Load Balancing in Oracle Database Real Application Cluster

Ms. Manju Sharma

(College of Computer Science & Information Systems, Jazan University, Jazan, KSA)

Abstract: *Clustering is architecture of joining set independent, interconnected computers to act as a single unit or on a single server. Presently the high availability, scalability, flexibility and ability combine the easy management of the successful infrastructure and cloud deployments. From more than a decade the Oracle Database with Real Application Clusters (RAC) has been the solution of choice for thousands of Oracle customers. Oracle Real Application Cluster 12c is a foundation of data centers, provides the significant enhancements in all of the areas of the business success. In this paper we will take an overview of load balancing, automatic failover and load balancing for Oracle real Application Clusters, Network Configuration , performance monitoring.*

Keyword*: Real Application Clusters, Advantages of RAC , Network Configuration, Oracle Cluster Registry File, Oracle Local Registry, Load Balancing , Voting Disk file.*

I. Introduction

In Real Application Clusters environments, all nodes concurrently execute transactions against the same database. Real Application Clusters coordinates each node's access to the shared data to provide consistency and integrity. Oracle RAC is a cluster database with a shared cache architecture that overcomes the limitations of traditional shared-nothing and shared-disk approaches to provide a highly scalable and available database solution for all your business applications. Oracle RAC provides the foundation for enterprise grid computing.

Oracle's Real Application Clusters (RAC) option supports the transparent deployment of a single database across a cluster of servers, providing fault tolerance from hardware failures or planned outages. Oracle RAC running on clusters provides Oracle's highest level of capability in terms of availability, scalability, and low-cost computing. Real Application Clusters joins two or more interconnected, but independent servers as one instance per node. Multiple instances can access the same database. Database files stored on disks physically or logically connected to each node, so that every instance can read from or write to them. Oracle RAC is one of the important in clustered oracle databases and uses oracle cluster ware software for the infrastructure to bind multiple servers so that they operate as a single system. A cluster comprises of multiple interconnected computers or servers that appear to be one server to end users and applications [3].

II. RAC (Real Application Clusters)

Oracle RAC architecture option provides a single system image for multiple servers to access one Oracle database. In Oracle RAC, each Oracle instance usually runs on a separate server. The combined processing power of the multiple servers can provide greater throughput and Oracle RAC scalability than is available from a single server. Figure 1 depicts line diagram of Oracle database with Oracle RAC architecture [1].

Reliability – if one node fails, the database won't fail.

Availability – nodes can be added or replaced without having to shut down the database.

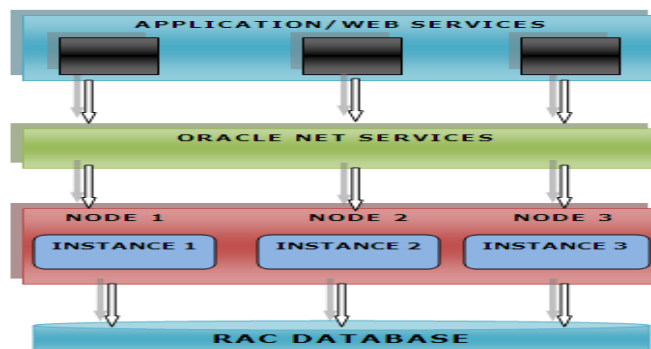**Scalability** – more nodes can be added to the cluster as the workload increases.

Figure 1: Oracle database with Oracle RAC architecture

Oracle Cluster ware is the software which enables the nodes to communicate with each other, allowing them to form the cluster of nodes which behaves as single logical server. Oracle Cluster ware is run by Cluster Ready Services (CRS) consisting of two key components: Oracle Cluster Registry (OCR) and Voting Disk.

Oracle Real Application Clusters 9i (Oracle RAC) used the same IDLM and relied on external cluster ware software (Sun Cluster, VERITAS Cluster, etc). It provides the basic clustering services at the operating system level that enable Oracle software to run in clustering mode. In earlier versions of Oracle (version 9i and earlier), RAC required a vendor supplied cluster ware like Sun Cluster or VERITAS Cluster Software with the exception of Linux and Windows.

Oracle RAC 10*g*Release 2 for Linux on zSeries® was introduced in 2008 with version 10.2.0.2, which is not an Oracle certified release. 10.2.0.2 was superseded with version 10.2.0.3, which is Oracle certified. However 10.2.0.3 became upgradeable to 10.2.0.4, also Oracle certified.

A RAC cluster includes one database, one or more instances, a database is a set of files, Located on shared storage, Contains all persistent resources. An instance is a set of memory structures and processes, Contain all temporal resources, Can be started and stopped independently.

III. Load Balancing

The Oracle RAC system can distribute the load over many nodes this feature called as load balancing.

There are two methods of load balancing :

3.1. Client Load Balancing distributes new connections among Oracle RAC nodes so that no one server is overloaded with connection requests and it is configured at net service name level by providing multiple descriptions in a description list or multiple addresses in an address list. For example, if connection fails over to another node in case of failure, the client load balancing ensures that the redirected connections are distributed among other nodes in the RAC.

3.2. Server Load Balancing distributes processing workload among Oracle RAC nodes. It divides the connection load evenly between all available listeners and distributes new user session connection requests to the least loaded listener(s) based on the total number of sessions which are already connected. Each listener communicates with the other listener(s) via each database instance's PMON process.
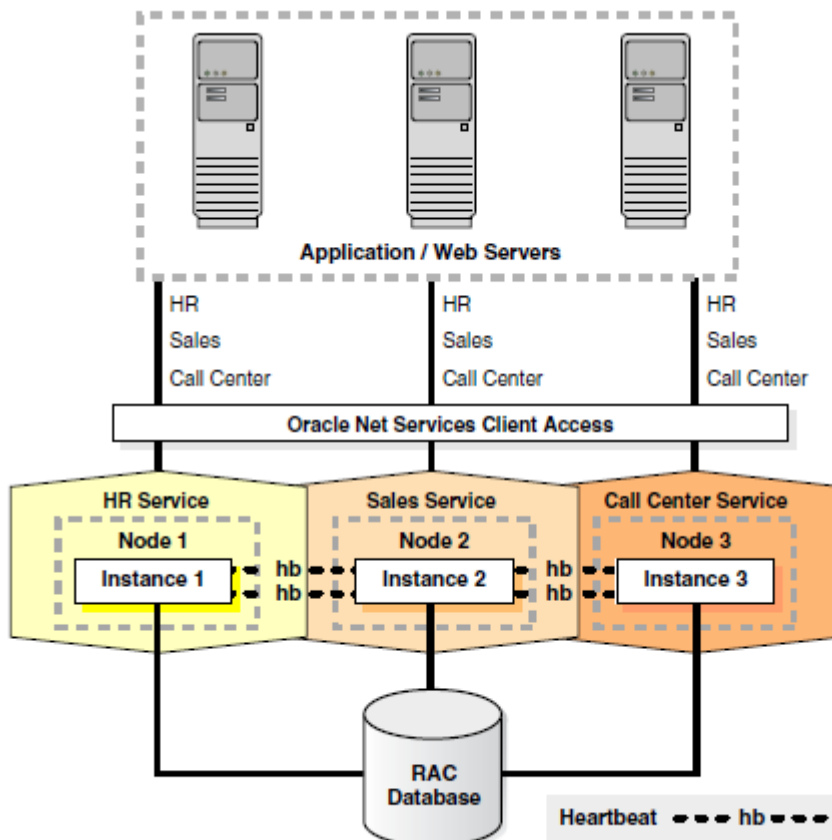

Figure 2: Oracle Database with Oracle RAC Architecture

IV. Automatic Failover

Automatic failover is supported in database for high safety mode. In high-safety mode with automatic failover, once the database is synchronized, if the principal server becomes unavailable, an automatic failover occurs. An automatic failover causes the secondary server to take over the role of principal server and bring its copy of the database to the user. Requiring that the server be synchronized prevents loss to the user during failover, because every transaction committed on the principal server can also be committed on the secondary server.

Automatic failover requires the following conditions:

1. The secondary server must be running in highsafety mode

2. The secondary server must have access to the main database How Automatic Failover Works Under the preceding conditions automatic failover initiates the following sequence of actions:

1. If the principal server fails, it changes the state of the principal server to DISCONNECTED and disconnects all clients from the principal server.

2. The secondary server register that the principal server is unavailable.

3. All clients from the principal server are shifted to the secondary server by clusterware [2].

•Grid Naming Service (GNS) introduced in Oracle RAC 11g R2. With GNS, Oracle Cluster ware (CRS) can manage Dynamic Host Configuration Protocol (DHCP) and DNS services for the dynamic node registration and configuration.
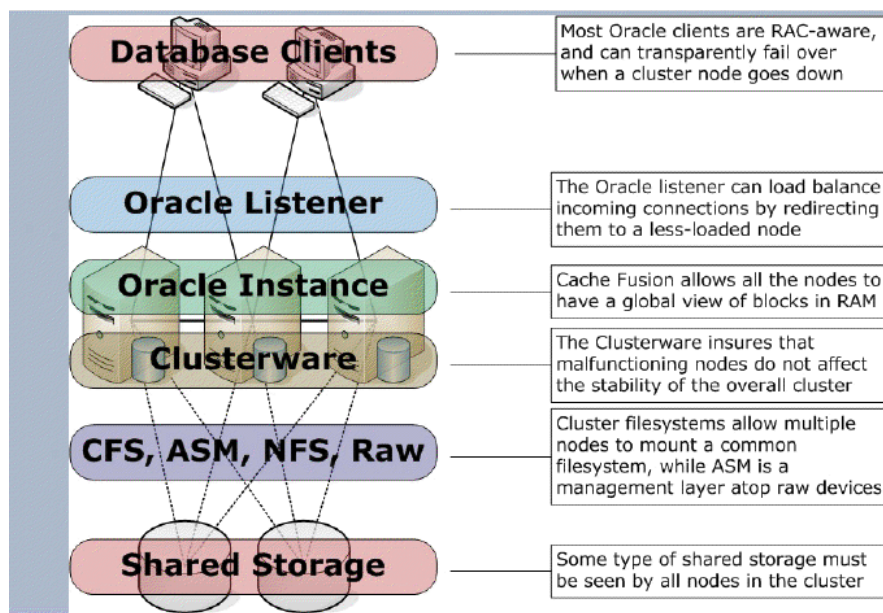
V. Interconnect

Instances communicate with each other over the interconnect (network). Information transferred between instances includes data blocks, locks SCNs. For checking lag/Traffic over inter connect we can follow below steps:-

Below views provide the current hardware configuration:

SELECT* FROMgv$configured_interconnects ORDERBYinst_id,name;
SELECT* FROMgv$cluster_interconnects ORDERBYinst_id,name;

There are multiple views providing the amount of blocks (data, undo, …) exchanged between cluster instances:

SELECTinst_id,class,cr_block,current_block **FROM** gv$instance_cache_transfer

WHEREinstance **IN**(1,2) **ORDERBY**inst_id,class;

For checking the current on going traffic we have the below formula derived from script **sprepins.sql**file located in $ORACLE_HOME/rdbms/admin directory.

Estd Interconnect traffic = ((Global Cache blocks received + Global Cache blocks served)*db_block_size + (GCS/GES messages received + GCS/GES messages sent)*200)/elapsed time

If you want to recompute what you find in an AWR report you can use DBA_HIST_DLM_MISC and DBA_HIST_SYSSTAT hsitory tables.

Checking Interconnect Used:-Identify the interconnect used lan902 172.17.1.0 global cluster_interconnect lan901 10.28.188.0 global public.

VI. Manual Network Configuration

Identity	Home Node	Host Node	Given Name	Type	Address	Address Assigned By	Resolved By
Node 1 Public	Node 1	racnode1	racnode1[1]	Public	192.0.2.101	Fixed	DNS
Node 1 VIP	Node 1	Selected by Oracle Clusterware	racnode1-vip	Virtual	192.0.2.104	Fixed	DNS, hosts file
Node 1 Private	Node 1	racnode1	racnode1-priv	Private	192.168.0.1	Fixed	DNS, hosts file, or none
Node 2 Public	Node 2	racnode2	racnode2[1]	Public	192.0.2.102	Fixed	DNS
Node 2 VIP	Node 2	Selected by Oracle Clusterware	racnode2-vip	Virtual	192.0.2.105	Fixed	DNS, hosts file
Node 2 Private	Node 2	racnode2	racnode2-priv	Private	192.168.0.2	Fixed	DNS, hosts file, or none
SCAN VIP 1	none	Selected by Oracle Clusterware	docrac-scan	virtual	192.0.2.201	Fixed	DNS
SCAN VIP 2	none	Selected by Oracle Clusterware	docrac-scan	virtual	192.0.2.202	Fixed	DNS
SCAN VIP 3	none	Selected by Oracle Clusterware	docrac-scan	virtual	192.0.2.203	Fixed	DNS

[1] Node host names may resolve to multiple addresses.

VII. Network Configuration

After configured the network, we need to perform verification tests to make sure it is configured properly. If there are problems with the network connection between nodes in the cluster, then the Oracle Cluster ware installation fails.

To verify the network configuration on a two-node cluster that is running Oracle Linux:
1. As the root user, verify the configuration of the public and private networks. Verify that the interfaces are Configure on the same network (either private or public) on all nodes in your cluster. In this example, eth0 is used for the public network and eth1 is used for the private network on each node.

```
# /sbin/ifconfig
eth0 Link encap:Ethernet HWaddr 00:0E:0C:08:67:A9
inet addr: 192.0.2.100 Bcast:192.0.2.255 Mask:255.255.240.0
UP BROADCAST RUNNING MULTICAST MTU:1500 Metric:1
RX packets:270332689 errors:0 dropped:0 overruns:0 frame:0
TX packets:112346591 errors:2 dropped:0 overruns:0 carrier:2
collisions:202 txqueuelen:1000
RX bytes:622032739 (593.2 MB) TX bytes:2846589958 (2714.7 MB)
Base address:0x2840 Memory:fe7e0000-fe800000


eth1 Link encap:Ethernet HWaddr 00:04:23:A6:CD:59
inet addr: 10.10.10.11 Bcast: 10.10.10.255 Mask:255.255.240.0
UP BROADCAST RUNNING MULTICAST MTU:1500 Metric:1
RX packets:21567028 errors:0 dropped:0 overruns:0 frame:0
TX packets:15259945 errors:0 dropped:0 overruns:0 carrier:0
collisions:0 txqueuelen:1000
RX bytes:4091201649 (3901.6 MB) TX bytes:377502797 (360.0 MB)
Base address:0x2800 Memory:fe880000-fe8a0000
```

2. As the root user, verify the network configuration by using the ping command to test the connection from each node in your cluster to all the other nodes. For example, as the root user, you might run the following commands on each node:
```
# ping -c 3 racnode1.example.com
# ping -c 3 racnode1
```

ping -c 3 racnode2.example.com
ping -c 3 racnode2

You should not get a response from the nodes using the ping command for the virtual IPs (racnode1-vip, racnode2-vip) or the SCAN IPs until after Oracle Cluster ware is installed and running. If the ping commands for the public
Addresses fail, and then resolve the issue before you proceed.

3. Ensure that you can access the default gateway with a ping command. To identify the default gateway, use the route command, as described in the Oracle Linux Help utility.

VIII.   Oracle Cluster Registry File

Oracle Cluster Registry (OCR) records cluster configuration information. If it fails, the entire cluster environment for Oracle11gRAC will be adversely affected and a possible outage may result if OCR is lost. It store information about Node membership information, Software active version, location of the11g voting disk, Server pools, Status for the cluster resources, Server, Network, Database, Instance, Listener-up/down, Dependencies, Management-policy(automatic/manual), Callout, scripts, Retries, ASM instance and Disk groups, CRS application resource profiles, Database services' characteristics, Details of the network interfaces, Information about OCR backups.

OCR during cluster set-up -to update the status of servers, CSS during node addition/deletion –to add/delete node names, CRSd about status of nodes during failure/reconfiguration, OUI, SRVCTL (used to manage clusters and RAC databases/instance), Cluster control utility –CRSCTL (to manage cluster/local resources), Enterprise Manager (EM), Database Configuration assistant (DBCA),Database Upgrade Assistant (DBUA), Network Configuration Assistant (NETCA) and ASM Configuration Assistant (ASMCA).

Purpose of OCR:-Oracle Clusterware reads the ocr.loc file for the location of the registry and to determine which applications resources need to be started and the nodes on which to start them, maintains and tracks information pertaining to the definition, availability and current state of the services, implements the workload balancing and continuous availability features of services, generates events during cluster state changes.

OCR Backup:-Oracle Clusterware11g Release2 backs up the OCR automatically every four hours on a schedule that is depend when the node started (not clock time). OCR backups are made to the GRID_HOME/cdata/<clustername>directory on the node performing the backups.

It is recommended that OCR backups may be placed on a shared location which can be configured during ocr config-backuploc<newlocation>command.
Oracle Cluster ware maintains the last three backups,over writing the older backups.Thus, you will have 24-hour backups, the current one, one four hours old and one eight hour sold.
#ocrconfig–showbackupauto
#ocrconfig–manualbackup
#ocrconfig–export<filename>
#ocrconfig–backuploc/u01/app/oracle/ocrloc

IX. Voting Disk File

➢   Contains information about cluster membership
➢   Used by CRS to avoid 'split-brain' scenarios if any node loses contact over the interconnect
➢   Mandatory to be located on shared storage.In Oracle 11gr2 it can be placed on ASM disks .
➢   Typically about 280MB in size
➢   CSSD process in each RAC node maintains its heart beat in a block of size 1 OS block, in the hot block of voting disk at a specific offset.
➢   Each node reads its kill block once per second, if the kill block is overwritten node commits suicide.
➢   Voting disks contain static and dynamic data.
➢   Static data : Info about nodes in the cluster
➢   Dynamic data : Disk heartbeat logging
➢   It maintains and consists of important details about the cluster nodes membership, such as
➢   which node is part of the cluster
➢   who (node) is joining the cluster, and
➢   who (node) is leaving the cluster.

➢ Contains information about cluster membership

We have an odd number of voting disks. each node should be able to access more than half the number of voting disks. A node not able to do so will have to be evicted from the cluster by another node that has more than half the voting disks, to maintain the integrity of the cluster .

The voting disk is not striped but put as a whole on ASM Disks. In the event that the disk containing the voting disk fails, Oracle ASM will choose another disk on which to store this data. Maximum 33 Voting disk in oracle 11gr2. Location of voting disk crsctl query css vote disk. Logs generated on location $ORACLE_HOME/log//cssd
Backup voting disk using DD command in linux/AIX and ocopy for Windows Configuration of various resources which need to be started on the node.

X. Oracle Local Registry
The Oracle Local Registry (OLR) is installed on each node in the cluster. The OLR is a local registry for node specific resources. The OLR is not shared by other nodes in the cluster.

Purpose of OLR:-It is the very first file that is accessed to start-up cluster-ware when OCR is stored on ASM.OCR should be accessible to find out there sources which need to be started on a node.

OCR is on ASM, it can't be read until ASM is up. To resolve this problem, information about their sources which need to best stored in an operating system file which is called Oracle Local Registry or OLR. When a node joins the cluster, OLR on that node is read, various resources, including ASM are started on the node. If OLR is missing or corrupted, cluster-ware can't be started on that node.

OLR Location:-The OLR file located in grid_home/cdata/<hostname>.olr . The location of OLR is stored in /etc/oracle/olr.loc. and used by OHASD.

XI. Conclusion
In this paper , we discussed the performance of load balance monitoring of Oracle RAC using will take an overview of load balancing, automatic failover and load balancing for Oracle real Application Clusters, performance monitoring. Oracle RAC system is a truly new approach based on which we can monitor as well as forecast the behavior of its load balancing act. This research work is performed on a two node RAC database, this can be extended more to incorporate in finding relative entropy based monitoring of load balance in multiple node Oracle RAC database.

References

[1].   International Journal of Computer Applications (0975 – 8887) Volume 100– No.1, August 2014 -Load Balance Monitoring in Oracle RAC Neha Chandrima, Sunil Phulre, Vineet Richhariya.
[2].   International Journal of Scientific & Engineering Research Volume 2, Issue 6, June-2011 1 ISSN 2229-5518   IJSER © 2011Oracle Real Application Clusters Deepali Kadam, Nandan Bhalwarkar, Rahul Neware, Rajesh Sapkale, Raunika Lamge.
[3].   docs.oracle.com/cd/E11882_01/rac.112/e41960/admcon.htm
[4].   R. Bianchini, L. I. Kontothanassis, R. Pinto, M. De Maria, M. Abud, C.L. Amorim. Hiding Communication Latency and Coherence Overhead in Software DSMs. Proc. 7th International Conference on Architectural Support for Programming Languages and Operating Systems (ASPLOS), October 1996.