

Course: CSCI660
Operating System Concepts & Theory

Lab 2 Report

Installation and Configuration of Lustre Filesystem

Submitted By:
Uzzal Das: 003650867
Nikhil Kasireddy: 003889950

Instructor: Dr. Tong Lai Yu

Date:05/10/2011

Purpose:

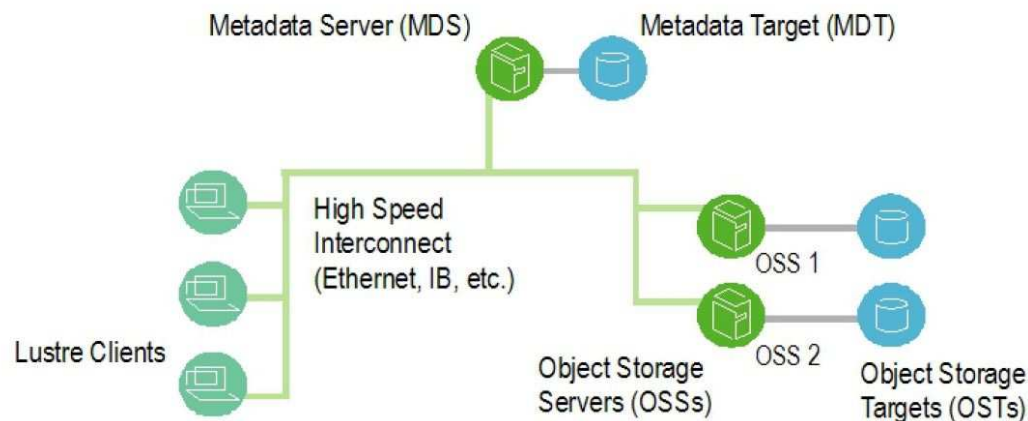
In this lab, we will install and configure the Lustre File System. Lustre is an open source and well defined storage management system widely used in a cluster. Lustre has many advantages such as maintaining a single data copy on multiple file systems, scalability, dynamic throughput, etc. It can handle a file system with large number of computer systems and provide a higher I/O throughput as needed.

Architecture:

Lustre file system has different basic components:

1. Metadata Server (MDS): Handles names and directories of a Lustre file system.
2. Metadata Target (MDT): usually stores filenames, directories, permissions and file layout on MDS.
3. Object Storage Servers (OSS): Handles file I/O service and network request between local OSTs.
4. Object Storage Target (OST): It helps to stores chunks of user files as data objects on one or more OSSs.
5. Lustre Clients: The system which runs Lustre software to mount the Lustre file system and operate.

The following diagram illustrates the concept:



Source: Lustre 1.8 Operations Manual

To make Lustre work in different server and clients, Lustre Networking (LNET) API is needed which act as a heterogeneous interface between different machines.

Downloading package:

Lustre 1.8.5 rpm package is available for download through www.lustre.org. One can register and download the file for a particular platform. For CentOS, we choose Red Hat Linux 5, i686 as the platform. The list of required rpm packages to be downloaded is:

```
e2fsprogs-1.41.10.sun2-0redhat.rhel5.i386.rpm
kernel-2.6.18-194.17.1.el5_lustre.1.8.5.i686.rpm
lustre-1.8.5-2.6.18_194.17.1.el5_lustre.1.8.5.i686.rpm
lustre-client-1.8.5-2.6.18_194.17.1.el5_lustre.1.8.5.i686.rpm
lustre-client-modules-1.8.5-2.6.18_194.17.1.el5_lustre.1.8.5.i686.rpm
lustre-ldiskfs-3.1.4-2.6.18_194.17.1.el5_lustre.1.8.5.i686.rpm
lustre-modules-1.8.5-2.6.18_194.17.1.el5_lustre.1.8.5.i686.rpm
```

The following rpms are to be installed on the server:

- kernel-2.6.18-194.17.1.el5_lustre.1.8.5.i686.rpm
- lustre-modules-1.8.5-2.6.18_194.17.1.el5_lustre.1.8.5.i686.rpm
- lustre-ldiskfs-3.1.4-2.6.18_194.17.1.el5_lustre.1.8.5.i686.rpm
- lustre-1.8.5-2.6.18_194.17.1.el5_lustre.1.8.5.i686.rpm

A directory in a server can also act as an client file system. In that case, there is no requirements to install any packages. But to use a remote machine as client following rpm packages should be installed.

- kernel-2.6.18-194.17.1.el5_lustre.1.8.5.i686.rpm
- lustre-client-1.8.5-2.6.18_194.17.1.el5_lustre.1.8.5.i686.rpm
- lustre-client-modules-1.8.5-2.6.18_194.17.1.el5_lustre.1.8.5.i686.rpm

Other optional packages and tools are also available for downloading from the website.

Installation:

Installing the server

Let's check the kernel version. The current version will be updated with Lustre based kernel after the Lustre is installed.

```
[root@localhost ~]# uname -r
2.6.18-194.el5
```

```
[root@localhost lustre-install]# ll /lib/modules/
total 8
drwxr-xr-x 6 root root 4096 Apr 19 16:09 2.6.18-194.el5
```

Server machine configuration:

```
[root@localhost ~]# ifconfig
eth0  Link encap:Ethernet  HWaddr 08:00:27:43:C9:0B
      inet addr:192.168.56.101  Bcast:192.168.56.255
      Mask:255.255.255.0
      inet6 addr: fe80::a00:27ff:fe43:c90b/64 Scope:Link
      UP BROADCAST RUNNING MULTICAST  MTU:1500  Metric:1
      RX packets:67 errors:0 dropped:0 overruns:0 frame:0
      TX packets:40 errors:0 dropped:0 overruns:0 carrier:0
      collisions:0 txqueuelen:1000
      RX bytes:10539 (10.2 KiB)  TX bytes:6916 (6.7 KiB)
      Interrupt:10 Base address:0xd020

eth1  Link encap:Ethernet  HWaddr 08:00:27:61:EA:A3
      inet addr:10.0.3.15  Bcast:10.0.3.255  Mask:255.255.255.0
      inet6 addr: fe80::a00:27ff:fe61:eea3/64 Scope:Link
      UP BROADCAST RUNNING MULTICAST  MTU:1500  Metric:1
      RX packets:6 errors:0 dropped:0 overruns:0 frame:0
      TX packets:42 errors:0 dropped:0 overruns:0 carrier:0
      collisions:0 txqueuelen:1000
      RX bytes:1442 (1.4 KiB)  TX bytes:6990 (6.8 KiB)
      Interrupt:9 Base address:0xd240

lo    Link encap:Local Loopback
      inet addr:127.0.0.1  Mask:255.0.0.0
      inet6 addr: ::1/128 Scope:Host
      UP LOOPBACK RUNNING  MTU:16436  Metric:1
      RX packets:0 errors:0 dropped:0 overruns:0 frame:0
      TX packets:0 errors:0 dropped:0 overruns:0 carrier:0
      collisions:0 txqueuelen:0
      RX bytes:0 (0.0 b)  TX bytes:0 (0.0 b)
```

Disk partitions:

```
[root@localhost lustre-install]# df -h
Filesystem                Size      Used Avail Use% Mounted on
/dev/mapper/VolGroup00-LogVol00
                          3.9G  1.7G  2.0G  46% /
/dev/hda3                  965M    18M   898M   2% /lus
/dev/hda2                  965M    18M   898M   2% /lus2
/dev/hda1                   99M    12M    82M  13% /boot
tmpfs                      506M     0   506M   0% /dev/shm
```

Two extra ext3 partitions lus and lus2 were created earlier. These two partitions will be used for MDS and OSS for the Lustre file system. At first, these two partitions should be unmounted.

```
[root@localhost Lustre]# umount /dev/hda2
[root@localhost Lustre]# umount /dev/hda3
```

The rpm packages should be installed in the server in a particular order as follows:

First the luster patched kernel is installed along with the luster-modules and luster-ldiskfs

```
[root@localhost Lustre]# rpm -ivh kernel-2.6.18-
194.17.1.el5_lustre.1.8.5.i686.rpm lustre-modules-1.8.5-
2.6.18_194.17.1.el5_lustre.1.8.5.i686.rpm lustre-ldiskfs-3.1.4-
2.6.18_194.17.1.el5_lustre.1.8.5.i686.rpm
Preparing...
##### [100%]
1:lustre-ldiskfs
##### [ 33%]
2:kernel
##### [ 67%]
3:lustre-modules
##### [100%]
Congratulations on finishing your Lustre installation! To register
your copy of Lustre and find out more about Lustre Support, Service,
and Training offerings please visit
http://www.sun.com/software/products/lustre/lustre_reg.jsp
```

Then the luster user tools and utilities should be installed:

```
[root@localhost Lustre]# rpm -ivh lustre-1.8.5-
2.6.18_194.17.1.el5_lustre.1.8.5.i686.rpm
Preparing...
##### [100%]
1:lustre
##### [100%]
```

At last, e2fsprogs package should be updated:

```
[root@localhost Lustre]# rpm -U e2fsprogs-1.41.10.sun2-
0redhat.rhel5.i386.rpm
```

/etc/grub.conf file can be checked to see whether Lustre was installed properly or not.

```
[root@localhost Lustre]# more /etc/grub.conf
# grub.conf generated by anaconda
#
```

```
# Note that you do not have to rerun grub after making changes to this
file
# NOTICE:  You have a /boot partition.  This means that
#           all kernel and initrd paths are relative to /boot/, eg.
#           root (hd0,0)
#           kernel /vmlinuz-version ro root=/dev/VolGroup00/LogVol100
#           initrd /initrd-version.img
#boot=/dev/hda
default=0
timeout=5
splashimage=(hd0,0)/grub/splash.xpm.gz
hiddenmenu
password --md5 $1$sUXs2uHe$1F9T9nbW8fqzpdX5nYYPy.
title CentOS (2.6.18-194.17.1.el5_lustre.1.8.5)
    root (hd0,0)
    kernel /vmlinuz-2.6.18-194.17.1.el5_lustre.1.8.5 ro
root=/dev/VolGroup00
/LogVol100
    initrd /initrd-2.6.18-194.17.1.el5_lustre.1.8.5.img
title CentOS (2.6.18-194.el5)
    root (hd0,0)
    kernel /vmlinuz-2.6.18-194.el5 ro root=/dev/VolGroup00/LogVol100
    initrd /initrd-2.6.18-194.el5.img
```

After installing all the rpm packages the system should be rebooted. Without rebooting no further necessary action can be taken.

Installing the client

For the client, 3 packages needed to be installed:

```
kernel-2.6.18-194.17.1.el5_lustre.1.8.5.i686.rpm
lustre-client-1.8.5-2.6.18_194.17.1.el5_lustre.1.8.5.i686.rpm
lustre-client-modules-1.8.5-2.6.18_194.17.1.el5_lustre.1.8.5.i686.rpm
```

A directory in a server can also act as an client file system. In that case, there is no requirements to install all the above packages. Just mounting the Lustre in a particular directory is enough.

Mounting and Configuration

On our MDS/OSS machine, we will configure the MDT on the disk /dev/hda2 and the OST on the disk /dev/hda3.

```
[root@localhost ~]# fdisk -l
```

```
Disk /dev/hda: 8589 MB, 8589934592 bytes
255 heads, 63 sectors/track, 1044 cylinders
Units = cylinders of 16065 * 512 = 8225280 bytes
```

Device	Boot	Start	End	Blocks	Id	System
/dev/hda1	*	1	13	104391	83	Linux

/dev/hda2	14	140	1020127+	83	Linux
/dev/hda3	141		267	1020127+	83 Linux
/dev/hda4	268	1044	6241252+	5	Extended
/dev/hda5	268	1044	6241221	8e	Linux LVM

In the /etc/modprobe.conf file "options lnet networks=tcp0" line should be added which will define and initiate LNET.

```
[root@localhost ~]# more /etc/modprobe.conf
alias eth0 pcnet32
alias scsi_hostadapter ata_piix
alias snd-card-0 snd-intel8x0
options snd-card-0 index=0
options snd-intel8x0 index=0
remove snd-intel8x0 { /usr/sbin/alsactl store 0 >/dev/null 2>&1 || : ;
}; /sbin/
modprobe -r --ignore-remove snd-intel8x0
alias eth1 pcnet32
options lnet networks=tcp0
```

Create the combined MDT/MGS file system on the device /dev/hda2

```
[root@localhost ~]# mkfs.lustre --mdt --mgs --fsname=lustre /dev/hda2

Permanent disk data:
Target:      lustre-MDTffff
Index:       unassigned
Lustre FS:   lustre
Mount type:  ldiskfs
Flags:       0x75
              (MDT MGS needs_index first_time update )
Persistent mount opts: iopen_nopriv,user_xattr,errors=remount-ro
Parameters: mdt.group_upcall=/usr/sbin/l_getgroups

checking for existing Lustre data: not found
device size = 996MB
2 6 18
formatting backing filesystem ldiskfs on /dev/hda2
      target name  lustre-MDTffff
      4k blocks    255031
      options      -i 4096 -I 512 -q -O dir_index,uninit_groups -F
mkfs_cmd = mke2fs -j -b 4096 -L lustre-MDTffff -i 4096 -I 512 -q -O
dir_index,uninit_groups -F /dev/hda2 255031
Writing CONFIGS/mountdata
```

Create a mount point and mount the MDT/MGS file system

```
[root@localhost ~]# mkdir -p /mnt/mdt
[root@localhost ~]# mount -t lustre /dev/hda2 /mnt/mdt
```

Create the OST file system on the device /dev/hda3

```
[root@localhost ~]# mkfs.lustre --ost --fsname=lustre
--mgsgnode=192.168.56.101@tcp0 /dev/hda3

Permanent disk data:
Target:      lustre-OSTffff
Index:       unassigned
Lustre FS:   lustre
Mount type:  ldiskfs
Flags:       0x72
              (OST needs_index first_time update )
Persistent mount opts: errors=remount-ro, extents, mballocc
Parameters: mgsgnode=192.168.56.101@tcp0

device size = 996MB
2 6 18
formatting backing filesystem ldiskfs on /dev/hda3
target name lustre-OSTffff
4k blocks      255031
options        -i 16384 -I 256 -q -O
dir_index,extents,uninit_groups -F
mkfs_cmd = mke2fs -j -b 4096 -L lustre-OSTffff -i 16384 -I 256 -q -O
dir_index,extents,uninit_groups -F /dev/hda3 255031
Writing CONFIGS/mountdata
```

Create a mount point and mount the OST file system

```
[root@localhost ~]# mkdir -p /mnt/ost
[root@localhost ~]# mount -t lustre /dev/hda3 /mnt/ost/
```

On the client machine, create a mount point and mount the Lustre file system in /mnt/lustre and /mnt/test which will work as two separate file systems.

```
[root@localhost ~]# mkdir -p /lustre
[root@localhost ~]# mount -t lustre 192.168.56.101@tcp0:/lustre /mnt/lustre
[root@localhost ~]# mkdir -p /mnt/test
[root@localhost ~]# mount -t lustre 192.168.56.101@tcp0:/lustre /mnt/test
```

With df -h command the whole Lustre system can be checked to see which partition was mounted to what.

```
[root@localhost ~]# df -h
Filesystem      Size  Used Avail Use% Mounted on
```



```

/dev/mapper/VolGroup00-LogVol00
      3.9G  1.8G  1.9G  50% /
/dev/hda1      99M   18M   76M  20% /boot
tmpfs          506M    0  506M   0% /dev/shm
/dev/hda2      872M   18M  805M   3% /mnt/mdt
/dev/hda3      981M   18M  914M   2% /mnt/ost
192.168.56.101@tcp0:/lustre
      981M   18M  914M   2% /mnt/lustre
192.168.56.101@tcp0:/lustre
      981M   18M  914M   2% /mnt/test

```

After all the above steps, the Lustre file system has been successfully installed and configured in the server.

Testing the file system

Let's create few files in one directory and see whether those files are automatically updated to other directory in the same machine or not.

```

[root@localhost /]# cd /mnt/lustre
[root@localhost lustre]# ls
[root@localhost lustre]# touch File1
[root@localhost lustre]# touch File2
[root@localhost lustre]# touch File3
[root@localhost lustre]# touch File4
[root@localhost lustre]# touch File5
[root@localhost lustre]# touch File6
[root@localhost lustre]# touch File7
[root@localhost lustre]# ls
File1 File2 File3 File4 File5 File6 File7
[root@localhost lustre]# cd ../
[root@localhost mnt]# cd ../
[root@localhost /]# cd /mnt/test
[root@localhost test]# ls
File1 File2 File3 File4 File5 File6 File7
[root@localhost test]# touch File8
[root@localhost test]# ls
File1 File2 File3 File4 File5 File6 File7 File8
[root@localhost test]# rm File1
rm: remove regular empty file `File1'? y
[root@localhost test]# cd ../
[root@localhost mnt]# cd ../
[root@localhost /]# cd /mnt/lustre
[root@localhost lustre]# ls
File2 File3 File4 File5 File6 File7 File8

```

Now, we can create a client in different machine and check whether Lustre file system working properly or not .

Let's check the configuration of the remote client machine first

```

[root@localhost lustre]# ifconfig eth0
eth0    Link encap:Ethernet  HWaddr 08:00:27:54:77:81

```

```

    inet addr:192.168.56.102 Bcast:192.168.56.255
Mask:255.255.255.0
    inet6 addr: fe80::a00:27ff:fe54:7781/64 Scope:Link
UP BROADCAST RUNNING MULTICAST MTU:1500 Metric:1
RX packets:141 errors:0 dropped:0 overruns:0 frame:0
TX packets:140 errors:0 dropped:0 overruns:0 carrier:0
collisions:0 txqueuelen:1000
RX bytes:51956 (50.7 KiB) TX bytes:31347 (30.6 KiB)
Interrupt:10 Base address:0xd020

```

Now we can create a directory and mount the Lustre file system to it and check for the files which we have created in the server different Lustre clients directory. If those files are here in new remote client, we can be sure that Lustre is working properly.

```

[root@localhost ~]# mkdir -p /mnt/lustre
[root@localhost ~]# mount -t lustre 192.168.56.101@tcp0:/lustre
/mnt/lustre
[root@localhost ~]# cd /mnt/lustre
[root@localhost lustre]# ls
File2 File3 File4 File5 File6 File7 File8

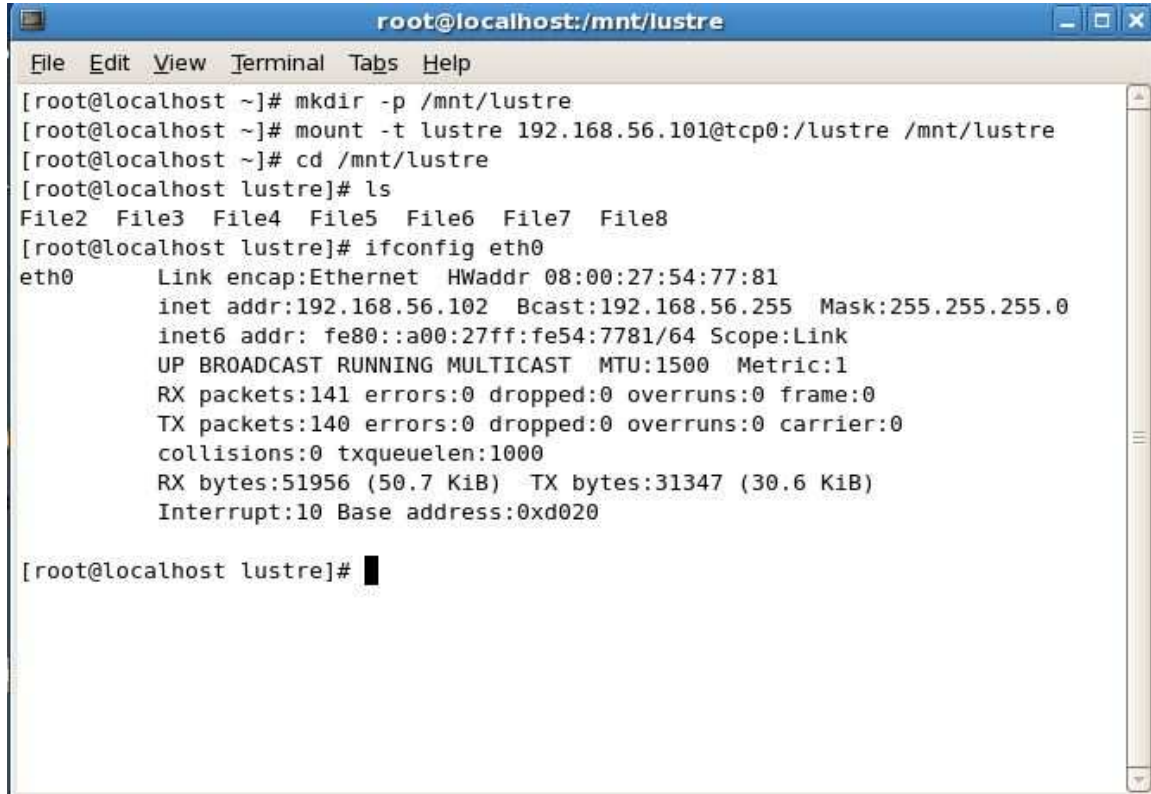
```

```

root@localhost:/mnt/lustre
File Edit View Terminal Tabs Help
[root@localhost /]# cd /mnt/lustre
[root@localhost lustre]# ls
[root@localhost lustre]# touch File1
[root@localhost lustre]# touch File2
[root@localhost lustre]# touch File3
[root@localhost lustre]# touch File4
[root@localhost lustre]# touch File5
[root@localhost lustre]# touch File6
[root@localhost lustre]# touch File7
[root@localhost lustre]# ls
File1 File2 File3 File4 File5 File6 File7
[root@localhost lustre]# cd ../
[root@localhost mnt]# cd ../
[root@localhost /]# cd /mnt/test
[root@localhost test]# ls
File1 File2 File3 File4 File5 File6 File7
[root@localhost test]# touch File8
[root@localhost test]# ls
File1 File2 File3 File4 File5 File6 File7 File8
[root@localhost test]# rm File1
rm: remove regular empty file `File1'? y
[root@localhost test]# cd ../
[root@localhost mnt]# cd ../
[root@localhost /]# cd /mnt/lustre
[root@localhost lustre]# ls
File2 File3 File4 File5 File6 File7 File8
[root@localhost lustre]#

```

fig: The snapshot of the testing of the Lustre system in server with client file directories.

A terminal window titled 'root@localhost:/mnt/lustre' with a menu bar (File, Edit, View, Terminal, Tabs, Help). The terminal shows the following commands and output:

```
[root@localhost ~]# mkdir -p /mnt/lustre
[root@localhost ~]# mount -t lustre 192.168.56.101@tcp0:/lustre /mnt/lustre
[root@localhost ~]# cd /mnt/lustre
[root@localhost lustre]# ls
File2 File3 File4 File5 File6 File7 File8
[root@localhost lustre]# ifconfig eth0
eth0      Link encap:Ethernet  HWaddr 08:00:27:54:77:81
          inet addr:192.168.56.102  Bcast:192.168.56.255  Mask:255.255.255.0
          inet6 addr: fe80::a00:27ff:fe54:7781/64 Scope:Link
          UP BROADCAST RUNNING MULTICAST  MTU:1500  Metric:1
          RX packets:141 errors:0 dropped:0 overruns:0 frame:0
          TX packets:140 errors:0 dropped:0 overruns:0 carrier:0
          collisions:0 txqueuelen:1000
          RX bytes:51956 (50.7 KiB)  TX bytes:31347 (30.6 KiB)
          Interrupt:10 Base address:0xd020

[root@localhost lustre]#
```

Fig: The snapshot of the testing of the Lustre system in remote client.

Stopping the Lustre client and server:

With following umount command we can stop lustre system.

```
# umount <mount point>
```

Conclusion

All the above steps should be executed carefully to install the complex Lustre file system. Every step is important. For the beginners, it can be a frustration to recognize the different parts of a command and write it properly for the correct executions in different machines. Also, firewall can be a issue when one want to connect remotely to a client. But once a clear idea is developed, the Lustre file system is easy to install and operate.

Source: Lustre 1.8 Operations Manual