Oracle Database 12*c* Release 2
Oracle Real Application Clusters

ORACLE WHITE PAPER | MARCH 2017

Executive Overview

Oracle Real Application Clusters (RAC) continues to be the premium solution utilized by Customers to provide High Availability and Scalability to the Oracle Database. There is simply no other solution in the market that combines all the features that Oracle RAC provides without requiring any changes to the application. The latest release of Oracle RAC improves these features and provides significant enhancements in all of the areas that matter the most to your business.

Originally focused on providing best of class database services, Oracle RAC has evolved over the years and now provides a comprehensive High Availability (HA) stack that can be used in Public, Private or Hybrid Clouds to ensure High Availability, Scalability, Flexibility and Agility for any application.

Availability and scalability requirements of Oracle Databases have become more stringent in recent years. Companies can simply not afford any downtime to their databases and applications. The Oracle RAC Family of Solutions provide a fine tuned product bundle to ensure the availability and scalability requirements of databases and applications are met.

Oracle Real Application Clusters 12*c* Release 2 continues to improve these features with massive engineering resources dedicated to make use of the latest hardware innovations and industry trends like cloud computing and Machine learning.

These features can be broadly classified as, features that provide

» Better Scalability
» Better Availability
» More Efficiency and Management of a pool of Clusters

Oracle Real Application Clusters – Overview

Oracle Database with the Oracle Real Application Clusters (RAC) option allows multiple instances running on different servers to access the same physical database stored on shared Storage. The database spans multiple hardware systems and yet appears as a single unified database to the application. This enables the utilization of commodity hardware to reduce total cost of ownership and to provide a scalable computing environment that supports various application workloads. If additional computing capacity is needed, customers can add additional nodes instead of replacing their existing servers. The only requirement is that servers in the cluster must run the same operating system and the same version of Oracle. They do not have to be of the same capacity. This saves on capital expenditure as this allows customers to buy servers with latest CPU and memory and use it alongside existing servers. This architecture also provides high availability as RAC instances running on different nodes provides protection from a node failure. It is important to note that (almost) all applications including Oracle Applications, Peoplesoft, Seibel, SAP etc run without any changes on Oracle RAC database.

Customer's requirement for database availability and scalability continue to increase. They cannot afford any downtime in their environments. These requirements are not isolated to just databases but include other critical components like servers, network, client connections etc. Furthermore, there is a need for an intelligent resource manager that is able to redirect incoming workloads dynamically to nodes which are idle or in some cases more capable in terms of computing power and memory.

The Oracle RAC Family of Solutions provides a fine tuned product bundle to ensure all these requirements are met. The Oracle RAC stack is comprised of the following components which are referred to as "Oracle RAC Family of Solutions".

Oracle RAC Family of Solutions

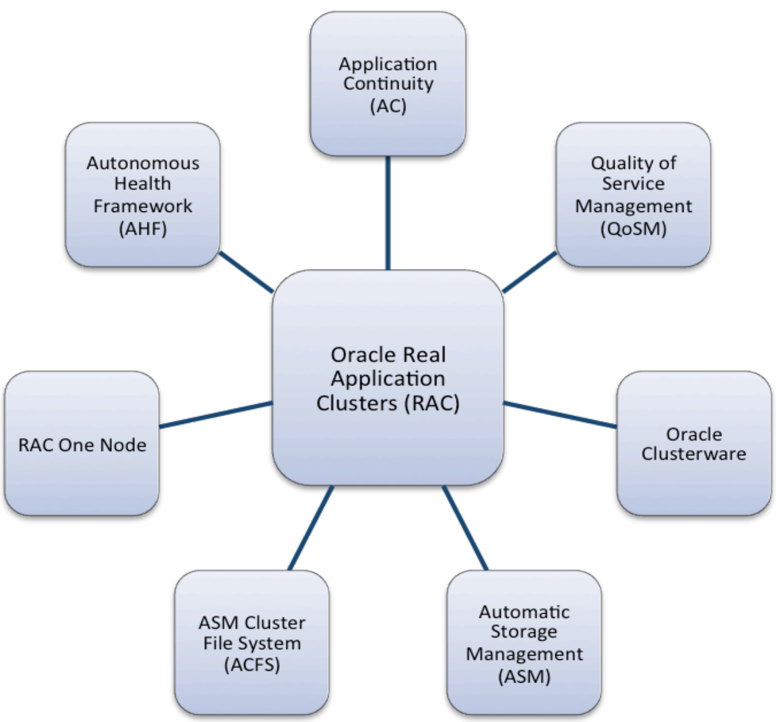

Figure 1: Oracle Family of Solutions

The functionality provided by Oracle RAC Family of Solutions can be used by licensed Oracle RAC or Oracle RAC One Node customers without any additional charge.

Oracle Clusterware

Oracle Clusterware is a technology that transforms a server farm into a cluster. Oracle Clusterware provides among other things node membership, node fencing and optimal resource placement functionality. Oracle Clusterware is a complete, free-of-charge clustering solution that can be used with Oracle RAC, RAC One Node and even Single Instance Oracle databases.

New in Oracle Clusterware 12*c* Release 2

Oracle Clusterware 12*c* Release 2 introduces new deployment options for easier management of large deployments of clusters. This new architecture is called Oracle Cluster Domain. Using a Cluster Domain deployment would free individual clusters to dedicate all its resources to the database or application while management tasks like deployment, storage management, performance monitoring etc is delegated to the Domain Services Cluster.

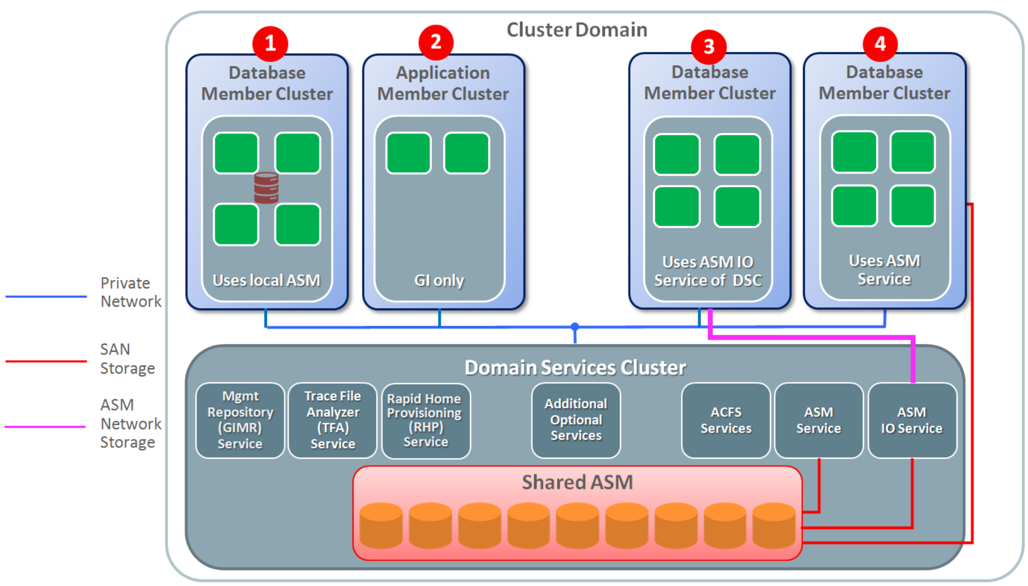

Figure 2: Oracle Cluster Domain

As shown in figure 2, a Cluster Domain consists of a Domain Services Cluster (DSC) and one or many member clusters. A Domain Services Cluster (DSC) offers services to the member clusters. Member clusters are the clusters on which databases or applications are deployed. There can be four types of member clusters. (1) Database member cluster(s) with high performance storage local to that member cluster, (2) Application member cluster(s) typically hosting applications without shared storage, (3) Database member cluster(s) accessing ASM storage on the DSC via ASM I/O Service and (4) Database member cluster(s) accessing ASM storage on the DSC directly via SAN storage. All the member clusters utilize the centralized Management Repository services, Trace File Analyzer services and other services provided by the Domain Services Cluster

Note that Oracle Clusterware 12*c* Release 2 can also be deployed as Standalone Cluster as in previous releases.

Choosing a deployment model

The choice of deployment depends on the installation type. New installations can choose between Cluster Domain based deployment model or Standalone Cluster deployment model. Upgrades from previous releases will continue to remain standalone as the deployment model is not changed during upgrades. Oracle Clusterware licensing remains the same for both deployment models. Some aspects to consider when choosing a deployment model are:

» Cluster Domain architecture delegates the management aspects of Member Clusters to the Domain Services Cluster. This optimizes the Member Cluster management in terms of both provisioning and performance management. Resources like CPU on the Member Cluster can now be dedicated to the database computing needs

» Cluster Domain architecture provides a unified consolidated storage solution via the Domain Services Cluster (DSC). This model makes it easier to provision new databases using the Oracle ASM cloning feature. Storage consolidation using the DSC deployment model benefits vastly from the new Database oriented storage management features introduced in Oracle ASM 12*c* Release 2

» Centralized data collection facilities provided by Autonomous Health Framework (AHF) in the DSC allow the Member Clusters' behavior to be analyzed using Machine learning capabilities used by AHF which continuously monitors the Member Clusters. This functionality can in many cases prevent a problem before it occurs. For example, AHF can detect anomalies between real time performance counters and expected values to notify system admin of impending performance issues while generating targeted diagnosis and corrective actions.

For more information about Oracle Clusterware, see www.oracle.com/goto/clusterware

Oracle Automatic Storage Management (ASM)

Oracle Automatic Storage Management (ASM) is the recommended volume manager which can be used for both, Oracle RAC and Single Instance Oracle Databases. Oracle ASM simplifies storage management through the principle of “stripe and mirror everything” (SAME). Intelligent mirroring capabilities allow administrators to define 2- or 3-way mirrors to protect vital data. When a read operation identifies a corrupt block on a disk, Oracle ASM automatically relocates the valid block from the mirrored copy to an uncorrupted portion of the disk.

New in Oracle ASM 12*c* Release 2

Oracle ASM 12*c* Release 2 introduces database-oriented storage management via the new ASM Flex Disk Group. An ASM Flex Disk Group provides enhanced management capabilities like (a) modifiable redundancy at individual database file level via File Groups (b) snapshot capabilities and (c) quota management at the database level for consolidated environments.

The ability to create snapshots without relying on the snapshot capabilities of the underlying storage enables DBAs to rapidly provision databases. ASM snapshots are executed at the database level without the need for downtime or any additional manual recovery steps. Modifiable redundancy allows database administrators to start with a conservative mirroring strategy and change the redundancy in future depending on business needs. Quota management allows storage administrators to set quota at the database level which helps in consolidation environments as it prevents one database from utilizing all the space in the Flex Disk Group.

For more information about Oracle Automatic Storage Management, see http://www.oracle.com/goto/asm

Oracle ASM Cluster File System ACFS

Oracle ACFS complements ASM file management capabilities by providing a POSIX-compatible file system to store general purpose and database files. Oracle database has for a long time provided column types to store blobs, XMLs and text files etc. However due to application or business requirements, customers needed a file system to store such data. Obviously storing this data outside the Oracle database requires customers to manually plan for data management activities like backup, synchronization across sites etc.

ACFS provides a Cluster file system which customers can use to store this data. They can additionally use ACFS features like Tagging, Replication and Snapshots to ease their data management activities. Customers can utilize ACFS tagging feature to add Tags to their data and retrieve tags using a command line or they can utilize Tagging API calls directly from their application. They can also use

ACFS snapshot capabilities utilizing copy-on-write (COW) on generic systems without relying on specialized storage. This means customers can take multiple snapshots while consuming minimal space. ACFS also provides replication facilities to replicate this data to other systems for site availability.

For more information about Oracle ACFS, see http://www.oracle.com/goto/acfs

Rapid Home Provisioning (RHP)

Oracle's Rapid Home Provisioning (RHP) helps Oracle database administrators to rapidly and efficiently provision Oracle Grid Infrastructure and Oracle Database(s). Rapid Home Provisioning (RHP) provides a centralized software deployment and maintenance functionality whereby software stored on the RHP server can be provisioned to any node or cluster.

For more information about Oracle RHP, see www.oracle.com/goto/rhp

Autonomous Health Framework (AHF)

Oracle Autonomous Health Framework (AHF) presents the next generation of tools which autonomously work 24x7 to keep database systems healthy and running while minimizing human reaction time. AHF daemons continuously collect and analyze performance and configuration data from the operating system and the database. AHF applies Machine learning concepts in real time to the data to anticipate events like impending component failures or change in workloads. Based on the outcome of its analysis, AHF can either automatically shift workload to another instance or notify the administrator about imminent failures that may require manual intervention.

For more information about AHF, see www.oracle.com/goto/ahf

SCAN

Single Client Access Name (SCAN) introduced in Oracle Grid Infrastructure 11g Release 2, acts as a cluster alias for databases in the cluster. The benefit is that the client's connect information does not need to change if nodes are added or removed from the cluster.

With Oracle Grid Infrastructure 12*c*, SCAN has been enhanced to (a) support IPv6 based IP addresses to provide a choice of IPv4 or IPv6 based addresses as needed, (b) support multiple subnets in the cluster to segregate incoming clients into different network subnets for fine grained management and (c) to restrict service registration to disallow arbitrary services that are not part of the RAC cluster from registering with SCAN to route user sessions to malicious nodes thereby providing enhanced security.

For information about SCAN, see www.oracle.com/technetwork/products/clustering/overview/scan-129069.pdf

Dynamic Database Services

Sessions connect to an instance using services. The service itself can be configured to run on one or more instances. An instance can host one or more services. Services allow breaking up workloads into manageable components providing more granularity. For example, OLTP users can be configured to use service(s) that redirect sessions to the Hub nodes, while DSS users can be configured to use (an)other service(s) to redirect sessions to read only instances. Once the services are configured, sessions are automatically routed to the appropriate node providing that service to ensure optimal use of resources without manual customer intervention.

Connection Load Balancing

Client-side "load balancing" balances connection requests across all SCAN listeners in the cluster by using the SCAN on the address list of the client connect string. SQL*NET will randomly select one of the SCAN IP addresses.

Server-side load balancing is achieved using the SCAN listener. Each SCAN listener is aware of all instances in the cluster providing each service. Based on the goal defined for the service, the listener chooses the instance that will best meet the goal and the connection is routed to that instance through the local listener.

Load Balancing Advisory

Load balancing distributes work across all of the available Oracle RAC database instances. Oracle recommends that applications use connection pools with persistent connections that span the instances that offer a particular service. When using persistent connections, connections are created infrequently and exist for a long duration. Work comes into the system with high frequency, borrows these connections, and exists for a relatively short duration. Load Balancing Advisory directs incoming user connections to the instances that will provide the optimal quality of service for that work. This minimizes the need to relocate the work later.

All the above components work cohesively to reroute work to provide the best service times globally, and more importantly work is re-routed gracefully depending on changing system conditions.

Better Scalability

The volume of data that customers have to manage is growing at a phenomenal rate. Businesses continue to face the uphill task to capture and analyze the data to react better to day-to-day operational or long term business events. Processing these large data volumes requires serious computing power. Oracle RAC is perfectly suited for such tasks as customers can start with a smaller footprint and scale out horizontally as needed.

figure 3 shows the result of SAP sales and distribution (SD) module benchmark with Oracle RAC, which shows how SD users increase as Oracle Instances are added to the system.

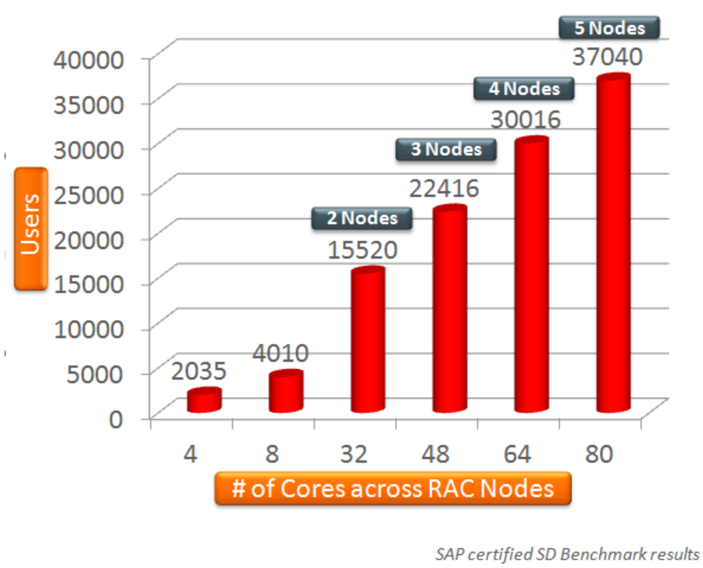

Figure 3: Scalability without Application Code changes

Oracle RAC provides all the availability and scalability features while providing ACID properties that most applications require. Customers who deploy their applications with Oracle RAC do not have to worry about stale reads. They can focus on adding business logic into their application while Oracle RAC database takes care of all their transactional needs. Moreover, they can utilize other Oracle Database features along with Oracle RAC. For example, Oracle Data Guard can be used for Disaster recovery, Oracle In-memory DB for Advanced Analytics, Oracle Multitenant for Consolidation etc can be used along with Oracle RAC with little or no additional configuration changes. In fact, the benefits provided by these database features are further augmented when used with Oracle RAC.

Improved Algorithms

Oracle RAC scalability is a result of optimized algorithms that keep up with the changes in technology. Oracle Cache Fusion which is a component of Oracle Real Application Clusters is the magic that works behind the scenes to synchronize the cache of all the Oracle RAC instances even when multiple users are running transactions that are actively updating these caches. There are absolutely no manual steps required for this concurrency. Until recently, Oracle Cache Fusion solely utilized the private network to keep the cache synchronized, as rotating disk performance has been traditionally slower. However, lately disk performance is improving as a result of storage vendors utilizing newer technologies like SSDs and NVME to reduce I/O latencies. So in certain circumstances, it may actually be beneficial to read blocks from disk rather than ship the block over the private network.

Cache Fusion's enhanced algorithm provides the best of both worlds. As shown in figure 4, Cache Fusion monitors the network and storage I/O statistics on an ongoing basis and will utilize the more efficient path as needed. Note that this is done automatically and on an ongoing basis, without any manual intervention. These optimizations ensure that customers continue to benefit from Oracle Real Application Clusters as their hardware changes without the need for manual environment specific adaptations.

The complexity of the algorithms in Oracle RAC are abstracted from the Application and do not require any change in application code.

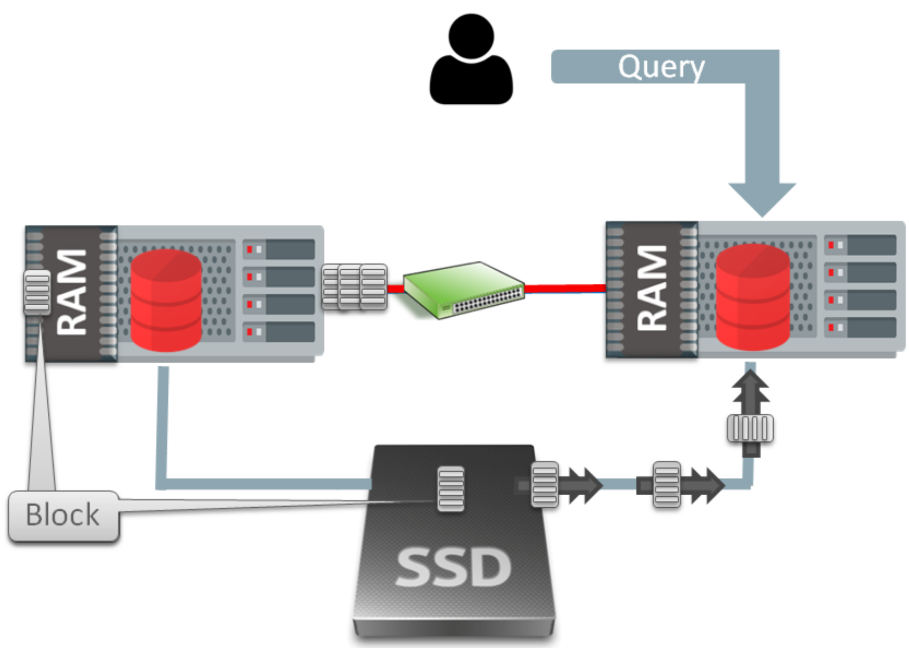

Figure 4: Scalability without Application Code changes

Service Oriented Buffer Cache

Service Oriented Buffer Cache as introduced in Oracle RAC 12*c* Release 2 essentially reduces and in many cases eliminates "physical reads" after a planned singleton service failover. Prior to this feature, singleton service failover affects response time as buffers that were cached on the failed instance have to be re-read from the disks incurring the overhead of physical reads. Starting with Oracle RAC 12*c* Release 2, Cache Fusion maintains an in-memory hash that tracks the blocks in the buffer cache and the service name used by the sessions to connect. Cache Fusion uses this hash in two ways:

(a) Resource Mastering optimization: Resource mastering of a resource cached in the buffer is only considered on the node where the service that the session used to access the resource is running, This results in improved performance as it eliminates the need for sending additional messages on the private network for resource change operations.

(b) Pre-Warm the Buffer Cache: During planned maintenance, when a singleton service is failed over, Cache Fusion will pre-warm the buffer of the instance to which the service is going to failover. This reduces the physical reads that the sessions would have otherwise incurred, resulting in consistent performance for those failed over sessions as shown in figure 5.

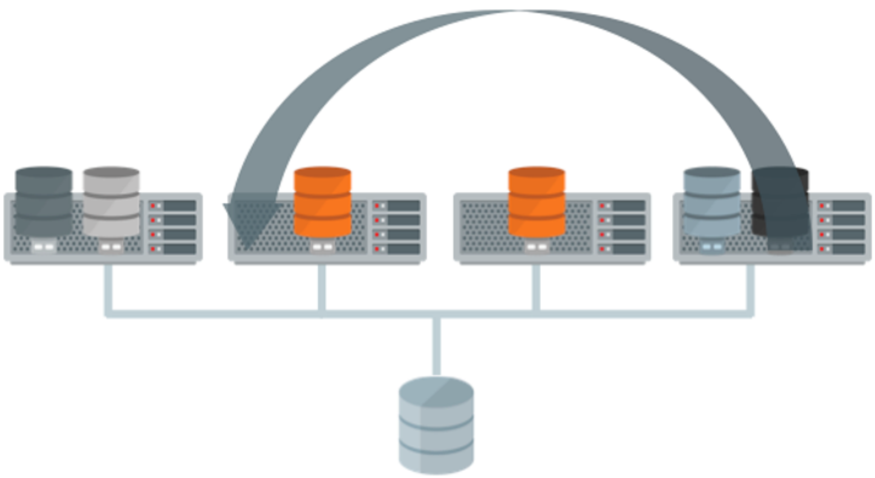

Figure 5: Service Oriented Buffer Cache

Pluggable Database and Service Isolation

Consolidation helps customers to use available computing power more efficiently while reducing overall costs. Oracle provides various solutions for consolidation that can be used together or individually as per customer needs. The Oracle Multitenant option is one of such options used by customers for consolidation. Consolidating databases without adequate planning for availability exposes customers to downtime as multiple databases running on one server results in that one server being a single point of failure. Oracle RAC can be used to seamlessly to provide high availability with Oracle Multitenant, as multiple Oracle RAC instances running on different nodes eliminate downtime on these consolidated environments caused by node failures. Additionally, customers benefit from the Service Isolation feature resulting in consistent performance achieved by reducing internal distributed lock manager (DLM) operations for “Pluggable Databases”/Services not offered in all instances and optimizing block operations based on in-memory block separation.

Better Availability

The cost to business due to application or database downtime has increased over the years. This is further exacerbated in consolidated environments as multiple applications and databases may be affected. Planning redundancy at multiple layers in the datacenter is critical to achieving availability. Tasks like firmware patching and OS patching require planned downtime. It is essential that events like nodes leaving or joining the cluster are (a) least disruptive to the underlying database and (b) notify the user sessions connected to the nodes affected by the planned maintenance operation so that those sessions can gracefully reconnected to another node in the cluster.

Smart Reconfiguration

Oracle RAC 12*c* Release 2 further reduces disruption from both planned and unplanned operations utilizing features like Recovery Buddy and service Isolation up to 4 times faster than previous releases.

Recovery Buddy

Recovery buddy introduces the concept of a Buddy instance for every Oracle RAC instance.

Figure 6: Recovery Buddy

Recovery Buddy tracks block changes in a hash table in the SGA of the buddy instance. So if an instance is evicted, the changes made by that instance are available in the SGA of its buddy instance. This optimization allows dramatically faster instance recover as this hash table is utilized instead of reading the redo log changes from disk. Recovery Buddy feature along with the optimized singleton workload scaling results in almost four times performance improvement during reconfiguration.

Oracle Read mostly instances on Reader Nodes

Oracle Grid Infrastructure 12*c* introduced the concept of Oracle Flex cluster in which nodes can be classified as a Hub node or a Leaf node. In Oracle Grid Infrastructure12*c* Release 1, database instances could only be installed on Hub nodes and Leaf nodes were used to host applications. Starting with Oracle Grid Infrastructure 12*c* Release 2, it is now possible to run read-only instances on Leaf nodes. This architecture vastly improves availability as database instances running on Hub nodes are not affected by the read only instances on Leaf nodes joining or leaving the cluster. Customers can configure their OLTP services to connect to the instances running on Hub nodes while DSS or read sessions can be directed to the instances running on Leaf nodes.

As Leaf nodes joining or leaving the cluster does not affect overall database availability, customers can scale up and improve the performance of DSS queries by adding additional read only instances running on Leaf nodes. Oracle Parallel Query slaves can be configured to run on these read only instances with additional memory to help improve typical parallel query activities like sorting or hash-joins. Additionally, instances running on Leaf nodes can be configured with the new “Local Temporary Tablespace”. Parallel Query slaves running massive sorts or hash joins can utilize these new Local Temporary Tablespace” if they run out of physical memory and need to spill to disk. figure 7 shows a four node RAC instance with two Hub nodes and two Leaf nodes with services configured.

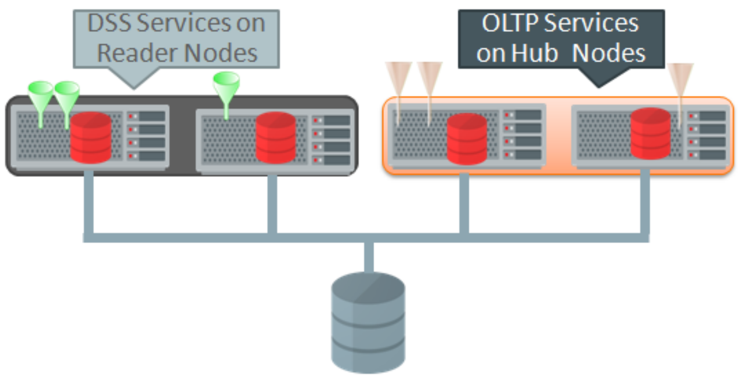

Figure 7: Read only instances on Reader nodes

Node Weighting

Prior to Oracle RAC 12c Release 2, node eviction caused by a split brain did not consider customer workload and followed a pattern to evict the higher node number leaving the lowest node number in the cluster. While this was effective, it did not take into account criticality of workloads and secondary failures.

Oracle RAC 12c Release 2 introduces Node Weighting feature which considers additional aspects like the criticality of the workload running on the node, number of services running on the node and even secondary failures like public network failure to reach a more informed decision on the node to be evicted during a split brain that results in the cluster splitting into equal half’s.

DBAs can optionally identify a node to be marked critical during configuration. The criticality attribute can be defined at the node level or at the service level using appropriate crsctl or srvctl commands. A node level definition can be used to mark a specific node that has say more CPU or memory than other node while a service level definition can be used if that service is deemed more important than other services running on the cluster.

Application Continuity

Application Continuity feature masks outages caused by nodes joining or leaving the cluster by replaying in-flight work for impacted database sessions. In Oracle RAC 12*c* Release 2, Application Continuity is available for applications using Java, OCI and ODP.NET and is integrated with most Application Servers including WebLogic and Tomcat.

For more information about Application Continuity, see http://www.oracle.com/technetwork/database/options/clustering/applicationcontinuity/overview/index.html

Efficient Management of a Pool of Clusters

The volume of data that the customer has to manage is growing and newer trends like analytics, Internet of things (IoT) continues to generate more and more data. Additionally, the need to mine data to improve various business services is becoming even more critical to organizations. Databases and applications to manage these ever increasing data volumes require an efficient and scalable deployment engine. Once deployed, these deployments need to be managed from both workload scheduling and performance management aspect. Oracle RAC 12*c* Release 2 features Rapid Home Provisioning (RHP) and Autonomous Health Framework (AHF) help manage these deployments in a scaleable and efficient manner.

Oracle Rapid Home Provisioning

Oracle RHP provides an efficient and non-disruptive method to provision, patch and upgrade various layers of the Oracle software infrastructure; these layers include but are not limited to Oracle Grid Infrastructure, Oracle Database (RAC, RAC One Node and Single Instances), Applications and Middleware.

RHP can be configured to easily provision, upgrade and patch Standalone clusters, Domain Services clusters and Member clusters as shown in figure 9. RHP itself can be provisioned in a standalone deployment model or as part of the new Cluster Domain deployment model. RHP can help standardize customer software install by the use of gold images. Essentially customers can create an environment or site specific gold image which can then be used as the standard image to be deployed.

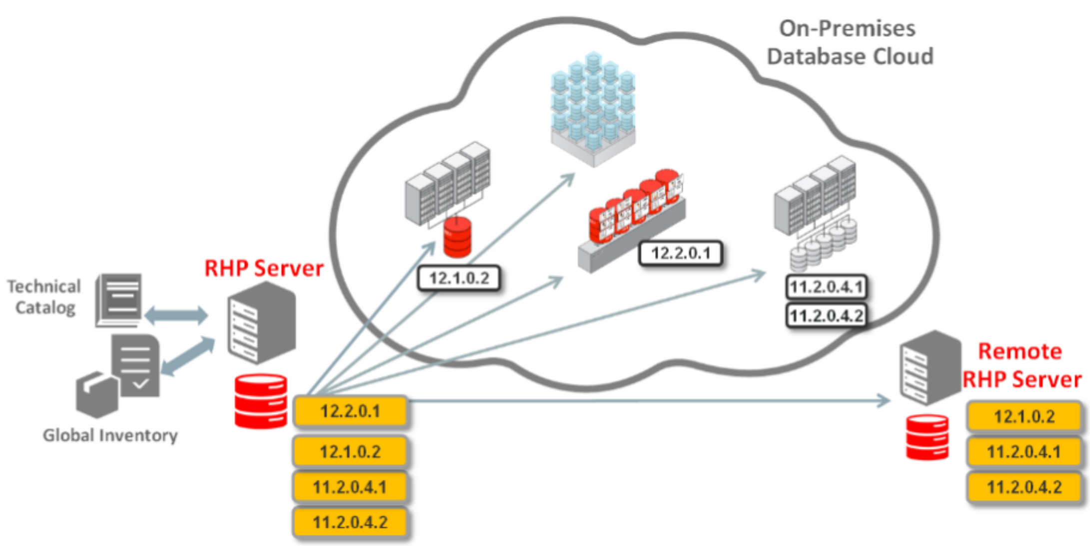

Figure 8: Rapid Home Provisioning

Oracle Autonomous Health Framework

Administrators are constantly overwhelmed with the increase in systems that need to be administered, monitored, managed and tuned and to keep them running efficiently within defined service level agreements (SLAs). Oracle Autonomous Health Framework (AHF) is a collection of next generation tools, which autonomously work together 24x7 to keep database systems healthy and running while minimizing human reaction time. These components include both existing tools as components in ORAchk, Cluster Verification Utility, Trace File Analyzer, Cluster Health Monitor, Quality of Service Management, Memory Guard, and new components in Cluster Health Advisor and Hang Manager as shown below

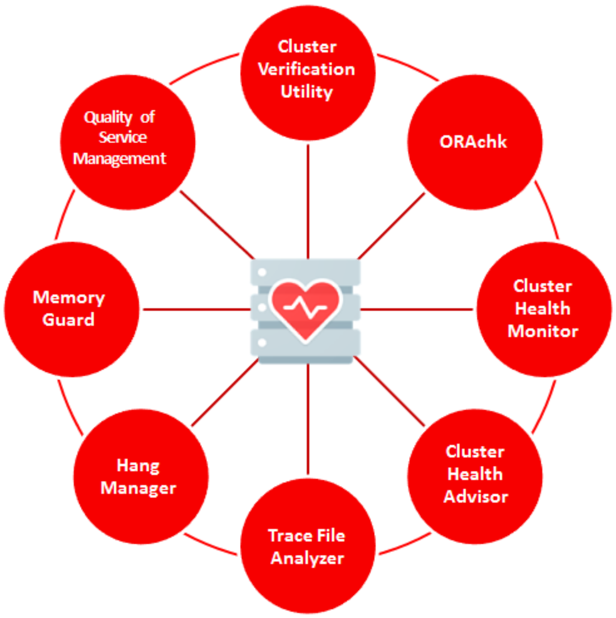

Figure 9: Autonomous Health Framework

Oracle RAC 11g introduced services as a concept that provides administrators means to split workloads into manageable chunks that can be directed to run on different nodes of the cluster. For example, an OLTP service can be defined for all OLTP users and an Analytics service can be defined for DSS workloads. In production systems, these workloads could be affected by runaway queries or hangs that may affect service level agreements of these services. Finding out the root cause of these runaway queries or hangs require manual intervention and is a function of System administrators and Database administrators reaction time. It is increasingly important to (a) find the issue quickly and (b) isolate the issue hopefully before it affects SLAs.

Oracle Autonomous Health Framework components work together in daemon mode to address such issues with the intent to minimize and reduce the impact of the reaction time on these systems. AHF components such as Cluster Verification Utility (CVU) and ORAchk provide lightweight and non-intrusive health checks for the Oracle stack of software and hardware components including Oracle RAC Databases. It proactively scans for known issues and recommends resolutions.

AHF components such as Cluster Health Monitor, Cluster Health Advisor, Hang Manager, Trace File Analyzer (TFA) and Quality of Service Management work together to provide higher availability and diagnosability by constantly monitoring the performance metrics on the nodes and efficiently directing workloads. Hang Manager automatically detects and resolves hangs that may otherwise cause database hangs and affect SLAs. TFA automatically collects relevant diagnostic information across multiple nodes that can be used by Oracle Support Services for problem resolution. Data collected by the Cluster Health Monitor is utilized by the Cluster Health Advisor as an early warning mechanism of impending performance issues. Machine learning concepts are applied to the data collected by the AHF daemons to constantly learn the usage patterns to (a) predict issues that would result in performance degradation or complete lack of service and (b) provide accurate cause of component, server, and Instance failures.

The components of AHF work together autonomously to identify issues that could affect the health of the system and provide corrective actions to resolve them quickly with minimal effort.

Conclusion

Scalability and availability requirements for OLTP systems are at all time highs. Customers simply cannot afford any downtime on these critical systems. It is becoming more and more critical to analyze the vast volumes of data quickly to react to changing customer demands. Oracle Real Application Clusters along with the Oracle RAC Family of Solutions makes it easier for customers to take advantage of the Business continuity, high availability, scalability, flexibility and agility features to rapidly adapt to changing business needs. Oracle Real Application Cluster 12*c* Release 2 continues on this path by providing significant enhancements in all of the areas that matter most to your continued business success. New features such as Autonomous Health Framework and Rapid Home Provisioning as well as Improved algorithms in all the components of the stack make Oracle RAC the database virtualization solution of choice for your IT infrastructure

Oracle Corporation, World Headquarters
500 Oracle Parkway
Redwood Shores, CA 94065, USA

Worldwide Inquiries
Phone: +1.650.506.7000
Fax: +1.650.506.7200

CONNECT WITH US

blogs.oracle.com/oracle

facebook.com/oracle

twitter.com/oracle

oracle.com

Integrated Cloud Applications & Platform Services

Oracle Real Application Clusters 12c Release 2 – Technical Overview
March 2017
Author: Anil Nair
Contributing Authors: Markus Michalewicz