Achieving Availability and Scalability with Oracle 12c Flex Clusters and Flex ASM

Kai Yu
Dell Inc.
Austin, Texas, USA

Keywords:

Oracle 12c RAC, Oracle 12 Clusterware, Oracle 12c ASM, Flex Clusters, Flex ASM, Pluggable Databases.

Introduction

Reliability, Availability, and Scalability (RAS) is becoming even more paramount now that application consolidation has become a key initiative in the data center. With Oracle Database 12c, Clusterware enables in the adherence to RAS. Flex Clusters along with Flex ASM enable database and application clusters to not only scale further horizontally, but also drive faster agility/ elasticity as the business demand changes. Along with these advantages, Oracle Database 12c Flex Clusters also inherently drives higher resource utilization. In this session attendees will learn the essentials of Flex Clusters and Flex ASM, understand business benefits and showcase the Flex Clusters/ Flex ASM deployment process and their management practices

Oracle 12c Grid Infrastructure and RAC

Oracle Real Application Clusters is an active-active cluster database solution. It is designed to
- Protect database availability against up to N-1 server failure
- Reduce planned downtime for hardware, OS, software upgrade
- be able to add node or remove node based on demand of capacity
- achieve the application load balancing
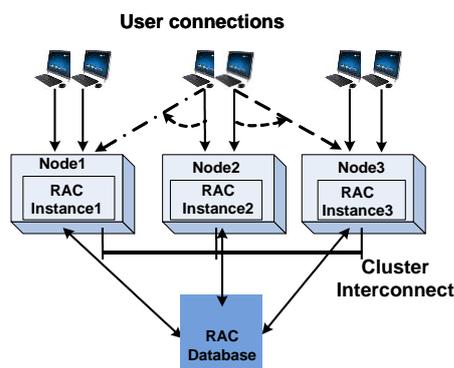- Provide high availability, scalability and flexibility

Illustration. 1: Oracle RAC for high availability, scalability and flexibility

The Oracle RAC Stack is composed of two major components
- Oracle Grid Infrastructure which includes Oracle Clusterware and ASM
- Oracle RAC coordinates and synchronize multiple DB instances through Cache Fusion technology

The following diagram illustrates the structure of these two components:
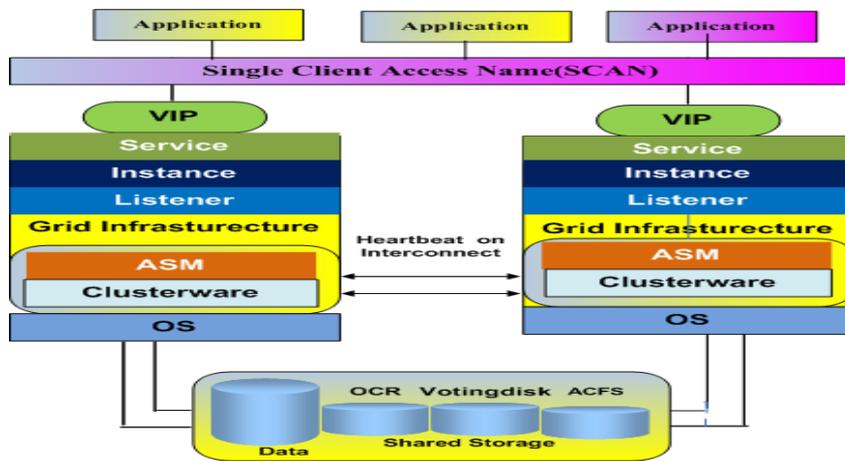

Illustration. 2: Oracle RAC Architecture and Components

Oracle 12c Clusterware is designed to
- Enable the communication between the cluster server
- Managing resources ASM instances, DB instances, Virtual IPs , SCAN, etc
- Foundation for RAC database and HA features
  - Manage failover of Virtual IP to other node
  - Restarts failed Oracle processes
  - Manages node memberships & prevent the split-brain syndrome
  – Installed into the same Grid Infrastructure home with ASM

Oracle 12c Clusterware requires two hardware components to support its operation: a) shared storage to storage Voting disk + OCRl b) Private Interconnects network for the cluster heart bit Oracle RAC.

Oracle 12c clusterware is composes of two major technology stack:Cluster Ready Service stack and high Availability Service stack. Each of these two stacks is composed of a set of clusterware processes as shown in following diagram:
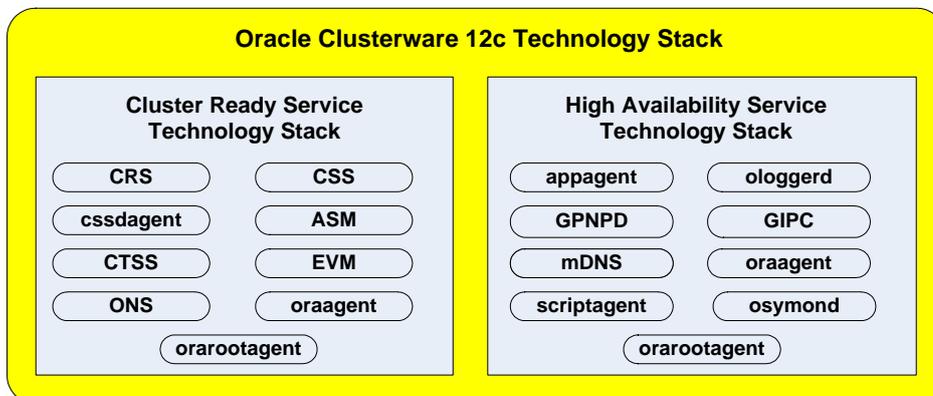
Illustration. 3: Oracle 12c Clusterware Technology Stack RAC

Whenever the server (RAC node) starts, the Clustweware will be started automatically. For example, in Linux environment, the Linux init process starts te init.ohash which starts the entire Oracle Clusterware stack through the multiple levels starting process as shown in the following diagram:
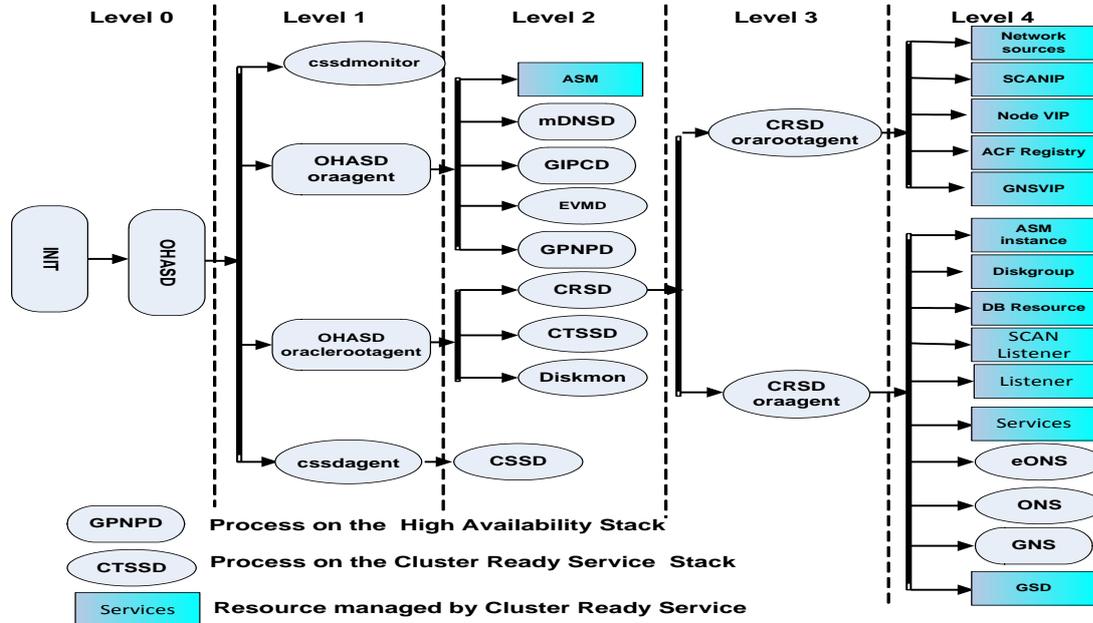

Illustration. 4: Oracle 12c Clusterware stack startup sequence

Please insert your pictures directly into the text and check if size and quality of the pictures is sufficient.

Oracle 12c Grid Infrastructure and RAC
 [Times New Roman 11 pt medium big]

Oracle 12c RAC introduced many new features . These new features and enhancements focus on business continuity , high availability, scalability, agility, cost-effective workload management. The following lists are some of the main Oracle 12c Clusterware and RAC new features and enhancements:

- Oracle Flex Clusters for cluster scalability
- Oracle Flex ASM for high availability scalability
- Oracle RAC support for Oracle 12c Pluggable databases
- What-If Command Evaluation
- Public Networks for RAC: IPv6 Support Added
- Application Continuity:
- Cluster Health Monitor (CHM) Enhancements
- Shared ASM Password File stored in ASM diskgroup
- IPv6 Support for Public Networks

The rest of the paper will examine three of most important new features: Oracle Flex Clusters for cluster scalability , Oracle Flex ASM for high availability scalability , Oracle RAC support for Oracle 12c Pluggable databases

Oracle 12cR1 Flex Clusters

Until Oracle 12c, the Oracle Clusterware requires that all RAC nodes tightly-connected and All nodes directly connected to storage. This leads to N * (N-1)/2 interconnect paths
and total N storage paths. These requirements prevent the cluster to go beyond of 100 nodes. As for a 100 node cluster, it will have 100 * 99/2 = 4950 interconnects and 100 storage connections which is very difficult and to scale out. To be able to scale out to more nodes much beyond 100, Oracle 12c RAC introduced a new two-layered hub-and–spoke topology called Flex Clusters. As shown in the following Illustration 4: A Flex Cluster consists of the type of nodes: Hub nodes and Leaf nodes:

Hub Node: all the nodes are interconnected and directly connected to the shared storage

Leaf Node are connected their dedicated Hub node, no storage connection is required. A Leaf node with storage access can be converted to a Hob node.
With this architecture, A Flex Cluster can be scaled to 64 Hub nodes and total 2000 nodes (Hub node and Leaf nodes together)
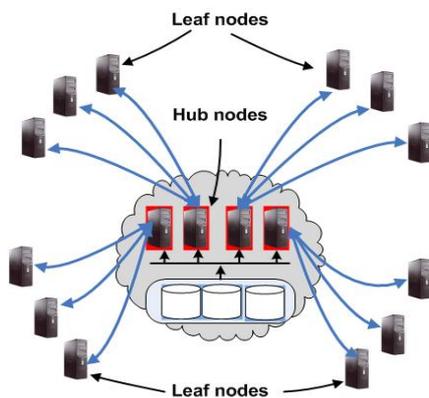

Illustration. 5: Oracle 12c Flex Cluster

A standard cluster is just like a 11gR2 cluster: all the nodes are Hub nodes. A standard cluster can be converted to a Flex Cluster, but it can't be changed back without a reconfiguration of a cluster.All Hub nodes function in same way as the standard cluster nodes using Flex ASM.

When you design the cluster initially, you may choose either standard cluster or Flex cluster. If you are not sure, choose the standard Cluster as it can be converted to Flex Cluster later on.

Configuring Flex Clusters

Flex Clusters can be configured on Oracle Grid Infrastructure OUI installer. You need to select the advanced configuration mode and then select the configure a Flex cluster option as shows on the following diagram:
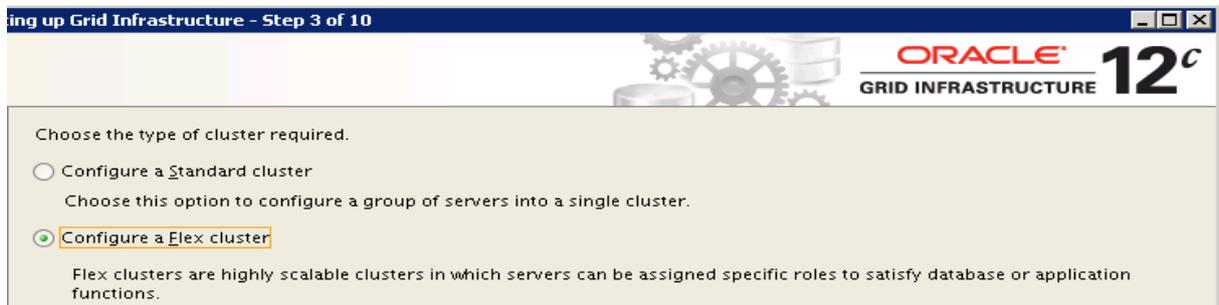

Illustration. 6: Select Flex Cluster option in OUI

Flex Clusters require to GNS VIP. So you need to select configure NS if you decided to configure Flex cluster during the Grid Infrastructure OUI:
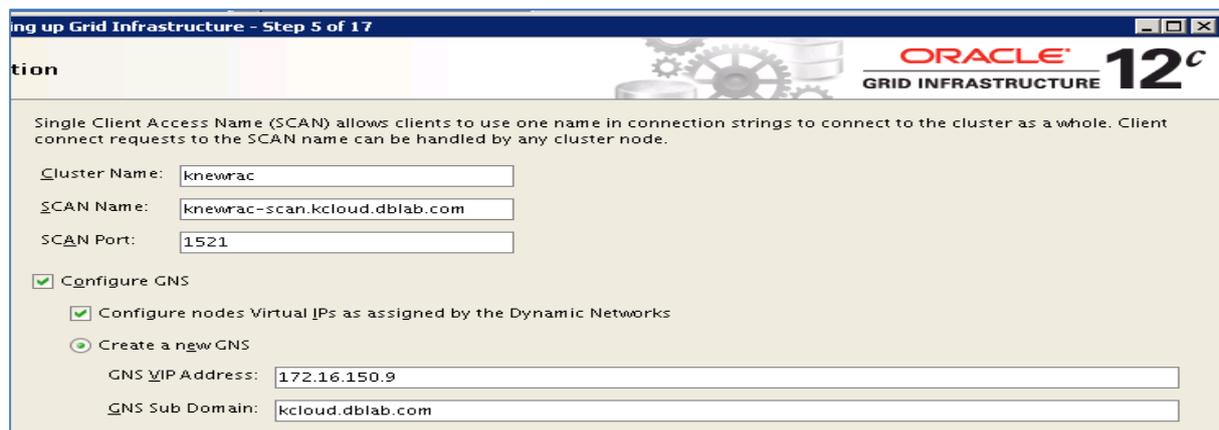

Illustration. 7: Specify GNS configuration for Flex Cluster

And you need to specify which nodes are Hub node and which nodes are Leaf nodes:
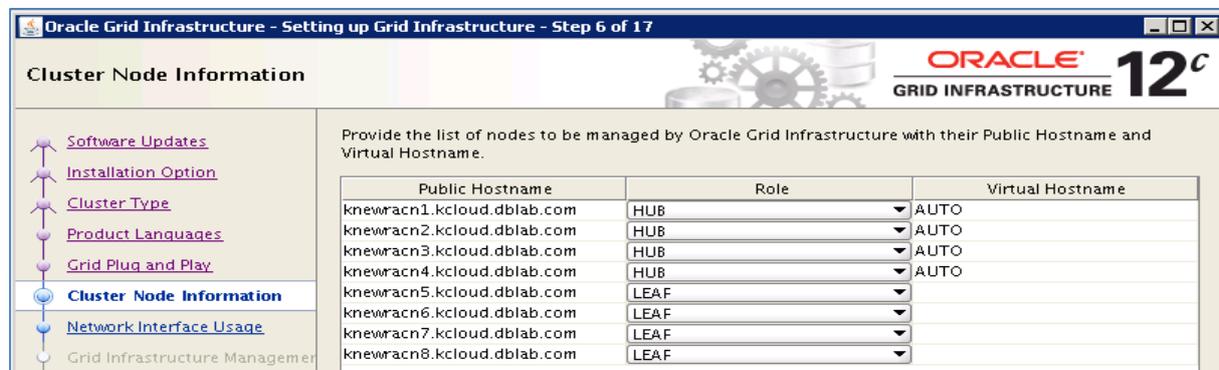
Illustration. 8: Specify Hub Nodes and Leaf nodes

After Grid Infrastructure installation, you can see the Hub node status with the commands like these:

Get the local node role:

```
[grid@knewracn1 ~]$ crsctl get node role status
Node 'knewracn1' active role is 'hub'
```

Get the roles of all the cluster nodes:

```
[grid@knewracn1 ~]$ crsctl get node role status -all
Node 'knewracn1' active role is 'hub'
Node 'knewracn2' active role is 'hub'
Node 'knewracn3' active role is 'hub'
Node 'knewracn4' active role is 'hub'
Node 'knewracn6' active role is 'leaf'
Node 'knewracn5' active role is 'leaf'
Node 'knewracn7' active role is 'leaf'
Node 'knewracn8' active role is 'leaf'
```

And check the clusterware stack:

```
[grid@knewracn1 ~]$ 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
```

Check the votedisks and VIP configuration:

```
[grid@knewracn1 ~]$ crsctl query css votedisk
##  STATE     File Universal Id              File Name Disk
group
-- -----     -----------------              --------- ------
---
 1.  ONLINE    af97e218f43c4fbfbfc5ba7e87f801f6  (ORCL:DATA1)
[DATA1]
Located 1 voting disk(s).

[grid@knewracn1 ~]$ srvctl config vip -n knewracn1
VIP exists: network number 1, hosting node knewracn1
VIP IPv4 Address: -/knewracn1-vip/172.16.150.155
VIP IPv6 Address:
```

And Leaf nodes such as knewran8:

Get the local node role:

```
[[grid@knewracn8 ~]$ crsctl get node role status
Node 'knewracn8' active role is 'leaf'
```

Get the roles of all the cluster nodes:

And check the clusterware stack on the node:

```
[grid@knewracn8 ~]$ 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
```

Check the votedisks and VIP configuration opn knewracn8

```
[grid@knewracn8 ~]$ crsctl query css votedisk
CRS-1668: operation is not allowed on a Leaf node
[grid@knewracn8 ~]$ srvctl config vip -n knewracn8
PRKO-2310 : VIP does not exist on node knewracn8.
```

Change Standard cluster to flex cluster

You can change astandard Cluster to a Flex Cluster with the following steps:
1. Meet the prerequisites: Add GNS service and add with a Fixed Virtual IP
     #srvctl add gns –vip <VIP_address> –domain <doman_name>
2. Enable Flex ASM (refer to next session)
3. Convert to Flex Cluster
      #crsctl set cluster mode flex
     # crsctl stop crs
      #crsctl start crs –wait
4.   Add Leaf listener: $srvctl add listener –leaflistener –skip

Oracle 12cR1 Flex ASM

Oracle Flex ASM Architecture:
 In pre-Oracle 12c such as Oracle 11gR2, the standard ASM infrastructure has two limitations:

- Each node has an ASM instance which cost CPU/memory
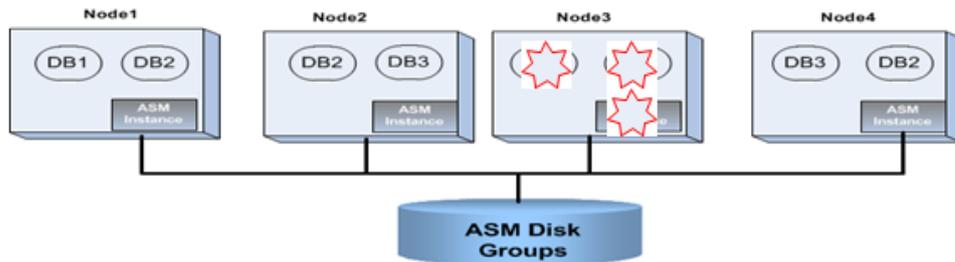- Local ASM instance failure will cause DB instance failure

Illustration. 8: Flex ASM architecture

As shown on the four node standard ASM above, if the ASM instance on node 3 fails, the database instances on node 3 both fail.

Oracle 12c introduced a new ASM structure called Flex ASM as the extension to the standard ASM which is same as the ASM in 11gR2 cluster. Flex ASM is an option on Oracle 12c which can be enabled or disabled by our choice. In Oracle Flex ASM structure, only a small number of nodes   run ASM instance. (default number is 3 and be specified by changed by admin). And each database instance is allowed to connect to ASM instances the local ASM or a remote ASM:
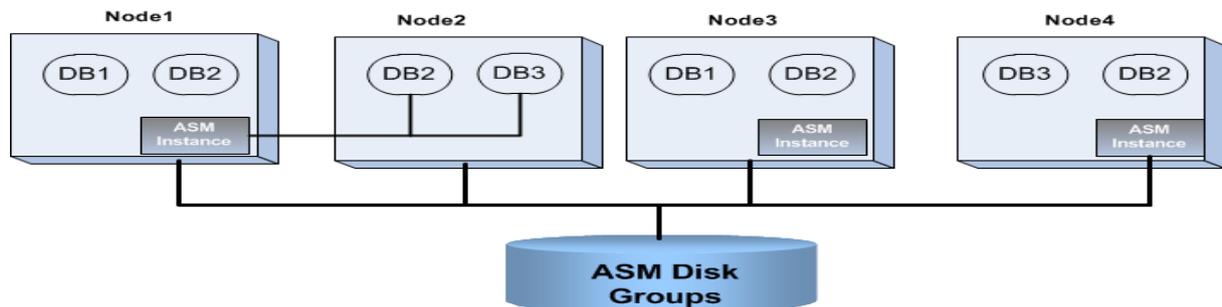

Illustration. 9: Flex ASM Configuration for Oracle 12c Databases

In a Flex ASM, there are two o kinds of Oracle ASM configurations:
- Local ASM clients connect to local ASM instance. For example, DB1 and DB2 instances on node1 connect to the local ASM instance on node1
- Flex ASM clients connect to a remote ASM instance. Foe example, DB2 and DB3 instances on node2 connect to the remote ASM instance on node1.

Flex ASM requires ASM network for communication between ASM clients and remote ASM instance. On Oracle 12cR1, ASM network can share the network with cluster private interconnect.
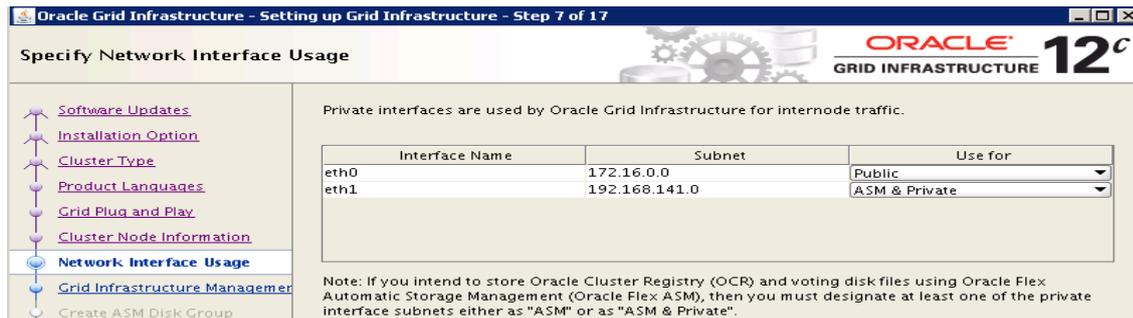
Illustration. 10: Select ASM network for Flex ASM

As shown above, during the Grid infrastructure OUI , you can select the same network interface for both ASM network and Cluster private interconnect traffic.

ASM client such as database instance uses ASM password file authentication. In Oracle 12c, ASM files is stored in the ASM diskgroup shared by all the nodes. This command shows the ASM diskgroup and directory where ASM password file orapwasm is stored:

```
ASMCMD> ls -l orapwasm
Type        Redund  Striped  Time           Sys  Name
PASSWORD_  UNPROT  COARSE   SEP 11 10:00:00  N    orapwasm => +DATA1/ASM/PASSWORD/pwdasm.256.825849121
```

Flex ASM and Flex Clusters

There are two new 12c features are related , but they are different. If you choose Flex Clusters during the OUI, ASM is automatically enable. And before you can change the standard cluster to Flex cluster, you have to enable Flex ASM first. In a flex cluster, only Hub nodes can run ASM instances as only hub nodes are required to have the direct storage access.

However you can run the Flex ASM in the standard cluster. In this case, since all the nodes in the standard cluster are Hub nodes, ASM instances can run on any nodes. But only a subset of nodes run ASM instance.

Configure Flex ASM

You can configure Flex ASM in two ways:
        . Configure ASM during the Grid Infrastructure OUI installation process
        . Convert the Standard ASM to Flex ASM.

To configure Flex ASM on Grid Infrastructure OUI , select Flex ASM option or select Flex Cluster
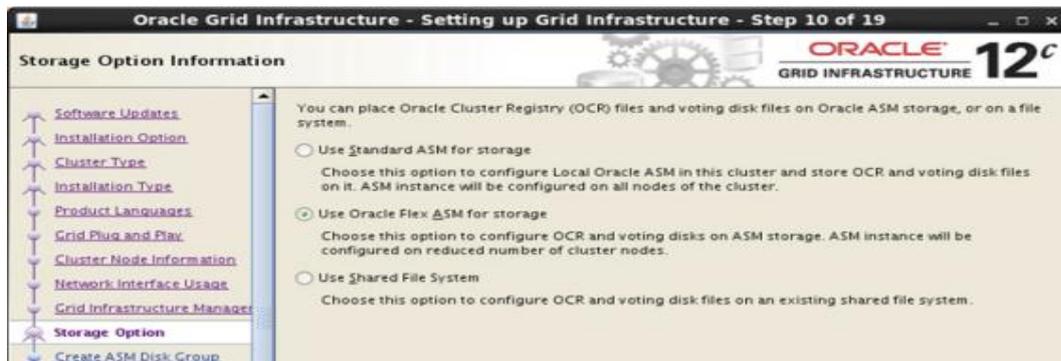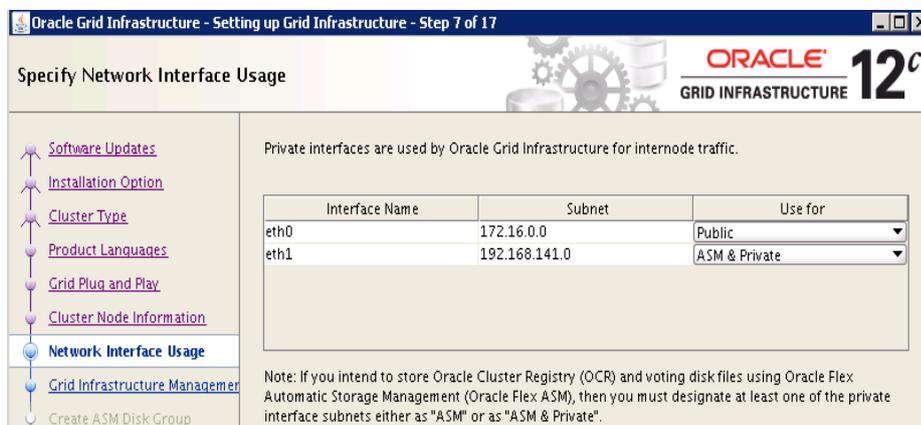
Illustration. 10: Select ASM network for Flex ASM

And the ASM network needs to be specified:

The second option to configure Flex ASM is to convert to standard ASM to Flex ASM. The steps are:
- Setup ASM network
- Convert ASM to Flex ASM with asmca tool such as:

```
$asmca -silent -convertToFlexASM  -asmNetworks
  eth1/192.168.141.0 -asmListenerPort 1521
```

- Run converttoFlexASM.sh as the root :on all nodes one at a time

Managing Flex ASM

Management of Flex ASM doesn't require any specific tasks . You can use asmcmd command line
And SRVCTL utility to manage the Flex ASM:

To check the current ASM mode:

```
[grid@knewracn1 ~]$ asmcmd showclustermode
ASM cluster : Flex mode enabled
```

To check the details of Flex ASM nodes:

```
[grid@knewracn1 ~]$ srvctl status asm -detail
ASM is running on knewracn1,knewracn4,knewracn3
ASM is enabled.
```

To check the Flex ASM configuration:

```
[grid@knewracn1 ~]$ srvctl config asm
ASM home: /u01/app/12.1.0/grid
Password file: +DATA1/orapwASM
ASM listener: LISTENER
ASM instance count: 3
Cluster ASM listener: ASMNET1LSNR_ASM
```

Database High Availability provided by Flex ASM

With Oracle flex ASM, the database instance no longer depends on the local ASM instance. In case of local ASM instance failure, the database instance can fail over their ASM connections to a remote ASM instance on other cluster node. For example, one ASM instance on node 3 fails, DB1 and DB3
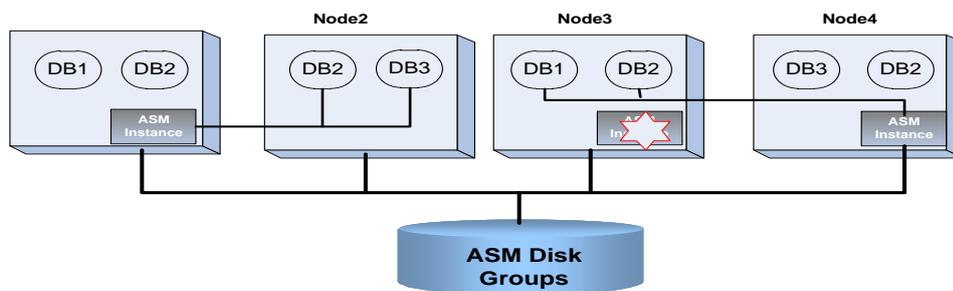

Illustration. 11: Database HA with Flex ASM

Instances on node3 fail over their ASM connections to ASM instance on node 4.

Deploy Oracle Flex ASM: Architecture Options

This session discusses various options of deploying Flex ASM .

If you deploy a new Oracle 12c database and decide to use Flex ASM option, you only need to run Flex ASM on a small number of the RAc nodes as shown the following architecture:
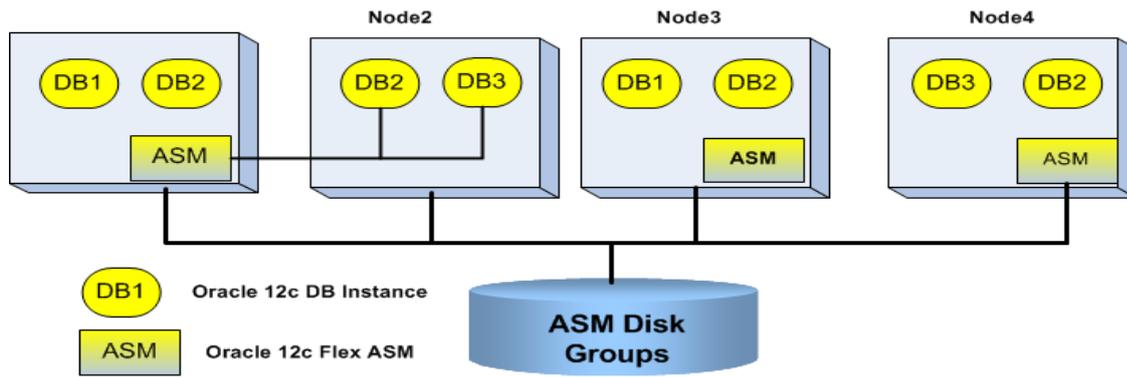
Illustration. 12: Deploying 12c database with Flex ASM

If you need to mix pre-12c databases and 12c databases on the same cluster, you can enable the Flex ASM option and make all the ASM instance count or cardinality as same as the cluster nodes, namely running ASM in all the cluster nodes:

```
$srvctl modify asm -count ALL
```

However in this architecture, in case ASM instance fails, only 12c database instances can failover their ASM connections to remote ASM instance.
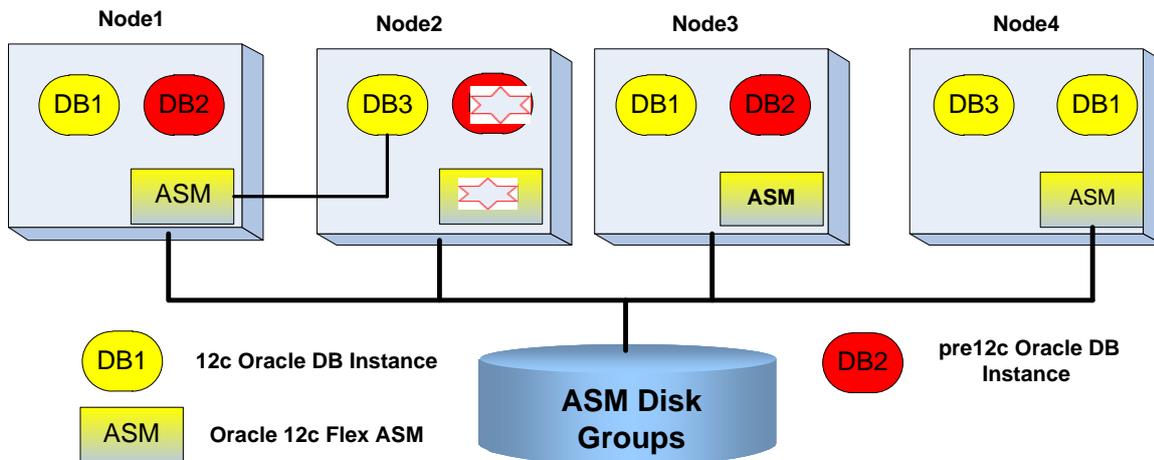

Illustration. 13: Deploying 12c and pre-12c databases with Flex ASM- ASM on all nodes

You also have an option to run few ASM instance. But you only can put pre-12c databases on the nodes that have ASM instance. And only 12c database instances can failover their ASM connections to remote ASM instance
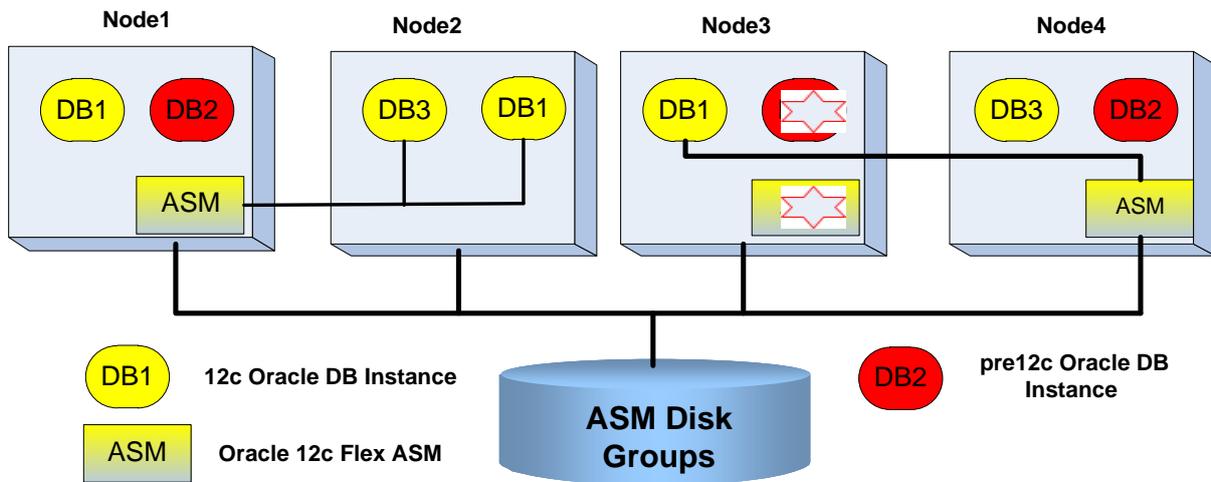
Illustration. 14: Deploying 12c and pre-12c databases with Flex ASM- ASM on part of nodes

How 12c Pluggable Database works on Oracle 12c RAC

Oracle 12c RAC supports Pluggable databases. PDBs can be created and configured on a multiple-node RAC environment as the architecture shown below:
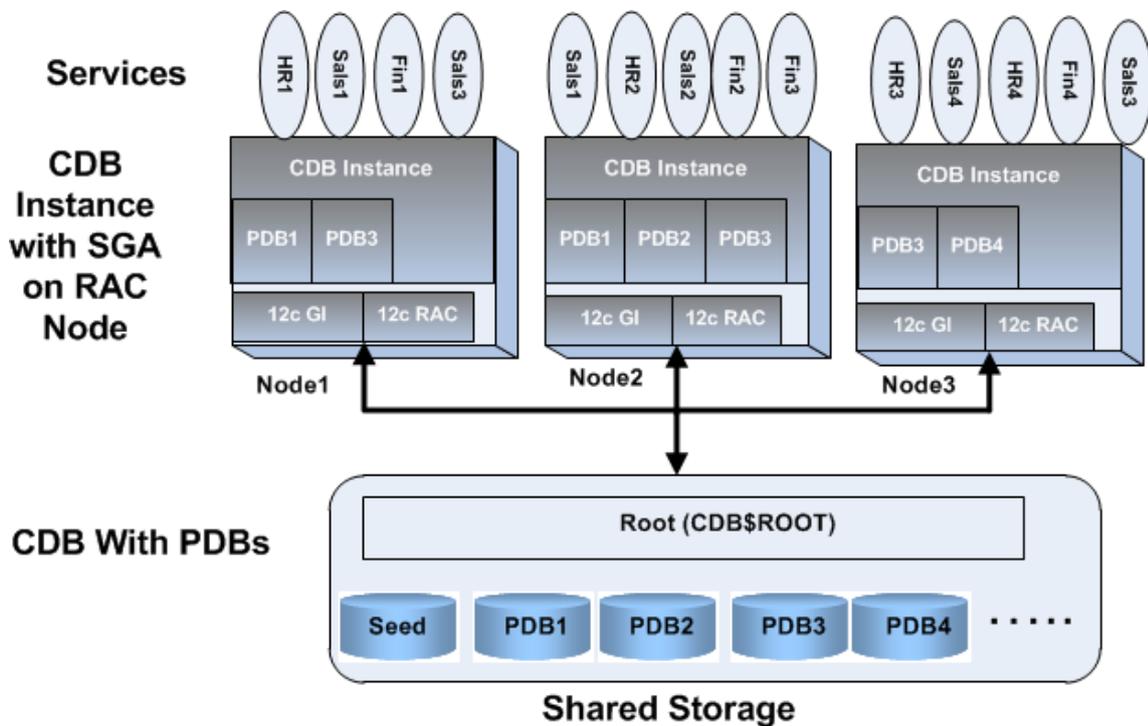

Illustration. 15: Pluggable databases on Oracle 12c RAC

Similar to the non-CDB, on each RACnode, a PDB can be on different open modes:
- Mounted
- Read only
- Read and write

You can check the open modes of all the PDBs on a RAC instance when you connect to the CDB root of the instance:

```
SELECT NAME, OPEN_MODE, RESTRICTED FROM V$PDBS;
NAME OPEN_MODE RESTRICTED
-------- ---------- ----------
PDB$SEED READ ONLY NO
PDB1 READ WRITE NO
PDB2 READ WRITE NO
PDB3 Mounted NO
```

To start a PDB, connect to the PDB with SQL*Plus and run the startup or the shutdown command with different:

```
startup open options:
STARTUP OPEN
STARTUP OPEN READ ONLY
STARTUP RESTRICT OPEN READ ONLY
SHUTDOWN IMMEDIATE
```

Create Pluggable databases :

Pluggable databases can be created with the container database during the DBCA database creation. You can specify how many pluggable databases to be created.
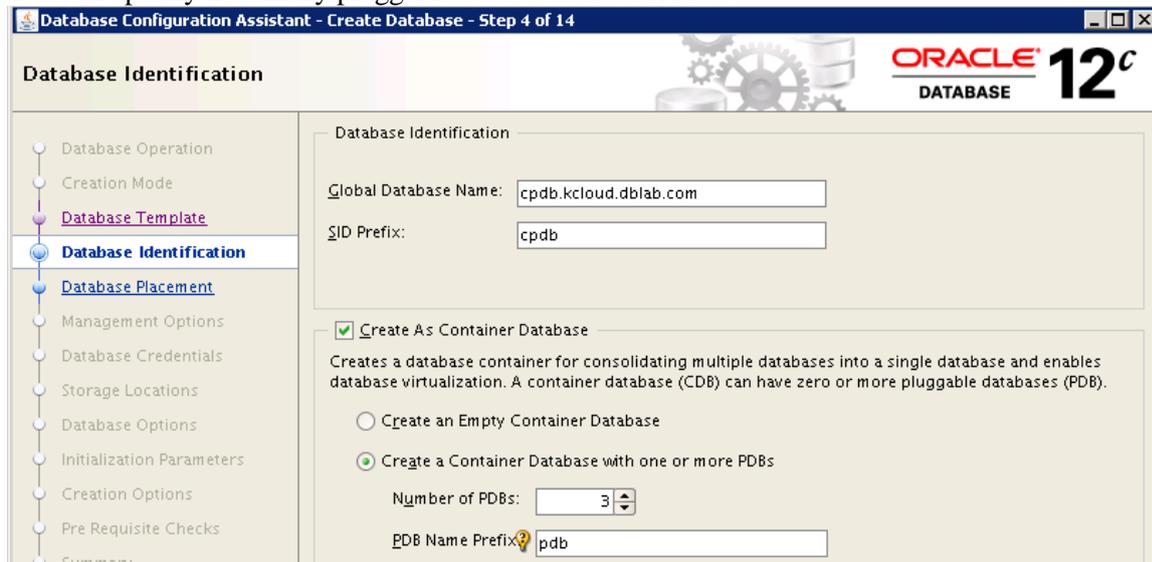

Illustration. 16: Create Pluggable databases with dbca

In order for applications to connect to the pluggable databases, you need to create dynamic services for each pluggable databases, for example,

```
$srvctl add service -db cpdb -service hr3 -pdb pdb3 -preferred cpdb3
-available cpdb2
$srvctl add service -db cpdb -service sales2 -pdb pdb3 -preferred
cpdb2 -available cpdb3
```

In order for applications to connect to a particular pluggable database, a net service in the tnsnames. ora file h need s to be created with the dynamic service. For example, net service name SALES_PDB2 is created using dynamic service sales2.kcloud.dblab.com .

```
SALES_PDB2 =
    (DESCRIPTION =
        (ADDRESS = (PROTOCOL = TCP)(HOST =
    knewracscan.kcloud.dblab.com)
        (PORT = 1521))
        (CONNECT_DATA =
        (SERVER = DEDICATED)
        (SERVICE_NAME = sales2.kcloud.dblab.com)
        )
    )
```
The sale application can use the net service name SALES_PDB2 as the connection string to connect PDB2

Contact address:

Name   Kai Yu
Company : Dell Inc.
Address: One Dell Way, Round Rock, Texas 78682, USA
Postal code, city

Phone:	+1(512)-728-6343
Fax:	+1(512)-728-6343
Email	kai_yu@dell.com
Internet:	http://kyuoracleblog.wordpress.com/