HP WORLD 2004
Solutions and Technology Conference & Expo

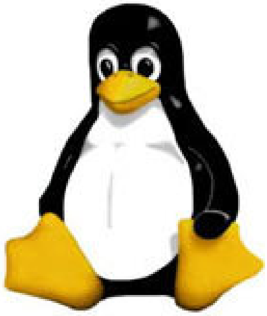
Oracle RAC on RedHat Linux
Best Practice guide and optimization

Yann Allandit
Oracle pre-sales consultant
Hewlett-Packard

Michael Aubertin
Linux consultant
ARES

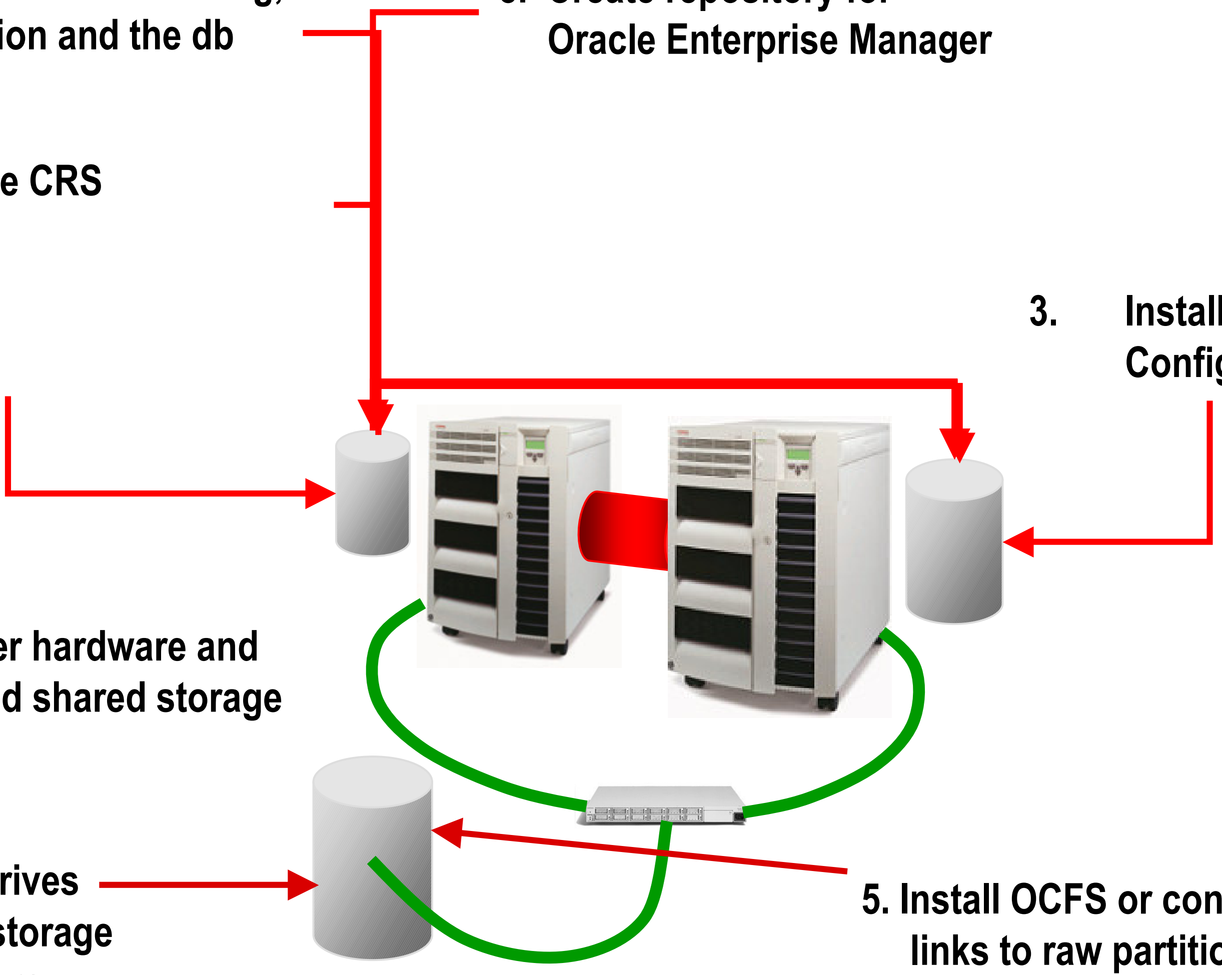
Summary

1. Hardware Pre-Requisites
2. Tune BIOS Configuration
3. Lun Definition
4. File System Features
5. I/O Handler in Linux Kernel
6. Presentation of the /proc Mechanism
7. Memory Handler
8. Cache Mechanism Overview
9. Presentation of NTPL
10. Kernel Parameters Tuning
11. RedHat Linux Installation For Oracle
12. HP Driver Installation
13. OS update Policy
14. Oracle IO and Settings
15. Failover/Failback tuning
16. Diagnostic Utilities
17. Linux Customization



Hardware Pre-Requisites

Installation overview



7. Run OUI to install Oracle 10g, the RAC option and the db

8. Create repository for Oracle Enterprise Manager

6. Install Oracle CRS

2. Install OS on NODE1

3. Install OS on NODE2 Configure interconnect

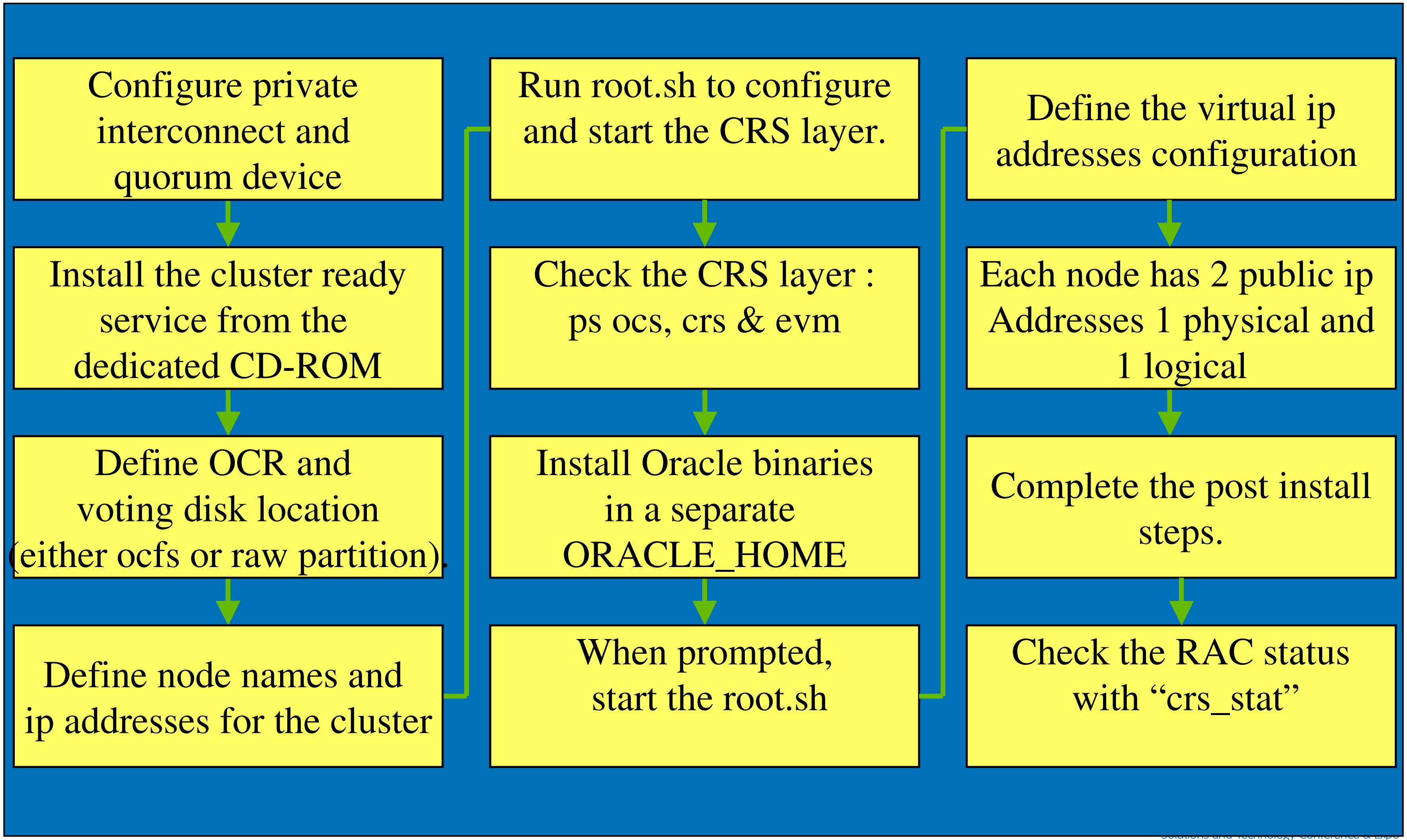
1. Configure cluster hardware and prepare local and shared storage

4. Create logical drives on the shared storage and leave them raw

5. Install OCFS or configure symbolic links to raw partitions

Click or press any key to begin slide animation.

Installation Flowchart for 10g RAC



Configure private interconnect and quorum device

Install the cluster ready service from the dedicated CD-ROM

Define OCR and voting disk location (either ocfs or raw partition).

Define node names and ip addresses for the cluster

Run root.sh to configure and start the CRS layer.

Check the CRS layer : ps ocs, crs & evm

Install Oracle binaries in a separate ORACLE_HOME

When prompted, start the root.sh

Define the virtual ip addresses configuration

Each node has 2 public ip Addresses 1 physical and 1 logical

Complete the post install steps.

Check the RAC status with “crs_stat”

Support consideration

Because tunning must be a custom optimization of a known system and ever be, an improvement of one system.

- Always use supported software parts
- Plan to use the last update for server bios.
- Run the last version of FC/AL adapter bios.

Overview of available products – ia32

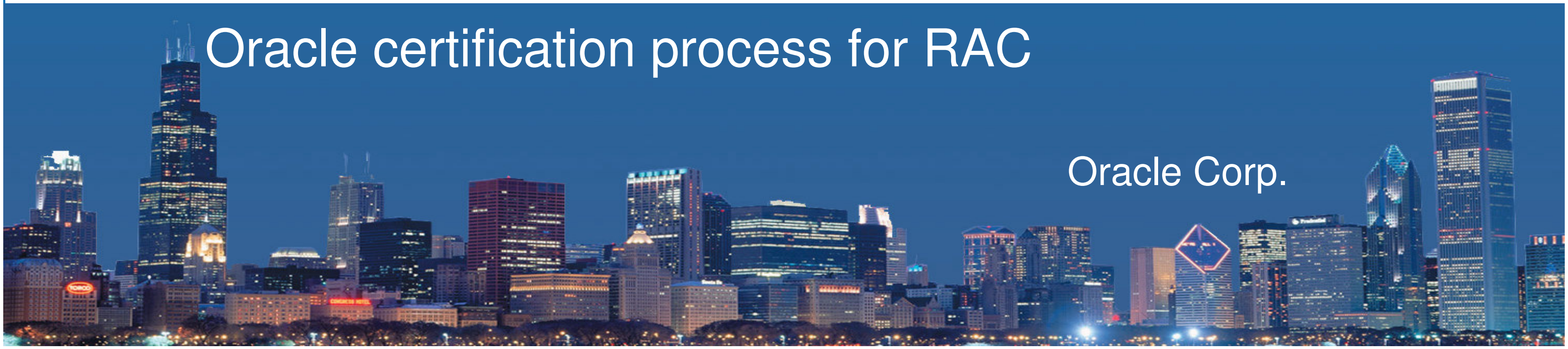
ORACLE	redhat RH 2.1	redhat RH 3	suse simply change SuSE SLES 8
Oracle DB + RAC	☑	☑	☑
Oracle e.business suite	☑	☑	☑
Oracle Application server	☑		☑
Oracle Collaboration Suite	☑	✓	☑

Simplified Approach

For RAC, Oracle took hardware out of the certification equation. They certify the Real Application Clusters, clusterware, and Operating System versions. Certification would now be awarded to a combination like Oracle RAC running against RedHat 2.1.

The specific hardware is no longer part of the certification. To communicate what customers need to use, Oracle provide a list of technologies that are compatible with RAC. Oracle will support the Oracle software on clusters that are comprised of compatible technology running on certified O/S and clusterware combinations.

Oracle recommends customers confirm that their vendors will support the hardware in their cluster, as not all vendors will choose to support all possible combinations. With these changes, Oracle is getting out the hardware business. Discussions about hardware support will no longer involve Oracle. Oracle is focusing on the O/S, cluster ware, and RAC combination.

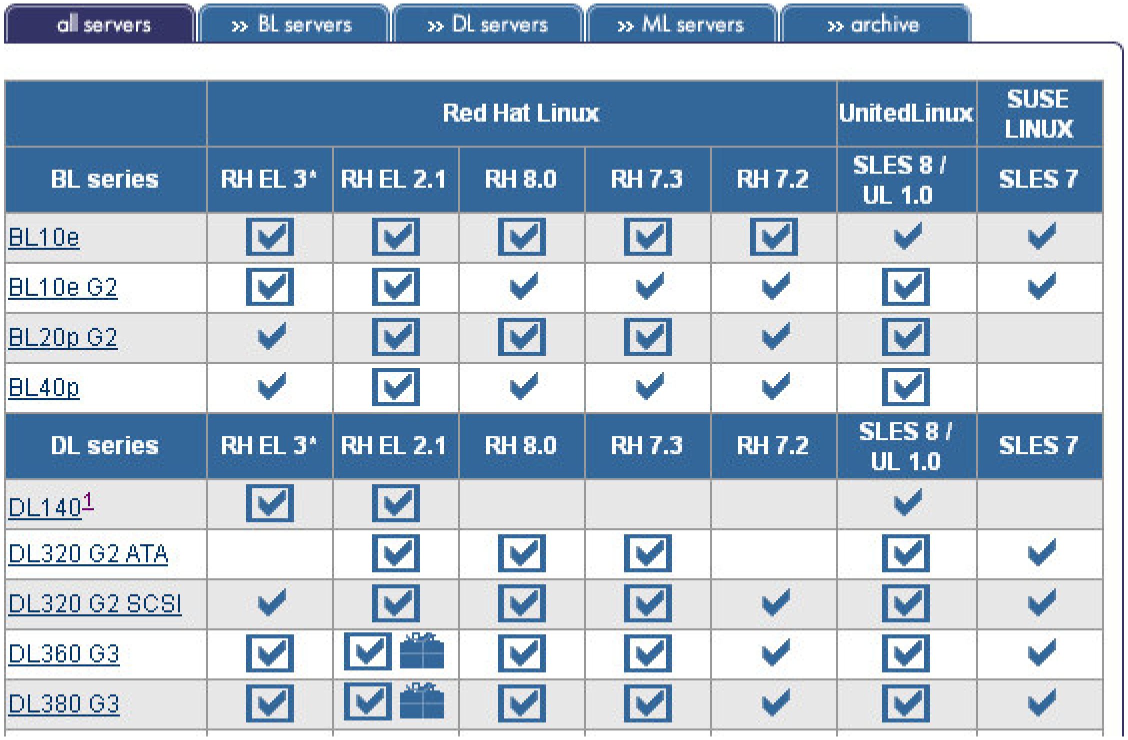


Oracle certification process for RAC

Oracle Corp.

HP Server Linux certification

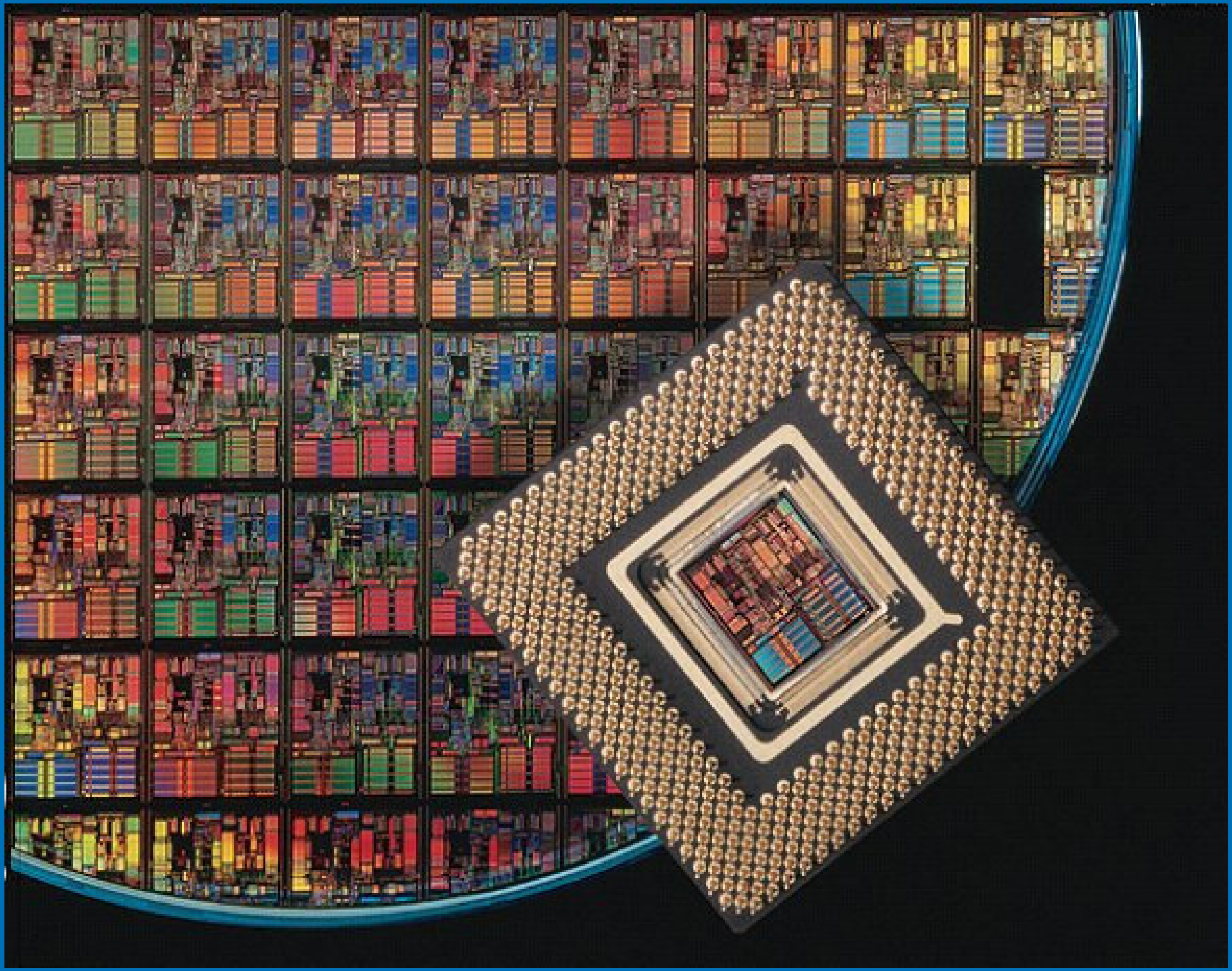
Check for the server certification with your HP representative or within the web site below.



	Red Hat Linux					UnitedLinux	SUSE LINUX
BL series	RH EL 3*	RH EL 2.1	RH 8.0	RH 7.3	RH 7.2	SLES 8 / UL 1.0	SLES 7
BL10e	☑	☑	☑	☑	☑	✓	✓
BL10e G2	☑	☑	✓	✓	✓	☑	✓
BL20p G2	✓	☑	☑	☑	✓	☑	
BL40p	✓	☑	✓	✓	✓	☑	
DL series	RH EL 3*	RH EL 2.1	RH 8.0	RH 7.3	RH 7.2	SLES 8 / UL 1.0	SLES 7
DL140[1]	☑	☑				✓	
DL320 G2 ATA		☑	☑	☑		☑	✓
DL320 G2 SCSI	✓	☑	☑	☑	✓	☑	✓
DL360 G3	☑	☑ 🎁	☑	☑	✓	☑	✓
DL380 G3	☑	☑ 🎁	☑	☑	✓	☑	✓

http://h18004.www1.hp.com/products/servers/linux/hpLinuxcert.html

Tune BIOS Configuration



Bios configuration

Reliability is a good part of an optimized cluster. So, according to the server matrix and SAN investigation, we have to:

- Hard set the FC/AL adapter throughput.
- Hard set SAN access mode (ie: loop, fabric...)
- Hard set any timeout.
- <troll>Disable hyper-threading</troll>.

Lun Definition

Lun definition

LUN definition should be made with hardware optimization mind set, neither with application specification nor human capabilities brain.

To release data jam in bottleneck, we have to use multiple way.

Don't forget that RAC is first a database server. That is to say his first job is serving data, not computing. So data access is the real battle field.

I/O architecture and usage investigation.

One success factor of an optimized Oracle RAC on Linux cluster is this investigation. In deed you have to know how data will be serve to:

- Determine writing and reading ratio.
- May be change SAN design.
- Define throughput and SAN access method.

Redundant Storage Sets (RSS)

- Sub-grouping of disks for failure separation
- Redundancy information contained within RSS
- Disk failure in one RSS does not affect other RSS groups
- Managed by controller firmware
- Target size = 8 disks
  - Minimum = 6
  - Maximum = 11

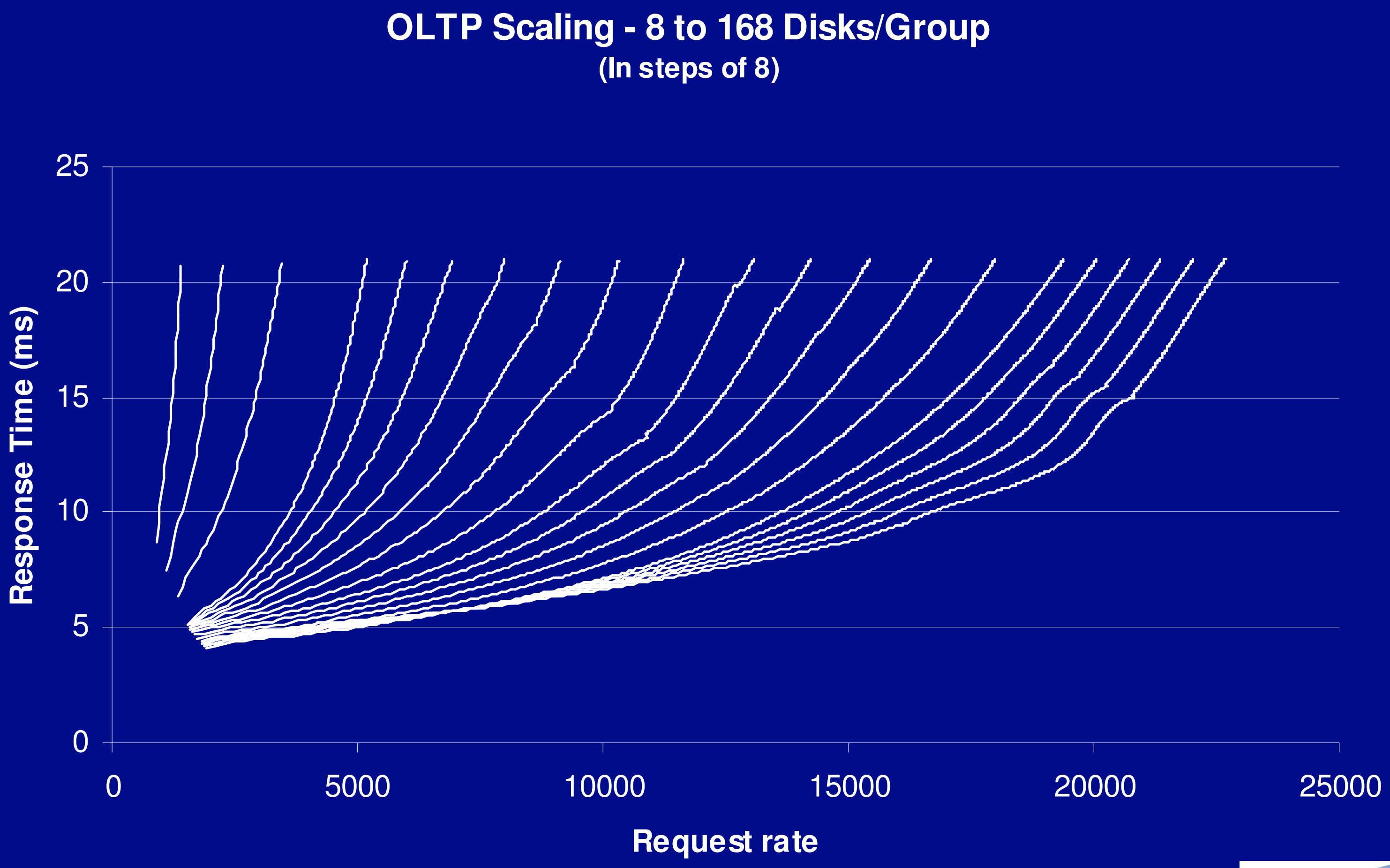
Application Workloads

I/O Profile	**VRaid 1 10K**	**VRaid 1 15K**	**VRaid 5 10K**	**VRaid 5 15K**
“Database”: 8 KB, 67% reads, random Req/Sec @ 30 ms	19,600	25,400	13,500	17,900
“Web server”: 8 KB, 100% reads, random Req/Sec @ 30 ms	21,100	29,900	21,100	29,700

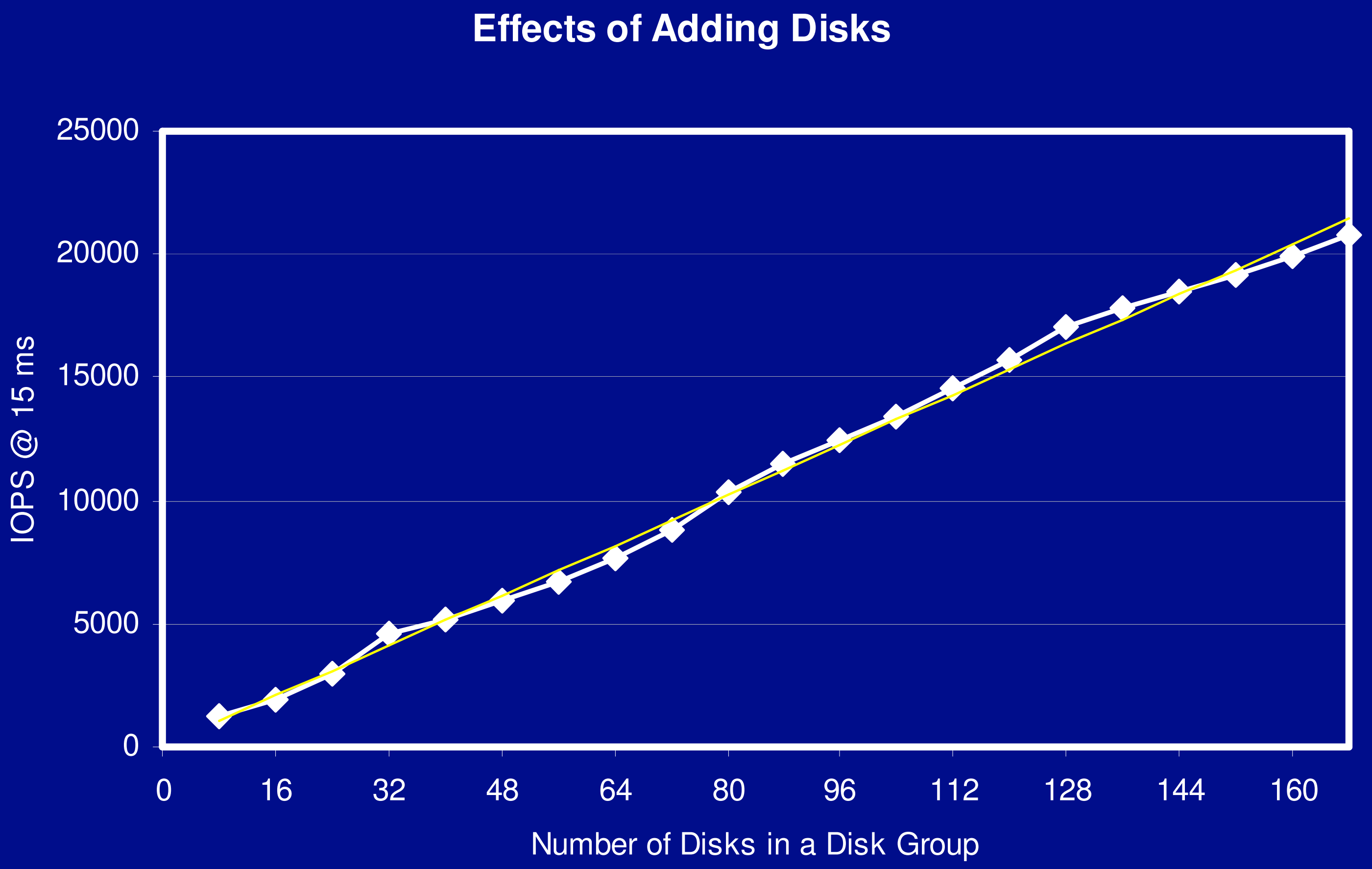
Performance Best Practices

- Single group is best
  - Spreads I/O evenly across all disks
- Use 15K RPM disks?
  - 30% to 40% faster on random access workloads
  - 15K disks may be more expensive…
    - More 10K disks for same cost

Effects of Larger Disk Groups



More Disks = More Performance



File System Features

Choosing the good file system

Redhat Linux supported file systems are :

- Ext2: “to finish”

Choosing the good file system

Redhat Linux supported file systems are :

- Ext3: “to finish”

Choosing the good file system

Redhat Linux supported file systems are :

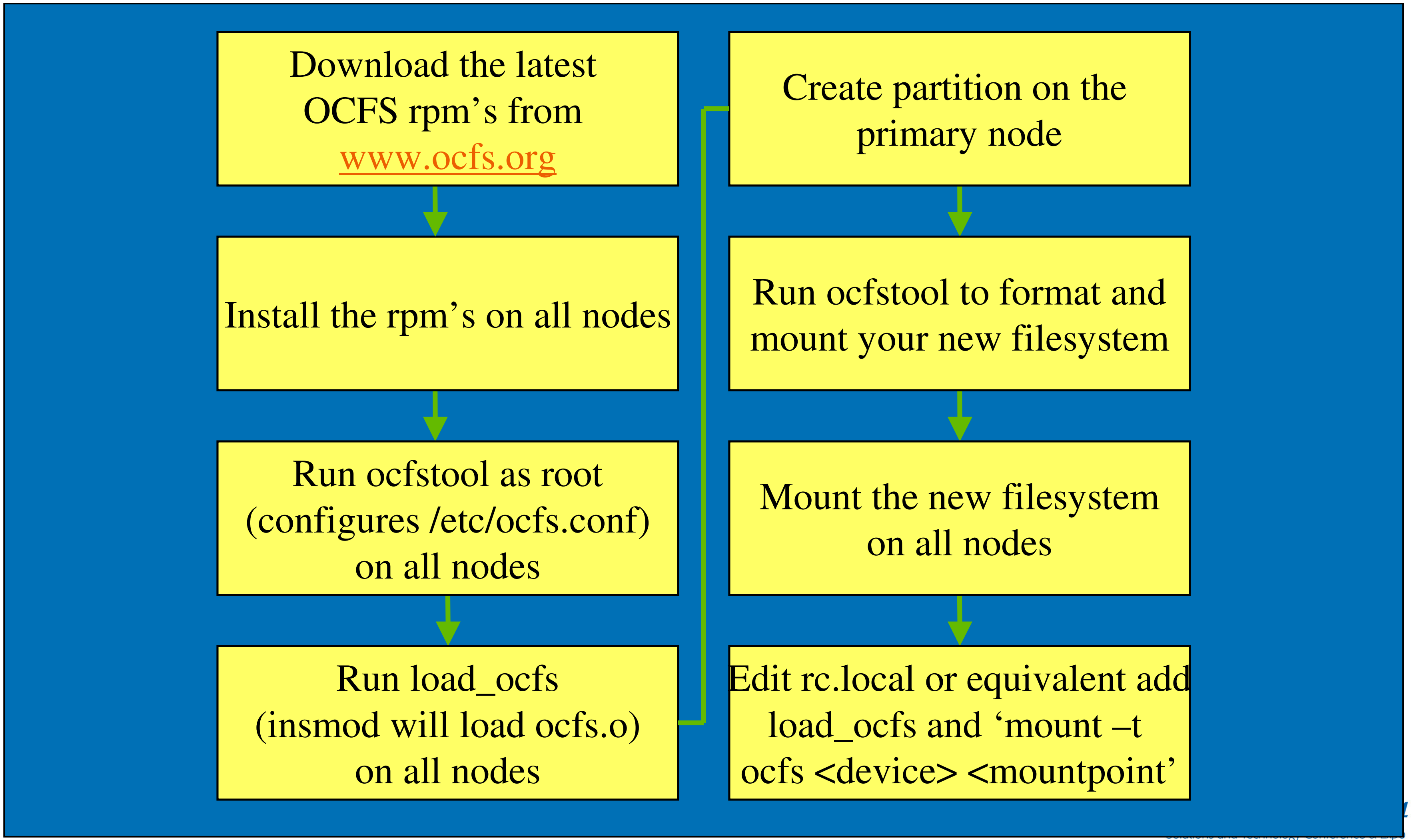
- Rawdevices: “to finish”

Choosing the good file system

Other often view file systems are :

- Reiserfs: “to finish”
- xfs: “to finish”

Installation Flowchart for OCFS



Download the latest OCFS rpm's from www.ocfs.org

Install the rpm's on all nodes

Run ocfstool as root (configures /etc/ocfs.conf) on all nodes

Run load_ocfs (insmod will load ocfs.o) on all nodes

Create partition on the primary node

Run ocfstool to format and mount your new filesystem

Mount the new filesystem on all nodes

Edit rc.local or equivalent add load_ocfs and 'mount –t ocfs <device> <mountpoint>'

OCFS and unbreakable Linux

Redhat

- Currently ships 4 flavors of the AS 2.1 kernel, viz., UP, SMP, Enterprise and Summit (IBM x440)
- Oracle provides a separate OCFS module for each of the kernel flavors
- Minor revisions of the kernel do not need a fresh build of ocfs
- e.g., ocfs built for e.12 will work for e.16, e.18, etc.

United Linux/Suse

- United Linux ships 3 flavors of its kernel, for the 2.4.19-64GB-SMP, the 2.4.19-4GB and the 2.4.19-4GB-SMP kernel
- OCFS 1.0.9 is supported on UL 1.0 Service Pack 2a or higher
- OCFS build is not currently upward compatible with kernel (pre SP3) ➔ must ensure OCFS build exists for each new Kernel version prior to upgrading kernel

OCFS and RAC

- Maintains cache coherency across nodes for the filesystem metadata only
- Does not synchronize the data cache buffers across nodes, lets RAC handle that OCFS journals filesystem metadata changes only
- Filedata changes are journalled by RAC (log files)
- Overcomes some limitations of raw devices on Linux
    - No limit on number of files
    - Allows for very large files (max 2TB)
    - Max volume size 32G (4K block) to 8T (1M block)
- Oracle DB performance is comparable to raw

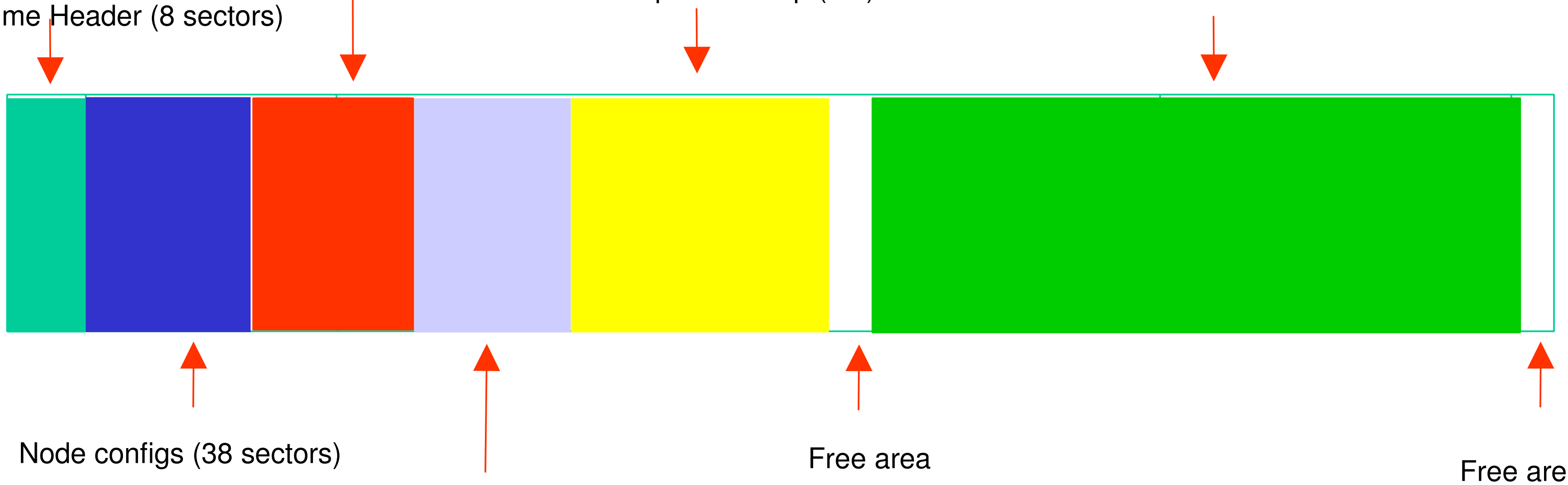
OCFS Best Practice

- Format with 128 KB block size. Sizes between 2 KB and 1 MB are supported. Smaller blocksize will have performance penality.
- Place archive log on a separate disk partition to avoid contention.
- OCFS requires contiguous space on disk for initial datafile creation or for any extension.
- OCFS recommend to have hardware RAID support.
- Avoid large numbers of mount points as this tends to create a performance bottleneck (typically < 50)
- OCFS will print error and debug information in system log (/var/log/messages & dmesg).
- Update fstab and modules.conf for automatic starting

Install tips for OCFS

- Ensure OCFS rpm corresponds to kernel version
    - uname –r (i.e. 2.4.19-4GB)
- Remember to also download rpm’s for OCFS “Support Tools” and “Additional Tools”
- Download the dd/tar/cp rpm that supports o_direct
- Use rpm –Uv to install all 4 rpm’s on all nodes
- Use OCFS for Oracle DB files only, not Oracle binaries (OCFS 1.0.x was not designed as a general purpose filesystem).

OCFS Linux – Volume layout



Volume Header (8 sectors)

Publish area (32 sectors)

Space bitmap (1M)

Data blocks

Node configs (38 sectors)

Vote area (32 sectors)

Free area

Free area

Note: Not drawn to scale

Performance ocfs vs raw

Database	Number of nodes	Number of Users	Average Think time	CPU utilisation	Transactions per minute
RAW	2	100	1-2 seconds	83%	13000
CFS	2	100	1-2 seconds	87%	13000
				Difference	0%
RAW	2	50	1-2 seconds	53%	8200
CFS	2	50	1-2 seconds	53%	8200
				Difference	0%
RAW	1	50	1-2 seconds	82%	6700
CFS	1	50	1-2 seconds	82%	6600
				Difference	1.5%

OCFS 2 - What’s New ?

- Disk format now deals in blocks and clusters, not bytes.
- Blocksize is now variable, decided at mkfs time (512B, 1K, 2K, or 4K).
- Filesystem is now on-disk compatible with multiple architectures.
- Extent metadata is reorganized.
- Allocation of data areas is now node-local.
- The on-disk inode is reorganized.
- In-memory inodes are directly connected to on-disk inodes.
- The superblock is completely new.
- System inodes are now dynamically located.

OCFS 2 - What's New ?

- No magic first mount
- Memory usage is greatly reduced.
- Vastly simpler DLM operation.
- Number of nodes is more flexible
- Journaling is done via the Journaled Block Device (JBD).
- Metadata and data can be cached.
- I/O is now asynchronous.
- More than one operation at a time.
- Physical sizing limits are massively increased.

OCFS 2 - What’s New ?

- Software sizing limits are also larger, though not by as much.
- Ext2/3-style directories.
- Proper, clean CDSL (Context Dependant Symbolic Link).

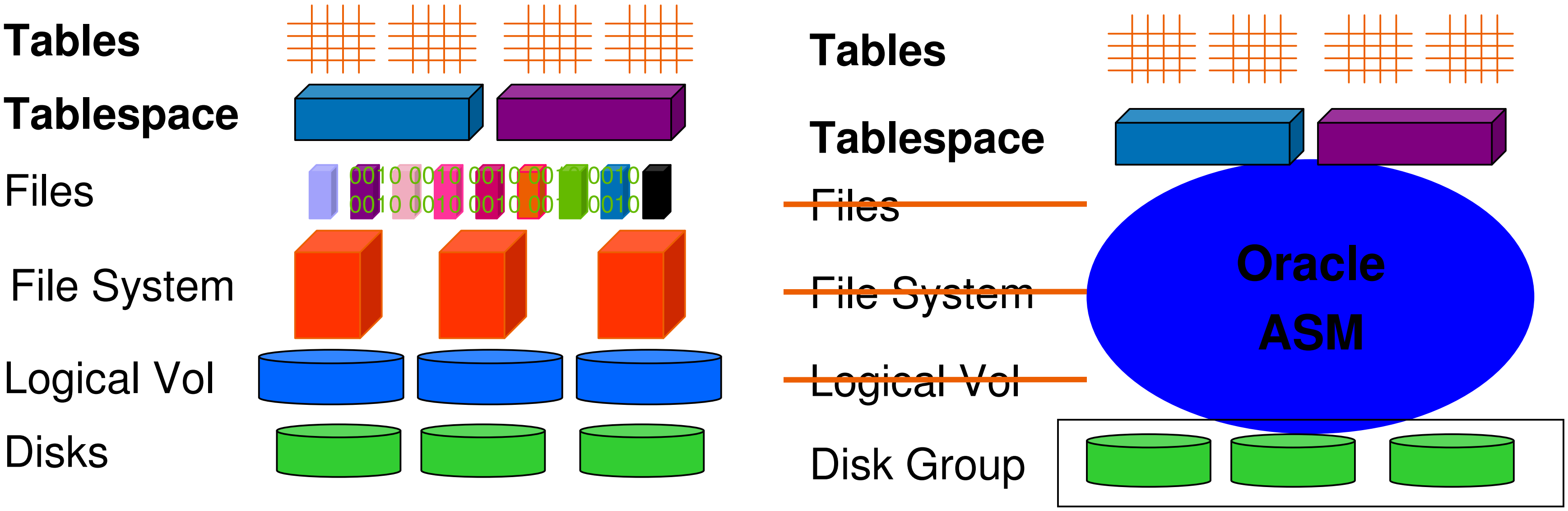
ASM - Overview

- ASM can be used to simplify the administration of Oracle database files. Instead of managing multiple database files, ASM requires to manage a small number of disk groups. A disk group is a set of disk devices that ASM manages as a single, logical unit.

- A particular disk group can be defined as the default disk group for a database and Oracle automatically allocates storage for and creates or deletes the files associated with the database object. When administering the database, there is only referring to database objects by name rather than by file name.

The Operational Stack

TODAY **ASM**



Tables

Tablespace

Files

File System

Logical Vol

Disks

Tables

Tablespace

~~Files~~

~~File System~~

~~Logical Vol~~

Disk Group

Oracle ASM

"The best way to lower mgmt costs is to remove complexity"

Traditional vs ASM - Setup

1. Determine required storage capacity
2. Install Volume Manager, File System
3. Architect data layout to avoid hot spot
4. Create logical volumes
5. Create file systems
6. Install database
7. Create database

1. Determine required storage capacity
2. Install ASM
3. Create Disk Groups
4. Install database
5. Create database

ASM – disk groups and failure groups

- A disk group can include any number of disk devices
- Each disk device can be :
  - an individual physical disk,
  - a multiple disk device such as a RAID storage array or logical volume
  - a partition on a physical disk.
- However, in most cases, disk groups consist of one or more individual physical disks.
- To enable ASM to balance I/O and storage appropriately within the disk group, all devices in the disk group should have similar, storage capacity and performance.

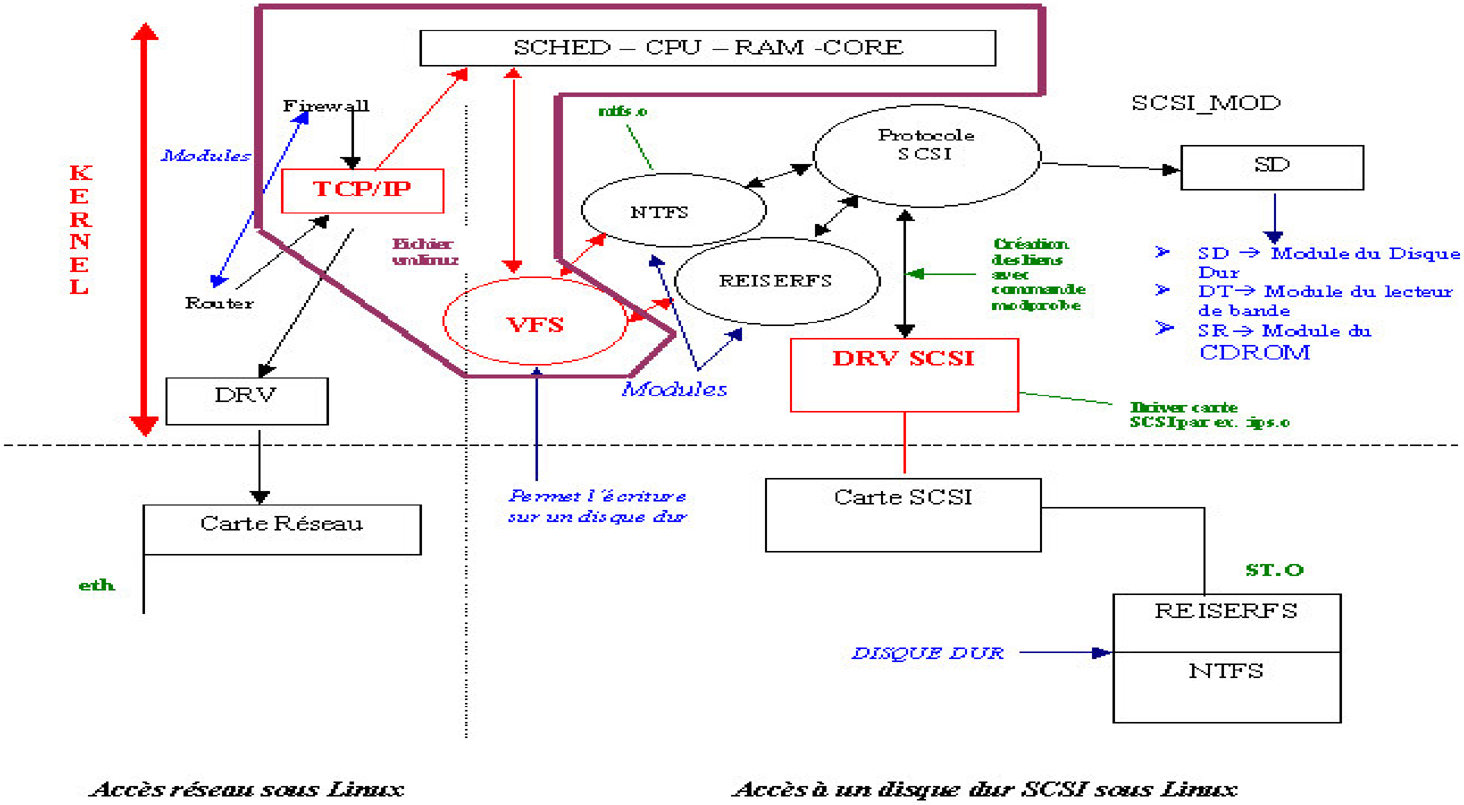
ASM – ASMlib API

- The ASMLIB API, developed by Oracle, provides five major feature enhancements over standard interfaces:
  - Disk discovery – Providing more information about the storage attributes to the Database and the DBA
  - I/O processing – To enable more efficient I/O
  - Usage hints – Providing more intelligence from the database to the storage
  - Write validation - Enable end to end checksum capability
  - Metadata validation – To provide highly efficient metadata update coordination

I/O Handler in Linux Kernel

The virtual file system

- Permit to handle many file system.... “to finnish and translate”



The virtual file system example

- How rm command works. “to finish”

- rm -> delete();
- glibc -> g_sysunlink();
- kernel -> sysunlink();
- ext2.o -> unlink();

Presentation of the /proc Mechanism

/proc

- What is /proc.
- For what's to do it was design.
- Proc API.
- How /proc can help us to optimize cluster.

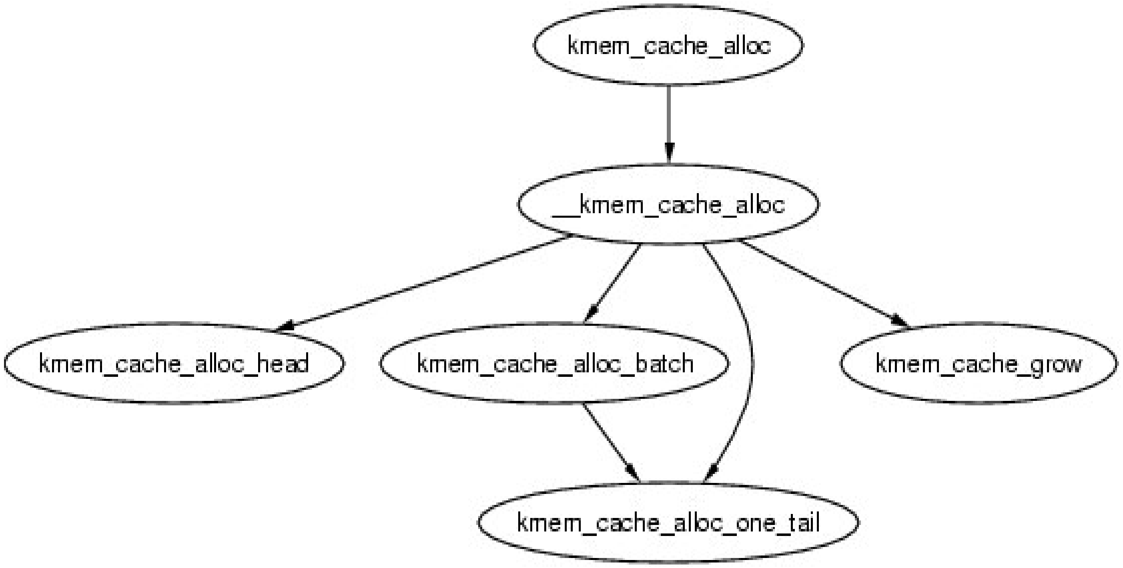
Memory Handler

Linux VM overview

- Philosophy of memory handling
- Concept and choice presentation.

Linux VM overview

- SMP cache mechanism.



Linux VM overview

- Redhat implementation.

RedHat memory limit

RHEL2.1 for ia32

- 2.4.9-e.XXUniprocessor kernel
- 2.4.9-e.XX-smpSMP kernel capable of handling up to 4GB of physical memory
- 2.4.9-e.XXenterprise-SMP kernel capable of handling up to about 16GB of physical memory

RHEL3 for ia32

- 2.4.21-4.EL Uniprocessor kernel
- 2.4.21-4.ELsmp SMP kernel capable of handling up to 16GB of physical memory
- 2.4.21-4.ELhugemem SMP kernel capable of handling beyond 16GB, up to 64GB

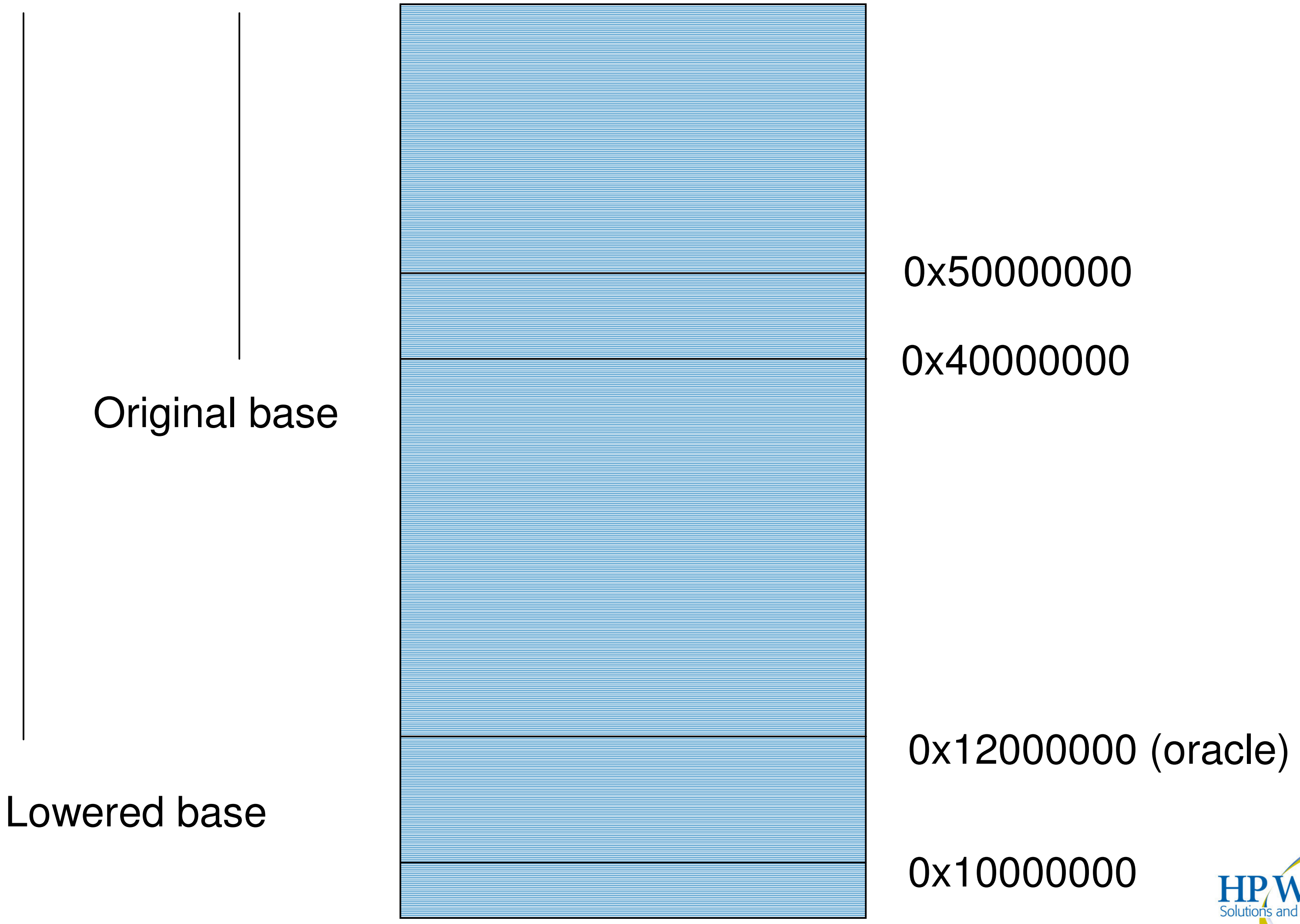
VLM Support

- Ability to use up to 64GB on a 32-bit system
- The PAE (page address extensions) mechanism allows addressing using 36 bits on IA-32 systems
- The entreprise kernel is able to set up to 64GB pagecache without any modifications

IA-32: memory map base address

- Available with Redhat 2.1 and United Linux 1.0
- Oracle running on Linux will, by default, be limited to a shared memory area of 1.7GB.
- Oracle running on Red Hat Advanced Server can allocate a shared memory area of 2.7GB by lowering Oracle’s memory map base address.
  - Relink Oracle with a modified base address
    - [oracle]$ cd $oracle_home/rdbms/lib
    - [oracle]$ genksms –s 0x15000000 > ksms.s
    - [oracle]$ make –f ins_rdbms.mk ksms.o
    - [oracle]$ make –f ins_rdbms.mk ioracle
  - Modify the base mmap address for the Oracle user before starting the Oracle processes
    - [root]# echo 268435456 > /proc/$pid/mapped_base
    - [root]# echo *3000000000* > /proc/sys/kernel/shmmax

IA-32: memory map base address



Oracle VLM using shmfs or ramfs

- For SGA sizes >2.7 GB, Oracle needs to allocate parts of the SGA through a memory mapped file system called shmfs.
  - The difference between both memory file system is that shmfs is pageable, ramfs is not
  - The init.ora parameter “use_indirect_data_buffers” determines whether Oracle will allocate memory through tmpfs. Set to true to enable the use of shmfs.
    - use_indirect_data_buffers=true
- Only database buffers may reside in the shmfs/ramfs memory area. All other parts of the Oracle SGA must still reside in regular Oracle shared memory.

Oracle VLM using shmfs

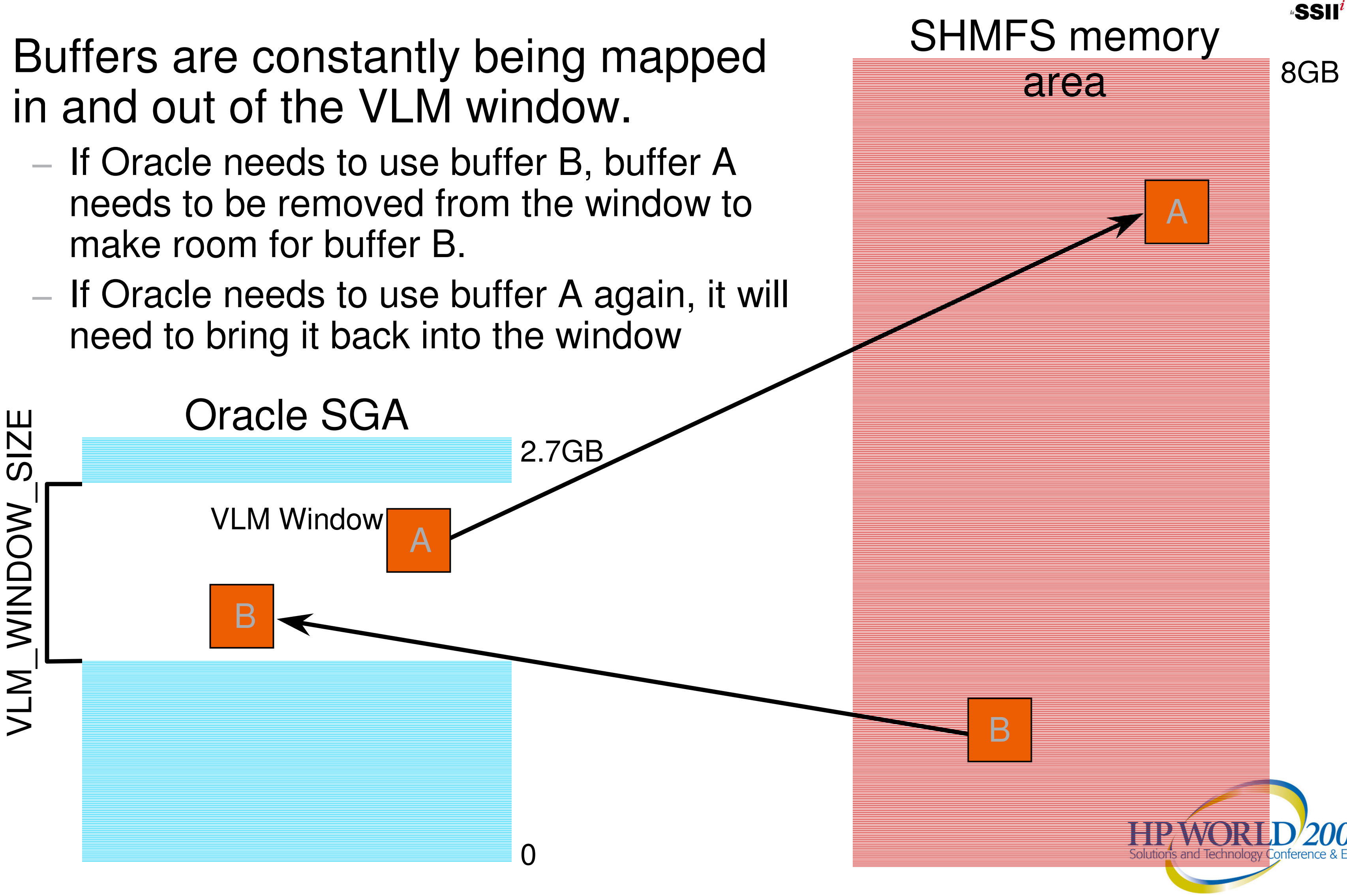
- The shmfs is available in both RH 2.1 and 3
- The mapping/unmapping happens relatively quickly via memory page table manipulation and does not result in memory copies.
- The non database buffer areas of the SGA must still fit under the 2.7GB SGA limit.
- Oracle Total SGA Size – Database Buffers + VLM_WINDOW_SIZE <= 2.7GB
- The amount of SGA space needed to track large amounts of Oracle memory may force you to use a larger Oracle block size to fit under the 2.7Gb limit.
  - Example: An Oracle database with a 2K block size won’t be able to allocate more than about 10GB of buffers. At least a 4K block size would be needed to fully utilize 16GB of memory.

Oracle VLM using shmfs or ramfs

- Oracle creates a “window” in its address space which it uses to map/unmap various buffers from the SGA.
  - The default size of this window is 512MB.
  - The window size can be increased or decreased by setting the environment variable VLM_WINDOW_SIZE to the size of the windows desired.
- The mapping/unmapping happens relatively quickly via memory page table manipulation and does not result in memory copies.
- The non database buffer areas of the SGA must still fit under the 2.7GB SGA limit.
- Oracle Total SGA Size – Database Buffers + VLM_WINDOW_SIZE <= 2.7GB
- The amount of SGA space needed to track large amounts of Oracle memory may force you to use a larger Oracle block size to fit under the 2.7Gb limit.
  - Example: An Oracle database with a 2K block size won’t be able to allocate more than about 10GB of buffers. At least a 4K block size would be needed to fully utilize 16GB of memory.

Oracle VLM using shmfs or ramfs

- Buffers are constantly being mapped in and out of the VLM window.
  - If Oracle needs to use buffer B, buffer A needs to be removed from the window to make room for buffer B.
  - If Oracle needs to use buffer A again, it will need to bring it back into the window



Oracle using shmfs in short

- Mount the shmfs file system as root using command:
- % mount -t shm shmfs -o nr_blocks=8388608 /dev/shm
- Set the shmmax parameter to half of RAM size
- $ echo 3000000000 >/proc/sys/kernel/shmmax
- Set the init.ora parameter use_indirect_data_buffers=true
- Startup oracle.

Oracle using ramfs in short

- Mount the shmfs file system as root using command:
- % umount /dev/shm
- % mount -t ramfs ramfs /dev/shm
- % chown oracle:dba /dev/shm
- - Increase the "max locked memory" ulimit (ulimit -l)
- Add the following to /etc/security/limits.conf:
- oracle soft memlock 3145728
- oracle hard memlock 3145728
- Set the init.ora parameter use_indirect_data_buffers=true
- Startup oracle.

large memory pages (bigpages)

- A separate memory area is allocated using 2MB or 4MB memory pages rather than the normal 4k.
- More efficient use of the processors limited memory map resources (TLB cache)
- Increased hit rates in the TLB cache cause less processor stalling and make the processors run more efficiently, especially in large memory configurations.
- This separate bigpage memory area is locked in memory and not swapped out.

Bigpages/Hugetlb

- Bigpages (RH 2.1) are called Hugetlb in RH3
- Pages are not swapable. This means that the SGA still in memory.
- Hugetlb is a backport from the kernel 2.6
- The pages are pre-allocated. that amount of physical memory can be used only through
- hugetlbfs or shm allocated with SHM_HUGETLB.
- Add “bigpages=xxxxMB” to the kernel boot line in grub.conf or lilo.conf (only with RH 2.1).
- echo the values to '/proc/sys/vm/hugetlb_pool' or Update the '/etc/sysctl.conf'. The value are in MB, and it allocates several 2MB pages.
- Values are viewable using '/proc/meminfo':
  - Hugepages_Total: 500
  - Hugepages_Free: 500
  - Hugepagesize: 2048K

IA-32: large memory pages (bigpages/hugetlb)

- If the kernel is booted with “bigpages =4000MB”

```
[oracle]$ cat /proc/meminfo
        total:    used:    free:  shared: buffers:  cached:
Mem:  16553721856 4332068864 12221652992   196608 13639680 33484800
Swap: 2093129728        0 2093129728
MemTotal:      16165744  kB
MemFree:       11935208  kB
MemShared:          192  kB
Buffers:          13320  kB
Cached:           32700  kB
SwapCached:           0  kB
Active:           42364  kB
Inact_dirty:       3848  kB
Inact_clean:          0  kB
Inact_target:   4177920  kB
HighTotal:     15532016  kB
HighFree:      11375900  kB
LowTotal:        633728  kB
LowFree:         559308  kB
SwapTotal:      2044072  kB
SwapFree:       2044072  kB
BigPagesFree:   4096000  kB
```

← The 4000MB of memory are subtracted from the free system memory…

← … and put into a separate allocation area just for bigpages.

IA-32: large memory pages (bigpages)

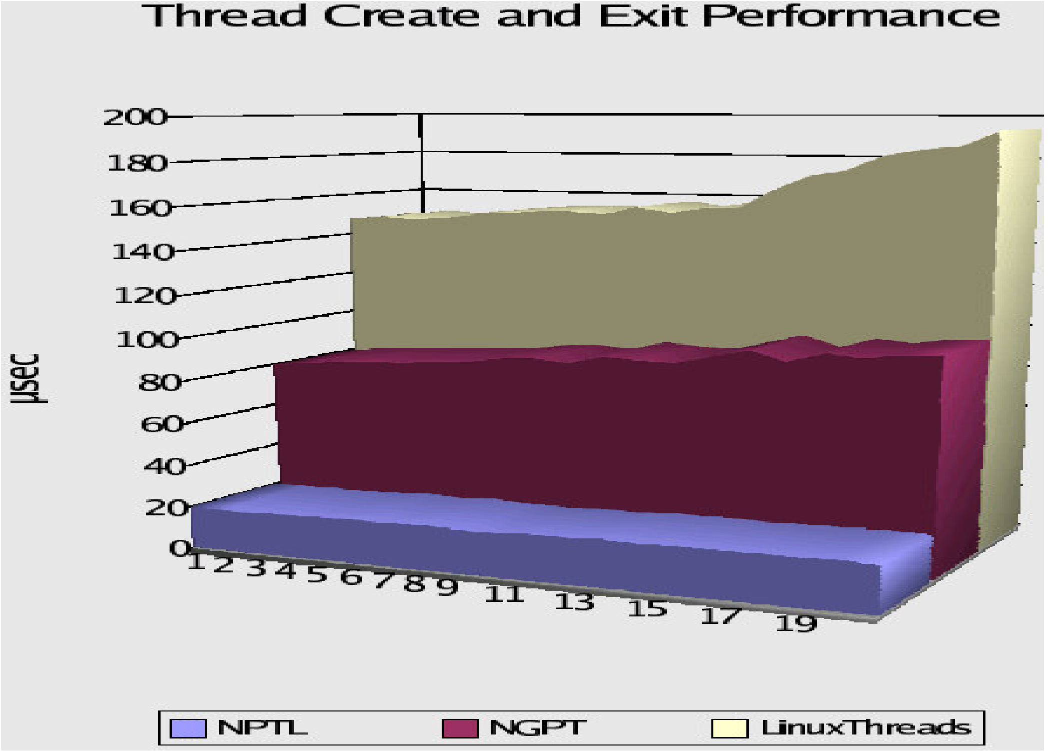
- Have the kernel allocate a pool of bigpage memory
- Add “bigpages=xxxxMB” to the kernel boot line in grub.conf or lilo.conf.
- Tell the kernel to use the bigpage pool for shared memory allocations
  - [root]# echo 2 > /proc/sys/kernel/shm-user-bigpages
- Size the bigpage memory pool to be only as large as needed because it is only used for shared memory allocations
- Unused memory in the bigpage pool will be not be available for general use, even if the system is swapping.

Cache Mechanism Overview

Presentation of NTPL

Native Posix Thread Linux

- NPTL Overview.



Kernel Parameters Tuning

Kernel parameters

- What parameters interesting us ?
- How to modify it ?
- Oracle prerequisite.
- Parameter by family

Kernel parameters - 1

Update the /etc/sysctl.conf and run “sysctl –p” as root.

Oracle Pre-requisites Parameters

- kernel.sem = 250 3200 100 128
- kernel.shmall = 2097152
- kernel.shmmax = 2147483648
- kernel.shmmni = 4096
- fs.file-max = 65536
- net.ipv4.ip_local_port_range = 1024 65000

Kernel parameters - 2

Swap optimization

```
vm.kswapd = 2048 128 32
vm.bdflush = 90 250 0 0 5000 10000 100 50 0
vm.page-cluster = 5
vm.pagetable_cache = 50 100
vm.pagecache = 10 20 30
vm.inactive_clean_percent = 100
vm.max-readahead = 256
vm.min-readahead = 6
```

UDP Sizing

```
net.core.rmem_max = 262144
net.core.wmem_max = 262144
net.core.rmem_default = 262144
net.core.rmem_default = 262144
```

RedHat Linux
Installation For Oracle

During installation...

- Take care about unsupported options.
- What packages are required ?
- Best practice of OS file system size.
- Pre-install oracle operations.

User and Group

Groups :

- Oinstall
- Dba
- Oper

User :

- Oracle

User equivalence :

- “uid” and “gid” have to be the same on all for a user or a group (see notes).
- Set the /etc/hosts.equiv file with nodename and oracle user name.

Oracle User – Environment Variable

Env. variable	*Usage*
ORACLE_BASE	Entry directory for oracle products.
ORA_CRS_HOME	CRS layer directory. Must be different of the ORACLE_HOME.
ORACLE_HOME	Directory to store Oracle binaries.
ORACLE_SID	Instance name.
ORA_NLS33	Store path to find the locale-specific NLS data (NLS is national language support).
NLS_LANG	Defines language, territory and character set used for databases and messages.
LD_LIBRARY_PATH	Path to find library files.
CLASSPATH	Path to find java classes.
TMPDIR and TEMP	Define temporary directory. This one as to be at least 400 Mb.
SRVM_SHARED_CONFIG	Path to raw device where Oracle server manager stores cluster configuration information. Used by srvconfig, srvctl, DBCA.
DBCA_RAW_CONFIG	Path to a file containing list of raw device locations. Used at database creation time with DBCA only if raw devices are choose for storage.
THREADS_FLAG	Allow JDK to use threads.
DISPLAY	Output value for X-window screen.
ORACLE_BASE	Main directory to store Oracle products.

System requirements

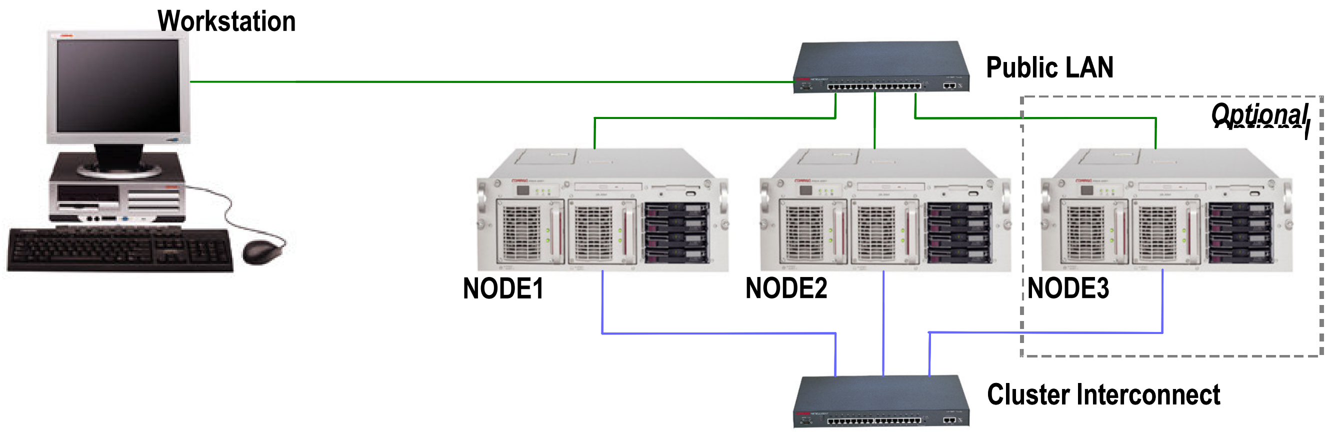
Disk space :	Check with :
512 MB RAM	`grep MemTotal /proc/meminfo`
4 GB of swap space	`# /sbin/swapon -s`
400 MB in /tmp	`# df -k /tmp`
2.5 GB for the ORACLE_HOME	`# df -k`
1.5 GB for the pre configured database	`# df -k`

Install ksh

- The korn shell is mandatory to install Oracle 10g
- ksh is not installed by default, it’s necessary to add a package after the Redhat installation.

  - Mount redhat cdrom #2
  - Cd /redhat/RPMS
  - Cp pdksh-5.2.14-13.i386.rpm /tmp
  - Rpm -ivh pdksh-5.2.14-13.i386.rpm

Network Configuration



3 ip addresses are necessary for each node

- One public lan (for customer access)
- One private interconnect (just for local usage)
- One virual ip address (known has public ip address)

Requiered Packages

Check the kernel sources and development tools are installed

```
rpm -q gcc cpp compat-libstdc++ glibc-devel kernel-headers binutils
```

Packages for RedHat 3

compat-db-4.0.14-5.i386.rpm
compat-gcc-7.3-2.96.122.i386.rpm
compat-gcc-c++-7.3-2.96.122.i386.rpm
compat-libstdc++-7.3-2.96.122.i386.rpm
compat-libstdc++-devel-7.3-2.96.122.i386.rpm
openmotif21-2.1.30-8.i386.rpm
setarch-1.3-1.i386.rpm
tcl-8.3.5-92.i386.rpm
pdksh-5.2.14-21.i386.rpm (if 10g)

Packages for RedHat 2.1

cpp-2.96-108.1.i386.rpm
glibc-devel-2.2.4-26.i386.rpm
kernel-headers-2.4.9-e.3.i386.rpm
gcc-2.96-108.1.i386.rpm
binutils-2.11.90.0.8-12.i386.rpm
pdksh-5.2.14-22.i386.rpm (if 10g)

gcc & g++

- If RedHat 3 is used, th defaults gcc and g++ released doesn’t work with Oracle.
- Perform the following step to use the 2.96.

Check the gcc g++ release with :

```
gcc –v or gcc –version
g++ -v or g++ -version
```

You should get:

```
Reading specs from /usr/lib/gcc-lib/i386-…
gcc version 2.96 20000731 (Red Hat Linux 7.3 2.96-123)
```

If Not,
As root user

```
# mv /usr/bin/gcc /usr/bin/gcc323
# ln -s /usr/bin/gcc296 /usr/bin/gcc
# mv /usr/bin/g++ /usr/bin/g++323
# ln -s /usr/bin/g++296 /usr/bin/g++
```

SSH setting

(SSH) is a program for logging into a remote machine and for executing commands on a remote machine. It is intended to replace rlogin and rsh, and provide secure encrypted communications between two untrusted hosts over an insecure network. X11 connections and arbitrary TCP/IP ports can also be forwarded over the secure channel.

HP Driver Installation

Storage Driver

- Update the QLA driver. The latest release available from HP is 6.06.50
- Available on http://h18007.www1.hp.com/storage/diskarrays-support.html

- For EVA, download the storage disk manaement from http://h18007.www1.hp.com/products/storageworks/softwaredrivers/enterprise/index.html

cciss (internal smart array) updated driver can be downloaded from

http://h18004.www1.hp.com/support/files/server/us/locate/101_4081.html

File system type

For security and boot performance reason, it is advised to use ext3 file system for local partitions.

But in some case in makes sense to use ext2 instead

OS update Policy

Operating system life cycle.

- Howto find and install HP certified driver.
- Update system consideration.
- Update kernel method.
- Emergency rescue mode

Oracle IO and Settings

Adjust Oracle Block Size

- A UNIX system reads entire operating system blocks from the disk. If the database block size is smaller than the UNIX file system block size, I/O bandwidth is inefficient. If you set the Oracle database block size to be a multiple of the file system blocksize, you can increase performance by up to five percent.
- The DB_BLOCK_SIZE initialization parameter sets the database block size. However, to change the value of this parameter, you must recreate the database.
- To see the current value of the DB_BLOCK_SIZE parameter, enter the SHOW PARAMETER DB_BLOCK_SIZE command in SQL*Plus

Enable asynchronous io.

Install “opatch” utility (2617419)
Apply the patch 3016968 which is necessary to enable asych_io.

Recompile the kernel to enable access to libaio and skgaioi.o.
cd to $ORACLE_HOME/rdbms/lib

```
make -f ins_rdbms.mk async_on
make -f ins_rdbms.mk ioracle
```

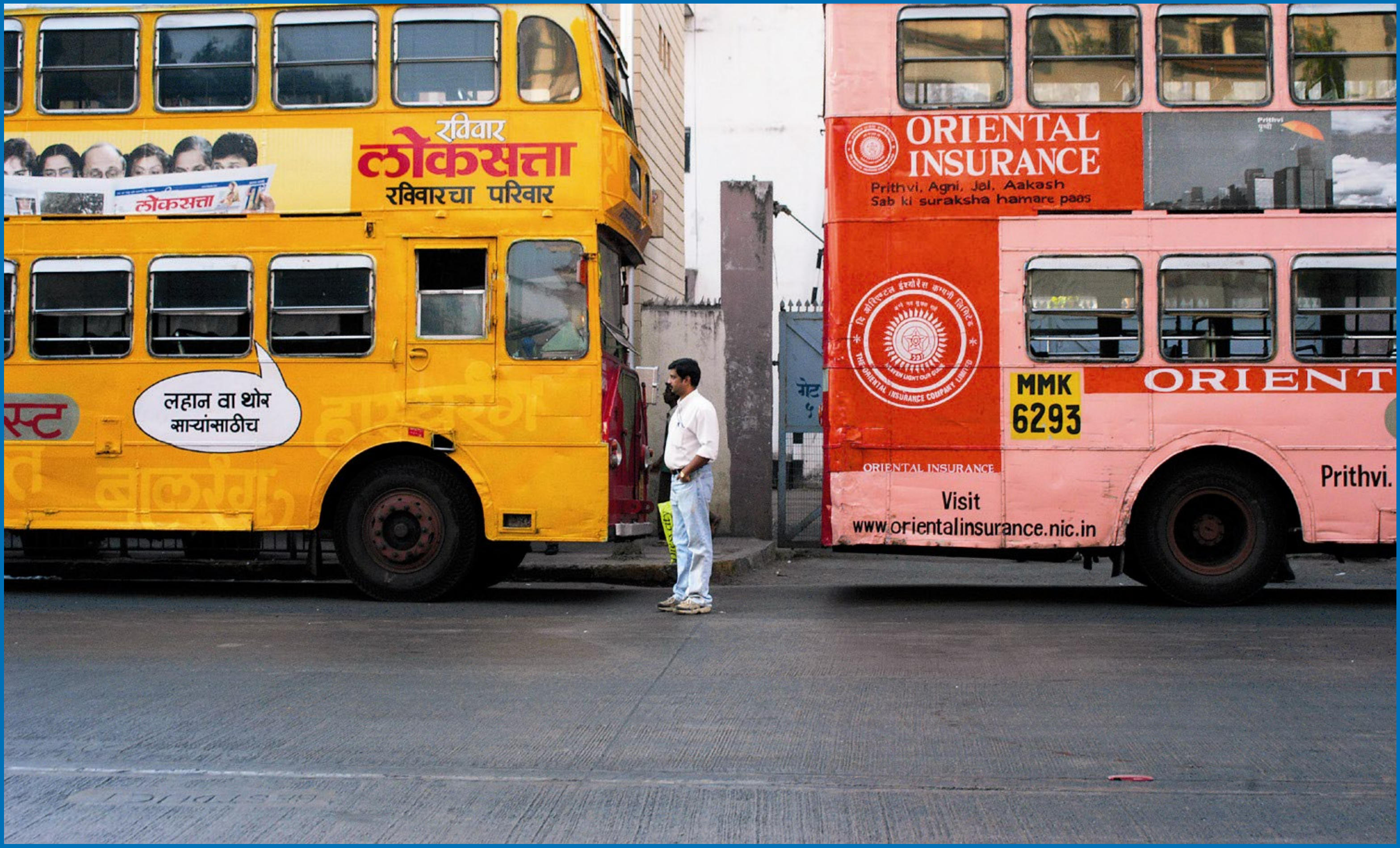
Make sure that all Oracle datafiles reside on filesystems that support asynchronous I/O. (For example, ext2, ext3, ocfs in RedHat 3) or on raw device.

Set init.ora file or spfile.ora
'disk_asynch_io=true'
‘filesystemio_options=asynch' (if dbf are on file system)

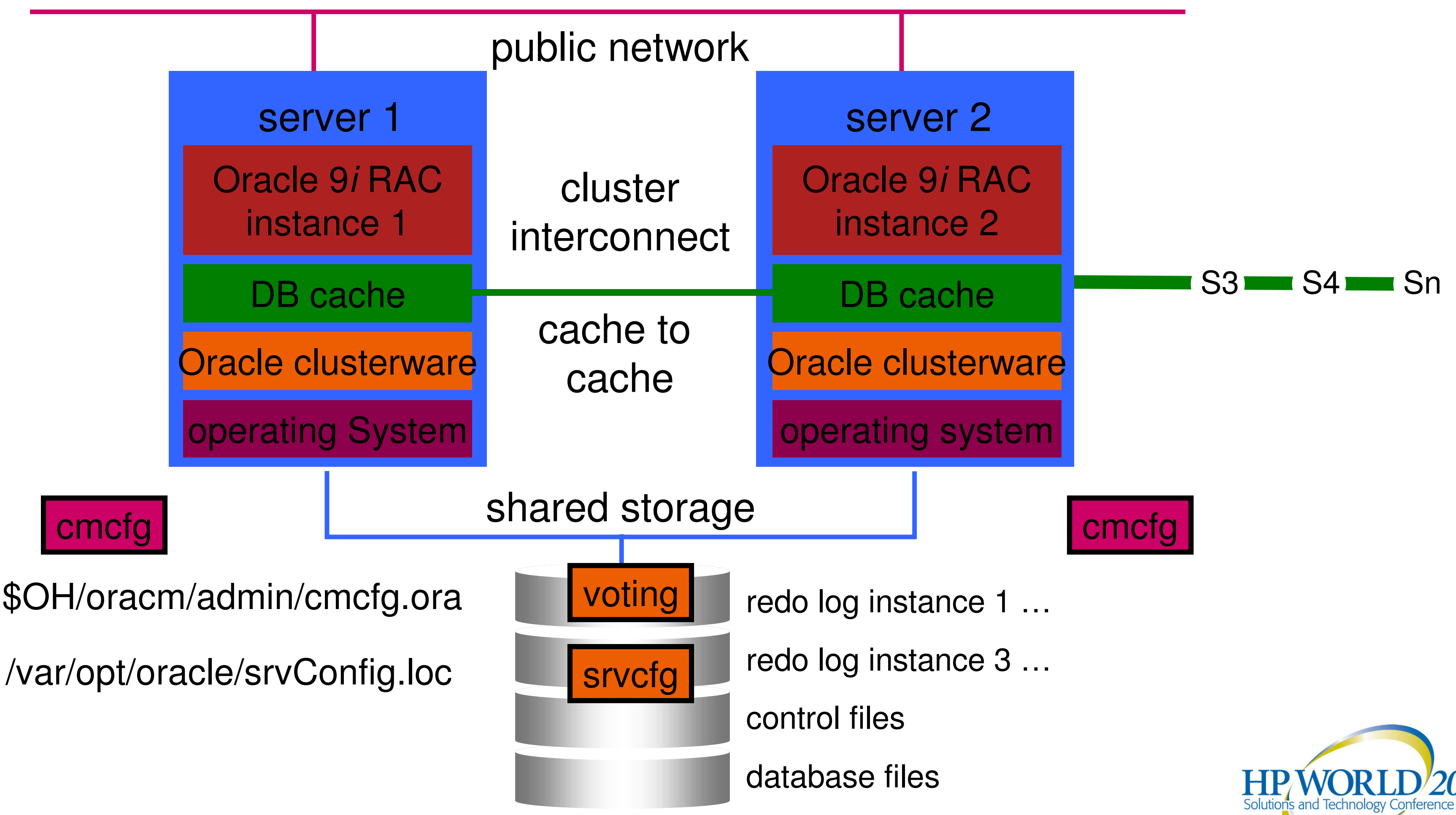
Direct I/O Support

- Direct I/O support is not available and is not supported on Red Hat Enterprise Linux 2.1 and SuSE Linux Enterprise Server 8. It is available and is supported on Red Hat Enterprise Linux 3 if the driver being used on the system supports varyio. To enable direct I/O support:
- Set the FILESYSTEMIO_OPTIONS initialization parameter to DIRECTIO.
- If you are using the asynchronous I/O option, set the FILESYSTEMIO_OPTIONS initialization parameter to SETALL.

Failover/Failback tuning



Oracle9*i* RAC Architecture



$OH/oracm/admin/cmcfg.ora

/var/opt/oracle/srvConfig.loc

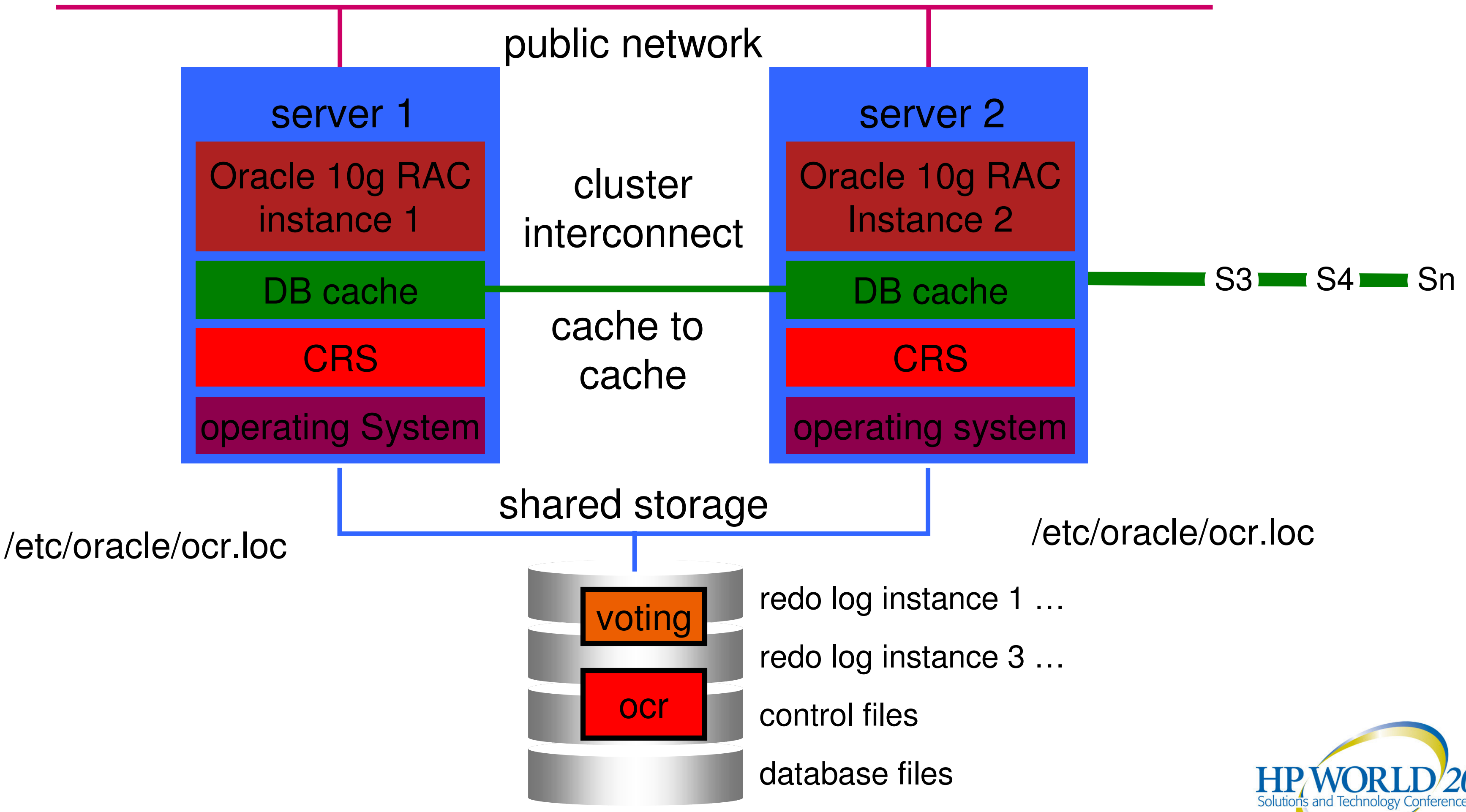
Oracle9*i* RAC Architecture

cmcfg.ora

```
HeartBeat=15000

ClusterName=Oracle Cluster Manager, version 9i

PollInterval=1000

MissCount=250

PrivateNodeNames=clusaptux6 clusaptux7

PublicNodeNames=saptux6 saptux7

ServicePort=9998

CmDiskFile=/ocfs1/voting

HostName=clusaptux6

KernelModuleName=hangcheck-timer
```

```
/sbin/insmod hangcheck-timer hangcheck_tick=30 hangcheck_margin=180
```

Oracle 10g RAC Architecture



Oracle10g RAC Architecture

ocr

RAC database definition
RAC instance definition
CRS definitions

- RAC service definition
    - prefered
    - available
- RAC node application definitions
    - listener
    - oem
    - ...

TNSNAMES Service Entry

```
OE =
  (DESCRIPTION =
    (ADDRESS = (PROTOCOL = TCP)(HOST = node1c_vip)(PORT = 1521))
    (ADDRESS = (PROTOCOL = TCP)(HOST = node2c_vip)(PORT = 1521))
    (LOAD_BALANCE = yes)
    (CONNECT_DATA =
      (SERVER = DEDICATED)
      (SERVICE_NAME = oe)
      (FAILOVER_MODE =
        (TYPE = SELECT)
        (METHOD = BASIC)
        (RETRIES = 180)
        (DELAY = 5)
      )
    )
  )
```

Diagnostic Utilities

Diagnostic utilities.

- ps.
- top.
- /proc.
- Linux bench source: http://lbs.sourceforge.net

Oracle Statspack Feature

- Statspack is a diagnostic tool for instance-wide performance problems
- It also supports application tuning activities by identifying high-load SQL statements.
- It can be used both proactively to monitor the changing load on a system, and also reactively to investigate a performance problem.
- To use Statspack you take a number of ‘snapshots’ of the Oracle performance data and you can then report on any pair of these snapshots.
- The greatest benefits are seen when there is ‘baseline’ performance data available for the system to compare with current data.

Using Oracle Statspack

Run the installation script using SQL*Plus from within the $ORACLE_HOME/rdbms/admin directory or the equivalent on your system:

```
SQL> connect / as sysdba
SQL> @spcreate
```

```
SQL> execute statspack.snap;
```

or

```
SQL> execute statspack.snap(i_session_id=>32);
```

run the spreport.sql report while being connected to the PERFSTAT user to generate a report.

Linux Customization

Packaging facility.

- Why it can be use ?
- How it can be use ?
- What about support ?
- What direct benefit ?

Using your own “master” cd set.

- Why it can be use ?
- How it can be build ?
- What about support ?
- What direct benefit ?

Anaconda overview.

- What is anaconda.
- How it work ?
- What is the step of cd creation ?

“If you ask nothing, you get nothing.”

Alan Cox

Gnome Hacker, Kernel Hacker, Hacker Hacker.

HP WORLD 2004

Solutions and Technology Conference & Expo

Co-produced by:

interex

shared knowledge • shared power

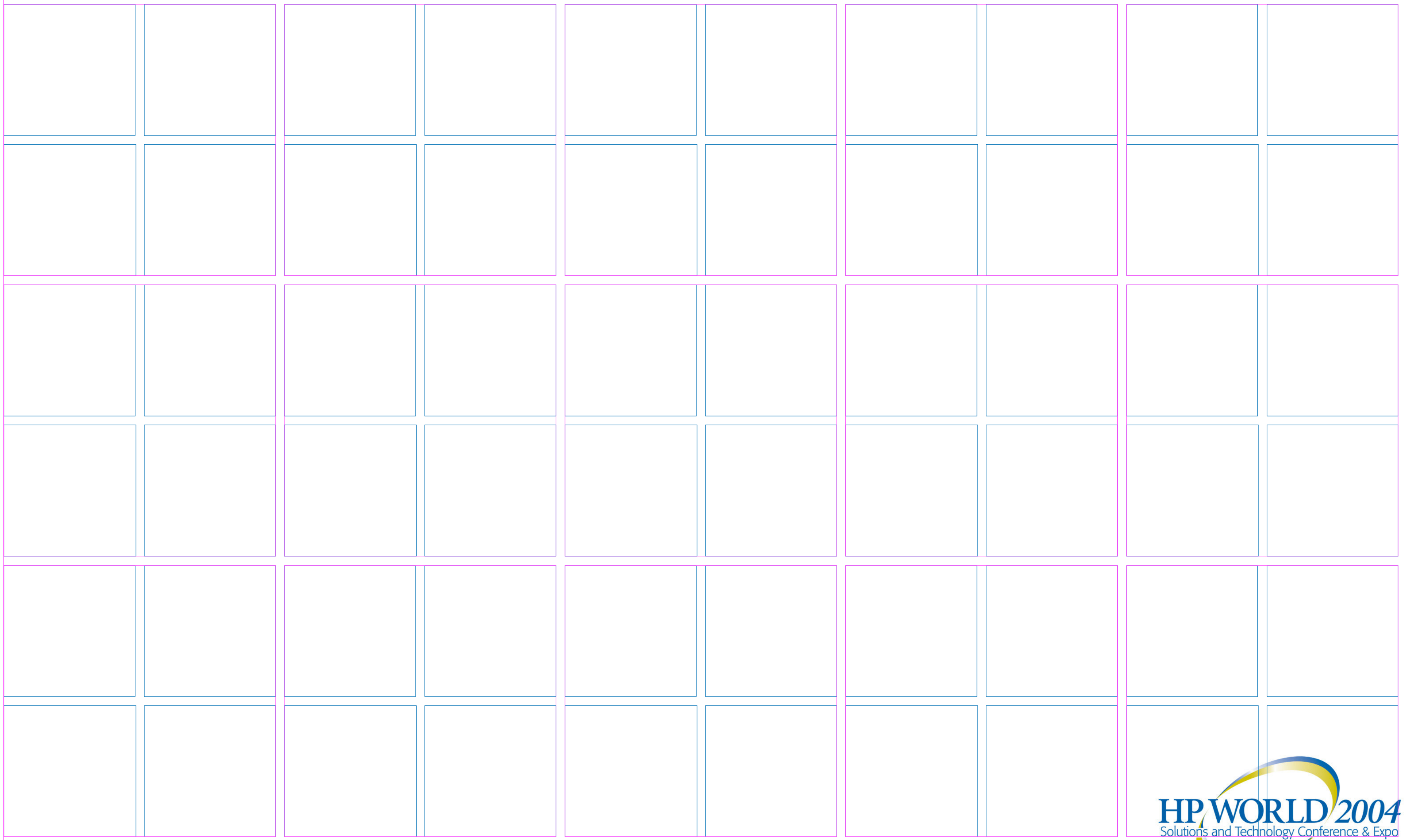
encompass

AN HP USER GROUP

RECOMMENDED TRAINING VENUE FOR THE

HP Certified Professional

Grid



“This is a standard quote slide. Type your quote inside the quotation marks.”

Jordan Doe

CIO, ABC Company

Transition slide headline text goes here

invent