

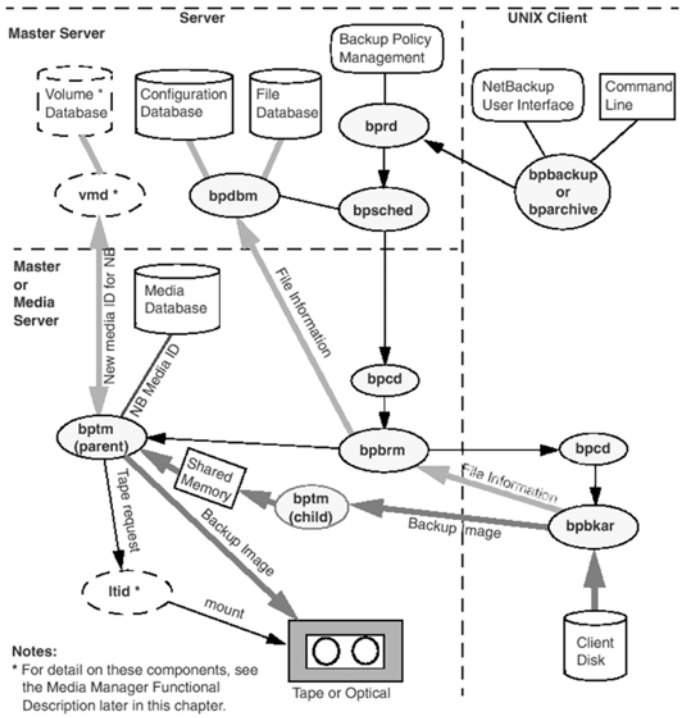
Master server is the central Server of all backup jobs, scheduling and configuration within a Netbackup Data-Zone. As mentioned master server mange's almost all jobs in a data-zone. Netbackup data-zone can contain many other media server with different operating systems set up. Job-scheduling will occur normally always from the master-server and mount or writing requests are then forwarded to the media server. By default master server is always media server too.
Always media-server sends metadata to the master server. This Master DB can get quite huge, depend if true image recover (move detection) is enabled and can represent about 0.5 % of saved Data x Version.

1.1.3 Master Server Performance sufficient?

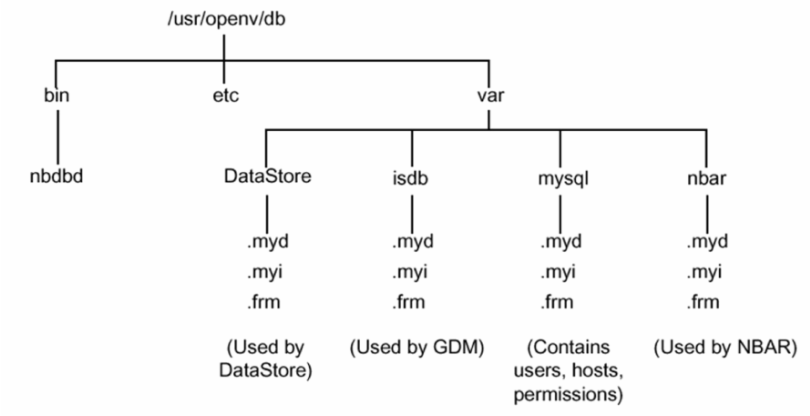
This point depend the number of installed master servers, we only discuss fundamentals.
To answer this question some theory about Veritas master server needs to be explained.
The Master Server is a real DB Server how writes for each file a record of save time, save volume, robot id and pointers/links to Full backups, if it was a incremental/differential backup.

Daemons writes to the master server > catalog.

Figure 8. Backup or Archive to Tape or Optical



In future the DB is called in Veritas term catalog, and has fallowing structure:



While most of the subdirectories in the NetBackup catalogs are relatively small, the images directory can grow to several tens, or even hundreds of gigabytes. (See "Determining Catalog Space Requirements" see in DC_Admin_Guide on page 174 for more information.

Determining Catalog Space Requirements

NetBackup requires disk space to store its error logs and information about saved files. The maximum amount of disk space that NetBackup requires at any given time varies according to the following factors:

- Number of files that you are backing up
- Frequency of full and incremental backups
- Number of user backups and archives
- Retention period of backups
- Average length of full pathname of files
- File information (such as owner permissions)
- Average amount of error log information existing at any given time
- Whether you have enabled the database compression option.

Veritas required per file (depend of path length) a average 150 byte per record and file.

Number of Files per System average	Nuber of Systems	Number of Versions	Bytes per record average path lenght	Total
			Bytes	MB
40000.00	100.00	1.00	150.00	572.20
40000.00	100.00	30.00	150.00	17166.14

The master server writes fallowing amount of data in the DB:

We presuming fallowing: (We only point up Full Backup on Weekend).

Backup-Window	Number of Files	Full/Incrementel	MB/sec for Master Server
	all Systems		
8 hour	4 mio (billion)	Full	19.86
16 hour	4 mio (billion)	Full	9.9

This calculation in not depend if we are doing LAN-Less, Network- backup or not, as all media-server writes always to the master server.

As log files are not taken in consideration we have to add about 10 % to these Values.

1.1.4 Media Server

The media-server is almost a master server itself. The only difference to the master server is the lack of the master db. The idea of media server was and is to do load-balancing and was implemented long time before SAN technology existed. In the early time of Veritas formerly "open vision" and in the newer 3 x version of NBU, Media Server cloud statically share tape libraries. Media-server can always access tape-device directly without to send data to master. But metadata for the Veritas DB are always send to the master server. Media Server can access Robotic and other devices directly but normally the global device database is on the master and therefore all requests for load/unload etc. comes from the master server.

1.1.5 SSO Option

The SSO or shared storage option is not a option itself as it is included in the default product, but needs additional licensing to activate. To use SAN infrastructure efficiently the already existing static library sharing from Veritas had to be changed to a more dynamically method. As in a SAN environment every Media Server can see all available tapes drives. Theoretically all media server cloud accesses all devices on the same time. This is unfortunately not possible and the device sharing is not a real sharing, but uses device sequentially one by one. Never two media server can access the devices concurrent in the same time.
If SSO is used all media server hosts must be in the same device sharing zone, as SSO uses a looking mechanism on SCSI bus. If WWN Ports are "zoned" from WNN to WNN SSO option would not work. Best is to setup a backup zone and to add all SAN backup WNN to this zone.

1.1.6 Conclusion

The Master Server needs just for catalog wring 10 MByte /sec, for full backups (within 16 hour). This means a 100 Mbit Interface should dedicated be available for catalog writing. If backup windows are even shorter, a 1 Gbit link is required.

But as today master server is in the meantime media server as well master is definitely overloaded, this is especially true for the Network interfaces.

1.2 Library / Tape Hardware

1.2.1 Tape Library

The used Compaq Tape Library ESL x (Manufactured by ATL) contains 16 SDLT Super DLT from Quantum connected to 2 x Fiber channel MDR (Modular Data Router) (Crossroads). This Library is accessed from two media server. Each media server accesses 8 SDLT Drives.

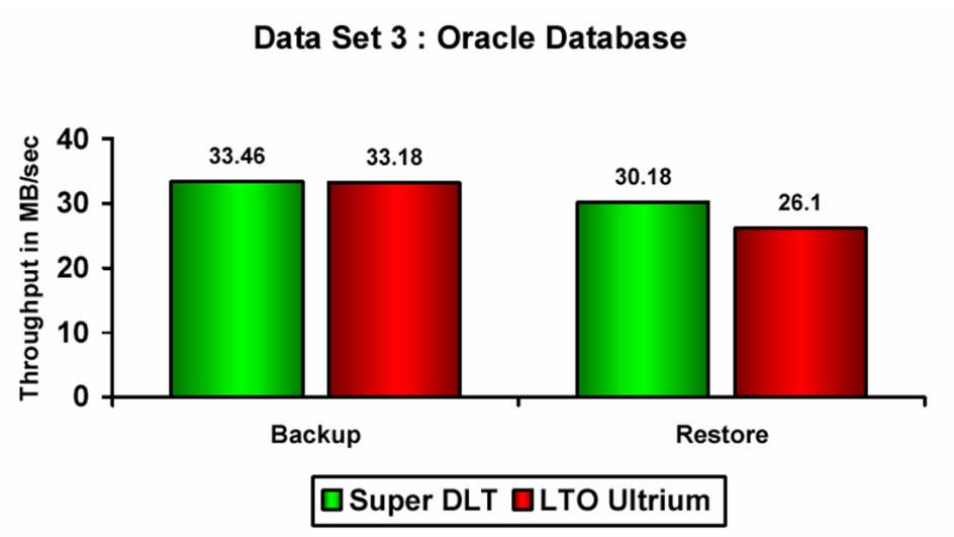
The Veritas master server is the robotic control server and handles all tape requests.

1.2.2 SDLT 220 Tape Drive Performance

Native sustained transfer rate	11 MB/s
2:1 compressed transfer rate	22 MB/s
Burst transfer rate SCSI bus	80 MB/s max
Native formatted capacity	110 GB
2:1 compressed capacity	220 GB
Cartridge load time to BOT, formatted tape	12 sec
to BOT, unformatted tape	40 sec
Average file access time	70 sec

Tape Performance with Oracle Data is about 25 – 35 MByte /sec per device, depend backup software etc.

Below backup performance analysis from Storage-Tek with arcserve shows the good SDLT drive compression algorithm on SDLT.



http://www.storagetek.com/pdfs/SDLT_tapetechWPaper.pdf

1.2.2.1 Compression Algorithm on SDLT for Oracle Data

Super DLTtape drives use the DLZ (Digital Lempel Ziv) compression algorithm. LTO Ultrium drives usethe ALDC (Adaptive Lossless Data Compression) algorithm. Both are so-called adaptive lossless compression techniques and variants of the LZ1 (Lempel-Ziv 1) class of data compression algorithms,first proposed by Abraham Lempel and Jacob Ziv in 1977. The patents to LZ1 are now owned by Hi/fn, the world's leading vendor of compression technology. DLZ is Hi/fn's implementation of the LZ1algorithm, based on improvements made to the original Lempel Ziv work by Stac Electronics [4] . ALDC is IBM's proprietary implementation of the LZ1 algorithm, which it licenses from Hi/fn [5] .

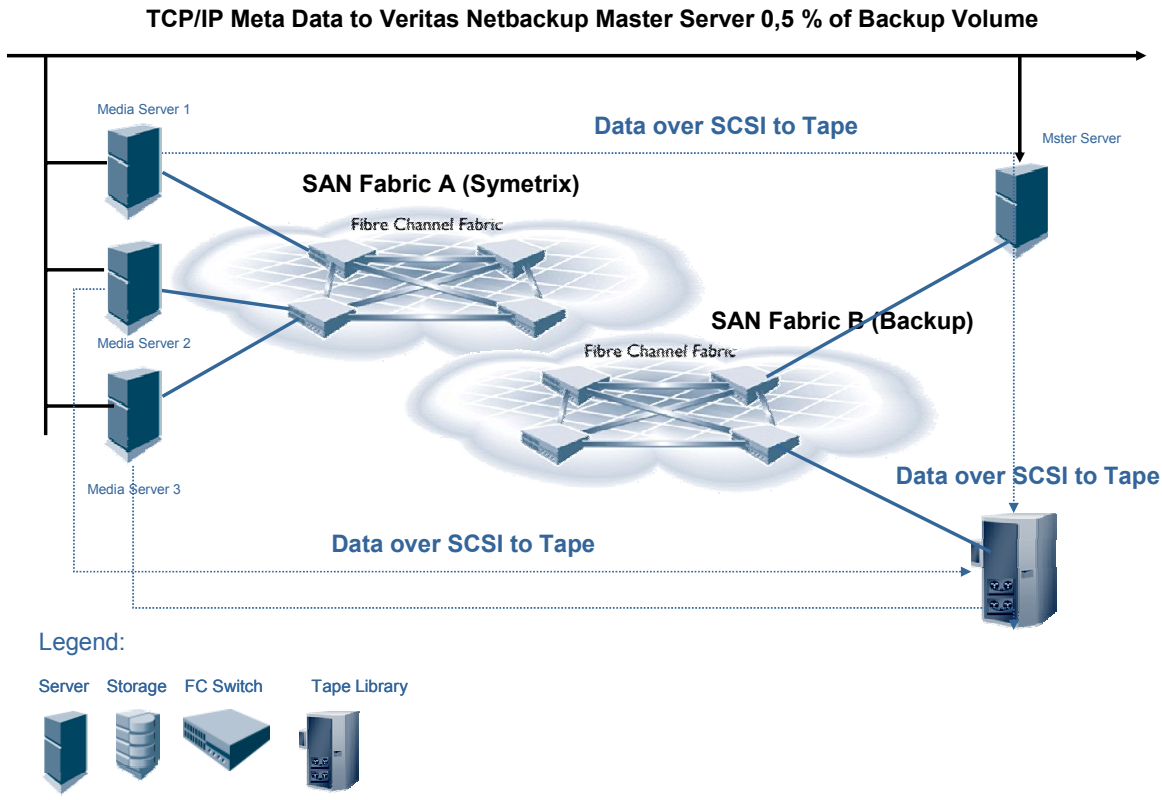
Hi/fn's research indicates that the DLZ (SDLT) compression algorithm is marginally better than ALDC (LTO Ultrium) at compression ratios of 3:1 and below, and slightly worse at higher compression ratios of 4:1 and above. DLZ is optimized for compression ratios of 3:1 or less since this represents the vast majority of real-world situations.

This is the reason that SDLT drives are better suited for Oracle Backup than LTO Ultrium (not applicable on the new 30/60 LTO generation)!

2 LAN LESS BACKUP PRINCIPE

Lan-Less Backup was promoted in the industry for about 2 – 3 Years ago. Only with the upcoming SAN infrastructure backup software can use this "relative new" technology to share devices. Devices can be shared in a SAN environment but never two Veritas Media Server can access the same device at the same time. Technically, device accesses devices sequential. Therefore other backup jobs have to wait until the device locking process releases device.

Veritas Media Server sends data like to a local attached tape device over SAN/SCSI protocol to the central tape device. Meta Data travels still trough LAN (TCP-IP) to the central Master Server where Catalog data is written.



3 PERFORMANCE RESULTS WITH LAN- LESS-BACKUP AND (*UNCACHED) UFS FILE SYSTEMS

(*File systems were originally mounted with no I/O cache and therefore suffered from high backup-speed degradation).

All test were done with a tar oracle dbf test file from a Oracle DB. The File Size was 5859 MB (6 GB). This File was copied on local SCSI attached Disk and as well on the EMC standard disk system attached trough fiber channel.

Unfortunately we lost one measurement with up to 22-25 MByte/s with vxfs and cache enabled.

3.1 Host Media Server lamir

Host_Name	Mount_Point	File	Command	time to compete
lamir	/data/oracle/u01#	backup_file	time bpbackup -w -L /app/mylog Backup_file	6m20.08s

3.1.1 Lan-Less Backup performance

5859 MB / 380 sec 15.4 MB /sec

File System was on lamir mounted with ufs *with* caching.

listing vfstab

```
/dev/vx/dsk/mispapp_dg/oradata02   /dev/vx/rdsk/mispapp_dg/oradata02   /data/oracle/u02   ufs   2   yes   largefiles
/dev/vx/dsk/mispapp_dg/oradata03   /dev/vx/rdsk/mispapp_dg/oradata03   /data/oracle/u03   ufs   2   yes   largefiles
/dev/vx/dsk/mispapp_dg/oradata04   /dev/vx/rdsk/mispapp_dg/oradata04   /data/oracle/u04   ufs   2   yes   largefiles
/dev/vx/dsk/mispapp_dg/oradata05   /dev/vx/rdsk/mispapp_dg/oradata05   /data/oracle/u05   ufs   2   yes   largefiles
/dev/vx/dsk/mispapp_dg/oradata06   /dev/vx/rdsk/mispapp_dg/oradata06   /data/oracle/u06   ufs   2   yes   largefiles
/dev/vx/dsk/mispapp_dg/oradata07   /dev/vx/rdsk/mispapp_dg/oradata07   /data/oracle/u07   ufs   2   yes   largefiles
/dev/vx/dsk/mispapp_dg/oradata08   /dev/vx/rdsk/mispapp_dg/oradata08   /data/oracle/u08   ufs   2   yes   largefiles
/dev/vx/dsk/mispapp_dg/oradata09   /dev/vx/rdsk/mispapp_dg/oradata09   /data/oracle/u09   ufs   2   yes   largefiles
/dev/vx/dsk/mispapp_dg/oradata10   /dev/vx/rdsk/mispapp_dg/oradata10   /data/oracle/u10   ufs   2   yes   largefiles
/dev/vx/dsk/mispapp_dg/oradata11   /dev/vx/rdsk/mispapp_dg/oradata11   /data/oracle/u11   ufs   2   yes   largefiles
/dev/vx/dsk/mispapp_dg/oradata12   /dev/vx/rdsk/mispapp_dg/oradata12   /data/oracle/u12   ufs   2   yes   largefiles
/dev/vx/dsk/mispapp_dg/oradata13   /dev/vx/rdsk/mispapp_dg/oradata13   /data/oracle/u13   ufs   2   yes   largefiles
/dev/vx/dsk/mispapp_dg/oradata14   /dev/vx/rdsk/mispapp_dg/oradata14   /data/oracle/u14   ufs   2   yes   largefiles
```

3.2 Host Media Server lafleur

Host_Name	Mount_Point	File	Command	time to compete	MB/sec
Eval Diskspeed Standard Volume					
lafleur	/data/oracle/u01	backup_file	cp backup_file - > /dev/null	2m49.36	34 MB/s
lafleur	data/oracle/u01	backup_file	time bpbackup -w -L /app/mylog Backup_file	5m24.80s	18 MB/s

3.2.1 Lan-Less Backup performance

SAN Backup performance = 18 MB/sec

File System was on lafleur was mounted with ufs and forced direct I/O (no cache).

listing from vfstab

```
/dev/vx/dsk/misprpt_dg/oradata01   /dev/vx/rdsk/misprpt_dg/oradata01   /data/oracle/u01  ufs  2  yes  largefiles,forcedirectio
/dev/vx/dsk/misprpt_dg/oradata02   /dev/vx/rdsk/misprpt_dg/oradata02   /data/oracle/u02  ufs  2  yes  largefiles,forcedirectio
/dev/vx/dsk/misprpt_dg/oradata03   /dev/vx/rdsk/misprpt_dg/oradata03   /data/oracle/u03  ufs  2  yes  largefiles,forcedirectio
/dev/vx/dsk/misprpt_dg/oradata04   /dev/vx/rdsk/misprpt_dg/oradata04   /data/oracle/u04  ufs  2  yes  largefiles,forcedirectio
/dev/vx/dsk/misprpt_dg/oradata05   /dev/vx/rdsk/misprpt_dg/oradata05   /data/oracle/u05  ufs  2  yes  largefiles,forcedirectio
/dev/vx/dsk/misprpt_dg/oradata06   /dev/vx/rdsk/misprpt_dg/oradata06   /data/oracle/u06  ufs  2  yes  largefiles,forcedirectio
```

3.3 Host Media Server lesert

Host_Name	Mount_Point	File	Command	time to compete	MB/sec
Eval Diskspeed Standard Volume					
lesert	/data/oracle/u06	backup_file	time cp backup_file - > dev/null	3m30.59s	28 MB/s
lesert	/data/oracle/u06	backup_file	time bpbackup -w -L /app/mylog Backup_file	5m22.80s	18 MB/s

3.3.1 Lan-Less Backup performance lesert

SAN Backup performance = 18 MB/sec

File System was on lesert was mounted with vxfs and forced direct I/O (no cache).

```
/dev/vx/dsk/mistrpt_dg/oradata01   /dev/vx/rdsk/mistrpt_dg/oradata01   /data/oracle/u01  vxfs  2  yes  convosync=direct
/dev/vx/dsk/mistrpt_dg/oradata02   /dev/vx/rdsk/mistrpt_dg/oradata02   /data/oracle/u02  vxfs  2  yes  convosync=direct
/dev/vx/dsk/mistrpt_dg/oradata03   /dev/vx/rdsk/mistrpt_dg/oradata03   /data/oracle/u03  vxfs  2  yes  convosync=direct
/dev/vx/dsk/mistrpt_dg/oradata04   /dev/vx/rdsk/mistrpt_dg/oradata04   /data/oracle/u04  vxfs  2  yes  convosync=direct
/dev/vx/dsk/mistrpt_dg/oradata05   /dev/vx/rdsk/mistrpt_dg/oradata05   /data/oracle/u05  vxfs  2  yes  convosync=direct
/dev/vx/dsk/mistrpt_dg/oradata06   /dev/vx/rdsk/mistrpt_dg/oradata06   /data/oracle/u06  vxfs  2  yes  convosync=direct
/dev/vx/dsk/mistrpt_dg/oradata07   /dev/vx/rdsk/mistrpt_dg/oradata07   /data/oracle/u07  vxfs  2  yes  convosync=direct
```

3.4 File system influence I/O

The performance can vary depend if file system is mounted with I/O caching or not. This is critical for file system backup and Oracle RMAN.

For normal file system-backup it cloud be critical, as for example ufs with no caching, close fiber channel port, and generate about 4000 I/O /sec. In this case, backup-speed decreases down to 600 KB/sec.

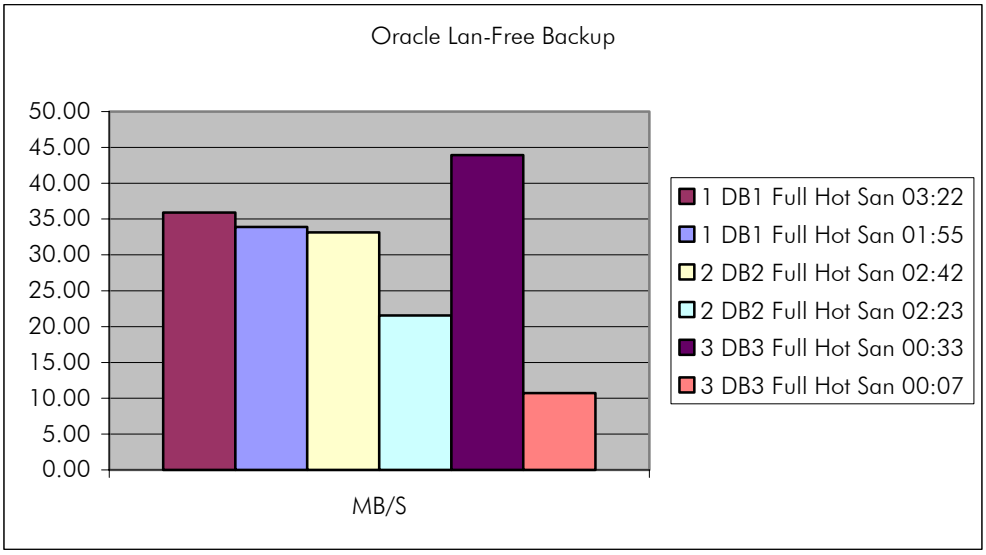
We recommend therefore mounting file system with ufs/vxfs cached.

4 ORACLE PERFORMANCE

After changing the file system from ufs/vxfs (no cached) to standard file system behavior (cache enabled) backup performance was much better.

4.1 Performance results

Some DB's are probably slower than 35 MB/sec because of nightly feeds/jobs. While testing isolated DB's per machine, we never had slower performance than 34-39 MB/sec. As can be see on Server 3 potential for even higher speed is there.



Server	*DB*	*Type*	*Dest*	*Duration*	*Size*	*Gbyte*	*MB/S*
1	DB1	Full Hot	San	03:22	435		35.89
1	DB1	Full Hot	San	01:55	234		33.91
2	DB2	Full Hot	San	02:42	322		33.13
2	DB2	Full Hot	San	02:23	185		21.56
3	DB3	Full Hot	San	00:33	87		43.94
3	DB3	Full Hot	San	00:07	4.5		10.71