

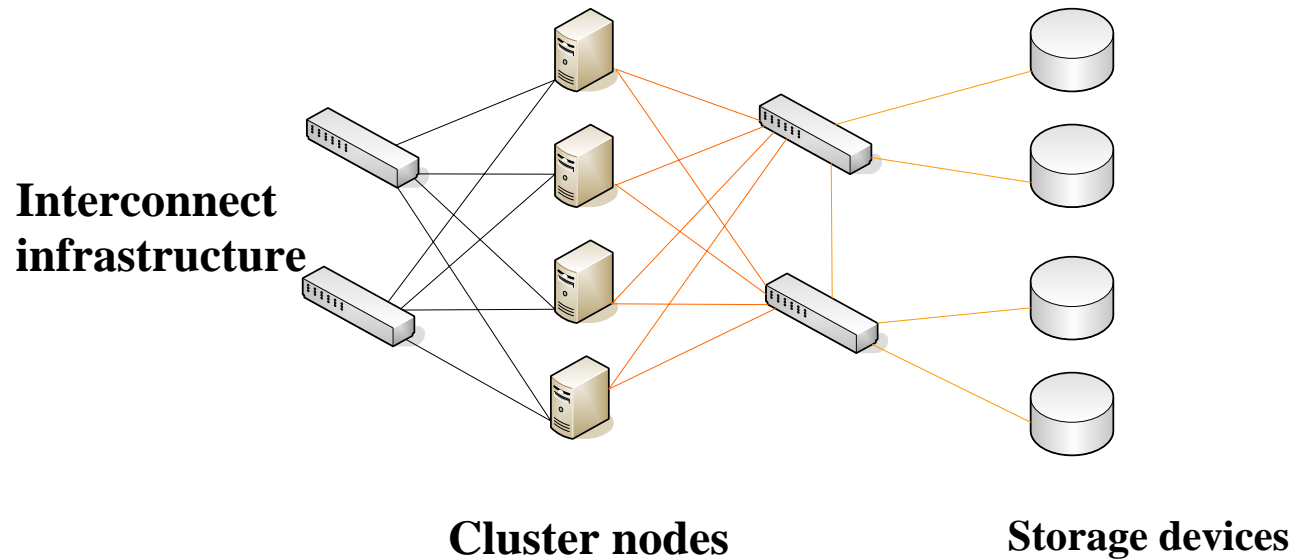
RAC 10g on Linux

Marta Jakubowska-Sobczak, OpenLab fellow



RAC overview

- RAC - a cluster database with a shared cache and a shared storage architecture
- Setup example





RAC components

- Two or more nodes (each running an Oracle instance)
- An interconnect infrastructure - a high-bandwidth, low latency communication link between the nodes
- Shared disk subsystem - can be either a cluster file system or raw devices
- The nodes cluster and its interconnect are linked to the storage devices by a storage area network



Specific software components

- Cluster Ready Services (CRS) - a complete and integrated clusterware management solution; two shared files required:
 - Oracle Cluster Registry(OCR)
 - CRS Voting Disk
- Automated Storage Management (ASM) - a thin layer between raw devices and database instance, provided by Oracle as a part of Oracle Database Server 10g
- Global Services Daemon (GSD) - coordinates with cluster manager to receive requests from RAC monitoring and management utilities to execute administrative tasks
- ...



Hardware and network requirements

- **Generally each node requires:**
 - access to the shared disks for storing database files and CRS (Cluster Ready Services) files
 - one private IP address for interconnect
 - one public IP address to serve as the Virtual IP address for clients connections and for connection failover (this is in addition to the operating-system managed public IP address)



Pre-installation tasks

check hardware requirements

- Check if each node meets the following requirements:
 - at least 512MB of physical RAM
 - 1GB of swap space or twice the size of RAM (for systems with 2GB of RAM and more, it can be between 1 and 2 times RAM size)
 - 400MB of disk space in the /tmp directory
 - Up to 4GB of disk space for the Oracle software



Pre-installation tasks check OS requirements

Operating system:

Red Hat Enterprise Linux AS/ES 2.1 (Update 3 or higher)

Red Hat Enterprise Linux AS/ES 3.0 (Update 2 or higher)

Kernel version:

Red Hat Enterprise Linux 3: 2.4.21-15.EL

Packages:

make-3.79.1 gcc-3.2.3-34
glibc-2.3.2-95.20 compat-db-4.0.14-5
compat-gcc-7.3-2.96.128 openmotif21-2.1.30-8 *
compat-gcc-c++-7.3-2.96.128
compat-libstdc++-7.3-2.96.128
compat-libstdc++-devel-7.3-2.96.128
setarch-1.3-1

* - we use openmotif-2.2.3-5.RHEL3.2



Kernel parameters

- `kernel.sem=250 32000 100 128`
- `kernel.shmall=2097152`
- `Kernel.shmmax` = half the size of physical memory
- `kernel.shmmni=4096`
- `fs.file-max=65536`
- `net.ipv4.ip_local_port_range=1024 65000`
- `net.core.rmem_default=262144`
- `net.core.rmem_max=262144`
- `net.core.wmem_default=262144`
- `net.core.wmem_max=262144`



Shell limits

- Add following line to `/etc/pam.d/login`

`session required /lib/security/pam_limits.so`

- Add the following lines to `/etc/security/limits.conf`

```
*          soft    nproc    2047
*          hard    nproc    16384
*          soft    nofile   1024
*          hard    nofile   65536
```



Pre-installation tasks

IP address requirements

Check if you have the following addresses for each node:

- an IP address & associated host name (registered in DNS) for each public network interface
- one unused virtual IP address & associated host name (registered in DNS) for the primary public network interface (associated with the same network interface on each node)
- a private IP address for each private interface, isolated from public network; must have the same network interface name on each node)

Commands:

```
/sbin/ifconfig
```



Network configuration example /etc/hosts

- 127.0.0.1 localhost.localdomain localhost
- #Public hostnames for eth0 interface (public network)
- 137.138.216.73 itrac13.cern.ch # RAC pub node 1
- 137.138.216.74 itrac14.cern.ch # RAC pub node 2
- #Private hostnames for eth1 interface (cluster interconnect)
- 192.168.13.1 atlr-priv1-13 # RAC priv1 node 1
- 192.168.13.2 atlr-priv1-14 # RAC priv1 node 2
- #Private hostnames for eth2 interface (cluster interconnect)
- 192.168.14.1 atlr-priv2-13 # RAC priv2 node 1
- 192.168.14.2 atlr-priv2-14 # RAC priv2 node 2
- #Virtual IP addresses eth0:1 interface (public Virtual IP)
- 137.138.216.84 itrac13-v # RAC virtual node 1
- 137.138.216.85 itrac14-v # RAC virtual node 2



Network configuration

- On all nodes make eth configuration permanent:
create /etc/sysconfig/network-scripts/ifcfg-eth[1-2] with the following contents:

```
DEVICE=eth[1-2]
BOOTPROTO=static
IPADDR="192.168.[13-14].xx"
NETMASK="255.255.255.0"
ONBOOT=yes
TYPE=Ethernet
```

- When the network configuration is done, it is important to make sure that the name of the public RAC nodes is displayed when you execute the following command: `$ hostname`



SSH configuration

- On all nodes repeat:

```
mkdir ~/.ssh
```

```
chmod 755 ~/.ssh
```

```
/usr/bin/ssh-keygen -t rsa ## leave phrase key empty
```

```
/usr/bin/ssh-keygen -t dsa ## leave phrase key empty
```

```
cd .ssh/
```

```
touch authorized_keys
```

```
chmod 644 authorized_keys
```

Then, on each RAC node copy the contents of
.ssh/id_rsa.pub and .ssh/id_dsa.pub to authorized_keys on
all RAC nodes (on this node and to others):

```
cat ~/id_rsa.pub ~/id_dsa.pub >> ~/.ssh/authorized_keys
```



Check ssh configuration

Do ssh from each RAC node to each RAC node (including the same node) without giving password.

```
ssh itrac13 hostname  
ssh itrac13.cern.ch hostname  
ssh atlr-priv1-13 hostname  
ssh atlr-priv2-13 hostname
```

You should be able to connect to the nodes without password specification. Repeat it from each node to each node.



UNIX groups & users required

For Oracle database installation:

- the OSDBA group (default dba) - users that have the SYSDBA privilege
- the OSOPER group (default oper) - optional, to separate users with limited database administrative privileges (SYSOPER)
- an unprivileged user (nobody)

For Oracle software installation:

- the Oracle Inventory group (oinstall) - it owns the Oracle inventory
- the Oracle software owner (oracle) - it owns all the software installed during installation & must have oinstall as its primary group & the dba and oper groups as secondary groups

The oracle software owner and the Oracle Inventory group, dba & oper groups must exist and be identical on all cluster nodes.

In our installations we don't distinguish the groups and use only one, called ci. See /etc/oraInst.loc



Directories structure

Oracle installations require:

- ORACLE_BASE directory - a top level directory for Oracle software installations; must have the same path on all nodes and shouldn't be on the same file system as OS
- A directory for Oracle inventory - a catalog of all Oracle software installed on the system
- A home directory for each product being installed (CRS, database server, etc.) - must be separated from each other and must be a subdirectory of ORACLE_BASE



Directories structure

Recommended configuration:

/ORA/dbs01/oracle

ORACLE_BASE directory

Disk space 4GB

/ORA/dbs01/oracle/product

/ORA/dbs01/oracle/product/10.1.0

/ORA/dbs01/oracle/product/10.1.0/crs

CRS home directory, at
least 1M

/ORA/dbs01/oracle/product/10.1.0/rdbms

Database server home
directory

/ORA/dbs01/oracle/oraInventory



Directories structure

- **Change ownership of oracle base directory:**

```
chown -R oracle:ci /ORA/dbs01/oracle
```

- **Check access rights:**

```
chmod 775 /ORA/dbs01/oracle
```

```
chmod 775 /ORA/dbs01/oracle/product
```

```
chmod 775 /ORA/dbs01/oracle/product/10.1.0
```

```
chmod 775 /ORA/dbs01/oracle/product/10.1.0/crs
```

```
chmod 775 /ORA/dbs01/oracle/oraInventory
```



Preparing disks for CRS & ASM

- Create the partitions on shared storage for OCR, CRS and ASM spfile:

Use: `sudo /sbin/fdisk /dev/sdb`

`/dev/sdb1` - 200M - for OCR file

`/dev/sdb2` - 50M - for CRS voting disk

`/dev/sdb3` - 10M - for ASM spfile

- The rest is going to be used as ASM disks (`/dev/sdb4`, `/dev/sdc`, ...)

Check with the command: `sudo /sbin/fdisk -l`



Disks preparing

Binding the created partitions to raw devices (on all nodes)

- Add to `/etc/sysconfig/rawdevices` files:

```
sudo sh -c 'echo "/dev/raw/raw1 /dev/sdb[1-3]" >>
/etc/sysconfig/rawdevices'
```

- Set permissions for these raw devices and restart `rawdevices` service

On all nodes do:

```
sudo chown oracle:ci /dev/raw/raw[1-3]
sudo chmod 660 /dev/raw/raw[1-3]
sudo /sbin/service rawdevices restart
```

Check partitions binding to raw devices: `sudo /usr/bin/raw -qa`



ASM installation

Install & configure ASM library (as root on all nodes)

- `sudo /etc/init.d/oracleasm configure (answers: oracle; ci; y; y)`
- Make sure all devices for ASM are clean (if fails to create disk):
`dd if=/dev/zero of=/dev/sd$letter bs=1024 count=1000`
- For all the storage disks available yet create ASM volumes on them (do it as root on primary node):

```
/etc/init.d/oracleasm createdisk STOR15_1 /dev/sdb4;  
/etc/init.d/oracleasm createdisk STOR15_2 /dev/sdc;  
/etc/init.d/oracleasm createdisk STOR15_3 /dev/sdd;
```



ASM installation ...

- **Enable the created ASM volumes (on all other nodes do):**
`sudo /etc/init.d/oracleasm enable`
`sudo /etc/init.d/oracleasm scandisks`
- **Reload FC switch controller (on all nodes):**
`sudo /sbin/rmmod qla2300;`
`sudo /sbin/modprobe qla2300;`



Hangcheck-timer module

Configure hangcheck-timer module on each node

- Verify that hangcheck-timer module is running on every node: `/sbin/lsmmod`
- If hangcheck-timer is not listed on a node do:
`sudo /sbin/insmod hangcheck-timer`
`hangcheck_tick=30 hangcheck_margin=180`

and

```
echo "options hangcheck-timer  
hangcheck_tick=30 hangcheck_margin=180" >>  
/etc/modules.conf'
```



Environment settings

As oracle

- Set environment adding following lines to ~/.bashrc (and rerun this file):

```
export ORACLE_BASE=/ORA/dbs01/oracle
```

```
export ORACLE_SID=test1 *
```

```
source ~/.bashrc
```

- If necessary, unset ORACLE_HOME and TNS_ADMIN (not set by default)

* "test" is the name of the cluster, "test1" will be the name of the instance on the primary node (from which installation runs)



Installing CRS

To lunch the installation do:

```
cd <CRS_installation_disks>/Disk1/  
./runInstaller
```

- **Inventory directory:** /ORA/dbs01/oracle/oraInventory
- **Unix Group Name:** ci.
- **File Locations:**
 - Destination Name:** OraCRSHome
 - Destination Path:** /ORA/dbs01/oracle/product/10.1.0/crs
- **Cluster Configuration:**
 - Cluster Name:** test
 - Cluster Nodes:** **Public Node Name:** itrac13 **Private Node Name:** atlr-priv1-13
 Public Node Name: itrac14 Private Node Name: atlr-priv1-14
- **Private Interconnect Enforcement:**
 - Interface Name:** eth0 **Interface Type:** Public
 - Interface Name: eth1 Interface Type: Private
 - Interface Name: eth2 Interface Type: Private
- **Oracle Cluster Registry:**
 - OCR Location:** /dev/raw/raw1
- **Voting Disk:**
 - Voting disk file name:** /dev/raw/raw2

If there are problems with CRS, it will create log files in CRS_HOME/log directory, but only if it exists. Therefore make sure to create this directory as oracle.

To verify the CRS installation, display all the nodes where CRS was installed:
oracle\$ CRS_HOME/bin/olsnodes -n



Installing Oracle 10g database software

To launch the installation:

```
cd <Oracle_server_installation_disks>/Disk1/  
./runInstaller
```

- File Locations:

Destination Name: OraRDBMS10gHome

Destination Path: /ORA/dbs01/oracle/product/10.1.0/rdbms

- **Hardware Cluster Installation Mode:** Cluster Installation (and select all nodes)

- **Installation Type:** Enterprise Edition

- **Database Configuration:** don't create database

- **VIP Configuration Assistant Tool:**

Network Interfaces: select all network interfaces

Virtual IPs for cluster nodes:

Node Name: itrac13

IP Alias Name: itrac13-v

IP address: 137.138.216.84

Subnet Mask: 255.255.255.0

(and similarly for other nodes)



Database creation

- Prepare the environment on all RAC nodes:
export
ORACLE_HOME=/ORA/dbs01/oracle/product/10.1.0/rdbms
export PATH=\$PATH:\$ORACLE_HOME/bin
- Lunch dbca (Database Configuration Assistant) to create a database



Create the database - some options

Some options to be chosen:

- Welcome Screen: **Oracle Real Application Clusters database**
- **Operations:** Create Database
- **Node Selection:** Select All (make sure all your RAC nodes show up and are selected)
- **Database Templates:** custom database
- Database Identification:
 - Global Database Name:** test.cern.ch
 - SID Prefix:** test
- **Management Option:** Don't check configuration with Enterprise Manager at all
- **Database Credentials:** Use the Same Password for All Accounts
- **Storage Options: ASM**
- **Create ASM Instance:**
 - enter the SYS password for the ASM instance.
 - create ASM SPfile:** /dev/raw/raw3
 - (dbca will create and start the ASM instance on all RAC nodes)
- **ASM Disk Groups:**
 - Create Disk Group(s):**
 - enter an arbitrary **Disk Group Name:** DG_DATA1
 - choose disks for this group:** choose from the list of created ASM disks
- **Database File Locations:**
 - use Oracle-Managed Files**
 - Database Area:** +DG_DATA1
- **Database Services:** one can create services
- **Initialization Parameters:** change settings as needed
- **Database Storage:** change settings as needed



Summary

- Interesting web pages:
 - <http://www.puschitz.com/InstallingOracle10gRAC.shtml>
 - <https://uimon.cern.ch/twiki/bin/view/ADCgroup/RacOnLinux>