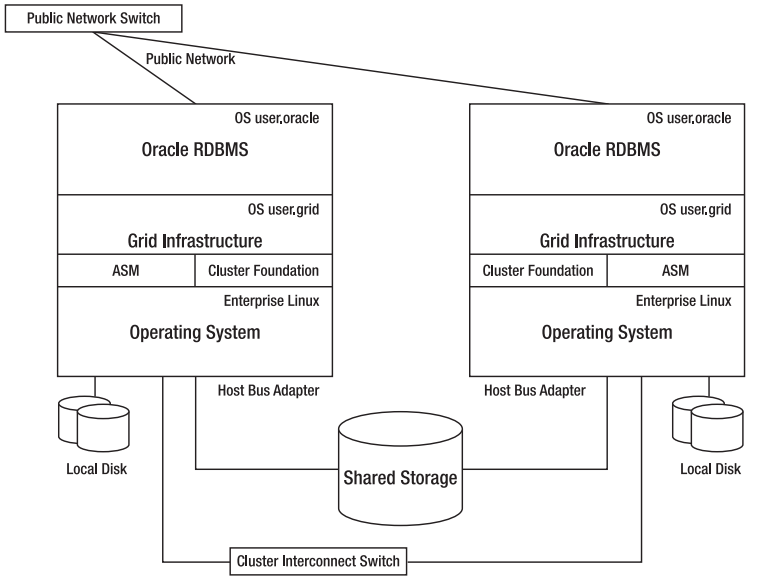
Oracle Real Applications Clusters 23ai (RAC)

# Oracle Rac Architecture Summary

## Architecture Summary



Oracle Rac include:

* Oracle Clusterware: setup infrastructure to be a cluster
* Oracle ASM: shared data storage
* Oracle RAC: shared data processing mechanism (global cache – fusion cache)

## Oracle Clusterware

### Oracle Clusterware concept

Cluster**:**

* A group of independent, but interconnected, computers that act as a single system
* Usually deployed to increase availability and performance or to balance a dynamically changing workload

What Is Clusterware:

Software that provides various interfaces and services for a cluster. Typically, this includes capabilities that:

* Allow the cluster to be managed as a whole (single entity)
* Protect the integrity of the cluster so that data is protected and the cluster continues to function even if communication with a cluster node is severed
* Maintain a registry of resources so that their location is known across the cluster and so that dependencies between resources is maintained
* Deal with changes to the cluster such as node additions, removals, or failures
* Provide a view of resources such as network addresses and files in a file system

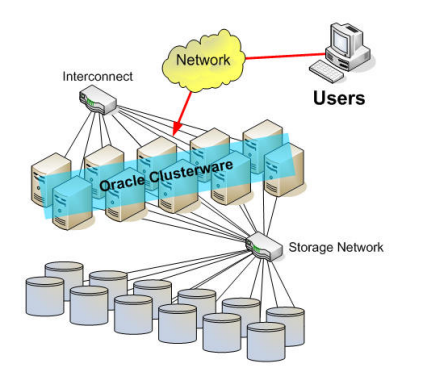
**Oracle Clusterware is:**

* A key part of Oracle Grid Infrastructure
* Integrated with Oracle Automatic Storage Management (ASM)
* The basis for ASM Cluster File System (ACFS)
* A foundation for Oracle Real Application Clusters (RAC)
* A generalized cluster infrastructure for all kinds of applications

**Oracle Clusterware Networking**

* Each node must have at least two network adapters.
* Each public network adapter must support TCP/IP.
* The interconnect adapter must support:
  + User Datagram Protocol (UDP) or Reliable Data Socket (RDS) for UNIX and Linux for database communication
  + TCP for Windows platforms for database communication
* All platforms use Grid Interprocess Communication (GIPc)

**Oracle Clusterware Architecture and Services**



Shared disk cluster architecture supporting application load balancing and failover

Services include:

* Cluster Management, which allows cluster services and application resources to be monitored and managed from any node in the cluster
* Node Monitoring, which provides real-time information regarding which nodes are currently available and the resources they support. Cluster integrity is also protected by evicting or fencing unresponsive nodes.
* Event Services, which publishes cluster events so that applications are aware of changes in the clusterEvent services
* Time Synchronization, which synchronizes the time on all nodes of the cluster
* Network Management, which provisions and manages Virtual IP (VIP) addresses that are associated with cluster nodes or application resources to provide a consistent network identity regardless of which nodes are available. In addition, Grid Naming Service (GNS) manages network naming within the cluster.
* High Availability, which services, monitors, and restarts all other resources as required

### Oracle Clusterware detail service

Oracle CRS (Cluster Ready Services) và OCR (Oracle Cluster Registry), Voting disk là các thành phần quan trọng trong kiến trúc Oracle Real Application Clusters (RAC).

#### Oracle CRS (Cluster Ready Services)

**Purpose**:

* CRS là một bộ dịch vụ và quản lý tài nguyên cho cụm Oracle RAC.
* CRS đảm bảo rằng các tài nguyên như cơ sở dữ liệu, phiên bản Oracle, dịch vụ, và các thành phần ứng dụng luôn có sẵn và chạy trong cụm.

**Function**:

* Quản lý việc khởi động và tắt máy của các thành phần cụm.
* Phát hiện và phục hồi từ các lỗi của các nút trong cụm.
* Phân phối và quản lý tài nguyên trong cụm.

**Component**:

* crsd: Daemon quản lý tài nguyên và dịch vụ của cụm.
* ocssd: Daemon phát hiện và quản lý cấu hình cụm.
* evmd: Daemon xử lý các sự kiện trong cụm.

#### OCR (Oracle Cluster Registry)

**Purpose**

* OCR là một kho lưu trữ thông tin cấu hình cụm cluster.
* Nó lưu trữ thông tin về cấu hình của tất cả các tài nguyên và dịch vụ được quản lý bởi CRS trong cụm Oracle RAC.

**Function**:

* Lưu trữ thông tin về cấu hình của các tài nguyên và dịch vụ trong cụm.
* OCR được sử dụng bởi CRS để biết cách cấu hình và quản lý các tài nguyên và dịch vụ trong cụm.

**Component**:

* Một tập tin hoặc bộ tập tin được lưu trữ trên đĩa chia sẻ trong cụm.
* Có thể có một bản sao OCR (OCR mirror) để đảm bảo tính sẵn sàng cao và khả năng phục hồi.

#### Voting Disk

**Purpose**

* Voting disk được sử dụng để đảm bảo rằng chỉ các nút trong cụm còn lại ít nhất một nửa (cộng một) nút có thể tiếp tục hoạt động trong trường hợp có sự cố xảy ra.
* Voting disk giúp tránh tình trạng "split-brain", nơi mà các nút cụm có thể bị phân chia và hoạt động độc lập, dẫn đến sự không nhất quán trong dữ liệu.

**Function**:

* Voting disk lưu trữ thông tin về các nút hiện tại và trạng thái của chúng trong cụm.
* **Bầu chọn (Voting)**: Trong trường hợp một hoặc nhiều nút trong cụm gặp sự cố, các nút còn lại sẽ "bỏ phiếu" để quyết định xem nút nào sẽ được giữ lại và tiếp tục hoạt động. Điều này giúp ngăn chặn tình trạng "split-brain".
* **Duy trì tính nhất quán**: Voting disk giúp đảm bảo rằng chỉ một tập hợp các nút nhất định có thể truy cập và cập nhật dữ liệu, duy trì tính nhất quán của cơ sở dữ liệu.

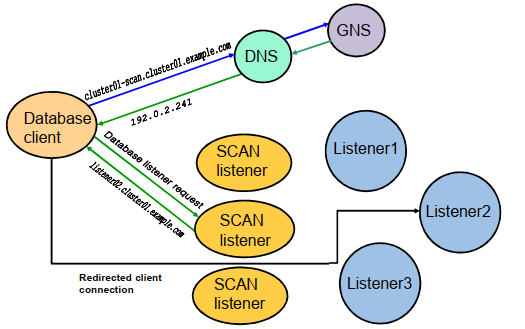
**Component**:

* Voting disk là một tập tin hoặc một tập hợp các tập tin trên đĩa chia sẻ có thể được truy cập bởi tất cả các nút trong cụm.
* Có thể có nhiều bản sao voting disk để đảm bảo độ tin cậy cao. Số lượng voting disk thường là số lẻ (tối thiểu là 3) để tránh tình trạng hòa (tie).

#### Single Client Access Name

The single client access name (SCAN) is the address used by clients connecting to the cluster.

The SCAN provides a stable, highly available name for clients to use, independent of the nodes that make up the cluster.



#### Managing Oracle Clusterware

**Resource Management Options**

Oracle Database provides two styles of resource management for RAC databases:

* Administrator-managed: Database administrators define on which servers a database resource should run, and place resources manually as needed. This is the management strategy used in previous releases:
  + Uses a subpool of the built-in GENERIC server pool
  + Places resources manually as defined by the database administrator
  + Is the management strategy used in previous releases
  + Is good for a small number of nodes
* Policy managed: Database administrators specify in which server pool (excluding generic or free) the database resource will run. Oracle Clusterware is responsible for placing the database resource on a server:
  + Divides servers in a cluster into sets of server pools
  + Controls placement of all services by using server pools
  + Is Workload Management compliant
  + Is best for a large number of nodes

**Server Pools**

Server pools are logical divisions of a cluster into pools of servers or nodes. They:

* Distribute a uniform workload over several servers in the cluster
* Are allocated to host databases or other resources
* Are managed using the crsctl and srvctl commands
* Support parent-child relationships among server pools
  + Top-level pools are mutually exclusive.
* Include two built-in server pools at Oracle Clusterware installation:
  + FREE: For servers that are not assigned to other pools
  + GENERIC: For administrator-managed fixed configuration and pre-11g Release 2 databases

## Oracle Automatic Storage Management (ASM)

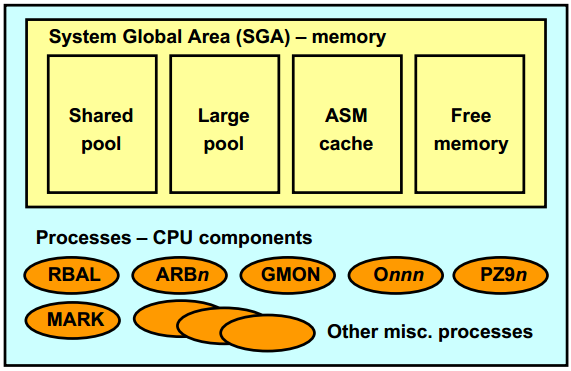
**What Is Oracle ASM?**

* In essence, Oracle Automatic Storage Management (ASM) is a volume manager and file system built into the Oracle Database server.
* Raw disk volumes are allocated to ASM for management and control in the same way that raw volumes are managed by a volume manager.
* ASM is highly integrated with, and highly optimized for, the Oracle Database. It has become the best practice standard for Oracle Database storage.

**ASM Key Features and Benefits**

* Stripes files rather than logical volumes
* Provides redundancy on a file basis
* Enables online disk reconfiguration and dynamic rebalancing
* Reduces the time significantly to resynchronize a transient failure by tracking changes while disk is offline
* Provides adjustable rebalancing speed
* Is cluster-aware
* Supports reading from mirrored copy instead of primary copy for extended clusters
* Is automatically installed as part of the Grid Infrastructure

### ASM Instance

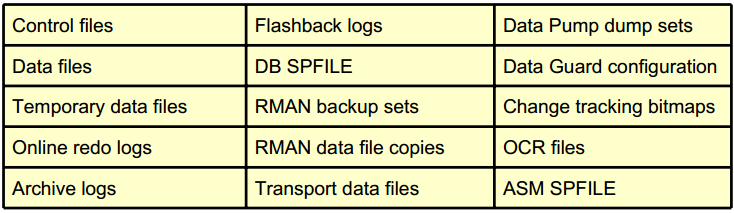


Component:

* **Shared Pool:** Used for metadata information
* **Large Pool:** Used for parallel operations
* **ASM Cache:** Used for reading and writing blocks during rebalance operations
* **Free Memory:** Unallocated memory available
* **Backgroud process:**
  + ARCn: The archiver processes
  + CKPT: The checkpoint process
  + DBWn: The database writer processes
  + DIAG: The diagnosability process
  + Jnnn: Job queue processes
  + LGWR: The log writer process
  + PMON: The process monitor process
  + PSP0: The process spawner process
  + QMNn: The queue monitor processes
  + RECO: The recoverer process
  + SMON: The system monitor process
  + VKTM: The virtual keeper of time process
  + MMAN: The memory manager process

### ASM File

ASM files are a limited set of file types stored in an ASM disk group:



* Are stored as a set or collection of data extents
* Are striped across all disks in a disk group

Use names that begin with a plus sign (+), which are automatically generated or from user-defined aliases

Oracle Managed Files (OMF)

Oracle Managed Files (OMF) là một tính năng trong Oracle Database giúp tự động hóa và đơn giản hóa việc quản lý các tập tin cơ sở dữ liệu. Khi sử dụng OMF, Oracle sẽ tự động tạo, đặt tên, và quản lý các tập tin dữ liệu, tập tin nhật ký (log files), và các tập tin khác mà không cần DBA phải can thiệp trực tiếp.

Purpose:

* **Đơn giản hóa quản lý tập tin**:
  + OMF giúp giảm bớt khối lượng công việc cho DBA bằng cách tự động hóa việc tạo, đổi tên, và xóa các tập tin cơ sở dữ liệu.
  + Tên và đường dẫn của các tập tin được Oracle tự động quản lý, giúp tránh lỗi do con người gây ra.
* **Tự động hóa**:
  + Oracle tự động quản lý không gian lưu trữ cho các tập tin cơ sở dữ liệu, giúp tối ưu hóa và sử dụng hiệu quả không gian đĩa.
* **Tích hợp với các tính năng khác của Oracle**:
  + OMF tích hợp tốt với các tính năng khác của Oracle như Oracle ASM (Automatic Storage Management) để cung cấp một giải pháp quản lý lưu trữ toàn diện.

## Oracle Global Cache (cache fusion)

Cơ chế chia sẻ và sử dụng dữ liệu đồng thời của Oracle Rac

Global cache management provides:

* A concurrency mechanism for multiple buffer caches
* An optimization of block access for reads
* An optimization of writes for dirty buffers
* A mechanism to optimize parallel queries

Global Cache Management Components

* The LMSn processes
* Buffers
* Buffer headers
* Global Cache Master Resources
* Global Cache Shadow Resources

# Oraclc RAC Admin action

crsctl manages clusterware-related operations:

* Starting and stopping Oracle Clusterware
* Enabling and disabling Oracle Clusterware daemons
* Registering cluster resources

srvctl manages Oracle resource–related operations:

* Starting and stopping database instances and services

## Controlling Oracle Clusterware

* When server start, crs is is automatically started by the /etc/init.d/ohasd startup script
* **Check status:**

# crsctl check crs  
CRS-4638: Oracle High Availability Services is online  
CRS-4537: Cluster Ready Services is online  
CRS-4529: Cluster Synchronization Services is online  
CRS-4533: Event Manager is online

# crsctl check cluster

* **Start /stop – enable/disable ở 1 node**

# crsctl start crs  
# crsctl stop crs

# crsctl enable crs  
# crsctl disable crs

* **To determine the location of the voting disk:**

# crsctl query css votedisk

## STATE File Universal Id File Name Disk group  
-- ----- ----------------- ---------- ----------  
1. ONLINE 8c2e45d734c64f8abf9f136990f3daf8 (ASMDISK01) [DATA]  
2. ONLINE 99bc153df3b84fb4bf071d916089fd4a (ASMDISK02) [DATA]  
3. ONLINE 0b090b6b19154fc1bf5913bc70340921 (ASMDISK03) [DATA]  
Located 3 voting disk(s).

* **To determine the location of the OCR:**

$ cat /etc/oracle/ocr.loc

ocrconfig\_loc=+DATA

local\_only=FALSE

$ ocrcheck

Status of Oracle Cluster Registry is as follows :

Version : 3

Total space (kbytes) : 262120

Used space (kbytes) : 3856

Available space (kbytes) : 258264

ID : 12663931

Device/File Name : +OCR\_VOTE

Device/File integrity check succeeded

Cluster registry integrity check succeeded

Logical corruption check bypassed due to non-privileged user

* **Check the ocssd.log for voting disks issues.**

$ grep voting <grid\_home>/log/<hostname>/cssd/ocssd.log

* **To add or delete one or more voting disks to non-ASM storage:**

# crsctl add css votedisk path\_to\_new\_voting\_disk

# crsctl delete css votedisk path\_to\_old\_voting\_disk

* **To add a voting disk to ASM:**

# crsctl replace votedisk +asm\_disk\_group

* **To migrate voting disks from non-ASM storage devices to ASM or vice versa, specify the ASM disk group name or path to the non-ASM storage device:**

# crsctl replace votedisk {+asm\_disk\_group | path\_to\_voting\_disk}

* **Locating the OCR Automatic Backups**
  + The OCR is backed up automatically.
  + Only one node performs the backup.
  + To determine the node and location of the backup:

$ ocrconfig -showbackup auto

ttdl-coredb1a 2015/11/24 15:12:50 /u01/app/11.2.0/grid/cdata/ttdl-coredb1/backup00.ocr

ttdl-coredb1a 2015/11/23 15:12:46 /u01/app/11.2.0/grid/cdata/ttdl-coredb1/day.ocr

ttdl-coredb1b 2015/11/13 14:56:59 /u01/app/11.2.0/grid/cdata/ttdl-coredb1/week.ocr

* + Files could be spread across nodes due to outages.
  + The backup frequency and retention policies are:
    - Every four hours: CRS keeps the last three copies.
    - At the end of every day: CRS keeps the last two copies.
    - At the end of every week: CRS keeps the last two copies.
  + Default directory:

[grid@...#](mailto:grid@...) $ORACLE\_HOME/cdata/<cluster name>

* **Adding, Replacing, and Repairing OCR Locations**
  + Add an OCR location to either ASM or other storage device:

# ocrconfig -add +DATA2

# ocrconfig -add /dev/sde1

* + To replace the current OCR location:

# ocrconfig -replace /dev/sde1 -replacement +DATA2

* + To repair OCR configuration, run this command on the node on which you have stopped Oracle Clusterware:

# ocrconfig -replace /dev/sde1 -replacement +DATA2

* **Removing an Oracle Cluster Registry Location**

# ocrconfig -delete +DATA2  
# ocrconfig -delete /dev/sde1

* **Migrating OCR Locations to ASM**
  + Ensure that Oracle Clusterware is upgraded to 11g Release 2

$ crsctl query crs activeversion  
Oracle Clusterware active version on cluster is [11.2.0.1.0]

* + Start ASM on all nodes and create a disk group that has at least 1 GB of space and has at least normal redundancy.
  + To add an OCR location to an ASM disk group, run the following command as root:

*#* ocrconfig -add +DATA2

* + To remove storage configurations no longer in use, run the following command as root:

# ocrconfig -delete /dev/raw/raw1

* **To perform a backup:**
  + Physical backup

# ocrconfig -manualbackup

* + To display a list of manual backups:

$ ocrconfig –showbackup manual

host02 2009/07/28 16:59:17

/u01/app/.../cdata/cluster01/backup\_20090728\_165917.ocr

* + To perform a logical backup:

# ocrconfig -export /home/oracle/ocr.backup

* **Recovering the OCR by Using Physical Backups**
  + Locate a physical backup:

$ ocrconfig –showbackup

* + Stop CRS on all nodes:

# crsctl stop crs -f

* + Restore the physical or logical OCR backup:

# ocrconfig –restore /u01/app/.../cdata/cluster01/day.ocr

or

# ocrconfig –import /shared/export/ocrback.dmp

* + Restart Oracle High Availability Services on all nodes:

# crsctl start crs

* + Check status

# ocrcheck

# RAC Database Monitoring and Tuning

## Most Common RAC Tuning Tips

Application tuning is often the most beneficial!

* Reduce long full-table scans in OLTP systems.
* Use Automatic Segment Space Management (ASSM).
* Increase sequence caches.
* Use partitioning to reduce interinstance traffic.
* Avoid unnecessary parsing.
* Minimize locking usage.
* Remove unselective indexes.
* Configure interconnect properly.