

Lustre

Desarrollo de un sistema de ficheros Lustre

Grado en Ingeniería Informática

Nombre: David Sánchez Montés

Asignatura: Servidores Web de Altas Prestaciones

Índice

- ¿Qué es Lustre?
 - Descripción
 - Arquitectura
 - Otras funcionalidades
- Instalación y configuración de Lustre.
 - Punto de partida
 - Instalación servidores
 - Configurar MGS y MDS
 - Configurar OSS
 - Configurar clientes
 - Demostración

¿Qué es Lustre?-Descripción

- Lustre es un sistema de archivos paralelos(E/S de datos en paralelo).
- El nombre viene de mezclar Linux y clúster.
- Presenta a los clientes una semántica POSIX
- Emplea una arquitectura de red cliente-servidor.
- Permite escalabilidad, alto rendimiento y alta disponibilidad.
- Lustre no implementa un almacenamiento redundante.

¿Qué es Lustre?-Descripción

- Capacidades:

Tamaño máximo del objeto: **16TB**

Tamaño máximo de archivo: **31.25PB**

Tamaño máximo del sistema de archivos: **512PB**

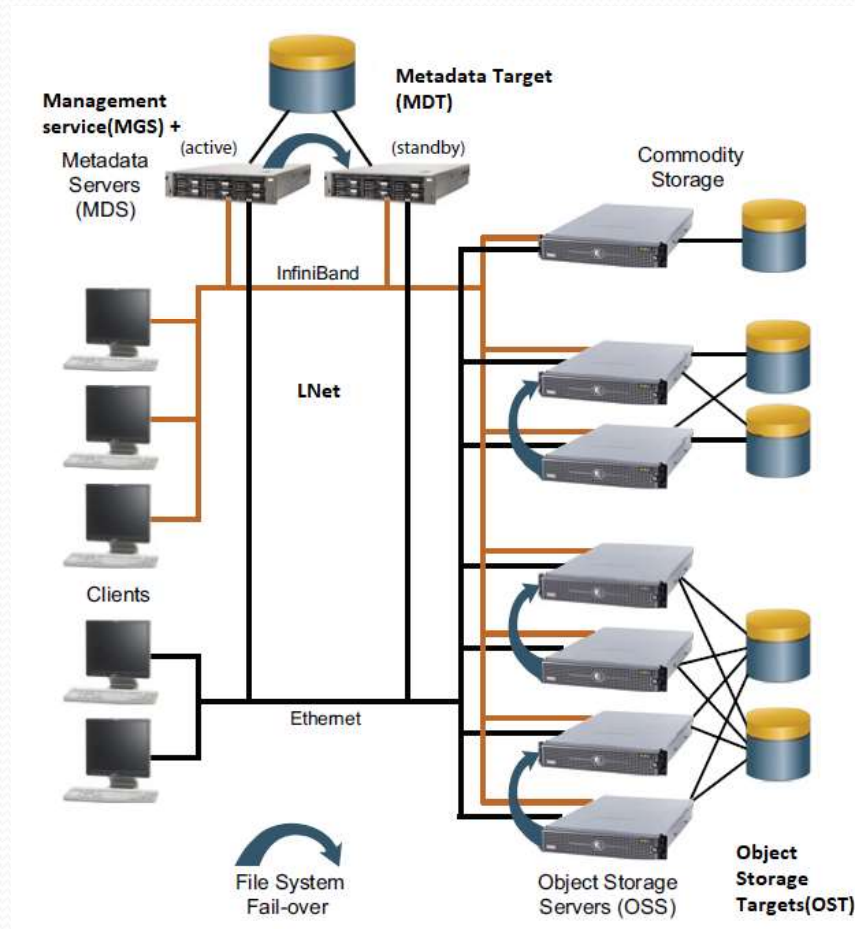
Nº máximo de archivos en el sistema de archivos:

4 mil millones por MDT

¿Qué es Lustre?-Descripción

- Un sistema de archivos Lustre tiene estas unidades a destacar que son:
 - Management service (MGS)
 - Metadata server (MDS)
 - Metadata targets (MDT)
 - Object storage server (OSS)
 - Object storage target (OST)

¿Qué es Lustre?-Arquitectura



¿Qué es Lustre?-Arquitectura

- Los componentes principales de un sistema de archivos Lustre son los servidores de metadatos(MDS) y los servidores de almacenamiento de objetos(OSS), que proporcionan operaciones con namespaces y servicios de E/S de forma masiva.
- Un MDS y un OSS no pueden estar en el mismo nodo, deben estar en nodos diferentes.
- El servidor de metadatos(MDS) guía las solicitudes de E/S del archivo real al servidor de almacenamiento(OSS).

¿Qué es Lustre?-Arquitectura

- **Management service (MGS)**

El MGS es el servidor de gestión de todo el sistema.
Sólo puede haber uno por sistema.

- **Metadata targets (MDT)**

Un MDT almacena metadatos del namespace, como nombres de archivos, directorios, permisos de acceso.

¿Qué es Lustre?-Arquitectura

- **Metadata server (MDS)**

El MDS controla la asignación de objetos de almacenamiento en los servidores de almacenamiento de objetos(OSS).

Administra todas las operaciones para un sistema de archivos Lustre.

- **Object storage target (OST)**

Son los que guardan los datos reales.

Sueles ser discos, particiones, RAIDS, volúmenes lógicos.

Son administrados por el OSS.

¿Qué es Lustre?-Arquitectura

- **Object storage server (OSS)**

Uno o más servidores de almacenamiento de objetos(OSS) almacenan archivos en uno o más objetivos de almacenamiento de objetos(OST).

Entre 2 y 8 OST.

- **Clientes**

Lustre se presenta como un punto de montaje separado en el sistema operativo del cliente.

¿Qué es Lustre?-Arquitectura

- *Red de Lustre(LNet)*

Protocolo de red de alta velocidad.

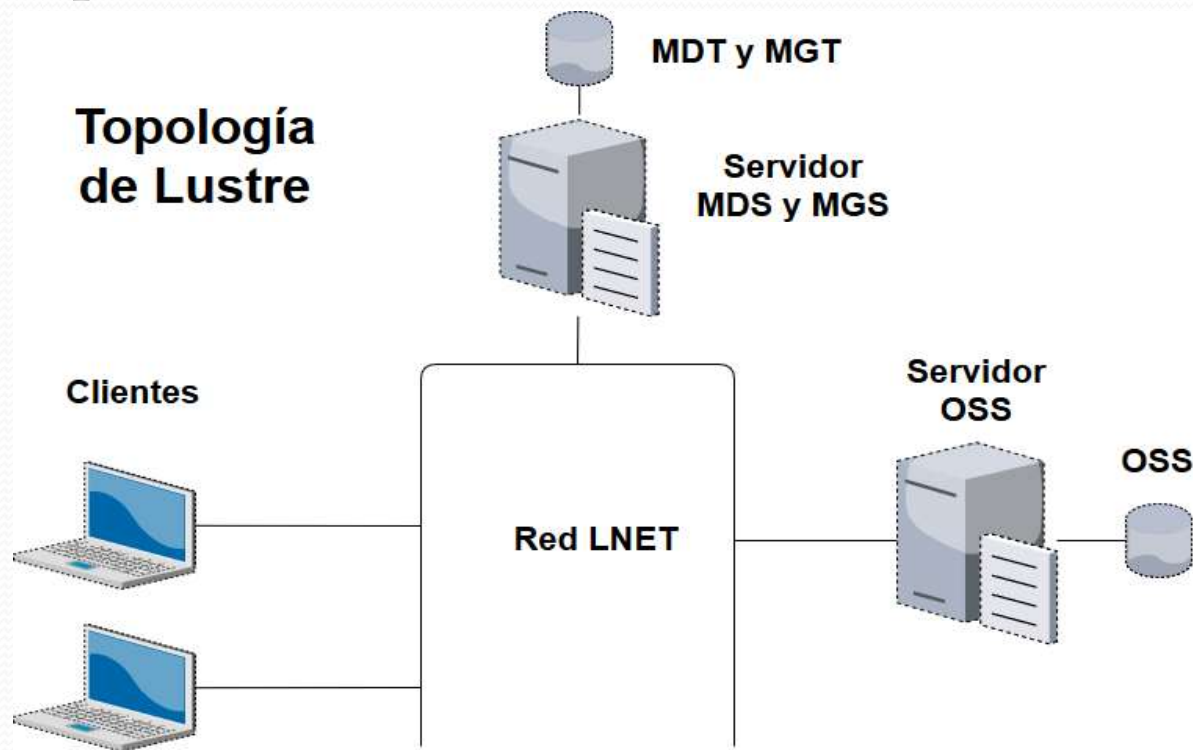
Abstrae los detalles de la red.

¿Qué es Lustre?-Funcionalidades

- Gran cantidad de funcionalidades:
 - Cuotas de disco
 - Lista de control de acceso
 - Failover
 - Striping (RAID O)
 - Administrar logs
 - Enrutamiento
 - Copias y restauración
 - Redundancia

Instalación – Punto de partida

- Diseño para la demostración:



Instalación – Instalación servidores

- Quitar firewall

\$ service iptables stop

- Desactivar SELinux

\$ vi /etc/sysconfig/selinux

```
# This file controls the state of SELinux on the system.
# SELINUX= can take one of these three values:
#     enforcing - SELinux security policy is enforced.
#     permissive - SELinux prints warnings instead of enforcing.
#     disabled - No SELinux policy is loaded.
SELINUX=disabled
# SELINUXTYPE= can take one of these two values:
#     targeted - Targeted processes are protected,
#     mls - Multi Level Security protection.
SELINUXTYPE=targeted
```


Instalación – Instalación servidores

- URL de paquetes:

https://downloads.hpdd.intel.com/public/e2fsprogs/1.42.12.wc1/el6/RPMS/x86_64/

https://downloads.hpdd.intel.com/public/lustre/lustre-2.7.0/el6/server/RPMS/x86_64/

- Paquetes necesarios:

```
[root@localhost rpm_servidor1]# ls
e2fsprogs-1.42.12.wc1-7.el6.x86_64.rpm
e2fsprogs-libs-1.42.12.wc1-7.el6.x86_64.rpm
kernel-2.6.32-504.8.1.el6_lustre.x86_64.rpm
kernel-firmware-2.6.32-504.8.1.el6_lustre.x86_64.rpm
libcom_err-1.42.12.wc1-7.el6.x86_64.rpm
libss-1.42.12.wc1-7.el6.x86_64.rpm
lustre-2.7.0-2.6.32_504.8.1.el6_lustre.x86_64.x86_64.rpm
lustre-modules-2.7.0-2.6.32_504.8.1.el6_lustre.x86_64.x86_64.rpm
lustre-osd-ldiskfs-2.7.0-2.6.32_504.8.1.el6_lustre.x86_64.x86_64.rpm
lustre-osd-ldiskfs-mount-2.7.0-2.6.32_504.8.1.el6_lustre.x86_64.x86_64.rpm
[root@localhost rpm_servidor1]# _
```

Instalación – Instalación servidores

- Instalación:

```
$ rpm -U kernel-firmware*
```

```
$ yum install kernel-2.6.*
```

```
$ yum install lustre-modules-*
```

```
$ yum install lustre-osd-ldiskfs-mount-*
```

```
$ rpm -Uhv e2fs* lib*
```

```
$ yum install lustre-osd-ldiskfs-2.7*
```

```
$ yum install lustre-2.7.*
```


Instalación – Instalación servidores

- Configuración de la LNet:

```
$ echo "options lnet networks=tcp" >  
/etc/modprobe.d/lnet.conf
```

```
[root@localhost ServidorMDS]# cat /etc/modprobe.d/lnet.conf  
options lnet networks=tcp  
[root@localhost ServidorMDS]#
```

Instalación – Instalación servidores

- Automatización:

```
$ chkconfig lnet --add
```

```
$ chkconfig lnet on
```

```
$ chkconfig lustre --add
```

```
$ chkconfig lustre on
```

- Y listo:

```
[root@localhost ServidorMDS]# uname -a
Linux localhost.localdomain 2.6.32-504.8.1.el6_lustre.x86_64 #1 SMP Thu Feb 26 2
2:39:27 PST 2015 x86_64 x86_64 x86_64 GNU/Linux
[root@localhost ServidorMDS]# _
```


Instalación – Configuración MGS

MDS

- Formato MDT:

\$ mkfs.lustre --fsname=DATA --mgs --mdt --reformat /dev/sdb

```
[root@localhost ~]# mkfs.lustre --fsname=DATA --mgs --mdt --reformat /dev/sdb
warning: /dev/sdb: for Lustre 2.4 and later, the target index must be specified
with --index

Permanent disk data:
Target:      DATA:MDT0000
Index:       0
Lustre FS:   DATA
Mount type:  ldiskfs
Flags:       0x65
              (MDT MGS first_time update )
Persistent mount opts: user_xattr,errors=remount-ro
Parameters:

device size = 5120MB
formatting backing filesystem ldiskfs on /dev/sdb
      target name  DATA:MDT0000
      4k blocks    1310720
      options      -J size=204 -I 512 -i 2048 -q -O dirdata,uninit_bg,^exten
ts,dir_nlink,quota,huge_file,flex_bg -E lazy_journal_init -F
mkfs_cmd = mke2fs -j -b 4096 -L DATA:MDT0000 -J size=204 -I 512 -i 2048 -q -O d
irdata,uninit_bg,^extents,dir_nlink,quota,huge_file,flex_bg -E lazy_journal_init
-F /dev/sdb 1310720
Writing CONFIGS/mountdata
[root@localhost ~]# _
```

Instalación – Configuración MGS

MDS

- Y montamos la unidad:

\$ mount -t lustre /dev/sdb /mnt/MDS

Comprobación:

```
[root@localhost ServidorMDS1# mount
/dev/mapper/VolGroup-lv_root on / type ext4 (rw)
proc on /proc type proc (rw)
sysfs on /sys type sysfs (rw)
devpts on /dev/pts type devpts (rw,gid=5,mode=620)
tmpfs on /dev/shm type tmpfs (rw)
/dev/sda1 on /boot type ext4 (rw)
none on /proc/sys/fs/binfmt_misc type binfmt_misc (rw)
/dev/sdb on /mnt/MDS type lustre (rw)
[root@localhost ServidorMDS1# df -h
Filesystem              Size  Used Avail Use% Mounted on
/dev/mapper/VolGroup-lv_root
                        6,5G  913M  5,3G  15% /
tmpfs                    246M    0  246M   0% /dev/shm
/dev/sda1                477M   61M  392M  14% /boot
/dev/sdb                 3,6G   24M   3,3G   1% /mnt/MDS
[root@localhost ServidorMDS1#
```


Instalación – Configuración MGS MDS

- Inicializar la LNet:
\$ modprobe lneth
\$ lctl network up
\$ lctl list_nids (*Este comando sirve para ver que está activo*)
- Y arrancamos Lustre:
\$ service lustre start

Instalación – Configuración OSS

- Formato MDT:

```
$ mkfs.lustre --fsname=DATA --ost --  
mgsnode=192.168.1.10 --index=1 --reformat  
/dev/sdb
```

```
[root@localhost ~]# mkfs.lustre --fsname=DATA --ost --mgsnode=192.168.1.10 --index=1 /dev/sdb  
  
Permanent disk data:  
Target:      DATA:OST0001  
Index:       1  
Lustre FS:   DATA  
Mount type:  ldiskfs  
Flags:       0x62  
              (OST first_time update )  
Persistent mount opts: errors=remount-ro  
Parameters:  mgsnode=192.168.1.10@tcp  
  
checking for existing Lustre data: not found  
device size = 5120MB  
formatting backing filesystem ldiskfs on /dev/sdb  
      target name  DATA:OST0001  
      4k blocks    1310720  
      options      -J size=204 -I 256 -q -O extents,uninit_bg,dir_nlink,quota,huge_file,flex_bg -G 256 -E resize=4290772992,lazy_journal_init -F  
mkfs_cmd = mke2fs -j -b 4096 -L DATA:OST0001 -J size=204 -I 256 -q -O extents,uninit_bg,dir_nlink,quota,huge_file,flex_bg -G 256 -E resize=4290772992,lazy_journal_init -F /dev/sdb 1310720  
Writing CONFIGS/mountdata  
[root@localhost ~]#
```


Instalación – Configuración OSS

- Y montamos la unidad:

\$ mount -t lustre /dev/sdb /mnt/OST1

Comprobación:

```
[root@localhost ServidorOSS1# mount
/dev/mapper/VolGroup-lv_root on / type ext4 (rw)
proc on /proc type proc (rw)
sysfs on /sys type sysfs (rw)
devpts on /dev/pts type devpts (rw,gid=5,mode=620)
tmpfs on /dev/shm type tmpfs (rw)
/dev/sda1 on /boot type ext4 (rw)
none on /proc/sys/fs/binfmt_misc type binfmt_misc (rw)
/dev/sdb on /mnt/OST1 type lustre (rw)
[root@localhost ServidorOSS1# df -h
Filesystem                Size      Used Avail Use% Mounted on
/dev/mapper/VolGroup-lv_root
                           6,5G    913M    5,3G   15% /
tmpfs                      246M         0   246M    0% /dev/shm
/dev/sda1                  477M     61M   392M   14% /boot
/dev/sdb                   4,8G     34M   4,5G    1% /mnt/OST1
[root@localhost ServidorOSS1# _
```

Instalación – Configuración MGS MDS

- Inicializar la LNet:
\$ modprobe lneth
\$ lctl network up
\$ lctl list_nids (*Este comando sirve para ver que está activo*)
- Y arrancamos Lustre:
\$ service lustre start

Instalación – Configuración cliente

- URL de los paquetes:

https://downloads.hpdd.intel.com/public/lustre/lustre-2.7.0/el6/client/RPMS/x86_64/

- Cuidado con el Kernel

```
[root@localhost rpm_client]# ls
kernel-2.6.32-504.8.1.el6.x86_64.rpm
lustre-client-2.7.0-2.6.32_504.8.1.el6.x86_64.x86_64.rpm
lustre-client-modules-2.7.0-2.6.32_504.8.1.el6.x86_64.x86_64.rpm
[root@localhost rpm_client]#
```

Instalación – Configuración cliente

- Instalación:

\$ yum install lustre-client-modules*

\$ yum install lustre-client-2.7*

```
[root@localhost rpm_client1]# uname -a
Linux localhost.localdomain 2.6.32-504.8.1.el6.x86_64 #1 SMP Tue Jan 27 13:39:10
CST 2015 x86_64 x86_64 x86_64 GNU/Linux
[root@localhost rpm_client1]# _
```


Instalación – Configuración cliente

- Montamos la unidad

**\$ mount -t lustre 192.168.1.10@tcp:/DATA
/mnt/cliente**

```
[root@localhost rpm_client1]# mount
/dev/mapper/VolGroup-lv_root on / type ext4 (rw)
proc on /proc type proc (rw)
sysfs on /sys type sysfs (rw)
devpts on /dev/pts type devpts (rw,gid=5,mode=620)
tmpfs on /dev/shm type tmpfs (rw,rootcontext="system_u:object_r:tmpfs_t:s0")
/dev/sda1 on /boot type ext4 (rw)
none on /proc/sys/fs/binfmt_misc type binfmt_misc (rw)
192.168.1.10@tcp:/DATA on /mnt/cliente type lustre (rw)
[root@localhost rpm_client1]# df -h
Filesystem                Size      Used Avail Use% Mounted on
/dev/mapper/VolGroup-lv_root
6,5G    922M    5,3G   15% /
tmpfs
246M         0    246M    0% /dev/shm
/dev/sda1
477M     56M   396M   13% /boot
192.168.1.10@tcp:/DATA
4,8G     34M   4,5G    1% /mnt/cliente
[root@localhost rpm_client1]#
```



Instalación – Demostración

- Procedemos a la demo