An introduction to 10g RAC

Christian Bilien
Enezis

Agenda

Planning Best Practices
- Understand and Plan the Architecture
- Manage Expectations
- Define objectives and success criteria
- Project plan

Implementation Best Practices
- Infrastructure considerations
- Installation/configuration
- Database creation
- Application considerations

Planning

Understand the Architecture
– Cluster terminology
– Functional basics

HA by eliminating node & Oracle as SPOFs

Scalability by making additional processing capacity available incrementally
– Hardware components

Private interconnect/network switch

Shared storage/concurrent access/storage switch
– Software components
OS, Cluster Manager, DBMS/RAC, Application

Differences between cluster managers

RAC Hardware Architecture

- Clustered Database Servers
- Mirrored Disk Subsystem
- High Speed Switch or Interconnect
- Hub or Switch Fabric
- Network Centralized
- Management Console
- Storage Area Network
- Low Latency Interconnect ie. GigE or Proprietary
- No Single Point Of Failure
- Shared Cache

RAC Software Architecture

- Shared Disk Database
- Shared Memory/Global Area
- Shared Data Model

Plan the Architecture

Eliminate SPOFs
- Cluster interconnect redundancy (NIC bonding/teaming, …)
- Implement multiple access paths to the storage array using 2 or more HBA's or initiators: multi-pathing software over these multiple devices to provide load balancing and failover.

Processing nodes – sufficient CPU to accommodate failure

Scalable I/O Subsystem
- Scalable as you add nodes

Workload Distribution (load balancing) strategy
- Net Services (SQL*Net)
- Oracle10g Services

Establish management infrastructure to manage to Service Level Agreements
- Grid Control

Cluster Hardware Considerations

Cluster interconnects
- FastEthernet, Gigabit Ethernet, Proprietary interconnects (SCI, Hyperfabric, memory channel, …)
- Dual interconnects, stick with GigE/UDP

Public networks
- Ethernet, FastEthernet, Gigabit Ethernet

Server Recommendations
- Minimum 2 CPUs per server
- 2 and 4 CPU servers normally most cost effective
- 1-2 GB of memory per CPU
- Dual IO Paths

Intelligent storage, or JBOD

Fiber Channel, SCSI, iSCSI or NAS storage connectivity

Future: Infiniband

Plan the Architecture

- Shared storage considerations (ASM, CFS, shared raw devices)
- Use S.A.M.E for shared storage layout
- Local ORACLE_HOME versus shared ORACLE_HOME
- Separate HOMEs for CRS, ASM, RDBMS
- OCR and Voting Disk on raw devices, unless using CFS

RAC Technology Certification

For more details on software certification and
compatible hardware:
http://technet.oracle.com/support/metalink/content.html

Discuss hardware configuration with your HW vendor

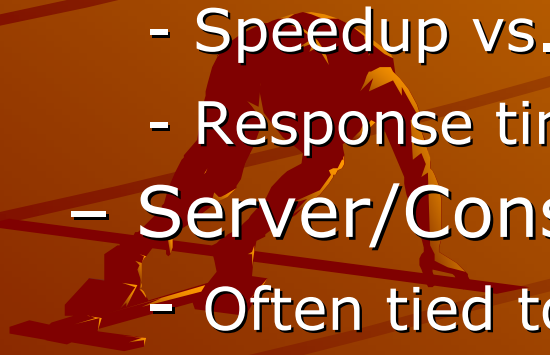
Try to stick to standard components that have been properly tested/certified

Set Expectations Appropriately

- If your application will scale transparently on SMP, then it is realistic to expect it to scale well on RAC, without having to make any changes to the application code.
- RAC eliminates the database instance, and the node itself, as a single point of failure, and ensures database integrity in the case of such failures

Planning: Define Objectives

- Objectives need to be quantified/measurable
- HA objectives
  - Planned vs. unplanned
  - Technology failures vs. site failures vs. human errors
- Scalability Objectives
  - Speedup vs. scaleup
  - Response time, throughput, other measurements
- Server/Consolidation Objectives
  - Often tied to TCO
  - Often subjective



Agenda

Planning Best Practices
- Understand and Plan the Architecture
- Manage Expectations
- Define objectives and success criteria
- Project plan

Implementation Best Practices
- Infrastructure considerations
- Installation/configuration
- Database creation
- Application considerations

Implementation Flowchart

- Configure HW
- Configure OS, Public Network, Private interconnect
- Configure Shared storage
- Install Oracle Software, including RAC and ASM
- Run VIPCA, automatically launched from RDBMS root.sh
- Create database with DBCA
- Install Oracle CRS Validate cluster/RAC configuration

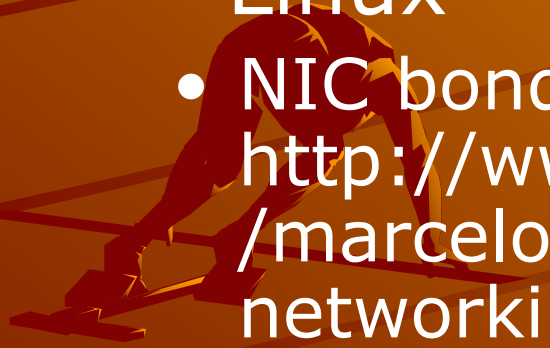
Operating System Configuration

- Confirm OS requirements from
  - Platform-specific install documentation
  - Quick install guides (if available) from Metalink/OTN
  - Release notes
- Follow these steps on EACH node of the cluster
  - Configure ssh
    10g OUI uses ssh, not rsh
  - Configure Private Interconnect
    . Use UDP and GigE
    . Non-routable IP addresses (eg 10.0.0.x)
    . Redundant switches as std configuration for ALL cluster sizes.
    . NIC teaming configuration (platform dependant)
  - Configure Public Network
    . VIP and name must be DNS-registered in addition to the standard static IP information
    . Will not be visible until VIPCA install is complete

NIC Bonding

Enezi

- Required for private interconnect resiliency.
- Various 3rd party vendor solutions available:
  - Linux
    - NIC bonding in RHEL 3.0 ES http://www.kernel.org/pub/linux/kernel/people/marcelo/linux-2.4/Documentation/networking/ bonding.txt
    - Intel® Advanced Network Services (ANS)
    - HANIC http://oss.oracle.com/projects/hanic



NIC Bonding cont.

- Solaris IPMP
- HP-UX APA
- AIX Etherchannel
- Windows

Shared Storage Configuration

- Configure devices for the Voting Disk and OCR file.
  - Voting Disk >= 20MB, OCR >= 100MB.
  - Use storage mirroring to protect these devices
- Configure shared Storage (for ASM)
  - Use large number of similarly sized “disks”
  - Confirm shared access to storage “disks” from all nodes
  - Use storage mirroring if available
  - Include space for flash recovery area
- Configure IO Multi-pathing
  - ASM must only see a single (virtual) path to the storage
  - Multi-pathing configuration is platform specific (e.g. Powerpath, SecurePath, …)
- Establish file system or location for ORACLE_HOME, CRS & ASM HOME

CRS Installation

- Create two raw devices for OCR and voting disk
- Install CRS/CSS stack with OUI
- Start stack with $CRS_HOME/root.sh

(inserted into /etc/inittab)

CRS Installation

- CRS is REQUIRED to be installed and running prior to installing 10g RAC.
- CRS must be installed in a different location from the ORACLE_HOME, (e.g. ORA_CRS_HOME).
- Shared Location(s) or devices for the Voting File and OCR file must be available PRIOR to installing CRS.
  – Reinstallation of CRS requires re-initialization of devices, including permissions.
- CRS and RAC require that the private and public network interfaces be configured prior to installing CRS or RAC
- Specify virtual interconnect for CRS communication

CRS Installation

- Only one set of CRS daemons can be running per RAC node.
- On Unix, the CRS stack is run from entries in/etc/inittab with 'respawn'.
- 10gR2:
  - Crsctl start crs
  - Crsctl stop crs

Installation flowchart

- Install Oracle software
- Root.sh (all nodes)
- Vipca: define vip
- Netca
- Dbca

VIP Installation

- The VIP Configuration Assistant (vipca) starts automatically from $ORACLE_HOME/root.sh
- The VIP must be a DNS known IP address because we use the VIP for the tnsnames connect.
- After finishing this you will see a new VIP interface eg:eth0:1. Use ifconfig (on most platforms) to verify this.

VIP Installation cont.

Enezi

- If a cluster is moving to a new datacenter (or subnet) it is necessary to change IPs. The VIP is stored withinthe OCR and any modification or change to the IPrequires additional administrative steps
  - See Metalink Note:276434.1 for details

Create RAC database using DBCA

- Set MAXINSTANCES, MAXLOGFILES, MAXLOGMEMBERS, MAXLOGHISTORY, MAXDATAFILES (auto with DBCA)
- Create tablespaces as locally Managed (auto with
- DBCA)
- Create all tablespaces with ASSM (auto with DBCA)
- Configure automatic UNDO management (auto with DBCA)
- Use SPFILE instead of multiple init.ora's (auto with DBCA)

Validate Cluster Configuration

- Query OCR to confirm status of all defined services: crsstat –t

Use script from Note 259301.1 to improve output formatting/ readability

```
HA Resource Target State
ora.BCRK.BCRK1.inst                    ONLINE ONLINE on sunblade-25
ora.BCRK.BCRK2.inst                    ONLINE ONLINE on sunblade-26
ora.BCRK.db ONLINE                     ONLINE on sunblade-25
ora.sunblade-25.ASM1.asm               ONLINE ONLINE on sunblade-25
ora.sunblade-25.LISTENER_SUNBLADE-25.lsnr ONLINE ONLINE on sunblade-25
ora.sunblade-25.gsd                    ONLINE ONLINE on sunblade-25
ora.sunblade-25.ons                    ONLINE ONLINE on sunblade-25
ora.sunblade-25.vip                    ONLINE ONLINE on sunblade-25
ora.sunblade-26.ASM2.asm               ONLINE ONLINE on sunblade-26
ora.sunblade-26.LISTENER_SUNBLADE-26.lsnr ONLINE ONLINE on sunblade-26
ora.sunblade-26.gsd ONLINE             ONLINE on sunblade-26
ora.sunblade-26.ons ONLINE             ONLINE on sunblade-26
ora.sunblade-26.vip ONLINE             ONLINE on sunblade-26
```

Validate RAC Configuration

- Instances running on all nodes

SQL> select * from gv$instance

- RAC communicating over the private Interconnect

SQL> oradebug setmypid
SQL> oradebug ipc
SQL> oradebug tracefile_name
/home/oracle/admin/RAC92_1/udump/rac92_1_ora_1343841.trc

- – Check trace file in the user_dump_dest:

SSKGXPT 0x2ab25bc flags info for network 0
socket no 10 IP **10.0.0.1** UDP 49197
sflags SSKGXPT_UP
info for network 1
socket no 0 IP 0.0.0.0 UDP 0
sflags SSKGXPT_DOWN

Validate RAC Configuration

- RAC is using desired IPC protocol: Check Alert.log
  cluster interconnect IPC version: Oracle UDP/IP
  IPC Vendor 1 proto 2 Version 1.0
  PMON started with pid=2
- Use cluster_interconnects only if necessary
  – RAC will use the same "virtual" interconnect selected during CRS install
  – To check which interconnect and is used and where it came from use select INDX ,PICKED_KSXPIA ,IP_KSXPIA from x$ksxpia;

INDX	PICKED_KSXPIA	IP_KSXPIA
0	OCR	172.16.193.1
1	OCR	10.252.25.143

- Pick: OCR … Oracle Clusterware
  OSD … Operating System dependent
  CI … indicates that the init.ora parameter cluster_interconnects is specified

SRVCTL

- srvctl status nodeapps -n <nodename> will show all services running on a node
- srvctl commands are documented in Appendix B of the RAC Admin Guide
- srvctl uses information from the OCR file

What is a Service?

- In Oracle10g services are built into the database.
- Divides work into logical workloads which share common functions, service level thresholds, priority & resource needs.
- Examples:
  - OLTP & Batch
  - ERP, CRM, HR, Email
  - DW & OLTP
  - Affinity Group 1,2,3,4,5,6,7,8,9,10