ORACLE

Best Practices for Oracle Real Application Clusters 10g on Microsoft Windows

Kevin Ashby
Enterprise Technology Strategist
Microsoft

Philip Newlan
Principal Product Manager
Oracle Server Technologies - RAC

The following is intended to outline our general product direction. It is intended for information purposes only, and may not be incorporated into any contract. It is not a commitment to deliver any material, code, or functionality, and should not be relied upon in making purchasing decisions.
The development, release, and timing of any features or functionality described for Oracle's products remain at the sole discretion of Oracle.

ORACLE

© 2006 Oracle Corporation – Proprietary and Confidential

Abstract

You love Microsoft Windows but want more availability and scalability? Come hear Oracle and Microsoft experts together extol the virtues of Oracle Real Application Clusters (RAC) 10g on Microsoft Windows operating systems. This session discusses the unique features that make Oracle RAC 10g and Microsoft Windows a winning combination for Oracle-on-Windows customers.

The session discusses implementation from the ground up, including topics ranging from the operating system, storage, interconnect, Oracle Clusterware, Oracle Cluster File System, and Oracle Automatic Storage Management, through to installation of the latest release of Oracle RAC 10g. Learn the best practices for implementing an Oracle RAC 10g cluster on Windows.

ORACLE

© 2006 Oracle Corporation – Proprietary and Confidential

Microsoft / Oracle Relationship

ORACLE

© 2006 Oracle Corporation – Proprietary and Confidential

Commitment to Windows

- Oracle7 was the first database to ship on Windows
  - in 1994
- Windows Server is a base development platform for Oracle
- Windows Server is Tier 1
  - fully supported platform for Oracle
- Focused engineering and product management groups for Windows-specific integration projects
- Supported both 32-bit and 64-bit Windows Server 2003 on day of O/S release

ORACLE

© 2006 Oracle Corporation – Proprietary and Confidential

Partnership

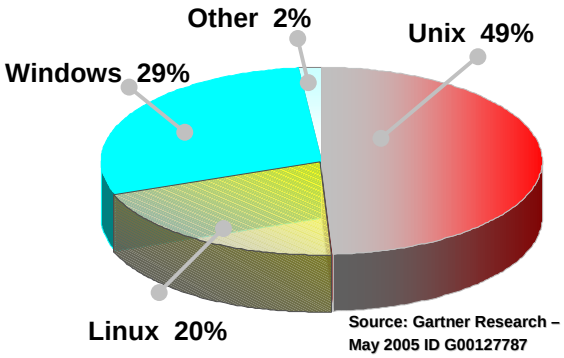
- Oracle sponsors Microsoft events
  - (PDC 2003, TechEd 2004, 2005)
- Microsoft sponsors Oracle events
  - OracleWorld 2005, UKOUG
- Oracle RAC Training Sessions at MS Redmond campus
- Communication channels open at various levels in the organizations
- Joint support contacts for customer escalations
  - How we jointly solve problems
- Conference calls, briefings, etc.

ORACLE

© 2006 Oracle Corporation – Proprietary and Confidential

Oracle on Windows
Widely Deployed

2004 Oracle Revenues by OS for new license sales



ORACLE

© 2006 Oracle Corporation – Proprietary and Confidential

Architecture

© 2006 Oracle Corporation – Proprietary and Confidential

Integration with MS Windows

- Most Oracle components run as Windows services
  - Database
  - Listener
  - Cluster Components
- Basic Instance Configuration is held in the Registry
- Events
  - Oracle writes to both the System & Application Event logs

© 2006 Oracle Corporation – Proprietary and Confidential

Supported 32-bit Architectures

Microsoft Windows Server 2003

- For 32-bit x86 systems
- Well-suited for mainstream deployments where 64-bit applications or drivers are unavailable
- Most common on servers with 1-4 processors

ORACLE

© 2006 Oracle Corporation – Proprietary and Confidential

Supported 64-bit Architectures

Microsoft Windows Server 2003 x64 Editions

- Gradually replaces 32-bit Windows as mainstream offering
- Suitable for all workloads
- Well-suited for combining 32-bit and 64-bit software

Microsoft Windows Server 2003 for Itanium-based Systems

- Intended for the most demanding Database and LOB deployments on 8-way+
- Designed for pure 64-bit software stacks

ORACLE

© 2006 Oracle Corporation – Proprietary and Confidential

32 bit Vs 64bit Memory Limits

General Memory Limits	32-bit	64-bit
Total Virtual Address Space	4 GB	16 TB
Virtual Address Space per 32-bit process	2GB (3 GB if system is booted with /3gb switch)	4GB if compiled with /LARGEADDRESSAWARE 2GB otherwise
Virtual Address Space per 64-bit process	Not applicable	8 TB
Paged Pool	470 MB	128 GB
Non-Paged Pool	256 MB	128 GB
System Cache	1 GB	1 TB
Physical Memory and CPU Limits[1]	32-bit	64-bit
Windows XP Professional	4 GB / 1-2 CPUs	32 GB / 1-2 CPUs
Windows Server 2003 Standard Edition	4 GB / 1-4 CPUs	32 GB / 1-4 CPUs
Windows Server 2003 Enterprise Edition	64 GB / 1-8 CPUs	1 TB / 1-8 CPUs
Windows Server 2003 Datacenter Edition	64 GB / 1-32 CPUs	1 TB / 1-64 CPUs

(1) Product names listed are for general reference only and do not reflect actual product names. 64-bit Windows Server Standard Edition will be available for x64 only.

ORACLE

© 2006 Oracle Corporation – Proprietary and Confidential

Refresher - Oracle Architecture

- On Linux & Unix
  - Process based
- OS/2 – some history
  - Our implementation was like our Unix versions
- On Windows
  - Thread Based
- Oracle is available on both 32bit & 64bit platforms
  - Move to 64bit only on some.

ORACLE

© 2006 Oracle Corporation – Proprietary and Confidential

Architecture – Thread Model

- Unlike UNIX ports of Oracle Database 10*g*, Oracle Database 10*g* on Windows is implemented as a single operating system process
- Typical “processes” such as PMON and LGWR have been converted to native Windows threads running in the single large Oracle process
- One process exists per instance on a Windows machine

ORACLE

© 2006 Oracle Corporation – Proprietary and Confidential

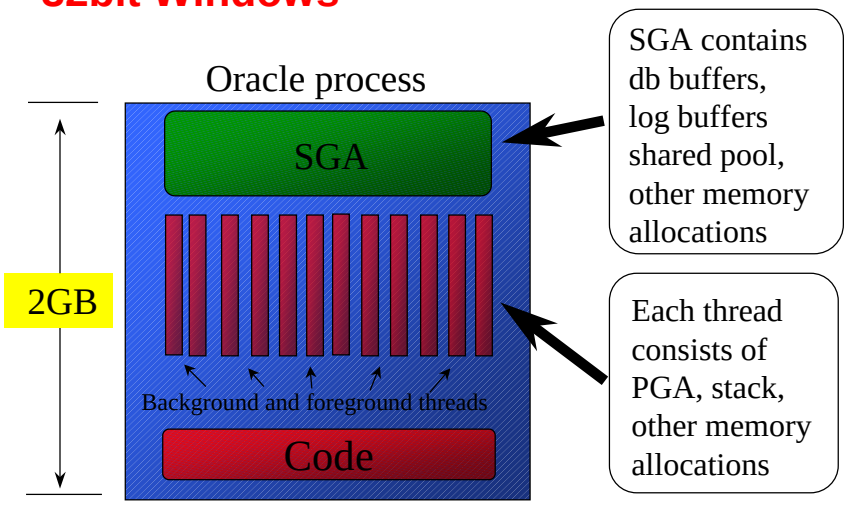
Architecture - Memory Model

- Process Memory restrictions
  - 32bit 4GB (Total User & Kernel) + extensions
  - 64bit 8TB
  - On 32bit Windows
    - All an instances memory requirement must fit in the 4GB limit
    - * with a few caveats
  - On 64bit Windows
    - A process has 8TB to ‘play with’

ORACLE

© 2006 Oracle Corporation – Proprietary and Confidential

Architecture – Thread and Memory Model
32bit Windows



© 2006 Oracle Corporation – Proprietary and Confidential

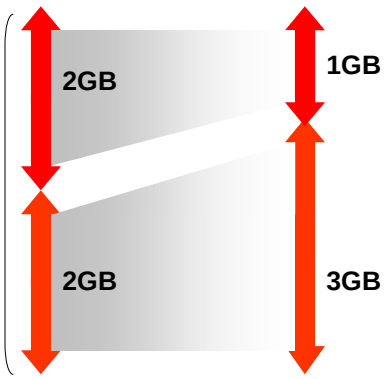
4GB RAM Tuning

- Increase addressable memory available to the Oracle process by adding /3GB switch to boot.ini file:
  multi(0)disk(0)rdisk(0)partition(1)\WINNT="Microsoft Windows 2000 Advanced Server" /fastdetect /3GB
- Reboot server to enable
- Must monitor kernel memory closely to prevent instability of operating system
- See Metalink Notes 46001.1 and 297498.1
- See Microsoft KB article 297812

© 2006 Oracle Corporation – Proprietary and Confidential

What's wrong with 32 bits ?

Nothing… but 32bits = 4Gb per process



Use /3G switch in boot.ini

Doing this restricts the OS non-paged pool

Certain OS pools get cut in half

ORACLE

© 2006 Oracle Corporation – Proprietary and Confidential

Key Items to Monitor for Memory Usage

- Perfmon - Virtual Bytes for oracle.exe to see total memory used by the process
- Total Pool Non-Paged Bytes – Memory Counter
- Total Pool Paged Bytes – Memory Counter
- Free System Page Table Entries (PTE's) – Memory Counter
- In addition to Perfmon, the Process Explorer tool from http://www.sysinternals.com is invaluable

ORACLE

© 2006 Oracle Corporation – Proprietary and Confidential

Why 64 bits

- Each thread – dedicated user connection
  - By default takes 1 Mb of process memory
- All Oracle memory has to come out of the address space
- The typical question we get all the time
  - *What's the user limit on Windows ?*

Why 64 bit ?
- Address Space

ORACLE

© 2006 Oracle Corporation – Proprietary and Confidential

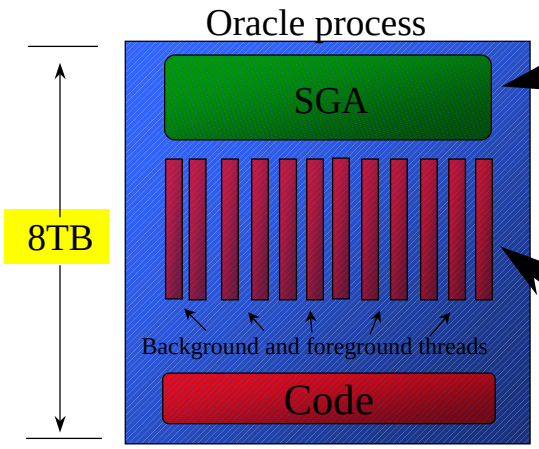
Adoption of 64bit OS's

- Options on other platforms
  - Oracle has supported 64 platforms for years
- Options on Windows
  - Windows Itanium
    - First 64 bit database on Windows (Oracle 9i)
  - Windows X64
    - EM64T/Opteron
  - For us it's a few #defines and then a recompile of our 32bit platform
- We develop on Linux32 and Windows32
  - port / recompile for everything else

ORACLE

© 2006 Oracle Corporation – Proprietary and Confidential

Architecture – Thread and Memory Model
64bit Windows



SGA contains db buffers, log buffers shared pool, other memory allocations

Each thread consists of PGA, stack, other memory allocations

© 2006 Oracle Corporation – Proprietary and Confidential

Installation

© 2006 Oracle Corporation – Proprietary and Confidential

Installation

- Pointers
  - Some installs fail due to incomplete pre-requisites
  - If it fails – clean it up properly
  - All nodes must run the same Operating System
- Pre-reqs
  - See following Slides
- Cleanup
  - Follow metalink note : 341214.1
- Maximum Nodes
  - 100

ORACLE

© 2006 Oracle Corporation – Proprietary and Confidential

MSCS

- Microsoft Cluster Server
  - Not required
  - If installed make sure it's quorum disk is not used by Oracle Clusterware
  - Make sure you do not use any of the MSCS shared disks
- Required if you want a cold failover cluster (not RAC)
  - Use in conjunction with Oracle Failsafe

ORACLE

© 2006 Oracle Corporation – Proprietary and Confidential

Installation Pre-reqs

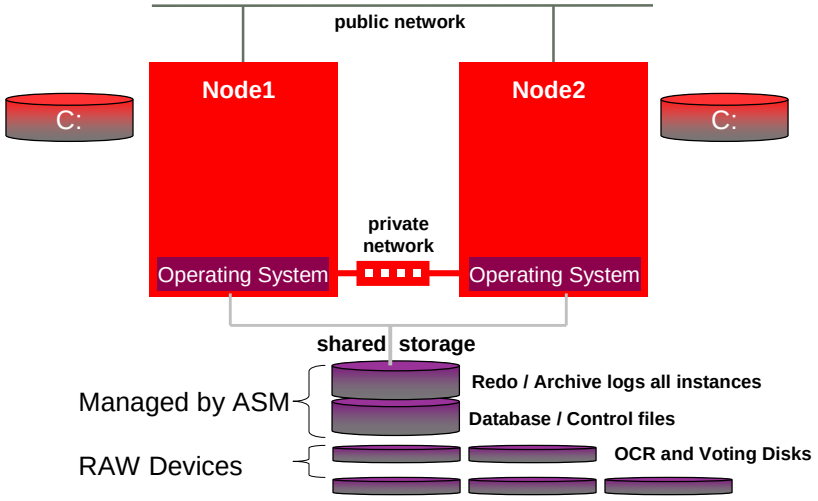
- Hardware is set up correctly
- Use the Cluster Verification Utility (CVU)
- User Accounts
- Remote Desktop Client
- Environment Variables
- Networking
- Disk Layout
- Enable Automount
- Stop troublesome Services
- Time Sync

ORACLE

© 2006 Oracle Corporation – Proprietary and Confidential

Install – Hardware

Start of Install



ORACLE

© 2006 Oracle Corporation – Proprietary and Confidential

Install - User Accounts

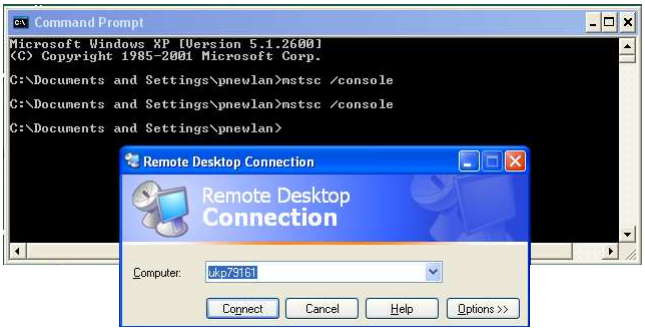
- Recommend Local Admin User accounts
- Whichever account you use must be a direct member of the Local Administrators group on all nodes
- The following command should complete successfully on all nodes
  - Net use IPC$
- Using regedit you must be able to
  - File->Connect Network Registry

ORACLE

© 2006 Oracle Corporation – Proprietary and Confidential

Install - Remote Desktop Client

- If you must use the remote desktop client
  - Invoke the client with the /console command line parameter

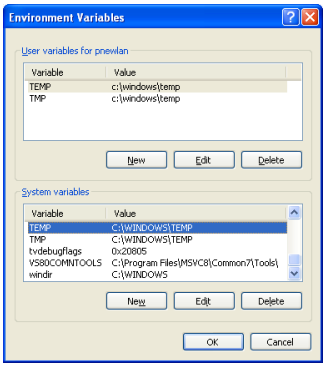


ORACLE

© 2006 Oracle Corporation – Proprietary and Confidential

Install - Environment Variables

- Ensure that both the TEMP & TMP environment variables point to the same directory on all nodes

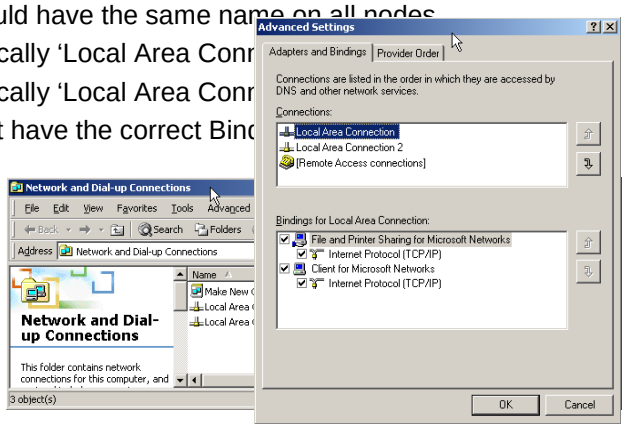


ORACLE

© 2006 Oracle Corporation – Proprietary and Confidential

Install - Networking

- Network Adapters
  - Should have the same name on all nodes
  - Typically 'Local Area Conn
  - Typically 'Local Area Conn
  - Must have the correct Binc



ORACLE

© 2006 Oracle Corporation – Proprietary and Confidential

Install - Networking

- You will need 3 IP address per node
  - Public Address *
  - Virtual IP Address #
  - Private Address *
- It is recommended that the Public & VIP addresses do NOT come from the Private IP Address range
- It is recommended that the Private address does come from the Private IP Address range.

- (*) In DNS and ping-able at start of install
- (#) In DNS but NOT ping-able at start of install

ORACLE

© 2006 Oracle Corporation – Proprietary and Confidential

Install - DiskLayout

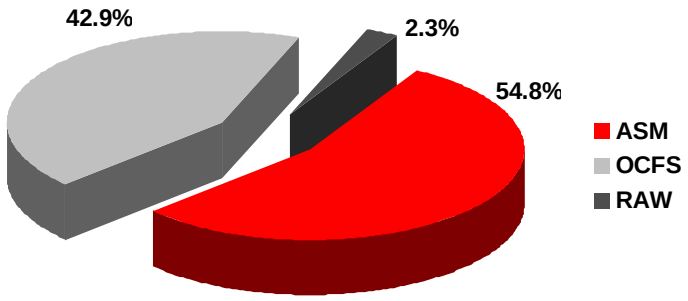
- OCR requires at least 100Mb
  - Recommend 2 devices
- Vote devices require at least 20Mb
  - Recommend 3 devices
- If using multiple devices
  - No common points of failure
- Start with an extended partition that fills the disk
  - Create smaller partitions for the OCR & vote devices
  - Create large / full disk partitions for ASM or OCFS
- No other FileSystem partitions on the same disk
- No OS Client side cache – Adapter cache is BAD

ORACLE

© 2006 Oracle Corporation – Proprietary and Confidential

Where do I locate my datafiles ?

- 10g Windows RAC tracked customers



© 2006 Oracle Corporation – Proprietary and Confidential

Install - Automount

- You must explicitly enable automount on all nodes in the cluster
- Use DISKPART

```
C:> DISKPART
DISKPART> automount enable
Automatic mounting of new volumes enabled.
DISKPART> exit
```

© 2006 Oracle Corporation – Proprietary and Confidential

Install - Stop Troublesome Services

- There is one service that typically causes issues during install
  - Microsoft Service: Distributed Transaction Coordinator
- Stop and set to manual start on all nodes in the cluster
- If you need the MSDTC you can start it again after the install has completed.

ORACLE

© 2006 Oracle Corporation – Proprietary and Confidential

Install - Time Sync

- The RAC nodes should be kept in time sync
- Install will complain if they are out by more than 10 seconds
- It's a lot easier to correlate log files from multiple nodes if the times are in sync.
- Use time service software to maintain the times on the nodes.

ORACLE

© 2006 Oracle Corporation – Proprietary and Confidential

Cluster Verification Utility - CVU

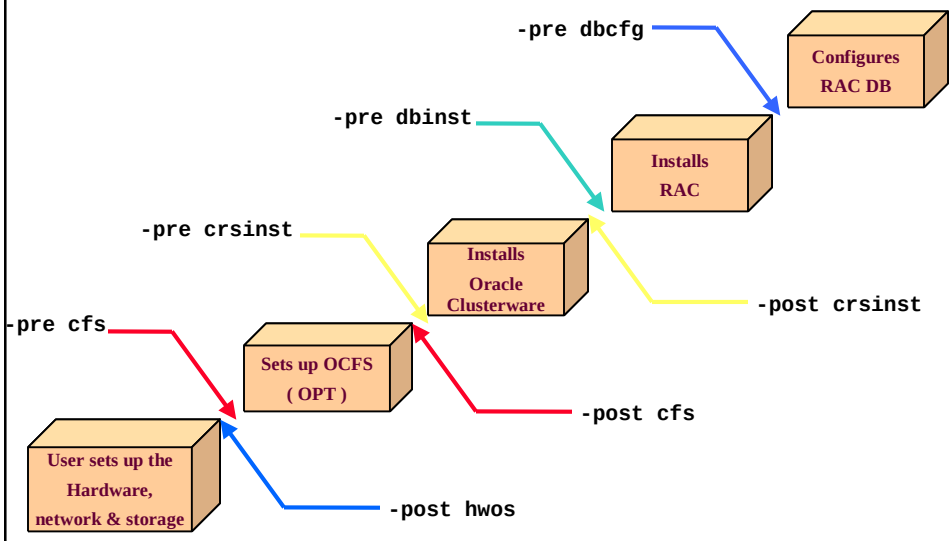
- CVU checks and reports on varies configuration items at various steps during the install
- Does not fix anything !
- Get the latest version from OTN
  - http://otn.oracle.com/rac - Right Hand Side - CVU

- Two steps to run prior to starting install
  - `runcluvfy stage -post hwos -n iwinrca01,iwinrca02`
  - `runcluvfy stage -pre crsinst -n iwinrca01,iwinrca02`

ORACLE

© 2006 Oracle Corporation – Proprietary and Confidential

Cluster Verification Utility



ORACLE

© 2006 Oracle Corporation – Proprietary and Confidential

CVU – post hwos – Part 1

```
D:\clusterware\cluvfy>runcluvfy stage -post hwos -n iwinrca01,iwinrca02

Performing post-checks for hardware and operating system setup

Checking node reachability...
Node reachability check passed from node "iwinrca01".

Checking user equivalence...
User equivalence check passed for user "Administrator".

Checking node connectivity...

Node connectivity check passed for subnet "139.185.150.0" with node(s) iwinrca02,iwinrca01.
Node connectivity check passed for subnet "10.10.10.0" with node(s) iwinrca02,iwinrca01.

Suitable interfaces for VIP on subnet "139.185.150.0":
iwinrca02 Local Area Connection:139.185.150.55
iwinrca01 Local Area Connection:139.185.150.54

Suitable interfaces for the private interconnect on subnet "10.10.10.0":
iwinrca02 Local Area Connection 2:10.10.10.3
iwinrca01 Local Area Connection 2:10.10.10.2
.
.
.
```

ORACLE

© 2006 Oracle Corporation – Proprietary and Confidential

CVU – post hwos – Part 2

```
.
.
.
Node connectivity check passed.

Checking shared storage accessibility...

  Disk Partition                        Sharing Nodes (2 in count)
  ------------------------------------  ------------------------
  \Device\Harddisk1\Partition1          iwinrca02 iwinrca01

  Disk Partition                        Sharing Nodes (2 in count)
  ------------------------------------  ------------------------
  \Device\Harddisk1\Partition2          iwinrca02 iwinrca01

  Disk Partition                        Sharing Nodes (2 in count)
  ------------------------------------  ------------------------
  \Device\Harddisk1\Partition3          iwinrca02 iwinrca01

  Disk Partition                        Sharing Nodes (2 in count)
  ------------------------------------  ------------------------
  \Device\Harddisk1\Partition4          iwinrca02 iwinrca01

  Disk Partition                        Sharing Nodes (2 in count)

 ------------------------------------  ------------------------
  \Device\Harddisk1\Partition5          iwinrca02 iwinrca01

Shared storage check was successful on nodes "iwinrca02,iwinrca01".

Post-check for hardware and operating system setup was successful.
```

ORACLE

© 2006 Oracle Corporation – Proprietary and Confidential

CVU – pre crsinst – Part 1

```
C:\stage\10gR2\clusterware\cluvfy>runcluvfy stage -pre crsinst  -n iwinrca01,iwinrca02
The system cannot find the file specified.
Performing pre-checks for cluster services setup

Checking node reachability...
Node reachability check passed from node "iwinrca01".
Checking user equivalence...
User equivalence check passed for user "Administrator".
Checking administrative privileges...

Checking node connectivity...
 Node connectivity check passed for subnet "139.185.150.0" with node(s) iwinrca02,iwinrca01.
Node connectivity check passed for subnet "10.10.10.0" with node(s) iwinrca02,iwinrca01.

Suitable interfaces for VIP on subnet "139.185.150.0":
iwinrca02 Local Area Connection:139.185.150.55
iwinrca01 Local Area Connection:139.185.150.54
.
.
.
```

© 2006 Oracle Corporation – Proprietary and Confidential

CVU – pre crsinst – Part 2

```
.
.
.
Suitable interfaces for the private interconnect on subnet "10.10.10.0":
iwinrca02 Local Area Connection 2:10.10.10.3
iwinrca01 Local Area Connection 2:10.10.10.2

Node connectivity check passed.

Checking system requirements for 'crs'...
Operating system version check passed.
Operating system patch check passed for "1".
Total memory check passed.
Swap space check passed.
System architecture check passed.
Free disk space check passed.

System requirement passed for 'crs'

Pre-check for cluster services setup was successful.
```

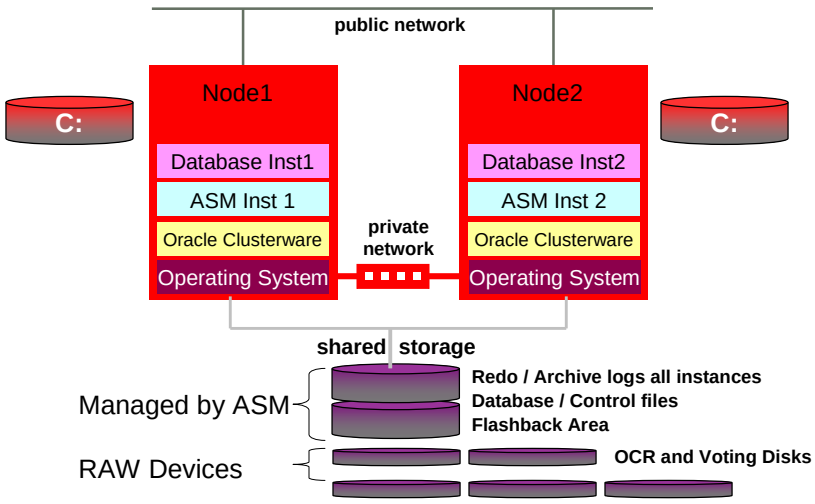
© 2006 Oracle Corporation – Proprietary and Confidential

Install - Continue with your Install

- Stage Software
- Install Oracle Clusterware
  - Own Home
- If using ASM
  - Install ASM Home and create first ASM diskgroup
- Install Oracle Database Home
  - Own Home
- Run dbca to create database

© 2006 Oracle Corporation – Proprietary and Confidential

Install - Environment Post Install



© 2006 Oracle Corporation – Proprietary and Confidential

Installation – Best Practices

- Use CVU to check at each stage of the install
- Eliminate Single Points of Failure
  - Network
  - Storage
  - Power Supply
- Use Fast Network Switches
  - GBE for Interconnect
- Use Static IP Addresses
  - Each node requires 3 IP Addresses
  - Do not use Private IP range for Public Address and VIP
  - Do not use Public IP range for Private Address

ORACLE

© 2006 Oracle Corporation – Proprietary and Confidential

Administration

ORACLE

© 2006 Oracle Corporation – Proprietary and Confidential

Admin Windows

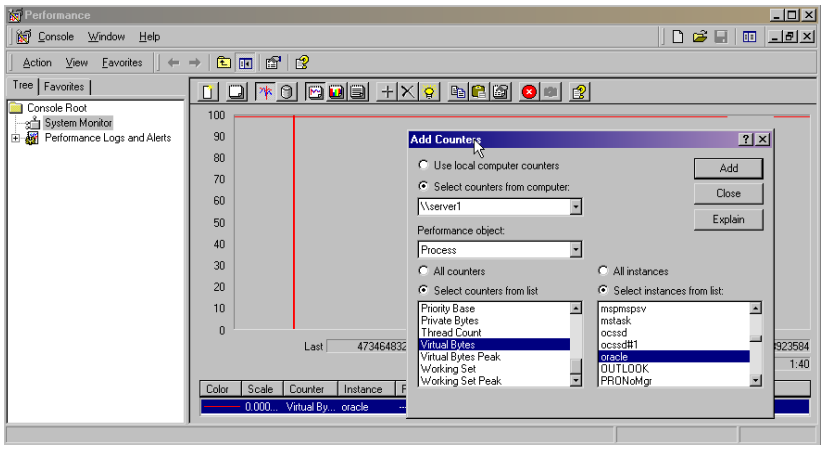
- Windows
  - Taskmanager
  - Event viewer
  - Performance monitor
    - Counters
    - Not loaded by default
- Custom Software Install -
- `Oracle Windows Interfaces 10.2.0.1.0`
  - `"Oracle Counters for Windows Performance Monitor 10.2.0.1.0".`

```
C:> operfcfg.exe -U SYSTEM -P password -D TNS_Alias
```

ORACLE

© 2006 Oracle Corporation – Proprietary and Confidential

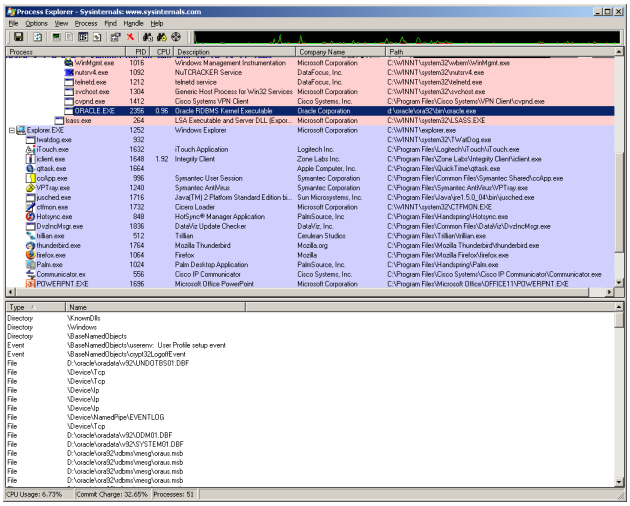
Performance Monitor



ORACLE

© 2006 Oracle Corporation – Proprietary and Confidential

Process Explorer



© 2006 Oracle Corporation – Proprietary and Confidential

Admin Oracle

- dbconsole
- EM Grid Control
- Command line tools
  - crsctl
  - srvctl
  - crs_stat
  - ocrconfig
  - Guioracleobjmanager
- Monitoring

© 2006 Oracle Corporation – Proprietary and Confidential

Startup

- RAC instances not set to automatic start
  - Unlike Single Instance
  - HKLM\software\oracle\<oraclehomekey>\
- Instances get started by Oracle Clusterware
- Use process delay (opmnd.exe) to wait for all drivers to load before loading the Oracle Services.
  - See Metalink Note: 358156.1

ORACLE

© 2006 Oracle Corporation – Proprietary and Confidential

Log Files

- Traditional alert logs etc… bdump udump
- Clusterware logs
  - ORA_CRS_HOME \log \<nodename>\
    - Clusterware Alert log
    - Various Clusterware component logs

ORACLE

© 2006 Oracle Corporation – Proprietary and Confidential

BACKUP

- Database files
  - Best practice is ASM using RMAN
  - OCFS – filesystem – hot backup / RMAN
- Clusterware Files
  - Best Practice is to use Oracle Clusterware redundancy
  - Cluster registry – OCR (1 or 2 devices)
  - Votedisk - (1 or 3 devices)
    - A file if using OCFS
  - Note path is a symbolic name
    - Use guioracleobjmanager.exe to setup

ORACLE

© 2006 Oracle Corporation – Proprietary and Confidential

Backup & Restore the OCR

- OCR is backed up automatically by Oracle Clusterware every 4 hours (on the OCR Master node) to ORA_CRS_HOME\cdata\crs
- Use ocrconfig –showbackup

```
C:\oracle\product\10.2.0\crs\cdata\crs>ocrconfig -showbackup
iwinrca02     2006/09/10 22:00:26     C:\oracle\product\10.2.0\crs10\cdata\crs
iwinrca02     2006/09/10 18:00:25     C:\oracle\product\10.2.0\crs10\cdata\crs
iwinrca02     2006/09/10 14:00:24     C:\oracle\product\10.2.0\crs10\cdata\crs
iwinrca02     2006/09/09 06:00:16     C:\oracle\product\10.2.0\crs10\cdata\crs
iwinrca02     2006/08/30 05:59:10     C:\oracle\product\10.2.0\crs10\cdata\crs
C:\oracle\product\10.2.0\crs\cdata\crs10>dir
09/10/2006  10:00p           6,123,520 backup00.ocr
09/10/2006  06:00p           6,123,520 backup01.ocr
09/10/2006  02:00p           6,123,520 backup02.ocr
09/09/2006  06:00a           6,123,520 day.ocr
08/30/2006  05:59a           6,123,520 week.ocr
```

- The crs10 subdirectory of cdata is the clustername

ORACLE

© 2006 Oracle Corporation – Proprietary and Confidential

Backup & restore the OCR

- ocrconfig –replace
  - ocrconfig -replace ocr *destination_file or disk*
  - ocrconfig -replace ocrmirror *destination_file or disk*
- ocrconfig –repair
  - Fixes location information for the OCR
    - In case or replacement with offline nodes
- Location of OCR
  - Held in registry
    - HKLM\software\oracle\ocr\ocrconfig_loc
    - HKLM\software\oracle\ocr\ocmirrorconfig_loc
    - Usually points to \\.\ocrcfg & \\.\ocrmirrorcfg on RAW partitions

ORACLE

© 2006 Oracle Corporation – Proprietary and Confidential

Backup & restore of votedevice

- Use OCOPY to backup the votedevice
- Usually located at \\.\votedsk1 on RAW partitions
  - Can be done at run-time
- Add & delete additional votedevices
  - Crsctl add css votedisk \\.*votedsk2*
  - Crsctl delete css votedisk \\.*votedsk2*
- Must use GuiOracleObjManager to create symbolic links first

ORACLE

© 2006 Oracle Corporation – Proprietary and Confidential

Migration

- You can use the same on-disk structure when migrating from 32bit to 64bit.
  - Disconnect SAN from 32bit machines
  - Connect 64bit machines to SAN
  - Install Oracle Software

ORACLE

© 2006 Oracle Corporation – Proprietary and Confidential

Summary

- Relationship
- Architecture
- Installation
- Administration

ORACLE

© 2006 Oracle Corporation – Proprietary and Confidential

Demonstrations in Exhibition Area

Moscone – West DEMOground

- 3 Demonstrations in DEMOground area
  - Oracle Real Application Clusters
  - Oracle Automatic Storage Management
  - Oracle Clusterware

ORACLE

© 2006 Oracle Corporation – Proprietary and Confidential

ORACLE

© 2006 Oracle Corporation – Proprietary and Confidential

ORACLE IS THE INFORMATION COMPANY