

11gR2 Grid Infrastructure

What is it ?



Sridhar Avantsa, Associate Practice Director
Rolta TUSC Infrastructure Services

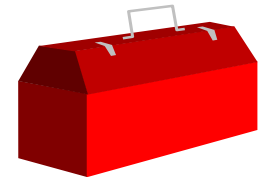
Copyright © 2010 Rolta International, Inc., All Rights Reserved

Introductions - About Me.

- ◆ Sridhar Avantsa
 - Associate Practice Director, Infrastructure Services
 - Manage the consulting arm of the practice.
 - Sridhar.avantsa@roltasolutions.com
- ◆ Working in IT with Oracle for much longer than I want to admit. 😊
 - First version of Oracle database I worked on was version 5
 - Forms 2.3 was still active.
- ◆ DBA, Systems and Solution Architect, Developer, tinkerer
 - Exadata

Introductions – About the audience

- ◆ Oracle6 and Earlier Experience?
- ◆ Oracle7 Experience ?
- ◆ Oracle8i Experience ?
- ◆ Oracle9i Experience ?
- ◆ Oracle10g Experience?
- ◆ Oracle Database 11g Experience?
- ◆ Oracle Database 11g R2 Experience?
- ◆ Goals
 - Introduce Grid Infrastructure and its concepts.
- ◆ Non-Goals
 - Become a 11gR2 master.



V\$ Views over the years

Version	V\$ Views	X\$ Tables
6	23	? (35)
7	72	126
8.0	132	200
8.1	185	271
9.0	227	352
9.2	259	394
10.1.0.2	340 (+31%)	543 (+38%)
10.2.0.1	396	613
11.1.0.6.0	484 (+23%)	798 (+30%)
11.2.0.1.0	496 (+25%)	945 (+54%)



11GR2 GRID INFRASTRUCTURE INTRODUCTION

Prior to 11gR2

- ◆ Oracle RAC configuration looked like (Best Practices):
 - One Oracle Home for the cluster ware.
 - One Oracle Home for the ASM instance
 - Oracle Home for the database itself, could be more than one based on version etc.
- ◆ Single Instance with ASM looked like (Best Practice):
 - One Oracle Home for the ASM instance.
 - Oracle Home for the database itself, could be more than one based on version etc.
- ◆ But:
 - ASM and clusterware are tied together at the hip, RAC or not.
 - The ASM Oracle Home was basically a DB Oracle Home, used to run the ASM instance alone.
 - ASM disk groups are managed solely by the ASM instance, practically synonymous.
 - ASM is a common, shared stack used by all databases on that server.
 - Management procedures varied between RAC and Non-RAC installations

With 11gR2 Grid Infrastructure

- ◆ The Grid infrastructure concept is truly about providing a unified tech stack for the grid.
- ◆ Applies to RAC, RAC one node and Non RAC.
- ◆ Loosely coupled dependency between ASM Disk groups and the ASM instance
- ◆ Integrates and Unifies Oracle clusterware, ASM disk groups and Oracle ASM into a single stack.
- ◆ A common interface to administer and manage the Grid layer, RAC or not.

Grid Infrastructure – Non RAC Mode

- ◆ Technically not mandatory, but:
- ◆ Required If the database storage will use ASM.
- ◆ Required for using Oracle AVM / ACFS functionality.
- ◆ Runs the background daemons required to support the ASM stack.
- ◆ Required for Oracle ReStart functionality.

11GR2 GRID INFRASTRUCTURE - NON CLUSTERED MODE

Terminologies

- ◆ HASD – Oracle High Availability Service Daemon. The “big dog” amongst the cluster ware daemons.
- ◆ CSSD – Oracle Cluster Synchronization Services Daemon.
- ◆ OCR – Oracle Cluster Registry. A registry of all the resources / processes under the management of the cluster.

Installation Overview

- ◆ Need to install the Grid Infrastructure Software component (before installing DB Software)
- ◆ Pre configuration requirements:
 - If using ASM, then make sure ASM disks have been created and assigned.
 - Create required users, groups and directories for Inventory / Home etc.
 - Set up kernel parameters / swap etc to meet minimum requirements.
 - Ensure we have X Windows requirement taken care.
- ◆ Grid Infrastructure software will have its own ORACLE_HOME.
 - Must not be placed under the Oracle base directory or the users home directory
 - Why?, Security & permissions reasons. Grid Home path ownership is changes to root.
 - If separation of duty is required, Grid Infrastructure owner can be de
 - Example:
 - ORACLE_BASE=/u01/app/oracle,
 - GRID_HOME=/u01/app/11.2.0/grid
 - INVLOC_PTR=\$ORACLE_BASE/orainventory;
 - DB_HOME==\$ORACLE_BASE/product/11.2.0/db
- ◆ Run thru runInstaller screens, about 15 or of them, Choose the non clustered option
- ◆ If we will not use the ASM option, choose the “install software only” option and configure later I would ask why not ? 😊.
- ◆ If we will use ASM then choose the install & Configure option.
 - You will be prompted to create at least one DISK GROUP.

Configuration Overview

- ◆ Uses the minimal required pieces of clustering software:
 - Still requires CSS to provide / manage synchronization between DB & ASM.
 - The root level HAS daemon is started via inittab
 - Still uses an OCR file to store configuration, but it is on local storage and run in “LOCAL” mode.
 - ASM instance SPFILE is stored in the Disk Group created during installation.
- ◆ The question is Why ?
 - Use the OCR to track resource dependency hierarchy and startup configuration.
 - A command interface similar in concept as use in a RAC environment.
 - Standardize on ASM meeting all Oracle Database Needs (ASM SPFILE etc).

Configuration Example

```
[oracle]$ cat /etc/oracle/ocr.loc  
ocrconfig_loc=/u01/app/11.2.0/grid/cdata/localhost/local.ocr  
local_only=TRUE
```

```
[oracle]$ crsctl check has  
CRS-4638: Oracle High Availability Services is online
```

```
[oracle]$ crsctl check css  
CRS-4529: Cluster Synchronization Services is online
```

```
[oracle]$ asmcmd ls -l +DATA/ASM/ASMPARAMETERFILE  
Type                Redund  Striped  Time                Sys  Name  
ASMPARAMETERFILE  UNPROT  COARSE  JAN 12             2011 Y  
REGISTRY.253.740230595
```

```
[oracle@s111gr2 bin]$ crs_stat -t  
Name                Type                Target             State             Host  
-----  
ora.DATA.dg         ora....up.type     ONLINE            ONLINE            s111gr2  
ora....ER.lsnr      ora....er.type     ONLINE            ONLINE            s111gr2  
ora.asm              ora.asm.type       ONLINE            ONLINE            s111gr2  
ora.cssd             ora.cssd.type      ONLINE            ONLINE            s111gr2  
ora.diskmon          ora....on.type     ONLINE            ONLINE            s111gr2  
ora.evmd             ora.evm.type       ONLINE            ONLINE            s111gr2  
ora.ons              ora.ons.type       OFFLINE           OFFLINE           s111gr2  
ora.s111gr2.db       ora....se.type     ONLINE            ONLINE            s111gr2
```

```
[oracle@s111gr2 bin]$ srvctl config asm -a  
ASM home: /u01/app/oracle/product/11.2.0/grid  
ASM listener: LISTENER  
Spfile: +DATA/asm/asmparameterfile/registry.253.740230595  
ASM diskgroup discovery string:  
ASM is enabled.  
[oracle@s111gr2 bin]$
```

A closer Look @ resources

HA Resource	Type	Target	State
-----	----	-----	-----
ora.DATA.dg	ora.diskgroup.type	ONLINE	ONLINE on s11gr2
ora.LISTENER.lsnr	ora.listener.type	ONLINE	ONLINE on s11gr2
ora.asm	ora.asm.type	ONLINE	ONLINE on s11gr2
ora.cssd	ora.cssd.type	ONLINE	ONLINE on s11gr2
ora.diskmon	ora.diskmon.type	ONLINE	ONLINE on s11gr2
ora.evmd	ora.evmd.type	ONLINE	ONLINE on s11gr2
ora.ons	ora.ons.type	OFFLINE	OFFLINE
ora.s11gr2.db	ora.database.type	ONLINE	ONLINE on s11gr2

What do we see:

- ♦ Resource Type of Diskgroup (Data Guard), separate from the ASM instance.
- ♦ Resource type of DISKMON /CSSd / EVMd & ONSd.
- ♦ Resources to cover database / Listeners etc.

And this means:

- ♦ The ASM diskgroup concept separated from ASM.
- ♦ The ASM Diskgroup can start up before ASM, that is ASM can read the SPFILE.
- ♦ Oracle Restart will use this information to activate Oracle Components

11GR2 GRID INFRASTRUCTURE – CLUSTERED MODE

Terminologies

- ◆ Already Covered earlier – HASD/ CSSD / OCR
- ◆ CRS
- ◆ VOTING DISKS – Quorum disks to help maintain cluster membership status and resolve split brain.
- ◆ Public IP – Public IP Address for the server.
- ◆ Virtual IP – A publicly routable floating IP address and name for each node.
- ◆ Private IP – IP Address for use @ clustering layer.
- ◆ SCAN – Single Client Access Name.

Installation Requirements

- ◆ Requirements covered Grid Infrastructure – Non Clustered install plus:
 - Ensure ssh equivalency requirements for the “oracle” user are met.
 - Specific Networking Requirements
 - Storage Specific Requirements
- ◆ Additional Storage from a database perspective
 - Recommended approach would be to use ASM.
 - Separate Disk Groups for database from the disk group used during clusterware installation.

Installation Requirements – Networking

- ◆ Public IP Address – managed and maintained by OS.
 - One per server in cluster. Resolved via DNS normally.
- ◆ Private IP Address - managed and maintained by OS.
 - One per server in cluster. Resolved via /etc/hosts most often.
 - Routable only within the cluster, since used by the clustering layer.
- ◆ Virtual IP Address (VIP) - Managed and maintained by Oracle HAS/CRS
 - One per server in the cluster. Resolved by DNS.
 - Publicly routable IP Address.
 - Ability to float between nodes.
- ◆ SCAN IP Address – managed and maintained by Oracle HAS/CRS.
 - One per cluster, SCAN Name matches the cluster name.
 - Resolved by DNS, but resolves to multiple IP addresses in a round robin fashion.
- ◆ Recommend:
 - Using NIC bonding in “Active Passive” mode to tolerate NIC failure. Especially for Private Interconnect.
 - Recommend the review and set up of Jumbo frames on the private interconnect, especially for large / busy databases.

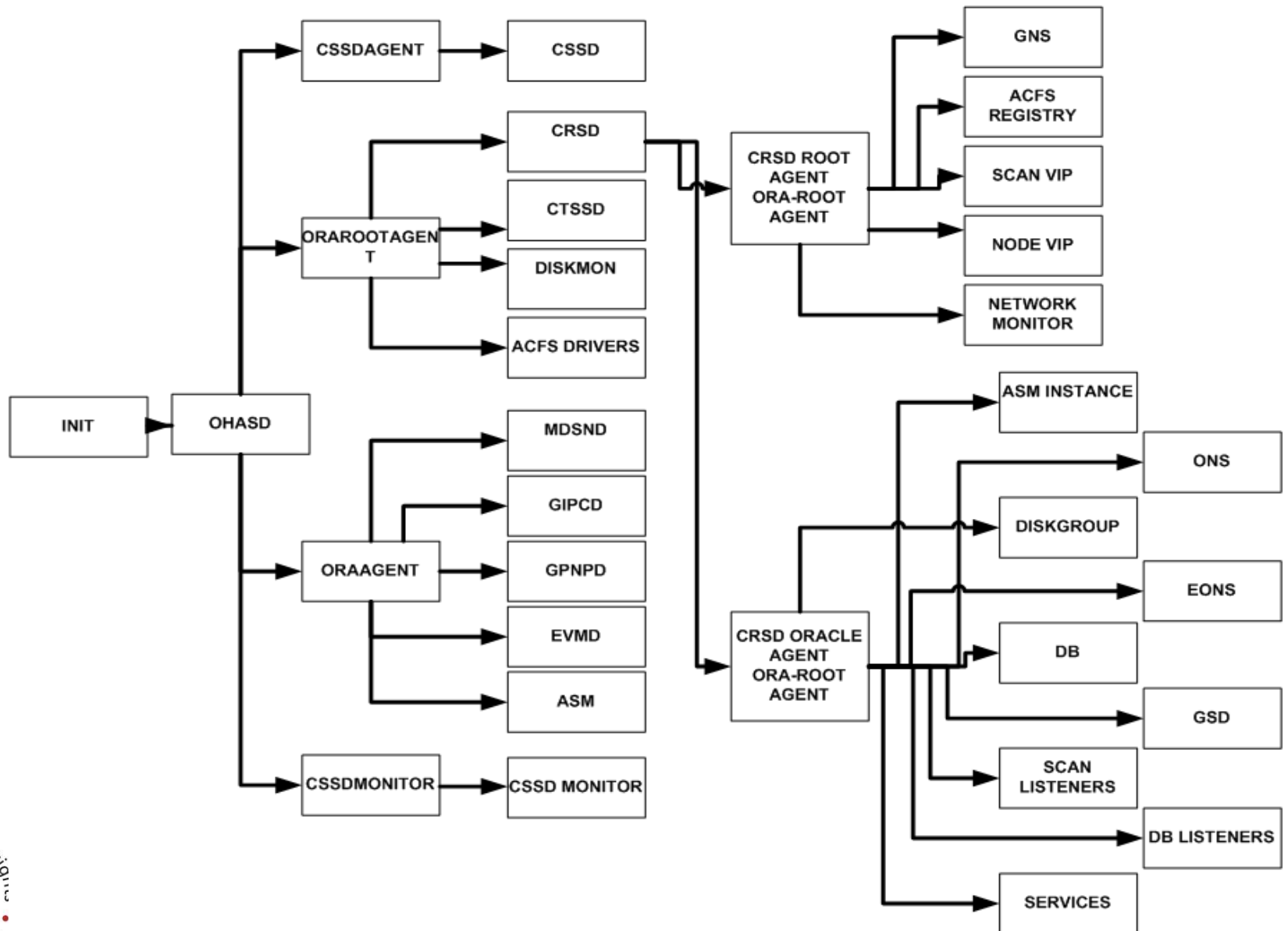
Installation Requirements – Storage Layer.

- ◆ With 11gR2, ASM disk group is where
 - OCR file will be stored.
 - Voting Disk Information is maintained.
 - ASM Instance SPFILE will be stored.
- ◆ The Disk group should be:
 - Maintained in at least normal redundancy or even high redundancy.
 - Should have at least 3 LUNS that comprise the disk group.
- ◆ Therefore, we recommend :
 - Creating disk group specifically for use by the cluster.
 - Name disk group as [CLUSTER_NAME]_DG.
 - Disk group does not need to be very large, use disks of 1 – 5 GB at most.

Installation Overview

- ◆ Oracle Grid Infrastructure software home requirements from the earlier slides applies.
- ◆ Run thru runInstaller screens, about 15 or of them, Choose the “install and Configure for a cluster” option
- ◆ In most cases, use the TYPICAL install.
 - Use Advance INSTALL for additional storage alternatives or to use of Oracle GNSs to manage the SCAN. I would have ask why? 😊.
- ◆ Remember SCAN NAME = CLUSTER NAME.
- ◆ Include all nodes in the cluster when prompted for nodes.

Clustering Daemons/Processes



RAC - Grid Install Components

- ◆ Grid Home mandatory for a RAC environment.
- ◆ What all runs from this ORACLE_HOME
 - Everything that runs in a non clustered environmental plus.
 - ALL CRS daemons EVMD / CSSD / CRSD / CTSSD etc.
- ◆ OCR & VOTING DISKS also now can be on ASM

Grid CRS – OCR File

- ◆ CRS OCR File can now be on a ASM Disk Group
 - Disk groups are brought online before ASM Instances are online
 - Raw Devices or Clustered File Systems not needed anymore

◆ Example:

```
[oracle@racnode1]$ cat /etc/oracle/ocr.loc
```

```
ocrconfig_loc=+DG01
```

```
local_only=FALSE
```

```
[oracle@racnode2]$ cat /etc/oracle/ocr.loc
```

```
ocrconfig_loc=+DG01
```

```
local_only=FALSE
```

```
[oracle@racnode1]$ asmcmd ls -l + DG01/racnode-  
cluster/OCRFILE/*
```

Type	Redund	Striped	Time	Sys	Name
OCRFILE	MIRROR	COARSE	NOV 09 14:00:00	Y	none =>
			REGISTRY.255.702483503		

Grid CRS – VOTING Disks

- ◆ CRS VOTING Disks now can utilize ASM as well.
 - Seems to work as the Disk Group Disk Level itself.
- ◆ Example:

```
[oracle@racnode1]$ crsctl query css votedisk
## STATE File Universal Id File Name Disk group
--  -----
1. ONLINE 90b0b490373e4f19bf640bdc23a7c043 (ORCL:DG01_DISK_01)
   [DG01]
2. ONLINE 56aa7975c8864f63bf3b645ab18818fa (ORCL:DG01_DISK_02)
   [DG01]
3. ONLINE 9229731611a94fccbf921b4aa8ef552b (ORCL:DG01_DISK_03)
   [DG01]
```

From SQL Plus connected to the ASM Instance

```
1* select g.name group_name, d.name disk_name,
       d.path disk_path, d.VOTING_FILE
       from v$asm_disk d, v$asm_diskgroup g
       where g.GROUP_NUMBER=g.GROUP_NUMBER

SQL> /
GROUP_NAME DISK_NAME DISK_PATH V
-----
DG01 DG01_DISK_01 ORCL:DG01_DISK_01 Y
DG01 DG01_DISK_02 ORCL:DG01_DISK_02 Y
DG01 DG01_DISK_03 ORCL:DG01_DISK_03 Y
```

CRS Resources Example

```
[oracle@racnode1 bin]$ crs-stat.sh
```

HA Resource	Type	Target	State
ora.ACFS.dg	ora.diskgroup.type	ONLINE	ONLINE on racnode1
ora.DATA.dg	ora.diskgroup.type	ONLINE	ONLINE on racnode1
ora.FRA.dg	ora.diskgroup.type	ONLINE	ONLINE on racnode1
ora.LISTENER.lsnr	ora.listener.type	ONLINE	ONLINE on racnode1
ora.LISTENER_SCAN1.lsnr	ora.scan_listener.type	ONLINE	ONLINE on racnode1
ora.asm	ora.asm.type	ONLINE	ONLINE on racnode1
ora.cvu	ora.cvu.type	ONLINE	ONLINE on racnode1
ora.gsd	ora.gsd.type	OFFLINE	OFFLINE
ora.net1.network	ora.network.type	ONLINE	ONLINE on racnode1
ora.oc4j	ora.oc4j.type	ONLINE	ONLINE on racnode1
ora.ons	ora.ons.type	ONLINE	ONLINE on racnode1
ora.rac11gr2.db	ora.database.type	ONLINE	ONLINE on racnode1
ora.rac11gr2.svcinst1.svc	ora.service.type	ONLINE	ONLINE on racnode1
ora.rac11gr2.svcinst2.svc	ora.service.type	ONLINE	ONLINE on racnode1
ora.racnode1.ASM1.asm	application	ONLINE	ONLINE on racnode1
ora.racnode1.LISTENER_RACNODE1.lsnr	application	ONLINE	ONLINE on racnode1
ora.racnode1.gsd	application	OFFLINE	OFFLINE
ora.racnode1.ons	application	ONLINE	ONLINE on racnode1
ora.racnode1.vip	ora.cluster_vip_net1.type	ONLINE	ONLINE on racnode1
ora.racnode2.ASM2.asm	application	ONLINE	OFFLINE
ora.racnode2.LISTENER_RACNODE2.lsnr	application	ONLINE	OFFLINE
ora.racnode2.gsd	application	OFFLINE	OFFLINE
ora.racnode2.ons	application	ONLINE	OFFLINE
ora.racnode2.vip	ora.cluster_vip_net1.type	ONLINE	ONLINE on racnode1
ora.registry.acfs	ora.registry.acfs.type	ONLINE	ONLINE on racnode1
ora.scan1.vip	ora.scan_vip.type	ONLINE	ONLINE on racnode1

11GR2 GRID INFRASTRUCTURE - SCAN

What is SCAN ?

- ◆ Single Client Address Name.
 - Access the entire RAC cluster by a single name.
 - No need for client to know about all nodes in the cluster

- ◆ How does CRS configure SCAN
 - DNS resolves SCAN Name into multiple IP's, in a round robin fashion.
 - NOTE: Oracle also provides an option where the scan IP addresses can be acquired from a DHCP server
 - CRS will take all the IP's and bring up listeners on these IP's.
 - The listeners are distributed across the nodes.
 - In the event of a node loss, the listeners move as well.
 - SCAN listener is independent and different from the local / regular database listener.

SCAN Configuration – DNS Example

```
[oracle]$ nslookup XXXracq-scan
```

```
Server:      10.10.10.3
```

```
Address:     10.10.10.3#53
```

```
Name: XXXracq-scan.network.int
```

```
Address: 10.1.0.23
```

```
Name: XXXracq-scan.network.int
```

```
Address: 10.1.0.22
```

```
Name: XXXracq-scan.network.int
```

```
Address: 10.1.0.24
```

SCAN Configuration – Clusterware Side

```
[oracle]$ srvctl config scan_listener
SCAN Listener LISTENER_SCAN1 exists. Port: TCP:1521
SCAN Listener LISTENER_SCAN2 exists. Port: TCP:1521
SCAN Listener LISTENER_SCAN3 exists. Port: TCP:1521
[oracle@]$ srvctl config scan
SCAN name: XXXracq-scan, Network: 1/10.1.0.0/255.255.255.0/bond0
SCAN VIP name: scan1, IP: /XXXracq-scan.network.int/10.1.0.23
SCAN VIP name: scan2, IP: /XXXracq-scan.network.int/10.1.0.22
SCAN VIP name: scan3, IP: /XXXracq-scan.network.int/10.1.0.24
```

- ◆ We see the 3 IP Addresses and that there are 3 scan listeners configured, one for each IP.
- ◆ Bit where are they running ?

SCAN Configuration – Clusterware Side

```
[oracle ~]$ srvctl status scan
SCAN VIP scan1 is enabled
SCAN VIP scan1 is running on node XXXdbs05
SCAN VIP scan2 is enabled
SCAN VIP scan2 is running on node XXXdbs06
SCAN VIP scan3 is enabled
SCAN VIP scan3 is running on node XXXdbs05
```

```
[oracle~]$ srvctl status scan_listener
SCAN Listener LISTENER_SCAN1 is enabled
SCAN listener LISTENER_SCAN1 is running on node XXXdbs05
SCAN Listener LISTENER_SCAN2 is enabled
SCAN listener LISTENER_SCAN2 is running on node XXXdbs06
SCAN Listener LISTENER_SCAN3 is enabled
SCAN listener LISTENER_SCAN3 is running on node XXXdbs05
```

- ◆ This is a 2 node cluster.
- ◆ 2 SCAN Listeners on one node, 2 on another.

SCAN & Load Balancing

- ◆ Database have “local_listener”, PMON register instance with the node specific SQL Net Listener.
- ◆ In RAC, there is the additional concept of “remote_listener”.
 - PMON for all instances register their services and instance information with the listeners on the other nodes.
 - PMON also sends load advisory information to the remote listener.
 - This “cross” registration allows the listeners to redirect DB connections to a less utilized load or a node servicing the requested service.
- ◆ With 11gR2, the remote registration happens against the SCAN listeners.

SCAN and DB Config:

- ◆ Remote_listener does not use a TNS alias, rather set to [SCAN_NAME:SCAN_PORT].
- ◆ Local listener can use a TNS alias or qualify it fully.

```
SYS@rac11gr21> show parameter listener
```

NAME	TYPE	VALUE
local_listener	string	(DESCRIPTION= (ADDRESS_LIST= (ADDRESS= (PROTOCOL=TCP) (HOST=192.168.116.40) (PORT=1521))))
remote_listener	string	racnode-cluster:1521

SCAN Remote Registration Example

- ◆ Set environment to GRID Infrastructure Oracle Home.
- ◆ Use “lsnrctl status SCAN LISTENER NAME” to get information.
- ◆ We see that the scan listener knows about all the nodes in the cluster.

```
[oracle@racnode1 admin]$ lsnrctl status LISTENER_SCAN1
LSNRCTL for Linux: Version 11.2.0.2.0 - Production on 18-JAN-2012 15:57:40
Copyright (c) 1991, 2010, Oracle. All rights reserved.
Connecting to (DESCRIPTION=(ADDRESS=(PROTOCOL=IPC) (KEY=LISTENER_SCAN1)))
STATUS of the LISTENER
-----
...
Listener Parameter File    /u01/app/11.2.0/grid/network/admin/listener.ora
Listener Log File         /u01/app/11.2.0/grid/log/diag/tnslsnr/racnode1/listener_scan1/alert/log.xml
Listening Endpoints Summary...
  (DESCRIPTION=(ADDRESS=(PROTOCOL=ipc) (KEY=LISTENER_SCAN1)))
  (DESCRIPTION=(ADDRESS=(PROTOCOL=tcp) (HOST=192.168.116.22) (PORT=1521)))
Services Summary...
Service "rac11gr2.myvm.com" has 2 instance(s).
  Instance "rac11gr21", status READY, has 1 handler(s) for this service...
  Instance "rac11gr22", status READY, has 1 handler(s) for this service...
Service "rac11gr2XDB.myvm.com" has 2 instance(s).
  Instance "rac11gr21", status READY, has 1 handler(s) for this service...
  Instance "rac11gr22", status READY, has 1 handler(s) for this service...
Service "svcinst1.myvm.com" has 1 instance(s).
  Instance "rac11gr21", status READY, has 1 handler(s) for this service...
Service "svcinst2.myvm.com" has 1 instance(s).
  Instance "rac11gr21", status READY, has 1 handler(s) for this service...
The command completed successfully
```

So what about the LOCAL Listener

- ◆ Only knows about the resources running on the local node !!!!!!!

```
[oracle@racnode1 admin]$ lsnrctl status LISTENER
LSNRCTL for Linux: Version 11.2.0.2.0 - Production on 18-JAN-2012 16:05:42
Copyright (c) 1991, 2010, Oracle. All rights reserved.
Connecting to (DESCRIPTION=(ADDRESS=(PROTOCOL=IPC)(KEY=LISTENER)))
STATUS of the LISTENER
-----
Alias                LISTENER
-----
Listener Parameter File  /u01/app/11.2.0/grid/network/admin/listener.ora
Listener Log File       /u01/app/oracle/diag/tnlsnr/racnode1/listener/alert/log.xml
Listening Endpoints Summary...
  (DESCRIPTION=(ADDRESS=(PROTOCOL=ipc)(KEY=LISTENER)))
  (DESCRIPTION=(ADDRESS=(PROTOCOL=tcp)(HOST=192.168.116.20)(PORT=1521)))
  (DESCRIPTION=(ADDRESS=(PROTOCOL=tcp)(HOST=192.168.116.40)(PORT=1521)))
Services Summary...
Service "+ASM" has 1 instance(s).
  Instance "+ASM1", status READY, has 1 handler(s) for this service...
Service "rac11gr2.myvm.com" has 1 instance(s).
  Instance "rac11gr21", status READY, has 1 handler(s) for this service...
Service "rac11gr2XDB.myvm.com" has 1 instance(s).
  Instance "rac11gr21", status READY, has 1 handler(s) for this service...
Service "svcinst1.myvm.com" has 1 instance(s).
  Instance "rac11gr21", status READY, has 1 handler(s) for this service...
Service "svcinst2.myvm.com" has 1 instance(s).
  Instance "rac11gr21", status READY, has 1 handler(s) for this service...
The command completed successfully
```

Client Connectivity Using SCAN

- ◆ TNS String on client servers looks like:
- ◆ With SCAN no need of using VIPs anymore to connect.
- ◆ Example:
 - “racnode-cluster” is the SCAN NAME in this case.
 - “rac11gr2.myvm.com” is the service and it can be changed

```
RAC11GR2 =  
  (DESCRIPTION =  
    (ADDRESS = (PROTOCOL = TCP) (HOST = racnode-cluster) (PORT =  
1521))  
    (CONNECT_DATA =  
      (SERVER = DEDICATED)  
      (SERVICE_NAME = rac11gr2.myvm.com)  
    )  
  )
```

SCAN Listener is a SPECIFIC CASE of TNS Listener

- ◆ Go to `$GRID_HOME/network/admin`.
- ◆ Look at the `listener.ora` file and u will see the SCAN LISTENERS defined in there. 😊
- ◆ `[oracle@racnode1 admin]$ more listener.ora`
- ◆ `LISTENER=(DESCRIPTION=(ADDRESS_LIST=(ADDRESS=(PROTOCOL=IPC)(KEY=LISTENER)))) # line added by Agent`
- ◆ `LISTENER_SCAN1=(DESCRIPTION=(ADDRESS_LIST=(ADDRESS=(PROTOCOL=IPC)(KEY=LISTENER_SCAN1)))) # line added by Agent`
- ◆ `ENABLE_GLOBAL_DYNAMIC_ENDPOINT_LISTENER_SCAN1=ON # line added by Agent`
- ◆ `ENABLE_GLOBAL_DYNAMIC_ENDPOINT_LISTENER=ON # line added by Agent`

11GR2 GRID INFRASTRUCTURE - ORACLE ACFS

ACFS - Introduction

- ◆ Stands for “Oracle Automatic Storage Management Cluster File System”.
 - Built on top and extends Oracle ASM.
- ◆ Optimized for Non Database Files, including:
 - Executables & Trace Files. Will support DB software Oracle Homes.
 - Supports audio, video, text and other generic “file” usage.
- ◆ Can not Use for:
 - Does not support having the Grid Infrastructure home (Chicken & Egg Issue). So obviously no OCR/VOTING disk files as well.
 - Anything that can be stored in ASM disk groups.
 - 11.2.0.3 onwards will support RMAN backupsets / Archive logs & Data Pump Dumps.

ACFS Feature / Functionality

- ◆ Oracle ACFS supports dynamic file system resizing.
- ◆ Also includes “SNAPSHOT” Functionality.
- ◆ Comes with the Oracle ADVM – Oracle ASM Dynamic Volume manager.
 - Provides volume management and disk device driver interfaces.
- ◆ Extending ASM implies some benefits:
 - ASM like performance .
 - ASM Striping / online redistribution will provide increase I/O parallelism.
 - Using ASM redundancy will provide for data protection.
- ◆ We can manage ACFS volumes via :
 - ASMCA
 - ASMCMD
 - SQL*Plus
 - OEM.

How does it Work:

- ◆ Must have loaded the ACSF Drivers first.
 - In a clustered Grid installation, HAS will do this for us.
 - In non clustered environment, we need to take care to set it up. We can add it to /etc/rc.local.
 - ◆ \$GRIDHOME/bin/acfsload start.
- ◆ First we create a ACFS Volume in a disk group.
- ◆ Diskgroup must have the following configuration:
 - compatible.asm: must be 11.2 or greater
 - compatible.rdbms: must be 11.2 or greater
 - compatible.advm: must be 11.2 or higher
- ◆ Then ACFS uses the “udev” interface to provide block/character device based access to the OS.
- ◆ Then create a ACFS file system using “mkfs” against the volume
- ◆ Register the ACFS mount point within the ACFS registry.
- ◆ Mount the file system: using “mount”

ACFS – Volume Management Commands

- ◆ Apart from the GUI options provided by OEM / ASMCA.
- ◆ ASMCMD provides the following commands to look at and manipulate ACFS Volumes:
 - volcreate: Create a volume
 - voldelete: Delete a volume
 - voldisable: Disable a volume.
 - volenable: Enable a Volume
 - volinfo: Volume Information,
 - volresize: Resize a volume
 - volset: Change existing attribute
 - volstat: Report volume I/O statistics.

ASMCMD Command examples

- ◆ **List Current Volume Group Attributes**

```
ASMCMD [+] > lsattr -G avm -l -m
```

Group_Name	Name	Value	RO	Sys
AVM	access_control.enabled	FALSE	N	Y
AVM	access_control.umask	066	N	Y
AVM	au_size	2097152	Y	Y
AVM	cell.smart_scan_capable	FALSE	N	N
AVM	compatible.asm	11.2.0.0.0	N	Y
AVM	compatible.rdbms	11.2.0.0.0	N	Y
AVM	disk_repair_time	3.6h	N	Y
AVM	sector_size	512	Y	Y

- ◆ **Set Volume Group Attribute fpor ADVm compatibility**

```
ASMCMD [+] > setattr -G avm compatible.advm 11.2.0.0.0
```

- ◆ **List Current Volume Group Attributes to see the change**

```
ASMCMD [+] > lsattr -G avm -l -m
```

Group_Name	Name	Value	RO	Sys
AVM	access_control.enabled	FALSE	N	Y
AVM	access_control.umask	066	N	Y
AVM	au_size	2097152	Y	Y
AVM	cell.smart_scan_capable	FALSE	N	N
AVM	compatible.advm	11.2.0.0.0	N	Y
AVM	compatible.asm	11.2.0.0.0	N	Y
AVM	compatible.rdbms	11.2.0.0.0	N	Y
AVM	disk_repair_time	3.6h	N	Y
AVM	sector_size	512	Y	Y

ASMCMD Command examples

- ◆ **Display information for any ACFS volumes in Disk Group "AVM"**

```
ASMCMD [+] > volinfo -G avm -a
diskgroup avm has no volumes or is not mounted
```

- ◆ **Creating an ACFS volume in Disk Group "AVM"**

```
ASMCMD [+] > volcreate -G avm -s 256M --redundancy unprotected --width 32K --column 4
myvol1
```

- ◆ **Display information for any ACFS volume "myvol1" in Disk Group "AVM"**

```
ASMCMD [+] > volinfo -G avm myvol1
Diskgroup Name: AVM
Volume Name: MYVOL1
Volume Device: /dev/asm/myvol1-495
State: ENABLED
Size (MB): 512
Resize Unit (MB): 512
Redundancy: UNPROT
Stripe Columns: 4
Stripe Width (K): 32
Usage:
Mountpath:
```

- ◆ **Disable ACFS volume "myvol1" in Disk Group "AVM" and display information.**

```
ASMCMD [+] > voldisable -G avm myvol1
ASMCMD [+] > volinfo -G avm myvol1
Diskgroup Name: AVM

Volume Name: MYVOL1
Volume Device: /dev/asm/myvol1-495
State: DISABLED
Size (MB): 512
Resize Unit (MB): 512
Redundancy: UNPROT
Stripe Columns: 4
Stripe Width (K): 32
Usage:
Mountpath:
```

ASMCMD Command examples

- ◆ Enable ACFS volume "myvol1" in Disk Group "AVM" and display information.

```
ASMCMD [+] > volenable -G avm myvol1
ASMCMD [+] > volinfo -Gavm myvol1
Diskgroup Name: AVM
```

```
Volume Name: MYVOL1
Volume Device: /dev/asm/myvol1-495
State: ENABLED
Size (MB): 512
Resize Unit (MB): 512
Redundancy: UNPROT
Stripe Columns: 4
Stripe Width (K): 32
Usage:
Mountpath:
```

- ◆ Delete ACFS volume "myvol1" in Disk Group "AVM" and display information.

```
ASMCMD [+] > voldelete -G avm myvol1
ASMCMD [+] > volinfo -G avm -a
diskgroup avm has no volumes or is not mounted
```

Create an ACFS File System on a volume

- ◆ Note: For these examples:

- UDEV Device associated with the volume is “/dev/asm/myvol1-495”.
- The mount-point for the file system is “ACFS_mount_point”

- ◆ Create ACFS Command:

```
/sbin/mkfs -t acfs /dev/asm/myvol1-495
```

- ◆ Register Mount Point Command:

```
/sbin/acfsutil registry -a -f /dev/asm/myvol1-495 /ACFS_mount_point
```

- ◆ Actual mounting the file system commands is: :

```
/bin/mount -t acfs /dev/asm/myvol1-495 /ACFS_mount_point
```

- ◆ Unmounting a File System:

```
sudo /bin/umount -t acfs /dev/asm/myvol1-495
```

- ◆ A df example:

```
oracle@sillgr2-64 oracle]$ df
Filesystem                1K-blocks      Used Available Use% Mounted on
/dev/mapper/VolGroup00-LogVol00
                          17330592    13506268    2929760    83% /
/dev/sda1                  101086      15886      79981    17% /boot
tmpfs                      771156      158800      612356    21% /dev/shm
/dev/asm/myvol1-495        524288      38152      486136     8%
  /ACFS_mount_point
```

A Note about Stripe Columns/Width etc.

- ♦ ADVM implements the concepts of extents and striping algorithm.
 - ADVM writes “stripe width” in a round robin fashion across the “stripe Columns”.
- ♦ Stripe width and columns will impact the extents as well as the sizing of the volume.
- ♦ To Understand the created the same volume with different stripe width/stripe column settings and size of 256M.

Stripe Width 32K, with a specified size of 256M:

# of Columns	Initial Size	Resize Unit	Notes
1	256M	128M	Stripe width over ridden to be 128K
2	256M	256M	Stripe width maintained at 32K.
3	384M	384M	Stripe width maintained at 32K.
4	512M	512M	Stripe width maintained at 32K.

Stripe Width 64K, with a specified size of 256M:

# of Columns	Initial Size	Resize Unit	Notes
1	256M	128M	Stripe width over ridden to be 128K
2	256M	256M	Stripe width maintained at 64K.
3	384M	384M	Stripe width maintained at 64K.
4	512M	512M	Stripe width maintained at 64K.

So what can we infer:

- ◆ With the numbers of columns = 1, stripe width defaults to 128k (131072).
- ◆ With the number of columns > 1, the stripe width specified is used.
- ◆ The stripe width does not seem to impact the sizing of the volume group.
- ◆ The # of stripe columns impacts the volume sizing & resize unit.
- ◆ The Resize unit is a factor of 128M. I would guess that this is platform specific.
- ◆ The multiplier effect seems to be a dependent on the the # of columns and the stripe size:
 - # of columns = 1 , resize unit is 128M.
 - # of columns > 1 , resize unit = NUM_STRIPE_COLUMN*128m.
- ◆ So the formula would seem to be something like:
 - $RESIZE_UNIT_IN_MB = MAX(2, NUM_STRIPE_COLUMN) * 128$
 - $MIN_VOL_SIZE = MAX(RESIZE_UNIT, SIZE_SPECIFIED-ROUNDED\ TO\ THE\ NEXT\ RESIZE\ UNIT).$

ACFS Utility Commands

- ◆ **[oracle@si11gr2-64 avm_myvol1]\$ acfsdriverstate
version**

ACFS-9205: OS/ADVM,ACFS installed version = 2.6.18-8.el5(x86_64)/090715.1

- ◆ **acfsutil help – will give the list options.**

- ◆ **Some examples:**

```
[oracle@si11gr2-64 temp]$ acfsutil  
registry -l
```

```
Device : /dev/asm/myvol1-495 : Mount  
Point :  
/u01/app/oracle/acfsmounts/avm_myvol1 :  
Options : none : Nodes : all : Disk  
Group : AVM : Volume : MYVOL1
```

Questions/Discussion



Thank You !!!