

# **Oracle 9i RAC**

## **RAC Deployment Best Practices**

**Marshall Presser**  
**Principal Technologist**  
**Oracle Government, Education, and Health**

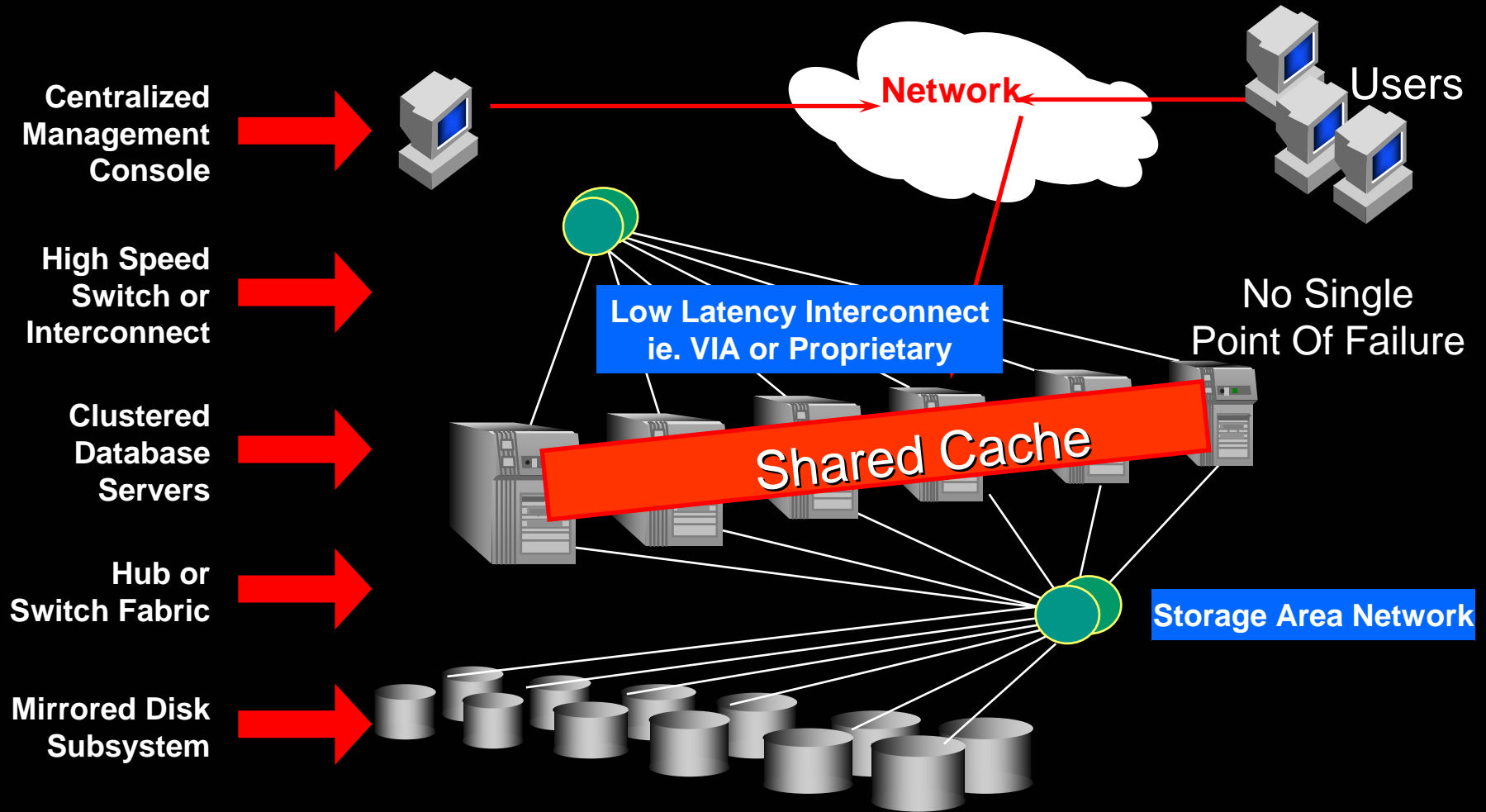
# Agenda

- Planning Best Practices
- Implementation Best Practices
- Production Migration Best Practices
- Customer Examples

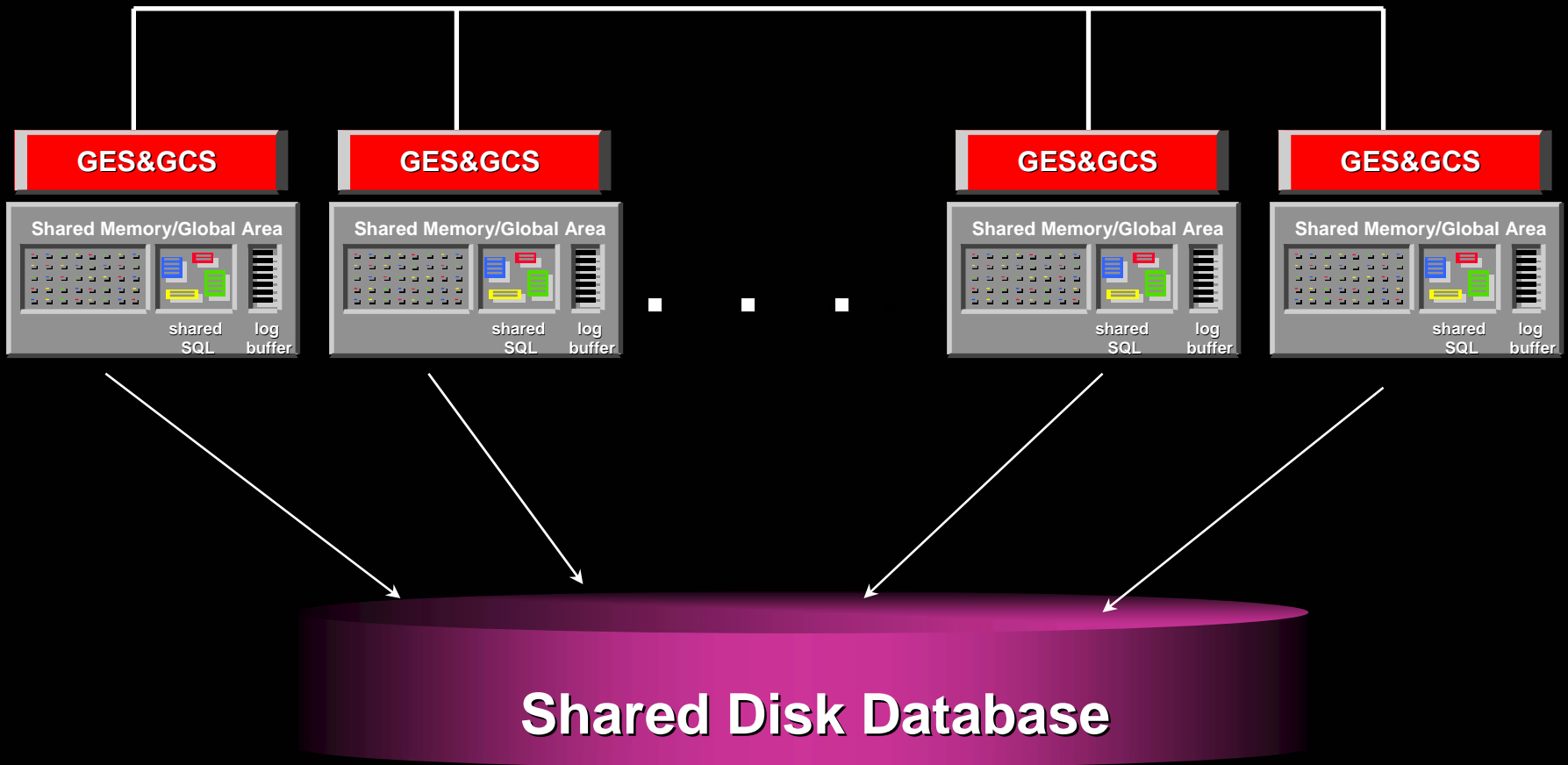
# Planning

- Understand the Architecture
  - Hardware components
  - Software components
  - Functional basics

# RAC Architecture



# RAC Architecture Shared Data Model



# Planning

- Set your 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 and quantify your objectives
  - 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

# Planning

- Build your project plan
  - Involve vendors as stakeholders
    - “one throat to choke”
  - Address knowledge gaps and training
    - Clusters, RAC, HA, Scalability, systems management
    - Leverage external resources as required
  - Establish Support mechanisms and escalation procedures

# Implementation

- Considerations for your infrastructure
  - I/O Subsystem
  - Processing nodes
  - Private Interconnect
    - Get the best interconnect available on Platform
      - High Bandwidth, Low Latency, failover
  - Cluster software
  - Eliminate SPOFs
  - Workload Distribution (load balancing) strategy
  - Systems management framework for monitoring and managing to SLAs

# Implementation

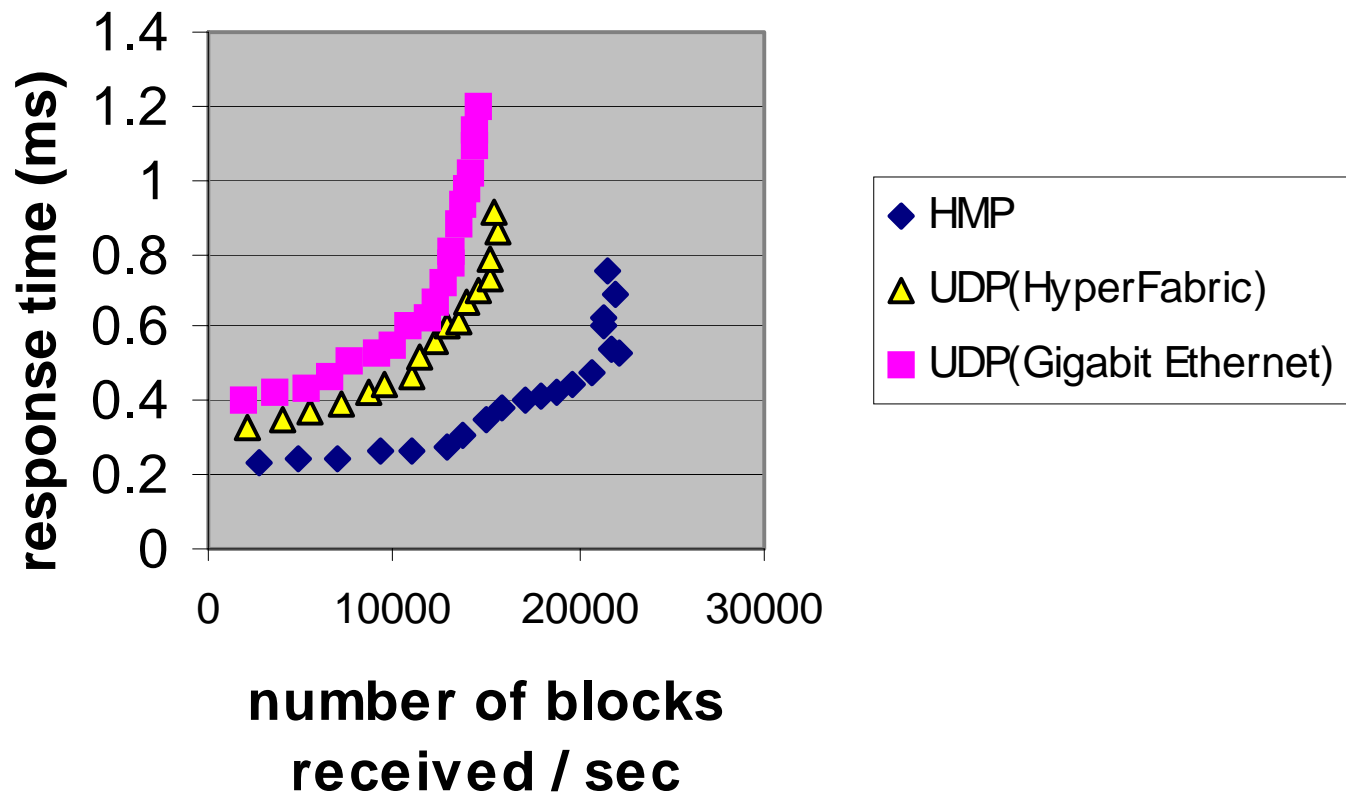
- Cluster Installation/configuration
  - Use Asynch IO
  - Use S.A.M.E. methodology
  - Raw device considerations
    - Small number of different sizes
    - Pool of predefined devices
  - CFS considerations
    - Availability vs managability tradeoffs
  - IPC considerations
    - If using TCP/UDP, set parameters to max value (usually 32K or 64K)
  - Test cluster functionality and shared storage access
  - Confirm cluster is communicating over the interconnect

# Implementation

- IPC Protocol
  - Messaging delays are highly dependent on IPC protocol.
  - The use of low-latency, user-mode IPC protocol, rather than UDP or TCP is strongly recommended
    - HP's Hyper Messaging Protocol (HMP)
    - HP/Compaq's Reliable Datagram (RDG)
    - SUN's Remote Shared Memory (RSM)
    - Virtual Interface Architecture (VIA) on Intel

# Scalability (2KB database block size)

## Throughput v.s. Response (2KB block)



# Implementation

- Oracle9i RAC Installation/Configuration
  - Use available documentation: Oracle9i ICG, Metalink “Step-by-Step” guides, OTN QuickStart Guides
  - Confirm prerequisites: disk space, memory, patch levels
  - Use current recommended Oracle software release (9.0.1.4, or 9.2.0.2)
  - Stage the Oracle CD’s to disk
  - Avoid NFS mounts for ORACLE\_HOME
  - Use OUI to install software across nodes

# Implementation

- Create your RAC database
  - Use DBCA to simplify DB creation
    - Ensure GSD is running on all nodes
  - 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)

# Implementation

- Automatic Segment Space Management (ASSM)
  - Eliminates complex process of computing PCTUSED, FREELISTS and FREELIST GROUPS
  - Allows dynamic affinity of space to instances and avoids hard partitioning of space inherent with free list groups.
    - Contention during concurrent access is removed and space usage optimized.
  - Doesn't need any maintenance.
  - Allows you to support any number of instances without any changes to the object.
  - Use OnLine rebuild features to move objects from Free List Groups to ASSM.

# Implementation

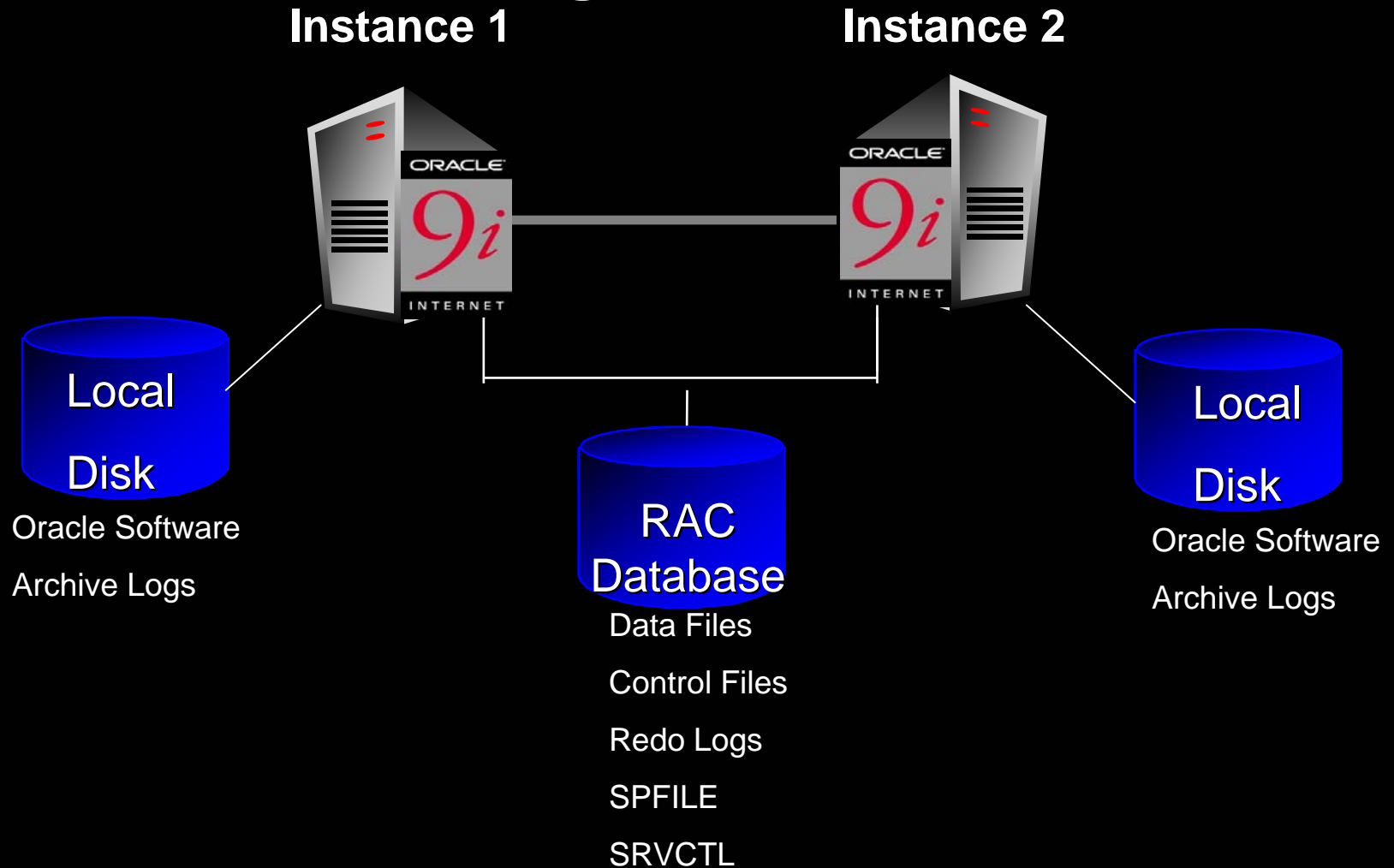
- Use automatic UNDO Management
  - Create one UNDO Tablespace for every instance
  - divide the 625MB among the instances
    - 50MB UNDO per instance minimum
  - Replaces Rollback Segments

# Implementation

- Ensure use of SPFILE
  - Replaces multiple, manually set init.ora, with one automatically managed file.
  - Syntax supports both global and instance-specific parameters
  - Used by default by DBCA.

# Implementation

## RAC Database Configuration



# Implementation

- Validate your RAC configuration
  - Instances running on all nodes
  - RAC communicating over the private Interconnect

```
SQL> oradebug setmypid
SQL> oradebug ipc
```

    - Check trace file in the user\_dump\_dest:

```
SSKGXPT 0x2ab25bc flags info for network 0
socket no 10 IP 204.152.65.33 UDP 49197
sflags SSKGXPT_UP
info for network 1
socket no 0 IP 0.0.0.0 UDP 0
sflags SSKGXPT_DOWN
```
  - 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
...
```

# Implementation

- Deploying your Application
  - No special application design or coding required for RAC
  - All applications that run well in a single instance environment should run well on RAC
  - But....
    - contention problems in single instance environment can be worse in a RAC environment.
    - Performance might be further improved if data dependent routing can be implemented in middle tier

# Implementation

- Deploying your Application
  - Same guidelines as single instance
    - SQL Tuning
    - Sequence Caching
    - Partition large objects
    - Use different block sizes
    - Tune instance recovery
    - Avoid DDL
    - Use LMT's and ASSM as noted earlier

# SQL Tuning

- Optimal Execution plan
- Shareable SQL
- Parsing
- Auditing
- Full table scans

# Sequences

- Sequence Number
  - Always use cache option
    - Sequence numbers can be lost
    - No guarantee of order
  - Alternate approach if no gap
    - pre-generation of numbers

# Partition Very Big Objects

- Use appropriate partitioning.
- Improves objects manageability.
- Hash, list, composite partitioning of DML intensive objects will help.

# Use Different Block Sizes

- Use large blocks for
  - Tables on which long scans are frequent.
  - For read-mostly Tables & Indexes.
  - Tables which are loaded with bulk load and no updates.
- Use small blocks for
  - At places not included above.

# Instance recovery

- Set `fast_start_mttr_target`.
- Size the buffer cache for single pass recovery.
- Ensure async I/O is used.
- Use recovery parallelism.

# Avoid DDL

- Do not create and delete tables as part of normal user applications – find alternatives
  - DDL accesses data dictionary and can create contention issues in both single instance and RAC

**“Switching from a single machine to Oracle9i Real Application Clusters was easier than upgrading from the last maintenance release!”**

**- Virgil Fernandez, CTO, Tomax Corporation**

# Implementation

- Monitoring and Tuning your database
  - Tune first for single instance 9i
  - Use Statspack:
    - snapshots at 10-20 min intervals during stress testing, hourly during normal operations
    - Run on all instances, staggered
  - Supplement with scripts/tracing
    - Monitor V\$SESSION\_WAIT to see which blocks are involved in wait events
    - Trace events like 10046/8 can provide additional wait event details
    - Monitor Alert logs and trace files, as on single instance
  - Oracle Performance Manager
  - RAC-specific views

# Performance Monitoring & Tuning

- Obvious application deficiency on a single node can't be solved by multiple nodes.
  - Single points of contention.
  - Not scalable on single instance DB
    - e.g. Real time transaction count update
  - I/O bound on single instance DB
- Tuning on single instance DB to insure applications scalable first

# Performance Monitoring & Tuning

- Deciding if RAC is the performance bottleneck
  - Amount of Cross Instance Traffic
    - Type of requests
    - Type of blocks
  - Latency
    - Block receive time
    - buffer size factor
    - bandwidth factor

# Production Migration

- Adhere to strong Systems Life Cycle Disciplines
  - Comprehensive test plans (functional and stress)
  - Rehearsed production migration plan
  - **Change Control**
    - **Separate environments for Dev, Test, QA/UAT, Production**
  - Backup and recovery procedures
  - Security controls

# RAC Reference Customers

[www.oracle.com/customers](http://www.oracle.com/customers)

- UPS - *Sun*
- FAA - *Linux*
- Lycos Europe – *Tru64*
- Dell Global IT - *Linux*
- Electronic Arts - *Linux*
- British Telecom - *Sun*
- Siemens ICM - *Sun*
- Korea Investment Trust Corporation - *HPUX*
- Axiom – *Tru64*
- NRW (German) Police - *Linux*
- Cern - *Linux*
- GM Vector SCM - *AIX*
- Austrian Railways – *Tru64*
- Freemarkets.com – *Tru64*
- Gas Authority of India (GAIL) - *Sun*
- Green Mountain Power – *Tru64*
- Nordac – *Tru64*
- Transports Mesguen – *Tru64*
- SITA – *OS390*
- Hite Brewery – *Win2K*
- Lithonia Lighting - *Linux*
- Oracle Global eMail - *HPUX*
- Oracle GSI - *HPUX*

# Electronic Arts (EA)

- World's leading independent developer and publisher of interactive entertainment software
- Plan on implementing 20 – 4 node clusters (2 CPU Dell/Linux).
- Performed ROI study:
  - Compared:
    - Linux cluster w/ RAC vs big Hardware
    - Small linux boxes ( 2 node 4 CPU) vs large linux boxes (4 node 2 CPU)
  - Results
    - Their application scaled very well on RAC
    - Choose RAC & 4 node system because of *lower cost* & ended up getting higher availability

# Green Mountain Power

- Power company, Burlington, Vt.
  - Numerous Applications, Including Field Service Management Application, Internal Accounting
  - PSFT, Banner migrations to follow by end of CY'02
  - Single Database, Multiple Schemas - 1 Per Application.
  - VMS Migration to Tru64, Oracle9i With Real Application Clusters.
  - Compaq TruCluster & RAC used as standard HA infrastructure for application, db, and server consolidation
- ***Successful Migration to Real Application Clusters in April, 2002 With No Application Changes or Problems.***
- Testimonial Video can be seen at:
  - <http://www.greenmountainpower.biz/whoweare/pressrel/cldsp/co.wvx>

# Axiom

- Based in Little Rock, Arkansas
- Provides data infrastructure, technology services, and data to help companies such as Visa, Capital One, Mercedes Benz and Palm understand customer behavior
- Manage a data warehouse of 25 - 50 terabytes
- Compaq Tru64 Clusters

*“With Oracle9i Real Application Clusters, we experienced a 38% performance improvement with no system tuning..”*

Tim Donar, Database Manager, Axiom

•Customer Profile:

<http://www.oracle.com/customers/profiles/PROFILE8118.HTML>

# Korea Investment Trust Corp. (KITC)

- Sells investment products, mainly Korean equity and bonds
- Home grown application developed for SMP
- RAC used for both HA and Scalability
- Fail over in 1 minute during stress test without special tuning or configuration
- 2-node HP Superdome cluster

*We depend on Oracle9i Real Application Clusters to provide the high availability and scalability that our stock market trading system requires."*

- Ryu, Si Hwan Deputy General Manager of Information System Strategy Department

- Customer Profile:

<http://www.oracle.com/customers/profiles/PROFILE6871.HTML>

# What our Customers Think

“RAC achieved failover times of between ten seconds and one minute.” - British Telecom

“Downtime could cost us \$1 million per day. That's why we use Oracle9*i* RAC.” - Schachter & Namdar

“We depend on Oracle9*i* RAC to provide the high availability and scalability that our stock market trading system requires.” - KITC

“We have been using Oracle's clustered database for over 17 months with close to zero downtime.” - VeriSign