Planning & Installing a RAC Database

Caleb Small, BSc, ISP
Caleb@Caleb.com
www.Caleb.com/dba
Who is Caleb?

- Lifelong IT career
- Oracle DBA since v7.0
- Former Instructor for Oracle Corp.
- Independent consultant
- Faculty appointment Camosun College
- U of W night courses
- Vice-president, VicOUG
- Director, PSOUG
What does he know about RAC?

- Installing Oracle Databases for 12yrs
- Studied RAC on 9i
- Re-wrote RAC curriculum for 10gR2
- Teaching RAC workshop for PSOUG
- Various RAC installations
- Setup PSOUG and VicOUG RAC labs
PSOUG Training Lab
PSOUG Server Room

- HP DL360 3G
- NetApp 270C
- Disk Shelf
- NetApp F720
- Disk Shelf
- Disk Shelf
So, you’re thinking of RAC…

Why?
Benefits of RAC

• Reliability
  – Eliminate the Instance (and server) as single point of failure

• Scalability
  – Exploit low cost commodity hardware
  – Add servers as necessary
Benefits of RAC

- Part of complete High Availability (HA) solution
- Application Failover (TAF and FCF)
- Load Balancing
- Incremental Scaling
- Lower Initial Cost / Higher Return On Investment
- Lower Maintenance Costs
- Server Consolidation
Challenges of RAC

- Complex architecture
- Setting and meeting realistic expectations
- Multiple technologies & skills required…
- Traditional DBA skills plus RAC specific
- Various personnel must understand the architecture and cooperate
Planning a RAC installation

People and process issues

– Understand the architecture
– Set expectations appropriately
– Assemble the technical team
– Plan and budget for training
– Cooperation and management
– Define and measure success factors
Planning a RAC installation

Technical issues

– Choose hardware / OS
– Design networks
– Design storage
– Install & configure software
– Modify applications
– Implement Services and Resource Plans
– Implement Load Balancing & Failover
– Performance monitoring & tuning
RAC Architecture

- RAC != “Clusters” the database object
- Database != Instance
- A database is a set of files containing all persistent resources
- An instance is a set of memory structures and processes
- Databases and Instances can be started and stopped independently
- In a stand-alone database the ratio is 1:1
- In a RAC database the ratio is 1:Many
What is RAC?

- Multiple instances running on separate servers (nodes)
- No instance is the parent or is privileged
- Single database on shared storage accessible to all nodes
- Instances exchange information over a private interconnect network
Cluster Topology

Node1
- Database instance 1
- ASM Instance 1
- CRS
- Operating System

Node2
- Database instance 2
- ASM Instance 2
- CRS
- Operating System

Node3
- Database instance 3
- ASM Instance 3
- CRS
- Operating System

Public network

Cluster interconnect

Storage network

Redo logs all instances
- Database files
- Spfile, control files
- OCR and Voting Disk

VIP1
VIP2
VIP3

Puget Sound Oracle Users Group
Caleb@Caleb.com
Cluster Topology

Centralized Management Console

High Speed Switch or Interconnect

Clustered Database Servers

Hub or Switch Fabric

Mirrored Disk Subsystem

Low Latency Interconnect

Network

Users

Storage Network (SAN or NAS)

No Single Point Of Failure

Puget Sound Oracle Users Group

Caleb@Caleb.com
Shared Cache Databases

- Oracle RAC is the only Shared Cache database
- Provide BOTH availability and scalability
- Maintained like a single database
- More reliable with more servers
Set Expectations

Understand what RAC is, and what it is NOT.
Scalability

• Required for performance and customer service
• Scaling up to another machine is expensive
• Plug-in scale-up and scale-down with the addition/removal of servers (nodes)
• Scale Up (the Old Approach)
  – Add CPU or memory to a server
    • Eg: Add 2 850 MHz procs and 2GB RAM quoted at $150K
• Scale Out (the New Approach)
  – Add more servers
    • Eg: Add dual 3.0 GHz server w/4GB RAM quoted at $24K
Value Proposition

• Exploit commodity hardware
• Lower maintenance costs
• Lower down-time costs
Pay and Scale Incrementally

- Workload
  - 100%
  - 200%
  - 300%

- Months
  - 3
  - 6
  - 9
  - 12
  - 15
  - 18
  - 21
  - 24

Puget Sound Oracle Users Group

Caleb@Caleb.com
E-Business Suite RAC Scalability

Oracle11i E-Business Suite Benchmark

84% Scalability

# Users

<table>
<thead>
<tr>
<th>Nodes</th>
<th>1 Node</th>
<th>2 Nodes</th>
<th>4 Nodes</th>
<th>5 Nodes</th>
<th>6 Nodes</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1,288</td>
<td>2,296*</td>
<td>4,368*</td>
<td>5,433</td>
<td>6,496</td>
</tr>
</tbody>
</table>

Puget Sound Oracle Users Group

Caleb@Caleb.com
Scalability Expectations

- 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.

- The “Spotlight Effect”
  - RAC is very good at putting a bright “spotlight” on poorly designed databases and applications.
Reliability

• 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

• RAC is **one part** a complete High Availability solution

• All the technology in the world will not make a poorly run shop reliable
Plan For Failures

• Eliminate Single Points of Failure
  – Cluster interconnect redundancy (NIC bonding / teaming)
  – Implement multiple access paths to the storage array
• Sufficient nodes to provide adequate CPU in the event of failover
• Scalable I/O subsystems
• Plan for downtime
• Establish realistic Service Level Agreements
Fault Tolerance

Unplanned Downtime
- System Failures
- Data Failures
- Human Error

Planned Downtime
- System Maintenance
- Database Maintenance
- Storage/Net Failures

Real Application Clusters
Continuous Availability for all Applications

Data Guard
Guaranteed Zero Data Loss

Flashback
Guaranteed Zero Data Loss

Dynamic Reconfiguration
Capacity on Demand without Interruption

Online Redefinition
Adapt to Change Online

ASM Mirroring
Storage Failure Protection
More Nodes = Higher Availability

- 2 Node Cluster
  - Retain 50% of CPU

- 4 Node Cluster
  - Retain 75% of CPU

- 6 Node Cluster
  - Retain 83% of CPU

There is no Convoy Effect
High-Availability (HA)

System HW
ECC, Hot Plug, Redundant PS and fans, etc.

Inc. Data Avail.
RAID, mirroring smart cont., Storage subsystems

Fault Recoverability
Image clusters, Cluster aware apps

Scalable Parallel Cluster

Wide Area Clusters
Geo. Sep. Replication Transaction Routing

Single System
Local Clusters
Distributed Clusters

Single Site
Multi-Site

88 hr/yr  8 hr/yr  60 min/yr  5 min/yr  0 min/yr

99.0% Normal Commercial Availability
99.9% High Availability
99.99% Fault Resilient Clusters Failover
99.999% Fault Tolerant
100.0% Continuous Processing

Puget Sound Oracle Users Group Caleb@Caleb.com
A Complete High-Availability Solution

- Recovery Manager (RMAN)
- Flashback Database / Query / Table
- Streams Replication
- Grid Control
- Real Application Clusters (RAC)
- Data Guard

RAC is one part of a complete HA solution
Server Consolidation

Single system image for competing workloads running across multiple systems

Puget Sound Oracle Users Group
Caleb@Caleb.com
Discrete Servers

Sales: 100%
Service: 20%
HR: 50%
Financials: 40%

Puget Sound Oracle Users Group

Caleb@Caleb.com
Discrete Servers

- Sales: 10%
- Service: 100%
- HR: 10%
- Financials: 80%
Discrete Servers

Sales: 40%
Service: 20%
HR: 100%
Financials: 40%
RAC Cluster

Finance
HR
Service
Sales
RAC Cluster After Failure

Finance
HR
Service
Sales
Assembling the Team

- Multiple skills are required
  - Traditional DBA plus RAC specific
  - System Administrator
  - Network Engineer
  - Storage Engineer
  - Application Developer
  - ...

Assembling the Team

• Unusual architectures are involved
  – Private networks
  – Virtual IP addresses
  – Network bonding
  – Cluster File Systems
  – RAW stripes on SAN
  – Root access to servers
  – Multiple instance connections
  – Load balancing & failover technologies
Assembling the Team

• Managing the Team
  – Who’s the boss?
  – The DBA “get’s it” when things break
  – Someone (DBA) must have working knowledge of all technical aspects:
    • Hardware
    • Operating System
    • Networks
    • Storage
    • Clusterware
    • Database
    • Application Design
Choosing Hardware & OS

- Existing expertise
- Relationships with vendors
- Political reasons
- Price and value
- Technical specifications
- Be sure it is certified for RAC!
Planning the Networks

- Three separate & distinct networks:
  - “Public” interface
  - Cluster interconnect
  - Storage network
- Gig ether
- Fiber channel
- Each network should be fully redundant and bonded.
General Networking Concepts

• Public Networks
  – Used for application communication
  – Not recommended for cluster communication

• Private Networks
  – Used for cluster communication
  – Can not be seen by machines not connected directly to the private network

• Storage Networks
  – Database I/O
Typical RAC Network Topology

Multi Path – dual redundant NICs, switches and cables for no single point of failure
Planning the Storage

• Hardware storage solution
  – SAN (storage area network) eg. EMC
  – NAS (network attached storage) eg. NetApp
  – JBOD (just a bunch of disks)

• Shared File System solution
  – NFS (certified Network File System)
  – ASM (Oracle Automatic Storage Management)
  – OCFS (Oracle Cluster File System)
  – RAW (raw devices)
  – Proprietary (Veritas, IBM GPFS, etc)
## Supported Storage Options

<table>
<thead>
<tr>
<th>Storage Option</th>
<th>OCR and Voting Disks</th>
<th>Oracle Software Installation</th>
<th>Database</th>
<th>Recovery</th>
</tr>
</thead>
<tbody>
<tr>
<td>Automatic Storage Management (ASM)</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>OCFS</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>OCFS2 (not cert.)</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>GPFS* for Linux on IBM POWER PC</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Local Storage</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>NFS File System</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Shared Raw Partitions</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
</tr>
</tbody>
</table>

* IBM General Parallel File System. Other CFS supported include DBE/AC (Veritas) and Tru64 CFS
Install and Configure Software

• OS install
• Oracle specific mods
• Network interface configuration
• CFS / ASM install
• Oracle Clusterware install
• Oracle Database binaries install
• Configure Oracle Net
• Create RAC database
• Configure Services, Resource Plans, Load Balancing & Failover…
OS Install

- Typically involves sys admin
- Select packages / components
- Install device drivers
- Recognize NICs / HBAs
- Configure NTP (network time protocol)
- Root access
Oracle Specific Mods

- Users & groups with consistent IDs
- Directories & permissions
- Kernel parameters
- Additional packages (eg. cvuqdisk)
- Configure hangcheck timer
- Environment variables
- Configure SSH
Network Interface Configuration

- **Public Network**
  - Physical (real) IPs
  - Virtual IPs

- **Private Interconnect Network**
  - Physical, private, non-routable IPs
  - Dual port, NIC bonding (different cards)

- **Storage Interconnect**
  - Dual gig E or fiber, bonded
  - Non-routable IPs
Oracle Clusterware Install

- Requires access to shared storage
- CFS or RAW but not ASM!
- Creates two critical shared files:
  - OCR (Oracle Cluster Registry)
  - Voting Disk
- Requires interconnect, storage and public networks in place
- Requires VIPs in place
- Will fail if any previous step missed
Clusterware Can Provide Application HA

Framework
- Please Start
- How are you?
- Please Stop

Agent
- 'Start'
- 'Check'
- 'Stop'

Puget Sound Oracle Users Group
Caleb@Caleb.com
Oracle Clusterware Install

- Clusterware installs on top of CFS/ASM
- Separate ORACLE_HOME
- Installed from primary node only
- Required for a RAC database
- Provides the HA framework
- Runs as root, managed as root
- Monitors & controls the node
- Can be configured to control other apps
Cluster Verification Utility

- **Setup Hardware**
- **Storage Network**
- **Set up OCFS (Opt)**
- **Installs CRS**
  - -pre crsinst
  - -post cfs
- **Installs RAC**
  - -pre dbinst
  - -post crsinst
- **Configures RAC DB**
  - -pre dbcfg
  - -post hwos

Puget Sound Oracle Users Group
Caleb@Caleb.com
CFS / ASM Install

- Installs on top of shared storage
- Provides read / **write** file sharing
- NFS is plug-n-play, no installation
- Others require installation / formatting
- ASM is a separate Oracle product install in a separate ORACLE_HOME
- ASM is another whole topic…. 
The “Black Line”

- Oracle RAC Database
- Clusterware Software
- Interconnect Software
- Cluster File System
- Cluster OS
- Clusterware Hardware
- Shared Disks
- Interconnects
- Clusterware Shared Disks
- Clusterware Hardware