Oracle RAC 10*g* Overview

An Oracle White Paper
November 2003

ORACLE®

Oracle RAC 10*g* Overview

TABLE OF CONTENTS

INTRODUCTION

Oracle Real Application Clusters (RAC) 10*g* is all about management. Management improvements in Oracle RAC 10*g* and in Oracle Database 10*g* itself dramatically reduce administration costs and provide new levels of flexibility that make business more adaptive, proactive, and agile. Oracle RAC 10g serves as a key foundation technology for enterprise Grids. Oracle Database 10*g* lowers the remaining barriers that have hindered the ubiquitous adoption of Oracle RAC technology, enabling the goal of "RAC for everyone".

Oracle RAC 10g serves as a key foundation technology for enterprise Grids and lowers the remaining barriers that have hindered the ubiquitous adoption of Oracle RAC technology, enabling the goal of "RAC for everyone".

ENTERPRISE GRIDS

Today, businesses are putting IT organizations under tremendous pressure to deliver the highest quality of service, in terms of maximum availability and scalability, at the lowest possible cost, and at the highest level of efficiency and flexibility. In short, IT is being asked to do much, much more with much, much less.

Enterprise Grids make these seemly impossible challenges achievable. Enterprise Grids, made up of large configurations of very low cost, commodity clusters, dramatically reduce the cost of computer hardware. Oracle RAC technology enables this low-cost hardware platform to deliver the highest quality of service that rivals and exceeds the levels of availability and scalability achieved by the most expensive, mainframe SMP computers. By dramatically reducing administration costs and providing new levels of administration flexibility, Oracle is enabling the enterprise Grid environment.

As enterprise Grids, data centers will have the ability to change their personality dynamically to meet the changing demands of business at each moment in time - just in time.

The enterprise Grid will have a profound impact, enabling business to be more adaptive, proactive, and agile. As enterprise Grids, data centers will have the ability to change their personality dynamically to meet the changing demands of business at each moment in time - just in time. Application workloads will be managed as services that must meet defined levels of quality. Processing resources and storage will be allocated to services in a fluid fashion to ensure that those quality levels are maintained. Each processing node or storage component in the Grid can change its individual personality almost instantaneously without any downtime for any application.

PART OF A LARGE SCALE EFFORT

The Oracle RAC 10*g* focus on manageability is part of a much larger Oracle Database effort. This effort began with the Oracle9*i* Database release where half of the operational cost of administering the Oracle Database was eliminated through improved management functions. This effort continues with Oracle Database 10*g* where the goal is to reduce the remaining database administration time and effort in half again. All of these enhancements directly improve the manageability of Oracle RAC environments because RAC is an integral component of the database.

A particularly significant achievement of this effort has been the success in automating many of Oracle's most advanced and unique technologies. It is easy to make functionality simple to use when the functionality is simplistic and of limited capability. The true challenge comes in making the most advanced capabilities work "out of the box" for everyone, every time. Major functional areas of Oracle's most advanced and unique technology have been totally automated, including:

- *Rollback Segment Management.* Rollback segments, which give the Oracle Database its unique ability to efficiently support mixed OLTP/DSS workloads, were difficult to administer but now require no tuning.

- *Backup & Recovery Management.* The Oracle Database's industry leading backup and recovery capabilities are now fully automated by Oracle Recovery Manager.

- *Memory Buffer Management.* The Oracle Database's sophisticated mechanism of both private area and shared global area memory buffers are a key enabler of Oracle's record setting scalability and performance. They now require no tuning.

- *Storage Management.* The Oracle Automatic Storage Manager 10*g* completely eliminates the need for I/O tuning by automatically balancing data placement across disks for optimal performance

WHAT IS RAC?

Oracle RAC enables the Oracle Database to run real applications on clusters. By real applications we mean that RAC supports mainstream business applications of all kinds. This includes popular packaged products such as SAP, PeopleSoft and Oracle E*Business Suite, as well as in-house developed applications. It includes both OLTP and DSS. And, it includes Oracle's unique ability to effectively support mixed OLTP/DSS environments. Oracle is

the only open systems database with this capability.

Oracle RAC running on a cluster provides Oracle's highest level of capability in terms of availability, scalability, and low-cost computing. If a node in the cluster fails, Oracle continues running on the remaining nodes. If more processing power is needed, new nodes can be easily added to the cluster. To keep costs low, even the highest-end systems can be built out of small, very low cost clusters made from standardized, commodity parts.

Oracle's key innovation is a technology called cache fusion, originally developed for Oracle9*i* Real Application Clusters. Cache fusion enables nodes on a cluster to synchronize their memory caches efficiently using a high-speed cluster interconnect so that disk I/O is minimized. The key, though, is that cache fusion enables shared access to all the data on disk by all the nodes on the cluster. Data does not need to be partitioned among the nodes. Oracle is the only open systems database with this capability.

Other database software that claims to run on clusters requires that the database data be partitioned. This is just not a practical possibility for real applications with complex data structures. It also makes changes to cluster systems all but impossible. If you add or remove nodes or storage resources the data needs to be totally re-partitioned.

Oracle RAC enables enterprise Grids. These are the data centers of the future. Enterprise Grids are built out of large configurations of standardized, commodity-priced components: processors, network, and storage. Oracle RAC's cache fusion technology provides the highest levels of availability and scalability. Oracle Database 10*g* and Oracle RAC 10*g* dramatically reduce operational costs and provide new levels of flexibility so that systems become more adaptive, proactive, and agile. Dynamic provisioning of nodes, storage, CPUs, and memory allow service levels to be easily and efficiently maintained while lowering cost still further through improved utilization.

NEW IN ORACLE REAL APPLICATION CLUSTERS 10*G*

The following are the key goals for Oracle RAC 10*g*. We will be discussing each of these in more detail below.

- Elimination of the need for 3rd party software components that add significantly to the complexity and cost of cluster environments by providing a full stack Oracle solution. Oracle RAC 10*g* integrated clusterware is discussed below. The other major component of this effort, Automatic Storage Manager, is discussed in other Oracle whitepapers.

- Management of cluster environments as a single, integrated system image.

- Automation of key functions such as workload management that previously had to be done manually.

- Improved integration with all other Oracle features and functions so that they fully and transparently support RAC environments.

- New and improved tools to verify the correctness of cluster configurations and help pinpoint the source of problems quickly.

- Performance improvements that will benefit many applications.

- Zero downtime patches for Oracle RAC cluster environments.

INTEGRATED CLUSTERWARE MANAGEMENT

Oracle RAC 10*g* offers a complete, integrated clusterware management solution on all platforms Oracle Database 10*g* runs on.

Oracle RAC 10*g* offers a complete, integrated clusterware management solution on all platforms Oracle Database 10*g* runs on. This clusterware functionality includes mechanisms for cluster connectivity, messaging and locking, cluster control and recovery, and a workload management framework (discussed below). No 3rd party clusterware management software need be purchased. Oracle will, however, continue to support select 3rd party clusterware products on specified platforms.
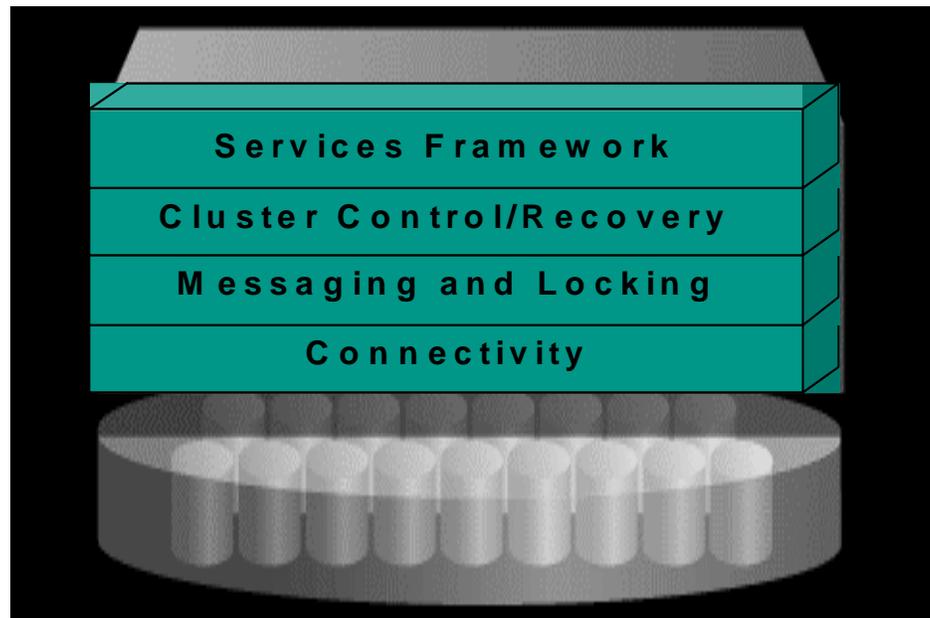

Figure 1. Integrated clusterware management in Oracle RAC 10g

Integrated clusterware management in Oracle RAC 10*g* offers the following benefits:

- *Low cost.* Oracle does not charge extra for this capability.

- *Single-vendor support.* No finger pointing is possible.

- *Simpler installation, configuration and ongoing maintenance.* The Oracle RAC 10*g* clusterware functionality is installed, configured and maintained using standard Oracle Database management tools. No extra integration steps are necessary.

- *Consistent high quality across all platforms.* Oracle is able to perform much more intensive testing of new software releases than is possible with 3rd party products.

- *Consistent functionality across all platforms.* For example, some 3rd party clusterware products limit the number of nodes that can be supported within a cluster. With Oracle RAC 10*g,* up to 64 nodes are supported on all platforms. Users can also expect a consistent response across all platforms to high availability challenges such as server node failures, interconnect failures, I/O fencing behaviors, and more.

- *Support for advanced functionality.* This includes integrated monitoring and notification capabilities to enable fast, coordinated recovery across both database and application tiers in case of failure (See "Fast Connection Fail-Over" section below).

SINGLE SYSTEM IMAGE MANAGEMENT

Oracle Enterprise Manager 10*g* has been significantly enhanced to enable true single system image management of cluster database deployments.

Oracle Enterprise Manager 10*g* has been significantly enhanced to enable true single system image management of cluster database deployments. Enterprise Manager's Cluster Database Page provides a single view of system status across multiple nodes. It also enables direct drill down to individual instances when needed. See Figure 2.

From the Cluster Database Page you can:

- View overall system status, e.g., the number of nodes in the cluster database and their current status

- View alerts aggregated across all instances with drill down to the source of each alert and additional detail

- Set threshold for alert generation on a cluster database-wide basis

- Monitor performance metrics aggregated across all instances or displayed side by side so that instances can be readily compared, with additional drill down as needed

- Monitor cluster cache coherency statistics (e.g., global buffer gets, etc.)

- Perform cluster database-wide operations including the ability to initiate backup & recovery operations, start/stop instances, and so on.

- Manage services (see "Automatic Workload Management" section below) by performing operations such as start/stop, enable/disable and relocate services as well as monitoring of service performance.

- Oracle Enterprise Manager 10*g* also provides a Cluster Page for viewing the cluster hardware and operating system platform as a whole. This is particularly useful when the cluster is supporting multiple databases. Overall cluster platform status can be readily accessed with easy drill down capabilities to individual databases when needed.
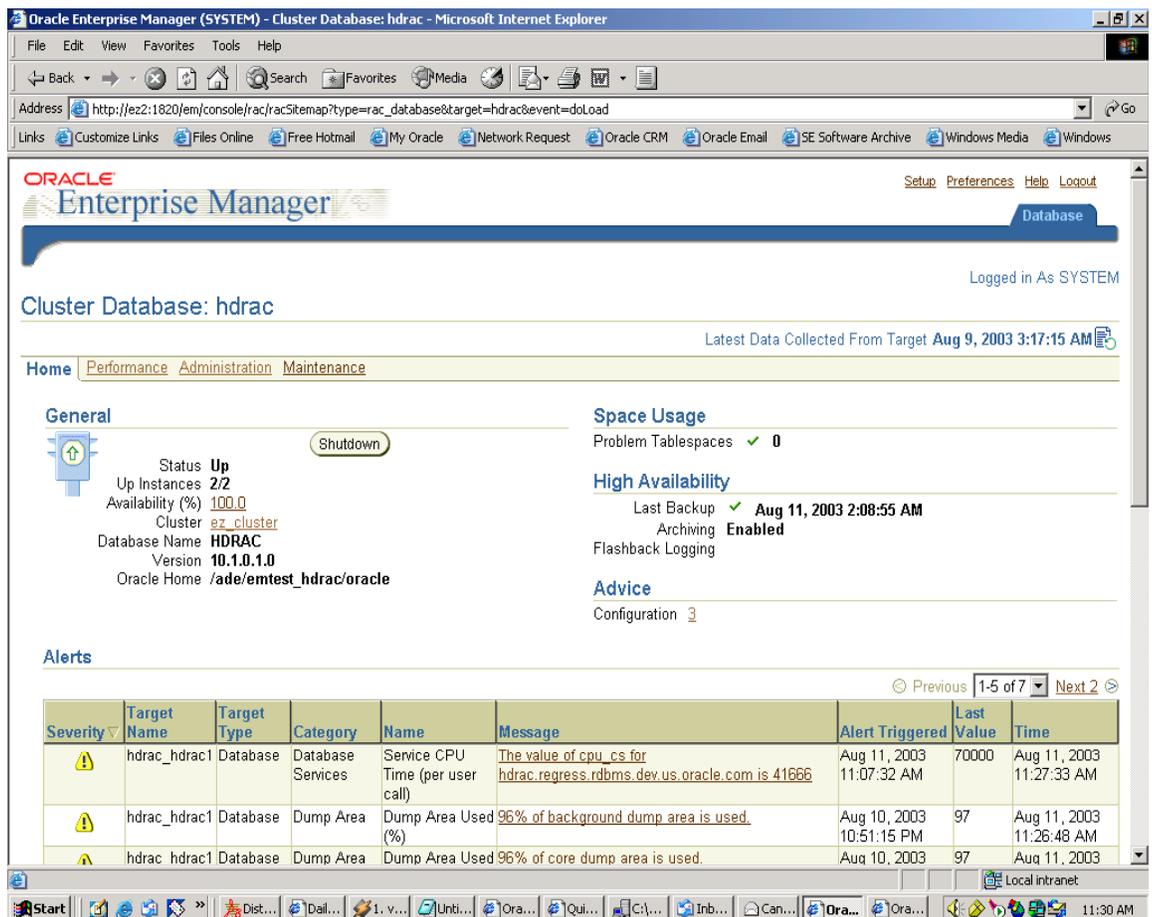

Figure 2. Single system image Cluster Database Page for Oracle RAC 10g

AUTOMATIC WORKLOAD MANAGEMENT

With Oracle Database 10*g*, application workloads can be defined as services so that they can be individually managed and controlled. DBAs control which processing resources are allocated to each service during both normal operations and in response to failures. Performance metrics are tracked by service and thresholds set to automatically generate alerts should these thresholds be crossed. CPU resource allocations and resource consumption controls are managed for services using Resource Manager.   Oracle tools and facilities such as Job Scheduler, Parallel Query, and Oracle Streams Advanced Queuing also use services to manage their workloads.

With Oracle Database 10*g*, rules can be defined to automatically allocate processing resources to services. Oracle RAC 10*g* instances can be allocated to process individual services or multiple services as needed. These allocation rules can be modified dynamically to meet changing business needs. For example, these rules could be modified at the end of quarter to ensure that there are enough processing resources to complete critical financial functions on time. Rules can also be defined so that when instances running critical services fail, the workload will be automatically shifted to instances running less critical workloads.

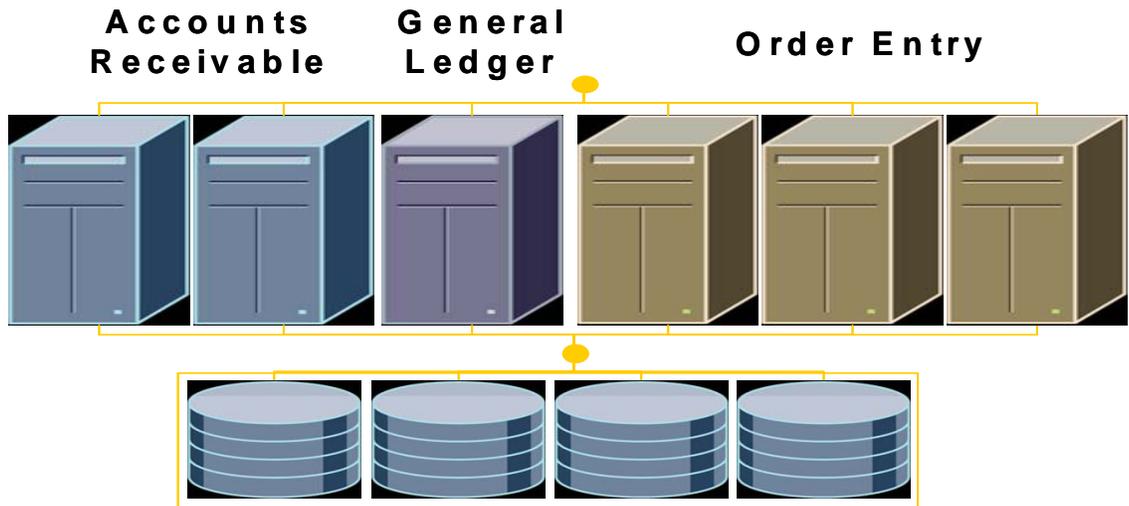### Accounts Receivable   General Ledger   Order Entry

Figure 3. Example of service provisioning across 6 node cluster

Services are identified by globally unique names. For example, an application suite can define a service for each application component, such as general ledger, accounts receivable, order entry, and so on. Middle tier applications and

clients select a service by specifying the service name in the TNS connect data when connecting to the database. No changes need be made to the applications themselves.

DBAs can use DBCA or SRVCTL to initially define services and assign them to a set of PREFERRED instances that the service will run on at startup and to set of AVAILABLE instances that will be automatically used should failures occur. These definitions are used by the Listener to assign connections to instances when new sessions are established. Using Oracle Enterprise Manager 10g, services can also be flexibly controlled. For example, DBAs can perform operations such as start/stop, enable/disable, and relocate on services.

Workload Monitoring

The Oracle Automatic Workload Repository 10*g* lets DBAs manage the performance of services for both RAC and single instance databases. Response times, CPU consumption, and other metrics are automatically collected by service. Applications can enable collection of finer grained performance data by instrumenting their code to set tags identifying operations within a service by their originating application MODULE and ACTION.

The Automatic Workload Repository maintains the metrics for service performance continuously. The views – V$SERVICE_METRICS and V$SERVICE_METRICS_HISTORY - contains measures for every service every 60 seconds for the past hour. Graphical displays within Oracle Enterprise Manager 10*g* make it easy to identify top services, top modules and top actions based on CPU consumption and other metrics.

Using Oracle Enterprise Manager 10*g*, thresholds can be set on call response times and other service level metrics so that alerts will be automatically generated when those thresholds are crossed. For example, DBAs can be alerted to allocate additional processing resources to a service should performance levels not meet required service levels. Application mid-tier components and products could also potentially make use of this data for load balancing.

Resource Management

Oracle Resource Manager 10*g* can automatically assign work requests that connect using a service to consumer groups for more fine tuned management of resource allocations. For example, a DBA could assign two-thirds of available CPU resources to a high priority service mapped to one consumer group and one-third of available CPU resources to a lower priority service mapped to a second consumer group. Other controls can be also placed on service workloads to limit the execution times of long running operations, the number of active sessions that are allowed to execute concurrently, the resources consumed by long running resource intensive sessions, and other resource

usages. Again, services can work with Resource Manager in both RAC and single instance environments.

Oracle Tools and Facilities

Other Oracle tools and facilities such as Job Scheduler and Parallel Query can make use of services by setting the service name as part of their workload definition. For the Job Scheduler, jobs are assigned to job classes that can run within services. For parallel query and parallel DML, the query coordinator connects to a service, and the parallel query slaves inherit the service for the duration of the execution. Work requests executing under a service inherit the performance thresholds of the service and are measured as part of the service.

FAST CONNECTION FAIL-OVER

Oracle RAC 10*g* enables fast, coordinated recovery between the database and application mid-tier components and products. Oracle RAC 10*g* will initially support Oracle JDBC Fast Connection Fail-Over with Oracle Application Server 10g and will support additional mid-tier components and products in the future.

Oracle RAC 10*g* detects when instances go down and when they come back up. A highly adaptable notification system sends UP and DOWN signals to the application mid-tier immediately so that appropriate self-correcting recovery procedures can be applied.   This is much more efficient than detecting failures of networking calls (e.g., through TCP/IP timeouts) and will reduce recovery time from many minutes to just a few seconds. It also enables a much more coherent and comprehensive response to both failure and recovery events than is possible with the use of Virtual IP (VIP) addresses alone. Use of VIP addresses, new with RAC 10g, enable failure messages to be returned more quickly.

With JDBC Fast Connection Fail-Over, the DOWN event initiates processing to clean up bad connections from the JDBC Implicit Connection Cache and prevents bad or invalid connections from being handed out to application work requests. If an application is in the middle of a transaction an appropriate SQL exception will be thrown and the transaction will be rolled back. The application or container can then retry the connection request and re-establish session state. The UP event initiates load balancing of connections in the cache. Connections are established and load balanced to all active RAC instances, without waiting for application connection requests or retries.

DATA GUARD INTEGRATION FOR DISASTER RECOVERY

With Oracle Enterprise Manager 10*g*, the management component of Oracle Data Guard, Data Guard Broker, is now completely integrated with RAC. Data Guard disaster recovery environments involving Oracle RAC databases can be as easily managed as those employing single instance databases.

Automatic workload management can be used to re-provision services to the standby system in case of disaster to ensure that the most critical services maintain required service levels while low priority services may be run at a diminished capacity or not at all. This can reduce costs by allowing the standby system to be configured with less total capacity than the primary.

CLUSTER VERIFICATION AND IMPROVED DIAGNOSTIC TOOLS

Oracle Database 10*g* introduces a new cluster configuration verification tool and improvements in the diagnostic tools first introduced in Oracle9i. Together these tools help users both avoid problems and resolve problems more quickly should they occur.

The cluster verification tool eliminates errors through pre and post validation of installation steps and/or configuration changes. It can also be used for ongoing cluster validation. The tool is invoked through a command line interface or through an API by other programs such as OUI, OEM, DBCA, and ASM. The cluster verification tool will be made available via OTN after Oracle Database 10*g* becomes generally available.

Diagnostic tools have been improved to further speed problem resolution times, reduce the need to reproduce problems, improve ease of use, and minimize database overhead in both RAC and non-cluster environments. In Oracle Database 10*g* much more of the Oracle Database executable has been instrumented to generate diagnostic trace data. Note that diagnostic tools should only be used under the direction of Oracle Support.

PERFORMANCE IMPROVEMENTS

Along with significant improvements in manageability, Oracle RAC 10*g* offers a number of performance improvements that will benefit many applications. These include improvements in instance failure recovery times, several optimizations that improve the performance key application operations, and dynamic re-mastering which automatically provides performance improvements for certain types of workloads.

Most applications will experience a speedup in instance failure recovery time due to a series of optimizations that reduce path length and allow key recovery functions to performed concurrently. Internal lab tests show reductions in recovery time of up to one half for applications configured for maximum

recovery speed.   The Fast Connection Fail-Over feature described above also enables fast, coordinated recovery between the database and application mid-tier components and products.

Other valuable optimizations provide performance benefits for certain types of application operations. These include reductions in transaction synchronization overhead that provide the most noticeable benefit to short transactions such as simple account balance updates. Message aggregation improves the efficiency of CacheFusion communication between nodes by combining multiple messages into one when appropriate. This provides the most noticeable benefit to operations that generate large numbers of messages between nodes such as check-pointing, online redo scans, and full table scans. Also, Oracle Streams Advanced Queuing now uses the fast CacheFusion communications layer to more efficiently deliver global events to subscribers on remote RAC instances.

Another important optimization is dynamic resource re-mastering. When applications consist of multiple workloads it is not unusual for one or more workload to exhibit a high degree of affinity to certain sets of data resources. If this occurs, RAC 10g will automatically detect it and re-master these resources to the instance displaying the affinity. As workloads change and are shifted between instances over time this mechanism adapts automatically to re-master resources optimally. This optimization improves performance without requiring any manual tuning or changes to applications or data layouts.

ZERO DOWNTIME PATCHING

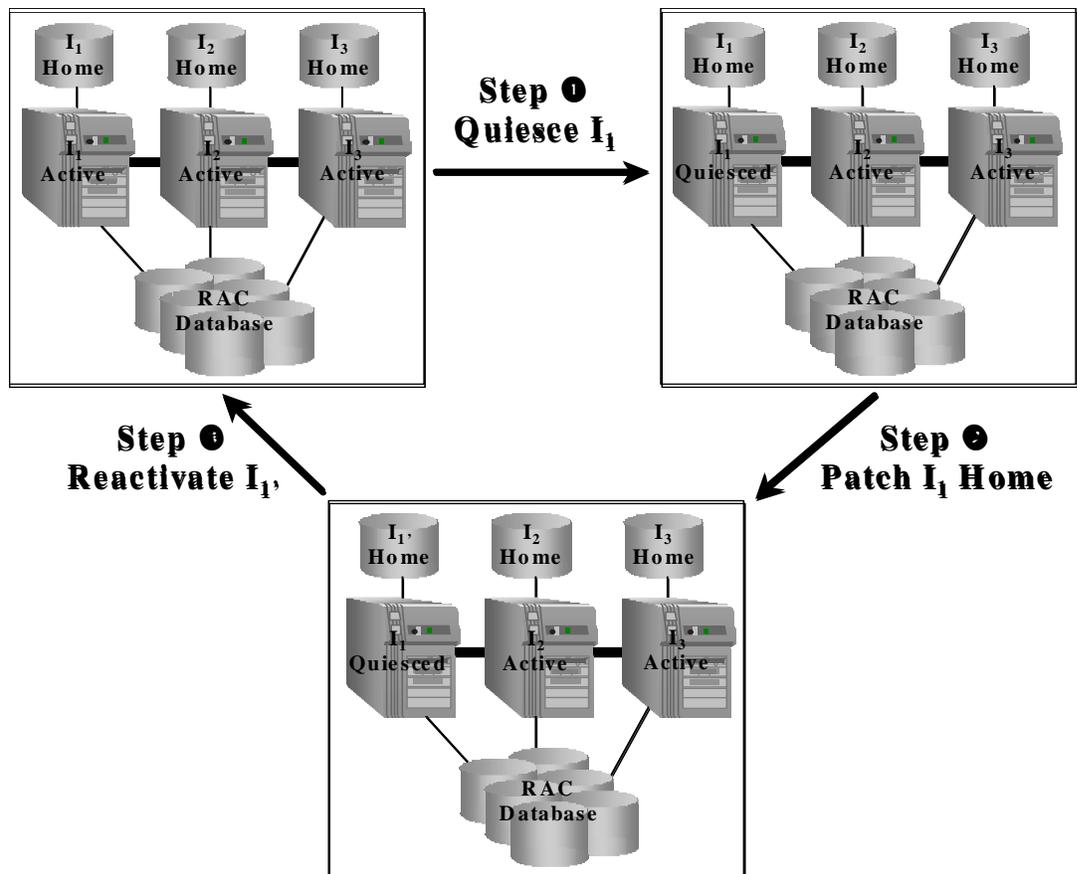Oracle now supports the application of patches to the nodes of a RAC system in a rolling fashion with no downtime. Patches are applied one node at a time while the other nodes in the RAC system are up and operational. See Figure 4. Patches will be labeled as being qualified for installation as rolling upgradeable, or not, depending on the changes being made by the patch. Oracle expects most patches will be rolling upgradeable but some that modify common structures shared between instances, or the contents of the database, will not be. In addition, only individual patches – not patch sets – will be rolling upgradeable. This capability is supported beginning with Oracle 9.2.0.2.

Step ❶
Quiesce I₁

I_1 Home I_2 Home I_3 Home

I_1 Quiesced I_2 Active I_3 Active

RAC Database

I_1 Home I_2 Home I_3 Home

I_1 Active I_2 Active I_3 Active

RAC Database

Step ❸
Reactivate I_1,

Step ❷
Patch I_1 Home

I_1, Home I_2 Home I_3 Home

I_1 Quiesced I_2 Active I_3 Active

RAC Database

Figure 4. Example of rolling patch upgrade with zero downtime in 3 node cluster

CONCLUSION

The Oracle Database 10*g*, and the specific new manageability enhancements provided by Oracle RAC 10*g*, enable Enterprise Grids and "RAC for everyone" - the ubiquitous adoption of Oracle RAC by all types of businesses for all types of applications. Enterprise Grids are built from large configurations of standardized, commodity-priced components: processors, network, and storage. With Oracle RAC's cache fusion technology, the Oracle Database adds to this the highest levels of availability and scalability. And now, with the Oracle Database 10*g* and Oracle RAC 10*g* dramatically reducing operational costs still further, new levels of flexibility become possible through dynamic provisioning of nodes, storage, CPUs, and memory to maintain service levels more easily and efficiently while lowering cost still further through improved utilization. Enterprise Grids are the data centers of the future and enable business to be adaptive, proactive, and agile.

ORACLE