OCFS (Oracle Cluster File System)

Para este ejemplo particular implementaré una solución de Cluster Filesystem llamada OCFS versión 2.0.

OCFS es una solución de Oracle 100% gratuita para plataformas Linux y Windows.

La implementación será realizada con VMWare Server 1.8 y Oracle Enterprise Linux 4.7.

La utilización de VMWare es netamente educativo, su finalidad en este ejercicio es simular el storage compartido entre dos servidores.

Paso 1.- Configurar adecuadamente los servidores.

a) Ambos servidores deben tener dos tarjetas de red, uno para la conexión de los usuarios al servidor y otra tarjeta para una conexión privada, la cual tiene como finalidad que cada servidor pueda comunicarse con el otro servidor y pueda además asegurarse que el otro servidor vive. **Si utilizamos OCFS para implementar Oracle RAC podemos utilizar la misma red privada para la interconnect del clusterware.**

En mi ejemplo: El servidor 1 tiene el IP Público (192.68.1.5) e IP Privado (20.0.0.5) y el servidor 2 tiene el IP Público (192.68.1.6) y el IP Privado (20.0.0.6).

b) Ambos servidores deben tener un storage común, es decir ambos servidores deben poder leer y escribir sobre un mismo conjunto de discos al mismo tiempo.

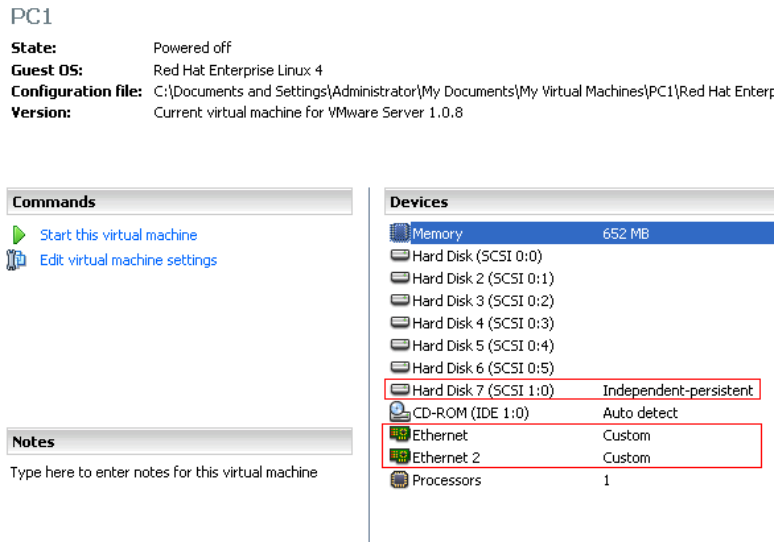

Figura 1

En la implementación de VMWare como muestro en la figura 1, he colocado 2 tarjetas de red y además un disco que será compartido entre ambos servidores, como característica este disco debe tener la propiedad de Independent-persistent y además debe estar configurado en otra tarjeta controladora (Ejemplo: SCSI 1) respecto a los demás discos internos (SCSI 0).

Adicional debo indicar que los discos serán compartidos y esto se realiza **en cada archivo de configuración de VMWare (*.vmx) de cada servidor.**

Ejemplo:

```
scsi1:0.shared = "TRUE"
scsi1.sharedBus = "virtual"

disk.locking = "FALSE"
diskLib.dataCacheMaxSize = "0"
diskLib.dataCacheMaxReadAheadSize = "0"
diskLib.dataCacheMinReadAheadSize = "0"
diskLib.dataCachePageSize = "0"
```

Colocamos scsi1:0 porque es el disco que estamos compartiendo entre los servidores.

Paso 2.- Particionamiento del disco compartido.

Vemos todas las particiones con el comando: **fdisk -l**, en nuestro ejemplo vemos que

tenemos asignado un nuevo disco "sdg" sin partición.

```
Disk /dev/sdg: 10.7 GB, 10737418240 bytes
255 heads, 63 sectors/track, 1305 cylinders
Units = cylinders of 16065 * 512 = 8225280 bytes

Disk /dev/sdg doesn't contain a valid partition table
```

Lo particionamos (fdisk /dev/sdg), **solo es necesario realizarlo en un servidor**:

```
[root@pcoracle1 ~]# fdisk /dev/sdg
Device contains neither a valid DOS partition table, nor Sun, SGI or OSF disklabel
Building a new DOS disklabel. Changes will remain in memory only,
until you decide to write them. After that, of course, the previous
content won't be recoverable.


The number of cylinders for this disk is set to 1305.
There is nothing wrong with that, but this is larger than 1024,
and could in certain setups cause problems with:
1) software that runs at boot time (e.g., old versions of LILO)
2) booting and partitioning software from other OSs
   (e.g., DOS FDISK, OS/2 FDISK)
Warning: invalid flag 0x0000 of partition table 4 will be corrected by w(rite)

Command (m for help): n
Command action
   e   extended
   p   primary partition (1-4)
p
Partition number (1-4): 1
First cylinder (1-1305, default 1):
Using default value 1
Last cylinder or +size or +sizeM or +sizeK (1-1305, default 1305):
Using default value 1305

Command (m for help): w
The partition table has been altered!

Calling ioctl() to re-read partition table.
Syncing disks.
[root@pcoracle1 ~]#
```

Paso 3.- Instalación del Software OCFS.

Debemos descargar el software OCFS de acuerdo a nuestra versión de kernel de Linux (**uname -a**) en la siguiente dirección: http://oss.oracle.com/projects/

En el caso de Oracle Enterprise Linux los rpm de OCFS se encuentran en la media del producto.

Para saber la plataforma con que estamos corriendo podemos ejecutar el siguiente comando:

rpm -qf /boot/vmlinuz-`uname -r` --queryformat "%{ARCH}\n"

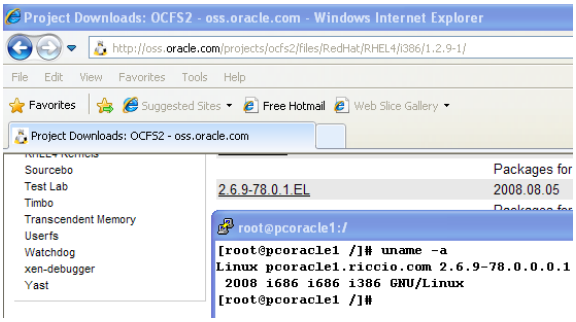

Figura 2

En nuestro escenario iniciaremos la instalación en ambos servidores:

```
[root@pcoracle1 RPMS]# ls *ocfs*
ocfs2-2.6.9-78.0.0.0.1.EL-1.2.9-1.el4.i686.rpm          ocfs2console-1.2.7-1.el4.i386.rpm
ocfs2-2.6.9-78.0.0.0.1.ELhugemem-1.2.9-1.el4.i686.rpm   ocfs2-tools-1.2.7-1.el4.i386.rpm
ocfs2-2.6.9-78.0.0.0.1.ELsmp-1.2.9-1.el4.i686.rpm       ocfs2-tools-devel-1.2.7-1.el4.i386.rpm
ocfs2-2.6.9-78.0.0.0.1.ELxenU-1.2.9-1.el4.i686.rpm
[root@pcoracle1 RPMS]# rpm -ivh ocfs2-tools-1.2.7-1.el4.i386.rpm
warning: ocfs2-tools-1.2.7-1.el4.i386.rpm: V3 DSA signature: NOKEY, key ID b38a8516
Preparing...                ########################################### [100%]
   1:ocfs2-tools            ########################################### [100%]
[root@pcoracle1 RPMS]# rpm -ivh ocfs2-2.6.9-78.0.0.0.1.EL-1.2.9-1.el4.i686.rpm
warning: ocfs2-2.6.9-78.0.0.0.1.EL-1.2.9-1.el4.i686.rpm: V3 DSA signature: NOKEY, key ID b38a8516
Preparing...                ########################################### [100%]
   1:ocfs2-2.6.9-78.0.0.0.1.########################################### [100%]
[root@pcoracle1 RPMS]# rpm -ivh ocfs2-2.6.9-78.0.0.0.1.ELsmp-1.2.9-1.el4.i686.rpm
warning: ocfs2-2.6.9-78.0.0.0.1.ELsmp-1.2.9-1.el4.i686.rpm: V3 DSA signature: NOKEY, key ID b38a8516
Preparing...                ########################################### [100%]
   1:ocfs2-2.6.9-78.0.0.0.1.########################################### [100%]
[root@pcoracle1 RPMS]# pwd
/mnt/Enterprise/RPMS
[root@pcoracle1 RPMS]#
```

Paso 4.- Formateamos la partición en formato OCFS.

```
[root@pcoracle1 /]# mkfs -t ocfs2 /dev/sdg1
mkfs.ocfs2 1.2.7
Filesystem label=
Block size=4096 (bits=12)
Cluster size=4096 (bits=12)
Volume size=10733957120 (2620595 clusters) (2620595 blocks)
82 cluster groups (tail covers 7859 clusters, rest cover 32256 clusters)
Journal size=67108864
Initial number of node slots: 4
Creating bitmaps: done
Initializing superblock: done
Writing system files: done
Writing superblock: done
Writing backup superblock: 2 block(s)
Formatting Journals: done
Writing lost+found: done
mkfs.ocfs2 successful
```

Paso 5.- Creamos el archivo de configuración del OCFS en ambos servidores.

```
[root@pcoracle1 ocfs2]# mkdir -p /etc/ocfs2
[root@pcoracle1 ocfs2]# cd /etc/ocfs2
[root@pcoracle1 ocfs2]# vi cluster.conf
[root@pcoracle1 ocfs2]# cat cluster.conf
cluster:
        node_count=2
        name=ocfs2

node:
        ip_port=7777
        ip_address=20.0.0.5
        number=1
        name=pcoracle1
        cluster=ocfs2

node:
        ip_port=7777
        ip_address=20.0.0.6
        number=2
        name=pcoracle2
        cluster=ocfs2
[root@pcoracle1 ocfs2]#
```

Esto debe ser realizado en ambos servidores.

Paso 6.- Iniciamos la configuración

Ejecutamos en ambos nodos: **/etc/init.d/o2cb configure**

```
[root@pcoracle1 /]# /etc/init.d/o2cb configure
Configuring the O2CB driver.

This will configure the on-boot properties of the O2CB driver.
The following questions will determine whether the driver is loaded on
boot.  The current values will be shown in brackets ('[]').  Hitting
<ENTER> without typing an answer will keep that current value.  Ctrl-C
will abort.

Load O2CB driver on boot (y/n) [n]:
Cluster to start on boot (Enter "none" to clear) [ocfs2]:
Specify heartbeat dead threshold (>=7) [31]:
Specify network idle timeout in ms (>=5000) [30000]:
Specify network keepalive delay in ms (>=1000) [2000]:
Specify network reconnect delay in ms (>=2000) [2000]:
Writing O2CB configuration: OK
[root@pcoracle1 /]#
```

Dejamos los valores por default (Todo indicamos con <ENTER>).

Paso 7.- Habilitamos el servicio de OCFS (Nos permite registrar el servicio O2CB).

Ejecutamos en ambos nodos: **/etc/init.d/o2cb enable**

```
[root@pcoracle2 ocfs2]# /etc/init.d/o2cb enable
Writing O2CB configuration: OK
Loading module "configfs": OK
Mounting configfs filesystem at /config: OK
Loading module "ocfs2_nodemanager": OK
Loading module "ocfs2_dlm": OK
Loading module "ocfs2_dlmfs": OK
Mounting ocfs2_dlmfs filesystem at /dlm: OK
Starting O2CB cluster ocfs2: OK
[root@pcoracle2 ocfs2]#
```

Paso 8.- Montando los discos en ambos servidores.

```
[root@pcoracle1 /]# mkdir -p /data
[root@pcoracle1 /]# date
Mon Dec 21 07:37:39 PET 2009
[root@pcoracle1 /]# mount -t ocfs2 /dev/sdg1 /data
[root@pcoracle1 /]#
```

```
[root@pcoracle2 /]# mkdir -p /data
[root@pcoracle2 /]# date
Mon Dec 21 07:38:05 PET 2009
[root@pcoracle2 /]# mount -t ocfs2 /dev/sdg1 /data
[root@pcoracle2 /]#
```

Probando:

Servidor 1 (Copiando algunos archivos al disco compartido).

```
[root@pcoracle1 /]# ls -lrt /data
total 12937
drwxr-xr-x  2 root root     4096 Dec 21 06:33 lost+found
-rw-r--r--  1 root root   330867 Dec 21 07:40 ocfs2-2.6.9-78.EL-1.2.9-1.el4.i686.rpm
-rw-r--r--  1 root root 11435972 Dec 21 07:40 ocfs2-2.6.9-78.EL-debuginfo-1.2.9-1.el4.i686.rpm
-rw-r--r--  1 root root   305344 Dec 21 07:40 ocfs2-2.6.9-78.ELsmp-1.2.9-1.el4.i686.rpm
-rw-r--r--  1 root root  1044312 Dec 21 07:40 ocfs2-tools-1.2.7-1.el4.i386.rpm
-rw-r--r--  1 root root   125810 Dec 21 07:40 ocfs2-tools-debuginfo-1.2.7-1.el4.i386.rpm
[root@pcoracle1 /]#
```

Servidor 2 (Revisando los archivos copiados)

```
[root@pcoracle2 /]# ls -lrt /data
total 12937
drwxr-xr-x  2 root root     4096 Dec 21 06:33 lost+found
-rw-r--r--  1 root root   330867 Dec 21 07:40 ocfs2-2.6.9-78.EL-1.2.9-1.el4.i686.rpm
-rw-r--r--  1 root root 11435972 Dec 21 07:40 ocfs2-2.6.9-78.EL-debuginfo-1.2.9-1.el4.i686.rpm
-rw-r--r--  1 root root   305344 Dec 21 07:40 ocfs2-2.6.9-78.ELsmp-1.2.9-1.el4.i686.rpm
-rw-r--r--  1 root root  1044312 Dec 21 07:40 ocfs2-tools-1.2.7-1.el4.i386.rpm
-rw-r--r--  1 root root   125810 Dec 21 07:40 ocfs2-tools-debuginfo-1.2.7-1.el4.i386.rpm
[root@pcoracle2 /]#
```

Como se puede apreciar ambos servidores pueden leer y escribir en un mismo storage, el cual es la limitante para un filesystem como EXT3 o NTFS. Además nos permite crear un cluster activo / activo.

Con esta implementación podríamos ya empezar la instalación de un cluster de Oracle (Clusterware).

Espero que este artículo pueda complementar lo aprendido en el artículo anterior y le sea útil si están comenzado un proyecto de alta disponibilidad.

Podemos encontrar la información oficial de OCFS en la siguiente página: http://oss.oracle.com/projects/ocfs2/