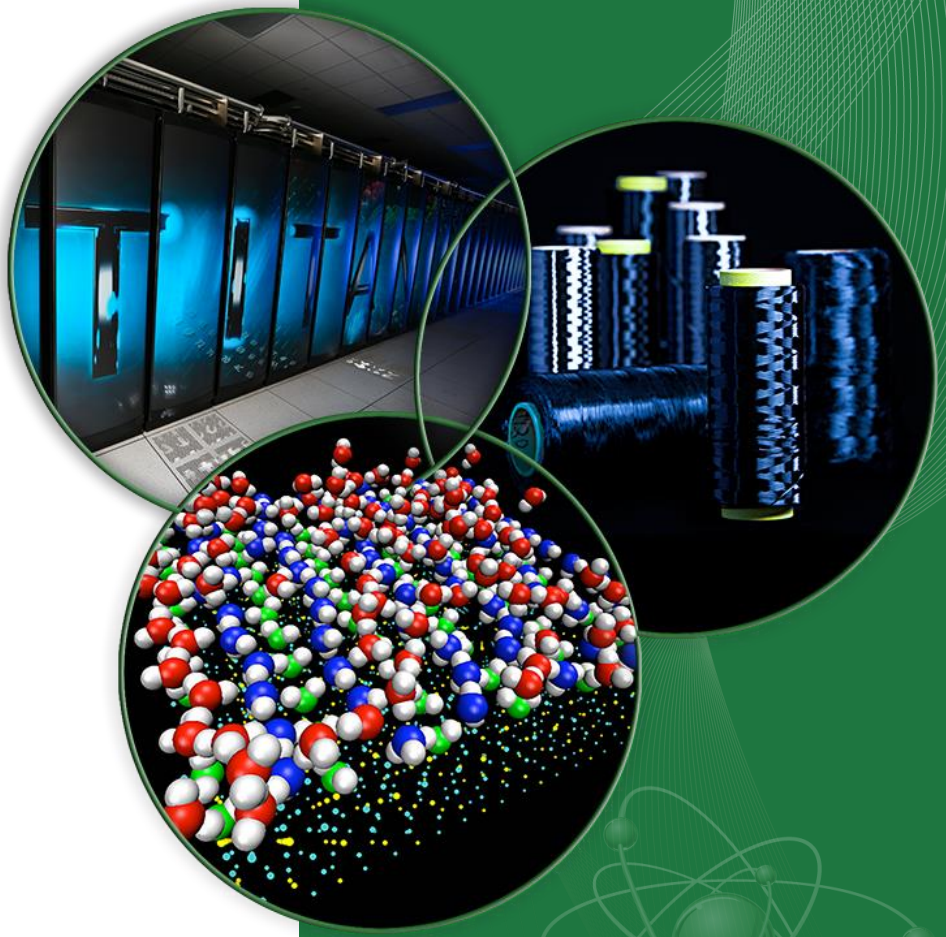
Oak Ridge National Laboratory

Computing and Computational Sciences Directorate



Basic Lustre Installation and Setup from Stock RPMs

Jeffrey Rossiter

Rick Mohr

Sarp Oral

Michael Brim

Jason Hill

Joel Reed

Neena Imam

ORNL is managed by UT-Battelle
for the US Department of Energy

OAK RIDGE National Laboratory

Introduction

- We will be setting up a basic Lustre system using stock RPMs
  - Lustre RPMs will be obtained/installed
  - Servers and clients will be configured
  - An example with multiple Lustre file systems will be given

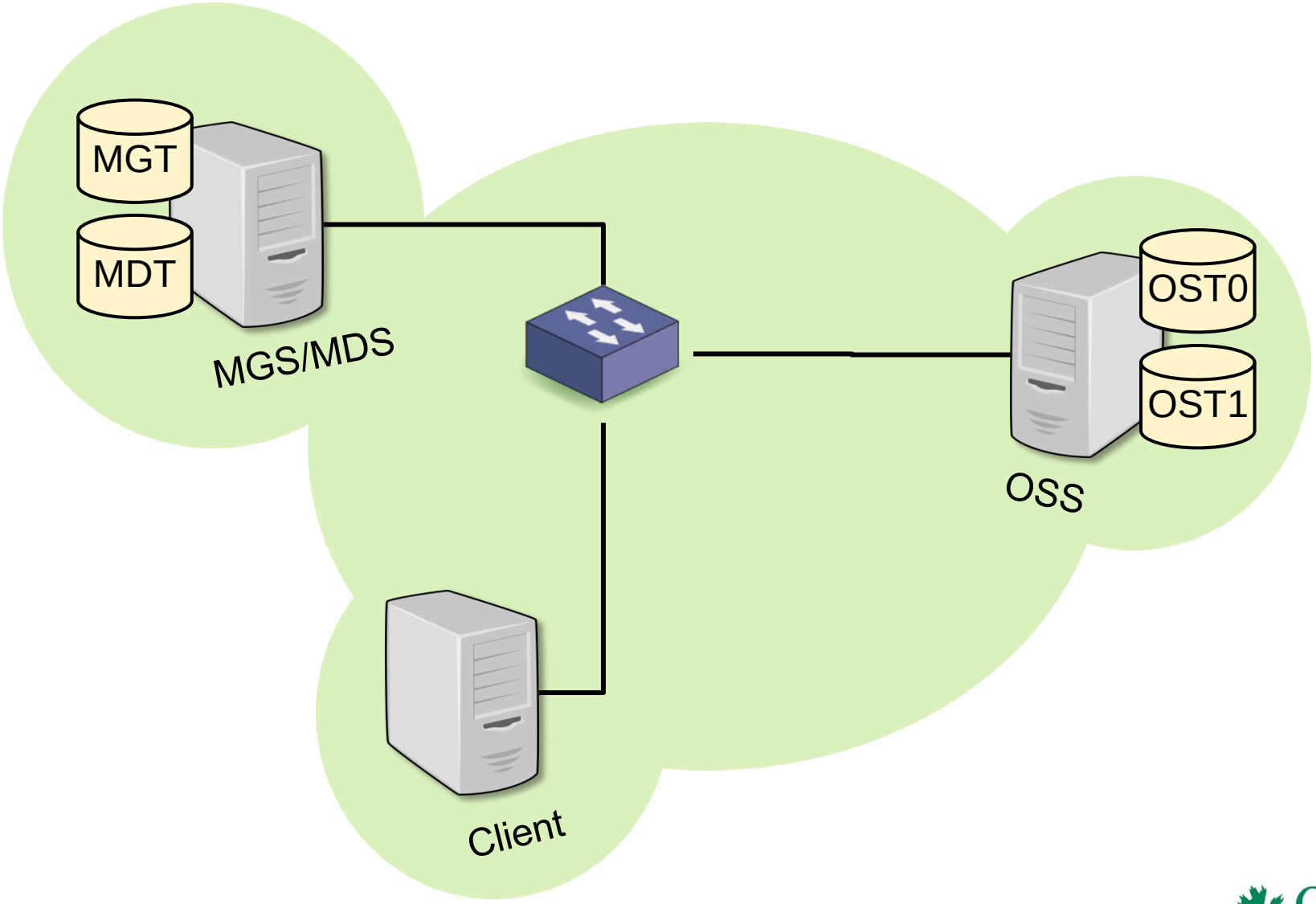
Outline

- System Overview
- Prepare Servers and Client
- Install Lustre on Servers
- Install Lustre on Client
- MGS/MDS Configuration / Starting the File System
- OSS Configuration
- Client Configuration
- Multiple File Systems

Up Next

- System Overview
- Install Lustre on Servers
- Install Lustre on Client
- MGS/MDS Configuration / Starting the File System
- OSS Configuration
- Client Configuration
- Multiple File Systems

System Overview



Up Next

- System Overview
- **Prepare Servers and Client**
- Install Lustre on Servers
- Install Lustre on Client
- MGS/MDS Configuration / Starting the File System
- OSS Configuration
- Client Configuration
- Multiple File Systems

Prepare Servers and Client (1)

- Install CentOS 6.6 for x86_64 architecture on every machine
  - Use the “minimal” ISO
    - http://www.centos.org/download/

Prepare Servers and Client (2)

- Disable SELinux
  - Set *SELINUX=disabled* in */etc/sysconfig/selinux*

SELinux should be disabled because the Lustre software does not currently support SELinux.

Prepare Servers and Client (3)

- Disable iptables

```
chkconfig --levels 345 iptables off
chkconfig --levels 345 ip6tables off
reboot
```

This is a quick and dirty approach. Lustre uses port 988 so if you want to keep the firewall up you can open port 988.

Prepare Servers and Client (4)

- Install dependencies on servers
  - wget, perl, libgssglue, net-snmp, libyaml, sg3_utils, openmpi, lsof, rsync

Wget is not a requirement. We will be using wget but you can use whatever http download utility you like.

Prepare Servers and Client (5)

- Update *kernel-firmware*

```
yum update kernel-firmware
```

You may be able to omit this step. You will be prompted later to update kernel-firmware if necessary.

Prepare Servers and Client (6)

- Ensure that each machine has a non-loopback entry for itself in */etc/hosts*

Lustre can only use non-loopback IP addresses.

Prepare Servers and Client (7)

- Create an entry in */etc/modprobe.d/lustre.conf*

```
options lnet networks=tcp # or networks=tcp(eth0)
```

The *options lnet networks* entry specifies which interfaces should be mapped to which LNET subnets.

Here we are saying that Lustre should use the LNET *tcp* subnet. Since we haven't specified which Ethernet interface to use, LNET will use the first available non-loopback interface.

The comment in the example specifies that eth0 should be part of the LNET tcp subnet. LNET also supports an InfiniBand subnet (o2ib).

Install Lustre on Servers (1)

- Download Lustre server kernel
  - https://downloads.hpdd.intel.com/public/lustre/lustre-2.7.0/el6.6/server/RPMS/x86_64/kernel-2.6.32-504.8.1.el6_lustre.x86_64.rpm

This is the Linux kernel with Lustre patches applied. All servers in your Lustre system must use the patched kernel.

Install Lustre on Servers (2)

- Install the kernel RPM

```
rpm -ivh kernel-2.6.32-504.8.1.el6_lustre.x86_64.rpm
```

- Install the kernel

```
/sbin/new-kernel-pkg --package kernel --mkinitrd \
--dracut --depmod \
--install 2.6.32-504.8.1.el6_lustre.x86_64
reboot
```

Install Lustre on Servers (3)

- Download the Lustre server modules listed on the next slide from https://downloads.hpdd.intel.com/public/lustre/lustre-2.7.0/el6.6/server/RPMS/x86_64/

Install Lustre on Servers (4)

Lustre RPM	Description
lustre-2.7.0	Provides user space tools and files for Lustre
lustre-iokit	Provides a collection of Lustre benchmark tools
lustre-modules	Server and network drivers for the kernel.
lustre-osd-ldiskfs	Provides an OSD API for using the ldiskfs file system on the Lustre servers
lustre-osd-ldiskfs-mount	Provides ldiskfs hooks for mount/mkfs into a dynamic library
lustre-tests	Provides binaries and scripts for Lustre testing framework

Install Lustre on Servers (5)

- Download e2fsprogs needed by Lustre from https://downloads.hpdd.intel.com/public/e2fsprogs/latest/el6/RPMS/x86_64/

RPM	Description
e2fsprogs-1.42.12.wc1-7.el6.x86_64.rpm	Provides utilities for working with ext2/ext3/ext4 file systems
e2fsprogs-libs-1.42.12.wc1-7.el6.x86_64.rpm	Provides e2fsprogs shared libraries
libcom_err-1.42.12.wc1-7.el6.x86_64.rpm	Provides an error description library for e2fsprogs
libss-1.42.12.wc1-7.el6.x86_64.rpm	Provides a command line interface parsing library for e2fsprogs

Install Lustre on Servers (6)

- Update / install e2fsprogs and related libraries

```
rpm -Uvh e2fsprogs-1.42.12.wc1-7.el6.x86_64.rpm e2fsprogs-
libs-1.42.12.wc1-7.el6.x86_64.rpm libcom_err-1.42.12.wc1-
7.el6.x86_64.rpm libss-1.42.12.wc1-7.el6.x86_64.rpm
```

Install Lustre on Servers (7)

- Install Lustre RPMs

```
rpm -ivh lustre-modules-*
rpm -ivh lustre-osd-ldiskfs-*
rpm -ivh lustre-2.7*
rpm -ivh lustre-iokit-2.7*
rpm -ivh lustre-tests-*
```

Up Next

- System Overview
- Install Lustre on Servers
- **Install Lustre on Client**
- MGS/MDS Configuration / Starting the File System
- OSS Configuration
- Client Configuration
- Multiple File Systems

Install Lustre on Client (1)

- Download Lustre client kernel RPM

```
wget \
http://mirror.centos.org/centos/6.6/updates/x86_64/Packages/kernel-2.6.32-504.8.1.el6.x86_64.rpm
```

This is a standard Linux kernel. Lustre clients are not required to use a Lustre-patched kernel. We are using this particular version because it must match the version of the kernel used on our Lustre servers.

Install Lustre on Client (2)

- Install the kernel RPM

```
rpm -ivh kernel-2.6.32-504.8.1.el6.x86_64.rpm
```

Install Lustre on Client (3)

- Install the kernel

```
/sbin/new-kernel-pkg --package kernel --mkinitrd \
--dracut --depmod --install 2.6.32-504.8.1.el6.x86_64
reboot
```

Install Lustre on Client (4)

- Download the Following Lustre client modules from https://downloads.hpdd.intel.com/public/lustre/lustre-2.7.0/el6.6/client/RPMS/x86_64/

Lustre RPM	Description
lustre-client-modules	Provides file system, server, and network drivers
lustre-client	Provides user space tools and files for Lustre

Install Lustre on Client (5)

- Install Lustre Client Modules

```
rpm -ivh lustre-client-modules-2.7.0-
2.6.32_504.8.1.el6.x86_64.x86_64.rpm
rpm -ivh lustre-client-2.7.0-
2.6.32_504.8.1.el6.x86_64.x86_64.rpm
```

Up Next

- System Overview
- Install Lustre on Servers
- Install Lustre on Client
- **MGS/MDS Configuration / Starting the File System**
- OSS Configuration
- Client Configuration
- Multiple File Systems

MGS/MDS Configuration / Starting the File System (1)

- Format the MGT

```
[root@mgs_mds]$ mkfs.lustre --mgs /dev/sdb
```

We are using the default file system for Lustre targets, ldiskfs. This was for the sake of simplicity. Alternatively you could use multiple devices configured as ZFS zpools. In that case your *mkfs* commands would need to be modified from our examples.

MGS/MDS Configuration / Starting the File System (2)

- Format the MDT

```
[root@mgs_mds]$ mkfs.lustre --fsname=lustre \
--mgsnode=mgs_mds@tcp --mdt --index=0 /dev/sdc
```

The value of fsname should be whatever you want to refer to this file system as. /dev/sdc must be the device you're using as the MDT. The value of mgsnode must be the NID of the MGS.

MGS/MDS Configuration / Starting the File System (3)

- Mount the MGT

```
[root@mgs_mds]$ mkdir /mnt/mgt
[root@mgs_mds]$ mount -t lustre /dev/sdb /mnt/mgt
```

- Mount the MDT

```
[root@mgs_mds]$ mkdir /mnt/mdt
[root@mgs_mds]$ mount -t lustre /dev/sdc /mnt/mdt
```

Up Next

- System Overview
- Install Lustre on Servers
- Install Lustre on Client
- MGS/MDS Configuration / Starting the File System
- **OSS Configuration**
- Client Configuration
- Multiple File Systems

OSS Configuration (1)

- Format the OSTs on the OSS

```
[root@oss]$ mkfs.lustre --fsname=lustre --ost \
--mgsnode=mgs_mds@tcp --index=0 /dev/sdb
[root@oss]$ mkfs.lustre --fsname=lustre --ost \
--mgsnode=mgs_mds@tcp --index=1 /dev/sdc
```

The value of *fsname* should be whatever you set it to on the MDT. The value of *mgsnode* should be the NID of the MGS. /dev/sd[b,c] should match the devices you are using for OSTs

OSS Configuration (2)

- Mount the OSTs

```
[root@oss]$ mkdir /mnt/ost0 /mnt/ost1
[root@oss]$ mount -t lustre /dev/sdb /mnt/ost0
[root@oss]$ mount -t lustre /dev/sdc /mnt/ost1
```

Up Next

- System Overview
- Install Lustre on Servers
- Install Lustre on Client
- MGS/MDS Configuration / Starting the File System
- OSS Configuration
- **Client Configuration**
- Multiple File Systems

Client Configuration (1)

- Mount the Lustre file system

```
[root@client]$ mkdir /mnt/lustre
[root@client]$ mount –t lustre mgs_mds@tcp:/lustre
/mnt/lustre
```

In the *mount* command we are mounting a lustre type file system called *lustre* that is being managed by *mgs@tcp*.

Client Configuration (2)

- Create a test file to be sure the file system is working

```
[root@client]$ touch /mnt/lustre/testFile
ls /mnt/lustre
```

Celebrate

- Congratulations! You should now have a working Lustre file system.
  - You can stop here or continue on to learn how to create multiple Lustre file systems.

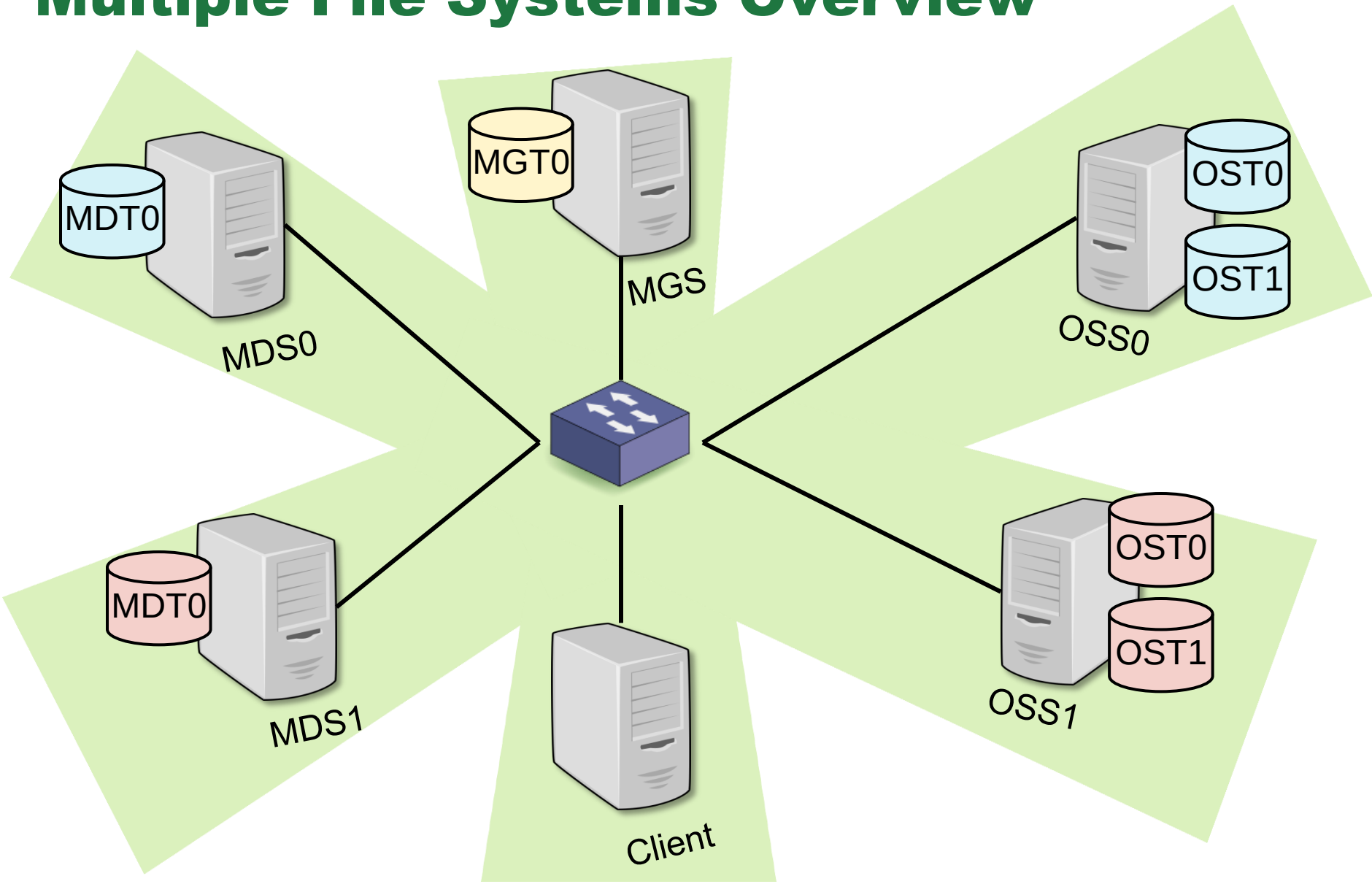
Up Next

- System Overview
- Install Lustre on Servers
- Install Lustre on Client
- MGS/MDS Configuration / Starting the File System
- OSS Configuration
- Client Configuration
- **Multiple File Systems**

Multiple File Systems

- Multiple Lustre file systems can be created by creating multiple MDTs
- The following example uses 1 MGS, 2 MDSs, and 2 OSSs to provide 2 Lustre file systems, lustre_0 and lustre_1

Multiple File Systems Overview



Multiple File Systems MGS/MDS Configuration

- Format and mount the targets

```
[root@mgs]$ mkfs.lustre --mgs /dev/sdb
[root@mgs]$ mount -t lustre /dev/sdb /mnt/mgt0
[root@mds0]$ mkfs.lustre --fsname=lustre_0 \
--mgsnode=mgs@tcp --mdt --index=0 /dev/sdb
[root@mds0]$ mkdir /mnt/mdt0
[root@mds0]$ mount -t lustre /dev/sdb /mnt/mdt0
[root@mds1]$ mkfs.lustre --fsname=lustre_1 \
--mgsnode=mgs@tcp --mdt --index=0 /dev/sdb
[root@mds1]$ mkdir /mnt/mdt0
[root@mds1]$ mount -t lustre /dev/sdb /mnt/mdt0
```

Multiple File Systems OSS Configuration (1)

- Format the OSTs

```
[root@oss0]$ mkfs.lustre --fsname=lustre_0 \
--mgsnode=mgs@tcp --ost --index=0 /dev/sdb
[root@oss0]$ mkfs.lustre --fsname=lustre_0 \
--mgsnode=mgs@tcp --ost --index=1 /dev/sdc
[root@oss1]$ mkfs.lustre --fsname=lustre_1 \
--mgsnode=mgs@tcp --ost --index=0 /dev/sdb
[root@oss1]$ mkfs.lustre --fsname=lustre_1 \
--mgsnode=mgs@tcp --ost --index=1 /dev/sdc
```

Multiple File Systems OSS Configuration (2)

- Mount the OSTs

```
[root@oss0]$ mkdir /mnt/ost0 /mnt/ost1
[root@oss0]$ mount -t lustre /dev/sdb /mnt/ost0
[root@oss0]$ mount -t lustre /dev/sdc /mnt/ost1
[root@oss1]$ mkdir /mnt/ost0 /mnt/ost1
[root@oss1]$ mount -t lustre /dev/sdb /mnt/ost0
[root@oss1]$ mount -t lustre /dev/sdc /mnt/ost1
```

Multiple File Systems Client Configuration

- Mount the file systems on client

```
[root@client]$ mkdir /mnt/lustre_0 /mnt/lustre_1
[root@client]$ mount -t lustre mgs@tcp:/lustre_0 \
/mnt/lustre_0
[root@client]$ mount -t lustre mgs@tcp:/lustre_1 \
/mnt/lustre_1
```

Celebrate

- Congratulations! You should now have two working Lustre file systems.

Acknowledgements

This work was supported by the United States Department of Defense (DoD) and used resources of the DoD-HPC Program at Oak Ridge National Laboratory.