Oracle Clusterware 10g

An Oracle Technical White Paper
April 2007

ORACLE

INTRODUCTION

Oracle Clusterware is a portable cluster software that allows clustering of single servers so that they cooperate as a single system. Oracle Clusterware also provides the required infrastructure for Oracle Real Application Clusters (RAC). In addition Oracle Clusterware enables the protection of any Oracle application or any other kind of application within a cluster. Oracle Clusterware is supported on all operating systems certified for Oracle RAC.

Within a Real Application Clusters environment Oracle Clusterware is responsible for the high availability management of the various components that such an environment consists of. This includes the monitoring and the process of restart actions for the nodes of the cluster itself, for the Oracle Database Instances, for the Oracle Listeners, and for the Oracle Database Services.

In combination with Oracle Real Application Clusters, Oracle Clusterware thereby contributes to the highest level of availability and most flexible scalability that Oracle Real Application Clusters provides.

Providing High Availability is an essential component in the world of business continuity

This paper explains how to use Oracle Clusterware to enable high availability for any kind of application.

In environments not including Oracle Real Application Clusters Oracle Clusterware still provides the level of high availability that is commonly known to be achievable for nearly all kind of applications and databases managed in cluster environments. Those environments can include Oracle Single Instance Databases, Oracle Application Servers, Oracle Enterprise Manager components, third party databases, or any other kind of application.

In accordance to Oracle's ambition to develop high quality software that adheres to common standards, all components managed by the Oracle Clusterware in those environments would be subject to the same conditions of how they are integrated into the cluster. At present only certain SAP applications make an exception here. For those applications Oracle provides a pre-configured clusterware agent.[1]

Like Oracle Real Application Clusters and other Oracle technology being used in grid environments, Oracle Clusterware is designed to provide high availability on commodity hardware. Oracle Clusterware therefore acts as a consolidated cluster infrastructure within Oracle's Enterprise grid computing architecture.

[1] Also see page 11 of this document for more information

WHAT IS ORACLE CLUSTERWARE 10G?

Oracle Clusterware is the technology that transforms a server farm into a cluster. A cluster in general is a group of independent servers that cooperate as a single system. Oracle Clusterware is the intelligence in this system that ensures the required cooperation.

Oracle Clusterware was introduced with Oracle Database 10g Release 1 as the general underlying clusterware that is required to run Oracle Real Application Clusters 10g. At this time Oracle Clusterware was known as Oracle CRS.

While CRS (Cluster Ready Services) is still an important part of the Oracle Clusterware, the term CRS does not cover the whole functionality of the Oracle clusterware and thus the product was renamed to Oracle Clusterware when the current Oracle Database version, Oracle Database 10g, was officially released.

Technically the usage of the term CRS for the whole product was in line with its limited application support in Oracle Database 10g Release 1. Although partially using proven DEC cluster technology, which was licensed by Oracle for the purpose of developing a general Oracle Clusterware, there was no support for non-Oracle processes with Oracle Database 10g Release 1 cluster software.

"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

Those restrictions were lifted when Oracle Database 10g Release 2 was released. Besides that, certain parts of the Oracle Clusterware were improved in a way that Oracle Clusterware would now support extended cluster environments and more complex cluster configurations that are commonly used.

At present Oracle Clusterware is mainly used to provide the basis to run Oracle Real Application Clusters. But there is a growing number of customer environments in which Oracle Clusterware is also being used to fulfill high availability requirements for other Oracle products or other applications.

ORACLE CLUSTERWARE ARCHITECTURE

In the event of a system failure, clustering ensures high availability to users. A redundant hardware component, 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.

Real Application Clusters Environments

In Real Application Clusters environments, 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." [2]

[2] Lundhild, Barb: "Oracle Real Application Clusters 10g"

In this sense Oracle Clusterware is the basis for Oracle Real Application Clusters. Therefore, there needs to be one incarnation of the Oracle Clusterware on every node of the cluster that an Oracle RAC Database Instance is supposed to run on. Nevertheless, the number of those nodes may be a subset of the cluster nodes that are managed by Oracle Clusterware.

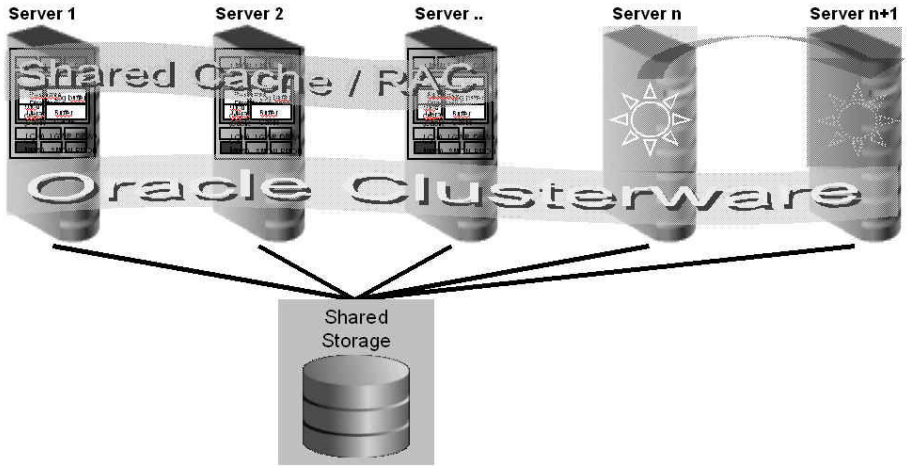

Figure 1 Oracle RAC / Oracle Clusterware relation

The ability to form a cluster in which nodes can be assigned to run a certain application or workload is one of the main key technologies when it comes to Workload Management.

At present this capability of the Oracle Clusterware is used by Oracle Real Application Clusters only. However, the open APIs available with Oracle Clusterware enables every application to make use of this feature directly.

The strong integration of Oracle Clusterware with Oracle Real Application Clusters has some impact on the components that come with Oracle Clusterware. But this does not limit its usage for any other kind of application. Instead, this integration enables better supportability, since there is usually only one support organization involved for the clusterware and the cluster database.

Before Oracle Database 10g third party clusterware was required on most operating systems in order to use RAC. Beginning with Oracle Database 10g Oracle Clusterware is required whenever RAC is going to be used. Oracle still supports the usage of third party clusterware in RAC environments.[3]

[3] Also see page 13 of this document for more information

In this case Oracle Clusterware ensures the interoperability between both clusterware versions used on the same system. Nevertheless, any interaction between the Real Application Clusters database and the cluster has to go through the Oracle Clusterware.

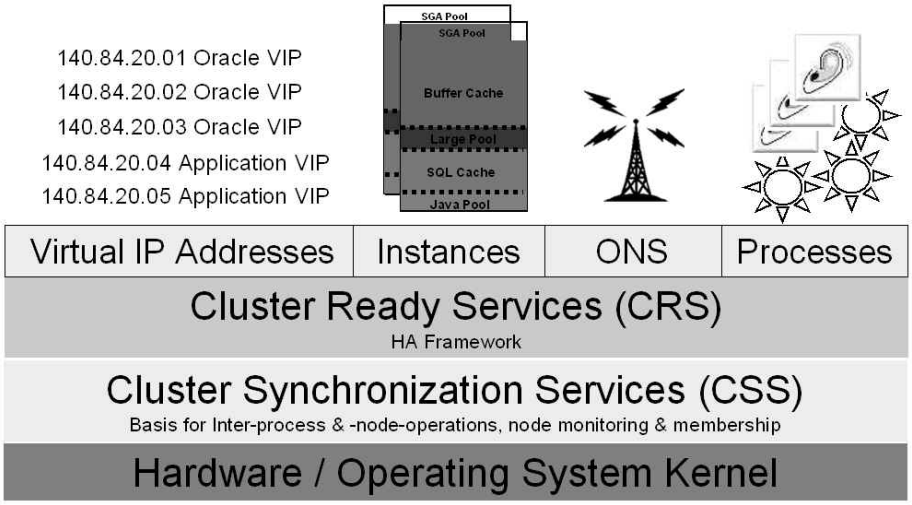

Figure 2: Oracle Clusterware layers

Oracle Clusterware mainly consists of three layers: Cluster Synchronization Services (CSS), Cluster Ready Services (CRS), and some default resources that are automatically available and being managed by Oracle Clusterware right after the installation and prior to any other configuration.

Oracle's Cluster Synchronization Services (CSS) are the basis for inter-process and inter-node-operations. In addition, CSS monitors all cluster nodes and handles the membership of each node within a certain cluster. Oracle Cluster Ready Services (CRS) is a synonym for the High Availability framework provided by Oracle Clusterware. All resources in the Oracle Clusterware are managed within this layer. This also applies to default resources.

Unlike in figure 2, there are basically only three resources automatically available and managed by Oracle Clusterware right after its installation. Those are usually referred to as Node Applications (nodeapps). Per default, there should be one resource of each kind on every node of the cluster:

- Virtual IP Address (VIP)
- Oracle Notification Service (ONS)
- Global Service Daemon (GSD)

The use of ONS and GSD will not be explained in this paper. For more information of those processes, refer to the official Oracle documentation.

In Oracle Real Application Clusters environments any RAC database or RAC instance is automatically registered in the Oracle Cluster Registry (OCR), which is the Oracle Clusterware repository. Any resource that is going to be managed by the Oracle Clusterware needs to be registered as a CRS resource, and then CRS stores the resource definitions in the OCR.

Whenever a RAC database is installed in a cluster managed by Oracle Clusterware, Oracle's configuration tools will perform those registration steps automatically. Apart from the Database Configuration Assistant (DBCA) this especially applies to the Network Configuration Assistant (NetCa), which therefore usually creates a cluster aware listener configuration. However, this tool can still be used to create local, non-clustered listener configurations in the same environment.

Clusters Without Real Application Clusters Databases

In environments not including a RAC database the principal architecture of the Oracle Clusterware remains the same. Therefore, no configuration change is required to use the Oracle Clusterware as a general purpose Clusterware.

However, it is not supported to modify the behavior of default resources to manage non-Oracle applications or processes. This applies to the Oracle VIP, the ONS, the GSD, and the Oracle default listener that might have been created in course of an Oracle Real Application Clusters installation.

Therefore and for more flexibility in the cluster, Oracle Clusterware makes use of so called Application VIPs. Technically those VIPs are derived from the Oracle VIP and provide similar high availability features. There is generally no limit in the number of Application VIPs being used on one node. Oracle VIPs are restricted to one per node, unless a failure in the cluster requires a temporary failover of one those VIPs to another node. In this case only the originally assigned Oracle VIP would be used on this node.

Installation and Configuration of Oracle Clusterware

The installation and configuration of the Oracle Clusterware is fully integrated into the Oracle Universal Installer, which the Oracle DBA should already be familiar with. Oracle Clusterware should be installed as the first Oracle software product in a cluster environment. Oracle Clusterware must reside in its own Oracle Home (but not placed under Oracle Base) and can be patched independently from any other Oracle product in the cluster.[4]

In Real Application Clusters environments no further direct configuration of Oracle Clusterware is usually required. The Oracle configuration assistants that are typically used to create an Oracle Real Application Clusters database based on the Oracle Clusterware will perform all necessary steps accordingly.

---

[4] Certain combinations of Oracle Clusterware versions and Oracle RAC or ASM versions may not be supported. Consult Oracle Metalink and page 12 of this document for more information.

In case the Oracle Clusterware is used to manage applications apart from an Oracle Real Application Clusters database, the required configuration steps can be individually performed by the user with the help of a self-explanatory set of command line interfaces. Those steps usually include the creation of a cluster group resource, mostly depending on an Application VIP, start, stop, check specifications, and an inter-resource dependency management.[5]

ORACLE CLUSTERWARE HARDWARE LAYOUT

As the basis for Oracle Real Application Clusters, Oracle Clusterware also requires a shared disk architecture. For the Oracle Clusterware it is therefore reasonable to assume that every node in the cluster is able to access a shared storage system. The type of disk storage used can be Network Attached Storage (NAS), Storage Area Network (SAN), or SCSI disks.

Generally, database servers require at least one network connection to the Local Area Network (LAN) for application connections. A cluster in addition requires a second, private network connection commonly known as the interconnect.

Oracle recommends using two network interfaces for the private 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 typically also used by RAC to implement the cache fusion technology. The usage of crossover cables as the interconnect is not supported.

Concluding, a typical Oracle cluster is made up of one to many servers each having a LAN connection, a private interconnect connection, and must be connected to a shared storage. 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 the Oracle Clusterware. In addition all servers must support the same hardware architecture. For example, all servers must either be 32bit or all servers must be 64bit.

At present there is no specific hardware or software certification for Oracle Clusterware only. Any certification or technology restriction related to Oracle Real Application Clusters can also be seen to be compulsory for the Oracle Clusterware. 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 System and Volume Management

Typically, every cluster software needs to store information about the cluster that it is managing. In addition every cluster software usually needs some kind of disk based area which is commonly used to adequately react to a communication problem on the interconnect resulting in a loss of coordination amongst the nodes.

[5] Also see page 10 of this document for more information

Cluster nodes working independently of each other in a cluster may lead to shared data being corrupted due to unsynchronized data access / writes. In order to prevent this behavior (Split Brain) a storage based resolution medium, e.g. Quorum Disk or Voting Disk, is typically used by most cluster software.

In Oracle Clusterware the main information about the cluster configuration is stored in the Oracle Cluster Registry (OCR). Since every node managed in an Oracle Clusterware cluster can be used to modify a cluster configuration, the OCR must be stored in a shared accessible storage.

The storage based resolution medium in order to prevent a Split Brain syndrome in the cluster used by the Oracle Clusterware is Voting Disk based. Since this disk is used as an alternative way to communicate in the cluster, it must also be accessible from every node in the cluster and must therefore reside on a shared disk.

Both devices are vital to run Oracle Clusterware. Oracle recommends using redundant copies of the OCR and the Voting Disk. Oracle Clusterware manages redundant copies of both files. They are only limited in the number of copies being supported. At present Oracle Clusterware can be configured to maintain one additional copy of the OCR automatically. However, 31 Voting Disks can be maintained. The minimum recommendation is to use 3 Voting Disks in order to tolerate a failure of one of those Voting Disks.

Unlike database files used by Oracle RAC, those files cannot be placed on disks managed by Automatic Storage Management (ASM), which is a feature included with Oracle Database 10g. At present Oracle only supports the use of Block Devices (Linux), RAW Devices, certified Cluster File systems, and certified NAS[6] systems accessed by NFS to host those Oracle Clusterware files. If volume managers are used to directly or indirectly manage any of those devices, those volume managers must be cluster aware.

Virtual Internet Protocol (VIP) Address

Initially a virtual IP is configured from an unused IP address on the same subnet as the Local Area Network (LAN). This address is used by applications to connect to e.g. a RAC database. If a node fails, the virtual IP is failed over to another node in the cluster to provide an immediate response to incoming connection requests.

Oracle Clusterware provides two types of virtual IP addresses: Application VIPs and Oracle VIPs. The latter ones are set up automatically in course of the Clusterware installation process. There is one Oracle VIP per node of the cluster. That is because Oracle Real Application Clusters 10g requires a virtual IP address for each server in the cluster.

There is no automatic setup of Application VIPs. Application VIPs are typically used whenever an application is going to be managed by Oracle Clusterware, so

[6] NAS support for Oracle Clusterware and Oracle RAC may be limited to certain operating systems. More information can be found in Oracle Metalink / Certify.

that those applications do not need to depend on the Oracle VIP. In addition, Application VIPs are useful as a basis for a cluster group resource within a cluster. Cluster group resources usually consist of processes that have a logical connection between each other.

For example, an application that accepts client requests should be available through a certain IP address, while the file system that this application is running on should be available on another node, once a failover in the cluster occurred. A cluster group for this scenario would consist of an Application VIP, the application processes, and the file system, unless a cluster file system is used.

The failover of the cluster group resource and especially the VIP increases the availability for applications, because an already connected client does not have to rely on network timeouts anymore to be notified about the failure. In addition new connection requests either for a RAC database or for a generic application are automatically routed to a VIP on one of the remaining nodes.[7]

This holds true for Application VIPs as well as for Oracle VIPs. The only difference would be that only the originally assigned Oracle VIP address is active on a node of the cluster regardless of how many Oracle VIPs are temporarily failed over to this node. There is no limitation on the number of active Application VIPs on a specific node.

Cluster Verification Utility

In Oracle Database 10g the use of the Cluster Verification Utility is highly recommended. Different Oracle tools will also use it internally. The cluster verification tool eliminates errors through pre and post validation of installation steps and/or configuration changes.

ORACLE CLUSTERWARE CONFIGURATIONS

Oracle Clusterware was originally designed to be the underlying clusterware for Oracle Real Application Cluster. Despite those intentions its usage as a general-purpose clusterware has never been neglected. Moreover, Oracle's ambition to develop grid compatible software products made Oracle Clusterware 10g one of the most flexible cluster infrastructures currently available on the market.

Generic Application Scenario

As with any other clusterware Oracle Clusterware is completely capable to be used independently of Oracle Real Application Clusters. In this case Oracle Clusterware still provides the level of high availability that is commonly known to be achievable for a database or any other kind of application managed in cluster environments.

---

[7] The automatic routing of connection requests for a generic application to a VIP on one of the remaining nodes may require further functionality not provided by Oracle.

Those applications can include Oracle Single Instance Databases, Oracle Application Servers, Oracle Enterprise Management components, third party databases, or any other kind of application. In accordance to Oracle's ambition to develop high quality software that adheres to common standards, all components managed by the Oracle Clusterware in those environments would be subject to the same conditions.

That means regardless of whether an Oracle product, an individual application or a third party application is going to be managed by Oracle Clusterware, there is a minimum set of steps required to achieve this kind of cluster management:

1. Create Application VIP
2. Write an action script that contains start, stop, and check actions
3. Create a resource profile for the application (define resource behavior)
4. Register the application as a resource with Oracle Clusterware

For all steps Oracle Clusterware provides self-explanatory command line interfaces that make it easy to individually perform the necessary steps and potential changes.

At present typically those steps must be performed individually for each application that is going to be managed by the cluster. Especially the script containing the appropriate start, stop, and check actions is usually to be developed by the user in regards to the application. Oracle is able to auto-generate this script utilizing an action script template as a basis for further customization.

In addition, Oracle provides a number of examples and documentation in terms of Technical White Papers to help developing those scripts. Those can be found under www.oracle.com/technology/products/database/clusterware/index.html[8]

An exception to this general policy can be found for certain SAP applications. Oracle has created an Oracle Clusterware tool, SAP Control (SAPCTL), to enable an easy management of SAP high availability resources.

SAPCTL provides an easy-to-use interface to administer the resources, scripts, and dependencies of Oracle Clusterware and SAP high availability resources. SAPCTL provides easy management for e.g. the SAP Enqueue Services, the SAP Replication Service, and the virtual IP address used by the SAP Enqueue Service. More information about this tool and its coverage can be found under: www.oracle.com/technology/products/database/clusterware/pdf/sap-availability-on-rac-twp.pdf

[8] Especially: Newlan, Philip:

- "Using Oracle Clusterware to Protect 3rd Party Applications"
- "Using Oracle Clusterware to Protect a Single Instance Oracle Database"
- "Using Oracle Clusterware to protect Oracle Application Server"

Oracle Clusterware in Extended / Stretched Clusters

Since it is a common configuration to set up a cluster in a way that the cluster nodes are not all located in one datacenter, Oracle Clusterware fully supports those types of configuration. Within Oracle terminology such an extended or stretched cluster is defined as "A cluster where all nodes are not located in the same room". [9]

This definition omits specifying a minimum or maximum distance between the nodes. And at present there is no official upper or lower limit on the supported distance between the nodes for Oracle Clusterware as a standalone product. Oracle has tested extended distance RAC clusters over distances of up to 100 kilometers. Nevertheless, the respective results may not be generally valid for the Oracle Clusterware as a standalone product.

Under www.oracle.com/technology/products/database/clustering/index.html[10] more information and considerations on setting up and using extended Oracle clusters can be found. Those considerations should generally be taken into account when an extended cluster is planned. If there is doubt whether or not Oracle Clusterware would still support a certain distance, consult Oracle Support through Oracle Metalink (http://metalink.oracle.com) for clarification.

Rolling Upgrade with Oracle Clusterware

Clusters in general are intended to provide certain high availability in case of unplanned downtime due to process, OS, or server failures. Oracle Clusterware moreover reduces planned downtime usually caused by patching.

Oracle Clusterware is generally Rolling Patch Upgradeable. "Rolling Patch Upgradeable" in this context means that Oracle Clusterware can be patched in a way that there is at least always one node of the cluster available.

Also other Oracle products can generally be patched in a rolling upgradeable manner. This for example applies to standard Oracle Database homes. Patches to be applied on those homes may be subject to certain restrictions.

All Oracle products that are typically used within an Oracle cluster stack generally support Minimal Downtime Patching. "Minimal Downtime Patching" in contrast to the Rolling Patch Upgrade scenario means that there is a certain downtime expected in course of the patching process. However, utilizing various techniques on the respective layer of the cluster can remarkably reduce this downtime. Minimal Downtime Patching should therefore always be used when Rolling Patch Upgrades are not applicable and minimal patch downtime is required. Examples of customers that have made use of those techniques can be found under: www.oracle.com/technology/deploy/availability/htdocs/HA_CaseStudies.html

[9] Oracle Database 10g Licensing Information, page 18 / chapter 2-4
[10] Especially: Peterson, Erik: "Oracle Real Application Clusters on Extended Distance Clusters"

Oracle Clusterware & Third Party Clusterware on One System

While generally not supported by most cluster software, Oracle does support Oracle Clusterware being used together with a third party Clusterware on the same system.[11] This again proves the flexibility in the usage of Oracle Clusterware and makes it the ideal cluster infrastructure for grid systems.

The Oracle Clusterware itself allows a simultaneous operation of Oracle Clusterware and a third party Clusterware that is certified by Oracle. In order to ensure the necessary changes within the Oracle Clusterware, the third party Clusterware needs to be installed first. The necessary changes to the Oracle Clusterware are then made in course of the consequent Oracle Clusterware installation. At this point Oracle Clusterware will be installed in a way that decisions about node membership are deferred to the third party cluster software.

Since the changes are made during the installation process a later removal of the underlying third party Clusterware would at present require a re-installation of the Oracle Clusterware.

While technically and generally possible as described to run Oracle Clusterware accompanied by a third party Clusterware on the same system, those configurations are not claimed to be best practice. If they are used, it is advisable to preserve a stringent configuration in order to ensure a smooth integration. More information on those configurations can generally be found in Oracle Metalink[12] or can often be obtained from the respective clusterware vendor.

SUMMARY AND CONCLUSION

Oracle Clusterware is still the basis for Oracle Real Application Clusters. But beginning with Oracle Database 10g Release 2, Oracle Clusterware is also a standalone cluster software fully capable to manage and thereby to protect any kind of application within a cluster environment.

Oracle Clusterware is supported on all operating systems certified for Oracle Real Application Clusters. That makes Oracle Clusterware one of the most portable clusterware solutions currently available.

In combination with its flexibility to be used together with third party clusterware solutions on the same system, Oracle Clusterware is the ideal cluster infrastructure for grid computing environments.

---

[11] Support may be restricted to certain third party clusterware versions and operating systems. More information can be found in Oracle Metalink / Certify

[12] Metalink Note 332257.1 "Using Oracle Clusterware with Vendor Clusterware FAQ"

ADDITIONAL INFORMATION

Additional information about Oracle Clusterware in general can be found under:

www.oracle.com/technology/products/database/clusterware/index.html

Material from this page used for this paper:

Newlan, Philip:

- "Using Oracle Clusterware to Protect 3rd Party Applications"
- "Using Oracle Clusterware to protect Oracle Application Server "
- "Using Oracle Clusterware to Protect a Single Instance Oracle Database"

Oracle documentation:

Oracle® Database Oracle Clusterware and Oracle Real Application Clusters Administration and Deployment Guide,10*g* Release 2 (10.2)
14 Making Applications Highly Available Using Oracle Clusterware

Additional information about Oracle Real Application Clusters can be found under

http://www.oracle.com/technology/products/database/clustering/index.html

Peterson, Erik: "Oracle Real Application Clusters on Extended Distance Clusters"

Additional material used for this paper

Lundhild, Barb: "Oracle Real Application Clusters 10g",
An Oracle Technical White Paper, May 2005

Additional as well as support and certification information can be found under

http://metalink.oracle.com

or http://www.oracle.com/technology/support/metalink/index.html

Material used for this paper:

Metalink Note 332257.1 "Using Oracle Clusterware with Vendor Clusterware FAQ"

Oracle Clusterware 10g
April 2007, Version 1.1
Author: Markus Michalewicz
Reviewers: Barb Lundhild, Jonathan Creighton, Philip Newlan

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