Oracle Real Application Clusters 10g

An Oracle Technical White Paper
May 2005

ORACLE

Introduction 3
What is Oracle Database 10g Real Application Clusters 3
Real Application Clusters Architecture 4
Oracle Clusterware 5
Hardware Architecture 5
File Systems and Volume Management 6
Virtual Internet Protocol Address (VIP) 6
Cluster Verification Utility 6
RAC on Extended Distance Clusters 7
Benefits of Oracle Real Application Clusters 7
High Availability 7
Scalability 8
Managing Your Oracle Real Application Clusters Database 9
Enterprise Manager 10g 9
Rolling Patch Application 11
Rolling Release Upgrade Support 11
Workload Management with REAL application clusters 11
Services 12
Connection Load Balancing 12
Fast Application Notification (FAN) 12
Load Balancing Advisory 13
Conclusion 13

"We're pushing out over a million page views on our Web site, and each one is dynamic, hosted up by hits to our database. We required something that could manage this with ease and the highest possible availability, so we chose Oracle Database with Real Application Clusters." Shawn Kernes, Vice President of Technology StubHub

INTRODUCTION

Oracle Real Application Clusters (RAC) allows Oracle Database to run any packaged or custom application, unchanged across a set of clustered servers. This provides the highest levels of availability and the most flexible scalability. If a clustered server fails, Oracle continues running on the remaining servers. And when you need more processing power, simply add another server without taking users offline. To keep costs low, even the highest-end systems can be built out of standardized, commodity parts.

Oracle Real Application Clusters provides a foundation for Oracle's Enterprise Grid Computing Architecture. Oracle RAC technology enables a 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.

This paper provides a technical overview of Oracle Real Application Clusters 10g with the emphasis on the features and functionality that can be implemented to provide the highest availability and scalability for enterprise applications.

WHAT IS ORACLE DATABASE 10G REAL APPLICATION CLUSTERS?

Oracle RAC enables the Oracle Database to run mainstream business applications of all kinds on clusters including popular packaged products (such as Oracle Applications, Peoplesoft, SAP), in-house developed applications, which can be either OLTP, DSS, or a mixed workload.

Oracle Real Application Clusters is an option of Oracle Database that was first introduced with Oracle 9*i*. Oracle Real Application Clusters is now proven technology used by thousands of customers in every industry in every type of application. Oracle RAC provides options for scaling applications beyond the capabilities of a single server. This allows customers to take advantage of lower cost commodity hardware to reduce their total cost of ownership and provide a scaleable computing environment that supports their application workload.

Project Mega Grid[1] is an example showing how a real world application workload can run either on a single SMP server or a cluster of servers and meet the same performance requirements. In addition, the clustered environment includes high availablity.

[1] Project MegaGrid is a joint project with Oracle, EMC, Dell and Intel http://www.oracle.com/megagrid

Oracle Real Application Clusters is a key component of the Oracle High Availability Architecture[2], which provides direction to architect the highest availability for applications. Oracle RAC provides the ability to remove the server as a single point of failure in any database application environment.

Real Application Clusters Architecture

A RAC database is a clustered database. A cluster is a group of independent servers that cooperate as a single system. Clusters provide improved fault resilience and modular incremental system growth over single symmetric multi-processor (SMP) systems. In the event of a system failure, clustering ensures high availability to users. Access to mission critical data is not lost. Redundant hardware components, such as additional nodes, interconnects, and disks, allow the cluster to provide high availability. Such redundant hardware architectures avoid single points-of-failure and provide exceptional fault resilience.

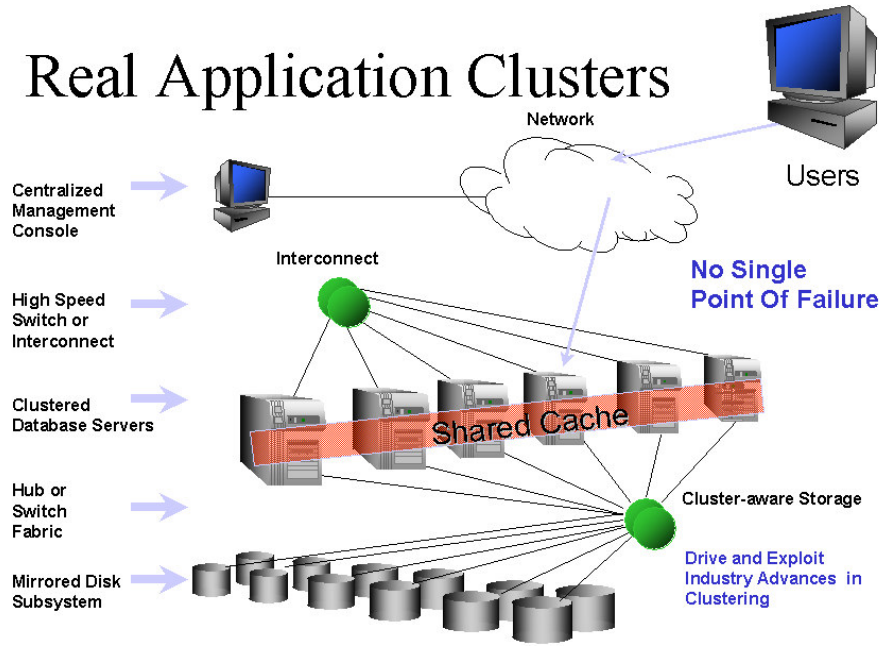

Figure 1 Real Application Clusters Architecture

"With Oracle we can add capacity without throwing out or replacing old computers."

John Kerin Executive Vice President, Chief Operating Officer, and Chief Technology Officer Chicago Stock Exchange

With Real Application Clusters, we de-couple the Oracle Instance (the processes and memory structures running on a server to allow access to the data) from the Oracle database (the physical structures residing on storage which actually hold the data, commonly known as datafiles). A clustered database is a single database that can be accessed by multiple instances. Each instance runs on a separate server in the cluster. When additional resources are required, additional nodes and instances can be easily added to the cluster with no downtime. Once the new instance is

[2] For more information on Oracle High Availability Architecture: http://download-west.oracle.com/docs/cd/B14117_01/server.101/b10726.pdf

started, applications using services can immediately take advantage of it with no changes to the application or application server.

Real Application Clusters is an extension of the Oracle Database and therefore benefits from the manageability, reliability and security features built into Oracle Database 10g.

Oracle Clusterware

Starting with Oracle Database 10g, Oracle provides Oracle Clusterware, a portable clusterware solution that is integrated and designed specifically for Oracle Database. You no longer have to purchase third party clusterware in order to have a RAC database. Oracle Clusterware is integrated with the Oracle Universal Installer, which the Oracle DBA is already familiar with. Support is made easier as there is one support organization to deal with for the clusterware and cluster database. You can choose to run Oracle RAC with selected third party clusterware, Oracle will work with certified third party clusterware however, Oracle Clusterware must manage all RAC databases.

Oracle Clusterware monitors and manages Real Application Cluster databases. When a node in the cluster is started, all instances, listeners and services are automatically started. If an instance fails, the clusterware will automatically restart the instance so the service is often restored before the administrator notices it was down.

With Oracle Database 10g Release 2, Oracle provides a High Availability API so that non-Oracle processes can be put under the control of the high availability framework within Oracle Clusterware. When registering the process with Oracle Clusterware, information is provided on how to start, stop, and monitor the process. You can also specify if the process should be relocated to another node in the cluster when the node it is executing on fails.

Hardware Architecture

Oracle Real Application Clusters is a shared everything architecture. All servers in the cluster must share all storage used for a RAC database. The type of disk storage used can be network attached storage (NAS), storage area network (SAN), or SCSI disk. Your storage choice is dictated by the server hardware choice and what your hardware vendor supports. The key to choosing your storage is choosing a storage system that will provide scaleable I/O for your application, an I/O system that will scale as additional servers are added to the cluster.

A cluster requires an additional network to the Local Area Network (LAN) that a database server is attached to for application connections. A cluster requires a second private network commonly known as the interconnect. Oracle recommends that you use 2 network interfaces for this network for high availability purposes. A network interface bonding external to Oracle should be used to provide failover and load balancing. The interconnect is used by the

cluster for inter-node messaging. The interconnect is also used by RAC to implement the cache fusion technology. Oracle recommends the use of UDP over GigE for the cluster interconnect. The use of crossover cables as the interconnect is not supported for a production RAC database.

The cluster is made up of 1 to many servers each having a LAN connection, an interconnect connection, and must be connected to the shared storage. With Oracle Database 10g Release 2, Oracle Clusterware and Real Application Clusters support up to 100 nodes in the cluster. Each server in the cluster does not have to be exactly the same but it must run the same operating system, and the same version of Oracle. All servers must support the same architecture E.G. all 32bit or all 64bit.

Current detailed information on certifications and technology restrictions related to Oracle Real Application Clusters can be obtained through Oracle Metalink (http://metalink.oracle.com).

File Systems and Volume Management

Since RAC is a shared everything architecture, the volume management and file system used must be cluster-aware. Oracle recommends the use of Automatic Storage Management (ASM), which is a feature, included with Oracle Database 10g to automate the management of storage for the database. ASM provides the performance of async I/O with the easy management of a file system. ASM distributes I/O load across all available resource to optimize performance while removing the need for manual I/O tuning.

Alternatively Oracle supports the use of raw devices and some cluster file systems such as Oracle Cluster File System (OCFS) which is available on Windows, Linux and Solaris (OCFS for Solaris will be released following Oracle Database 10g Release 2).

Virtual Internet Protocol Address (VIP)

Oracle Real Application Clusters 10g requires a virtual IP address for each server in the cluster. The virtual IP address is an unused IP address on the same subnet as the Local Area Network (LAN). This address is used by applications to connect to the RAC database. If a node fails, the Virtual IP is failed over to another node in the cluster to provide an immediate node down response to connection requests. This increases the availability for applications as they no longer have to wait for network timeouts before the connection request fails over to another instance in the cluster.

Cluster Verification Utility

Oracle Database 10g Release 2 introduces a new cluster configuration verification tool. 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 Oracle Universal Installer (OUI).

RAC on Extended Distance Clusters

RAC on Extended Distance Clusters is an architecture where nodes in the cluster reside in locations that are physically separate. RAC on Extended Distance Clusters provides extremely fast recovery from a site failure and allows for all nodes, at all sites, to actively process transactions as part of single database cluster. While this architecture creates great interest and has been successfully implemented, it is critical to understand where this architecture best fits especially in regards to distance, latency, and degree of protection it provides.

The high impact of latency, and therefore distance, creates some practical limitations as to where this architecture can be deployed. This architecture fits best where the 2 datacenters are located relatively close (<~100km) and where the extremely expensive costs of setting up direct cables with dedicated channels between the sites has already been taken.

RAC on Extended Distance Clusters provides greater high availability than local RAC but it may not fit the full Disaster Recovery requirements of your organization. Feasible separation is great protection for some disasters (local power outage, airplane crash, server room flooding) but not all. Disasters such as earthquakes, hurricanes, and regional floods may affect a greater area. Customers should do an analysis to determine if both sites are likely to be affected by the same disaster. For comprehensive protection against disasters including protection against corruptions and regional disasters, Oracle recommends the use of Data Guard with RAC as described in the Oracle High Availability Architecture documentation. Data Guard also provides additional benefits such as support for rolling upgrades across Oracle versions.

Configuring an extended distance cluster is more complex than a local cluster. Specific focus needs to go into node layout, voting disks, and data disk placement. Implemented properly, this architecture can provide greater HA than a local RAC database. The combination of Oracle Clusterware, Oracle Real Application Clusters and Automatic Storage Management can be used to create extended distance clusters.

BENEFITS OF ORACLE REAL APPLICATION CLUSTERS

High Availability

Oracle Real Application Clusters 10g provides the infrastructure for datacentre high availability. It is also an integral component of Oracle's High Availability Architecture, which provides best practices to provide the highest availability data management solution. Oracle Real Application Clusters provides protection against the main characteristics of high availability solutions.

Reliability – Oracle Database is known for its reliability. Real Application Clusters takes this a step further by removing the database server as a single point of failure. If an instance fails, the remaining instances in the cluster are open and active.

Recoverability – Oracle Database includes many features that make it easy to recover from all types of failures. If an instance fails in a RAC database, it is recognized by another instance in the cluster and recovery automatically takes place. Fast Application Notification, Fast Connection Failover and Transparent Application Failover make it easy for applications to mask component failures from the user.

"When Oracle announced 10*g*, it really captivated us. We were very excited to start leveraging the high availability capacity and the flexibility that 10*g* provides."
-- Laurence Grant, IT Director of Enterprise Computing Systems, Talk America

Error Detection – Oracle Clusterware automatically monitors RAC databases and provides fast detection of problems in the environment. Also it automatically recovers from failures often before anyone has noticed a failure has occurred. Fast Application Notification provides the ability for applications to receive immediate notification of cluster component failures and mask the failure from the user by resubmitting the transaction to a surviving node in the cluster.

Continuous Operations – Real Application Clusters provides continuous service for both planned and unplanned outages. If a node (or instance) fails, the database remains open and the application is able to access data. Most database maintenance operations can be completed without down time and are transparent to the user. Many other maintenance tasks can be done in a rolling fashion so application downtime is minimized or removed. Fast Application Notification and Fast Connection Failover assist applications in meeting service levels and masking component failures in the cluster.

Scalability

"Once we saw that we could get three Linux servers for the price of one UNIX server, we had to evaluate this new architecture to see if it could perform and scale," said Darryl Boone, Vanderbilt's assistant director for architecture and operations. "Our tests showed that we would get three times the server power and performance for the dollar, plus greater availability. Next we factored in Oracle and HP's commitment to Linux. The benefit to the university over the long term was too great to walk away from."

Oracle Real Application Clusters provides unique technology for scaling applications. Traditionally, when the database server ran out of capacity, it was replaced with a new larger server. As servers grow in capacity, they are more expensive. For databases using RAC, there are alternatives for increasing the capacity. Applications that have traditionally run on large SMP servers can be migrated to run on clusters of small servers. Alternatively, you can maintain the investment in the current hardware and add a new server to the cluster (or to create a cluster) to increase the capacity. Adding servers to a cluster with Oracle Clusterware and RAC does not require an outage and as soon as the new instance is started, the application can take advantage of the extra capacity. All servers in the cluster must run the same operating system and same version of Oracle but they do not have to be exactly the same capacity. Customers today run clusters that fit their needs whether they are clusters of servers where each server is a 2 cpu commodity server to clusters where the servers have 32 or 64 cpus in each server.

Oracle Real Application Clusters architecture automatically accommodates rapidly changing business requirements and the resulting workload changes. Application

users, or mid tier application server clients, connect to the database by way of a service name. Oracle automatically balances the user load among the multiple nodes in the cluster. The Real Application Clusters database instances on the different nodes subscribe to all or some subset of database services. This provides DBAs the flexibility of choosing whether specific application clients that connect to a particular database service can connect to some or all of the database nodes. Administrators can painlessly add processing capacity as application requirements grow. The Cache Fusion architecture of RAC immediately utilizes the CPU and memory resources of the new node. DBAs do not need to manually re-partition data.

Another way of distributing workload in an Oracle database is through the Oracle Database's parallel execution feature. Parallel execution (I.E. parallel query or parallel DML) divides the work of executing a SQL statement across multiple processes. In an Oracle Real Application Clusters environment, these processes can be balanced across multiple instances. Oracle's cost-based optimizer incorporates parallel execution considerations as a fundamental component in arriving at optimal execution plans. In a Real Application Clusters environment, intelligent decisions are made with regard to intra-node and inter-node parallelism. For example, if a particular query requires six query processes to complete the work and six CPUs are idle on the local node (the node that the user connected to), then the query is processed using only local resources. This demonstrates efficient intra-node parallelism and eliminates the query coordination overhead across multiple nodes. However, if there are only two CPUs available on the local node, then those two CPUs and four CPUs of another node are used to process the query. In this manner, both inter-node and intra-node parallelism are used to provide speed up for query operations.

MANAGING YOUR ORACLE REAL APPLICATION CLUSTERS DATABASE

Oracle Real Application Clusters provides a single system image for easy configuration and management. The RAC database can be installed, configured, and managed from a single location. All tools and utilities provided to manage the database are cluster-aware from the Oracle Universal Installer (OUI), to Enterprise Manager including the database configuration assistant (DBCA), the database upgrade assistant (DBUA), the network configuration assistant (NETCA), and the command line interfaces such as srvctl.

Enterprise Manager 10g

"Oracle Grid Control is becoming very important – eliminating tedious work and making our DBAs more productive" David Milne, Director Database Technologies, Chicago Stock Exchange

Enterprise Manager 10g Database Control is the GUI management tool provided by Oracle to manage your Oracle Database. Database Control is automatically configured by the DBCA when a database is created. Enterprise Manager 10g Grid Control is the GUI Management tool provided by Oracle to manage your enterprise. Grid Control is installed from a separate CD included in the Oracle

Database CD pack. Both these tools are cluster-aware and provide a centralized console to manage your cluster database.

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 by performing operations such as create, modify, start/stop, enable/disable and relocate services as well as monitoring of service performance.

Oracle Enterprise Manager 10*g* Grid Control provides a Cluster Page for viewing the cluster hardware and operating system 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.

Oracle Enterprise Manager 10g Release 2 Grid Control provides a utility that automates the conversion of a single instance Oracle Database to a RAC Database.

Oracle Enterprise Manager 10g Release 2 Grid Control provides additional capabilities to make the provisioning of Real Application Clusters databases easier. The initial creation of a cluster including lying down of Oracle home and the configuring of the clusterware can be easily done through Enterprise Manager. The Oracle Home software can be kept in Enterprise Manager as the known "Gold Image" or sourced from a known reference host. The "Gold Image" is created from a copy of a known good implementation of Oracle Clusterware 10g Release 2 or Oracle Real Application Clusters 10g Release 2 environment. In Grid Control 10g Release 2, the cloning application will support complete end-to-end creation of new RAC and Oracle Clusterware software including execution of superuser actions (root.sh) and customizable pre and post steps. This can also be used when adding a new node to an existing cluster.

For Linux operating systems, Oracle can also provision an "image" to a bare metal node. The image could consist of the Operating System, the Oracle Enterprise Manager agent, Oracle Clusterware, and Oracle Database with Real Application

Clusters. This image can be associated with a hardware profile. All the components for this image are stored as "Gold Images" in Enterprise Manager. A wizard allows for choosing of hardware and provisioning of the whole stack onto new hardware. The new node is automatically added to the cluster.

Rolling Patch Application

Oracle supports the application of patches to the nodes of a RAC database 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. This requires that each node has a separate Oracle Home. Patches will be labeled as being qualified for installation as rolling upgradeable, or not, depending on the changes being made by the patch. Some patches 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. All Oracle Clusterware patches can be applied in a rolling fashion.

Rolling Release Upgrade Support

Oracle Clusterware supports rolling upgrades from Release 1 (known as Cluster Ready Services) to Release 2. This provides the ability to upgrade the clusterware without taking the cluster out of service and therefore enables 24x7 operation of business.

Oracle RAC 10*g* Release 2 supports database software upgrades (from Oracle Database 10*g* Release 1 Patchset 1 onwards) in a rolling fashion – with near zero database downtime, by using Data Guard SQL Apply. The steps involve upgrading the logical standby database to the next release, running in a mixed mode to test and validate the upgrade, doing a role reversal by switching over to the upgraded database, and then finally upgrading the old primary database. While running in a mixed mode for testing purpose, the upgrade can be aborted and the software downgraded, without data loss. For additional data protection during these steps, a second standby database may be used.
By supporting rolling upgrades with minimal downtimes, Data Guard reduces the large maintenance windows typical of many administrative tasks, and enables the 24x7 operation of the business.

WORKLOAD MANAGEMENT WITH REAL APPLICATION CLUSTERS

Applications using a RAC database need to manage the workload across the cluster. Oracle Real Application Clusters 10g includes innovative technology to manage workloads providing the best application throughput given the configuration and high availability for the application.

Services

Workload Management relies on the use of Services, a feature of Oracle Database 10g. Services hide the complexity of a RAC database by providing a single system image to manage workload. Services allow applications to benefit from the reliability of a cluster. Traditionally a database provided a single service and this name was the connect data given to SQL*NET. With Oracle Database 10g, a DBA can define up to 100 database services to be provided by a single database. This allows you to breakup workloads from applications into manageable components based on business requirements such as service levels and priorities. Services are integrated with many features of Oracle Database 10g. Application users can be automatically assigned to a Resource Manager consumer group, which limits their resources such as cpu. Batch Jobs can be assigned to specific job classes based on their service. The use of services achieves location transparency for queues when using Oracle Streams Advanced Queuing.

A service can span one or more instances of an Oracle database and an instance can support multiple services. The number of instances offering a service is managed dynamically by the DBA independently of the application. When outages occur, services are automatically restored to surviving instances. When instances are restored, any services that are not running are restored automatically.

Connection Load Balancing

Oracle Net Services provides connection load balancing for database connections. Client side load balancing which balances connection requests across all listeners for the cluster, is achieved by listing all servers in the cluster in the address list of the client connect string. SQL*NET will randomly select one of the servers. If the server chosen is not available, the next server in the list is tried. Server side load balancing is achieved at the listener. Each listener is aware of all instances in the cluster providing each service. Based on goal defined for the service, the listener chooses the instance that will best meet the goal and the connection is made to that instance.

Fast Application Notification (FAN)

Fast Application Notification provides integration between the RAC database and the application. It allows the application to be aware of the current configuration of the cluster at any given time so that application connections are only made to instances that are currently able to respond to the application requests. The Oracle RAC 10g HA framework posts a FAN event immediately when a state change occurs within the cluster.

Integrated clients receive these events and immediately react. For down events, application interruption is minimized by cleaning up connections to the failed instance, in-flight transactions are interrupted with an error returned to the application. Applications making connections are directed to active instances only. Server side callouts can be used to log trouble tickets or page administrators

alerting them of the failure. For UP events, new connections are created to allow the application to immediately take advantage of the extra resources available. Oracle JDBC, ODP.NET and OCI clients are integrated with FAN. Other applications can take advantage of FAN by using the application-programming interface to directly subscribe to FAN events.

Load Balancing Advisory

Database workloads change over time as well as the cluster configuration can change, it is important to create and allocate database connections based on the most up to date information. Oracle Real Application Clusters 10g Release 2, provides a load balancing advisory. RAC constantly monitors the workload being executed for each service by each instance providing the service. This information is published to the Automatic Workload Repository and published to the application using FAN events. The FAN event includes the current service level provided and a recommendation of what percentage of connections to be directed to each instance.

The integrated Oracle Clients use these events to provide intelligent load balancing of application requests. Most connection pools use a random or round robin algorithm to select an idle connection from the pool when the application does a get connection. Using FAN events from the load balancing advisory, the connection pool will select the connection currently providing the best service. Oracle JDBC and ODP.NET provide runtime connection load balancing through integration with the load balancing advisory.

CONCLUSION

"If we have hardware or software problems with the cluster now, our users will not be aware of them." Gunnar Mikkelsen DBA and Project Leader University of Oslo

Oracle Real Application Clusters has been designed for high availability and scalability. By providing protection from hardware and software failures, Oracle Real Application Clusters provides systems availability ensuring continuous data access. Its scale out and scale up features offer a platform, which can grow in any direction allowing enterprises to grow their businesses. Existing applications as well as newly developed applications benefit from the transparency Oracle Real Application Clusters provides. Application development as well as administration and change management thus become much easier allowing reduction in total cost of ownership. Oracle Real Application Clusters is unique to the market with its offering and capabilities. RAC is used by thousands of customers worldwide in all industries in mission critical and many other application environments.

Oracle Real Application Clusters 10g
May] 2005
Author: Barb Lundhild
Contributing Authors: Peter Sechser

Oracle Corporation
World Headquarters
500 Oracle Parkway
Redwood Shores, CA 94065
U.S.A.

Worldwide Inquiries:
Phone: +1.650.506.7000
Fax: +1.650.506.7200
oracle.com