Upgrading a Lustre* File System to Intel® Enterprise Edition for Lustre* Software

Partner Guide

May 13, 2016
Software Version 2.4.0.0 and later
World Wide Web: http://www.intel.com

Disclaimer and legal information

Contents

About this Document

This document provides guidance to perform a static upgrade of a 'legacy' Lustre* file system to a current version of the Intel® Enterprise Edition for Lustre* software code base, with all the robustness and new features provided. This process updates Lustre, but does *not* install the Intel® Manager for Lustre* *management* software.

Intended Audience

The intended audience for this guide are partners who are supporting legacy Lustre* file systems and who want to upgrade them to storage solutions based on Intel® Enterprise Edition for Lustre* software. Readers are assumed to be full-time Linux system administrators or equivalent that have:

- Experience administering file systems and are familiar with storage components such as block storage, SAN, and LVM
- Proficiency in setting up, administering and maintaining networks. Knowledge of LNet is required. Knowledge of InfiniBand* is required if InfiniBand is to be used.
- Detailed knowledge of the overall configuration of the storage system and the ability to verify that the configuration matches the configuration requirements as defined in this guide.

This document is not intended for end users of storage solutions implemented using the Intel® Enterprise Edition for Lustre* software.

Conventions Used

Conventions used in this document include:

- # preceding a command indicates the command is to be entered as root
- $ indicates a command is to be entered as a user
- *<variable_name>* indicates the placeholder text that appears between the angle brackets is to be replaced with an appropriate value

Related Documentation

- *Intel Manager® for Lustre* Software User Guide*
- *Intel® Enterprise Edition for Lustre* Partner Installation Guide*
- *Creating a Scalable File Service for Windows Networks using Intel® EE for Lustre Software*
- *Creating a Monitored Lustre* Storage Solution over a ZFS File System*

- *Hierarchical Storage Management Configuration Guide*
- *Installing Hadoop, the Hadoop Adapter for Intel® EE for Lustre*, and the Job Scheduler Integration*
- *Creating an HBase Cluster and Integrating Hive on an Intel® EE for Lustre® File System*
- *Upgrading a Lustre file system to Intel® Enterprise Edition for Lustre* software (Lustre only)*
- *Configuring LNet Routers for File Systems based on Intel* EE for Lustre* Software*
- *Intel® EE for Lustre* Hierarchical Storage Management Framework White Paper*
- *Architecting a High-Performance Storage System White Paper*

Overview

This document provides guidance to perform a static upgrade of a 'legacy' Lustre* file system to a current version of the Intel® Enterprise Edition for Lustre* software code base, with all the robustness and new features provided. This process updates Lustre, but does *not* install the Intel® Manager for Lustre* management software.

Prerequisites

Operating System Requirements

Note: This document does not discuss changing your Lustre file system from one operating system to another. This means that if your starting Lustre file system is running under Red* Hat Enterprise Linux* (for example), then the resulting upgraded IEEL Lustre file system will be running under that OS.

Note: This procedures described herein apply to Intel® Enterprise Edition for Lustre* software, versions 2.4.0.0 and later. Please see the Release Notes for the version of Intel® Enterprise Edition for Lustre* software you are installing to learn the operating system requirements. All servers and clients must be running the same operating system and version to complete these procedures.

To carry out this procedure you will need to shut down your existing Lustre file system and take it out of service.

Required Software

If you do not already have a copy of the Intel® Enterprise Edition for Lustre** software, version 2.4.0.0 or later, you may register as an Intel Reseller Partner as described in Appendix B, and obtain a copy of this software.

File System Configuration Requirements for Lustre-only Software Upgrade

The Lustre file system to be updated with Intel EE for Lustre* software (excluding Intel® Manager for Lustre* software) should conform to the design and configuration constraints defined in this section.

The high-level configuration of the initial Lustre file system should include the following components:

- Management server (MGS): The MGS provides access to the Lustre management target (MGT) storage.

- Metadata server (MDS): The MDS provides access to the metadata target storage.
- Object storage server (OSS): At least one server provides access to the object storage targets, which store the file system's data. Typical Lustre file systems may have ten or twenty OSSs, or many more.
- (Optional) Management target (MGT): The MGT stores configuration information for all the Lustre file systems in a cluster and provides this information to other Lustre components. The MGT is accessed by the MGS.
- Metadata target (MDT): The MDT stores metadata (such as file names, directories, permissions, and file layout) for attached storage and makes them available to clients. The MDT is accessed by the MDS.
- Object storage targets (OSTs): Client file system data is stored in one or more objects that are located on separate OSTs. The number of objects per file is configurable by the user and can be tuned to optimize performance for a given workload. RAID 6 is recommended for OSTs.
- Management network - The Management network is 1-gigabit Ethernet, connecting every server in the file system. This network is used for SSH, and to manage the servers. This network also makes a separate connection to an IPMI port installed on each managed server.
- Lustre network – A high-performance Lustre network (LNET) is generally either 10-gigabit Ethernet, or Infiniband, and provides high-speed file system access to each of clients. The required data rate of this network is generally driven by the file system size, the number of clients, and the average throughput requirements for each client.

Upgrading for Lustre Version

Use this procedure to upgrade Lustre software release 1.8.6-wc1 or later, to the most current Intel® EE for Lustre* software. Please contact support for Intel® EE for Lustre* software (john.fuchs-chesney@intel.com) if you have a version earlier than 1.8.6-wc1.

Complete these steps as the root user:

1. Create a complete, restorable file system backup.

   Caution: Before installing the Intel® EE for Lustre* software, backup ALL data on the Lustre file system. The Lustre software included with the Intel® EE for Lustre* package contains kernel modifications that interact with storage devices. If not installed, configured, or administered properly, these modifications may introduce security issues and data loss. If a full backup of the file system is not practical, a device-level backup of the MDT file system is recommended. See Section 17.2 of the Lustre Manual: Backing Up and Restoring an MDS or OST (Device Level):

https://build.hpdd.intel.com/job/lustre-manual/lastSuccessfulBuild/artifact/lustre_manual.xhtml#dbdoclet.50438207_71633

2. Shut down the file system by unmounting all clients and servers in the order shown next. (Unmounting a block device causes the Lustre software to be shut down on that node.)

   a. Unmount the clients. On each client node, run:
   umount - a - t lustre

   b. Unmount the MDT. On the MDS node, run:
   umount - a - t lustre

   c. If you are using a separate MGS, unmount the MGS. On the MGS node, run:
   umount - a - t lustre

   d. Unmount all the OSTs. On each OSS node, run:
   umount - a - t lustre

3. On each server, upgrade the Linux operating system to the Linux kernel version provided with the Intel® EE for Lustre* software download:

   a. Log onto a Lustre server as the root user and change directory to the directory containing the kernel RPMs.

   b. Use the yum command to install the packages:
   # yum localinstall ./kernel*

   Note: If no yum command is available (in SLES for example) the rpm command may be used:
   # rpm – ivh ./kernel*.rpm

   c. Verify the packages are installed correctly:
   rpm - qa|egrep "kernel"

   d. Reboot the Lustre server.

   e. Repeat steps 3a – 3d on every Lustre server.

4. Upgrade e2fsprogs on all Lustre servers to the version provided with the Intel® EE for Lustre* software download.

   a. Log onto the Lustre server as the root user and change directory to the directory containing the e2fsprog RPMs.

b. Use the yum command to install the packages:
   # yum localinstall e2fsprogs* libcom* libss*

   Note: If no yum command is available (in SLES for example), the rpm command may be used:
   # rpm – Uvh e2fsprogs* libcom* libss*

c. Verify that the packages are installed correctly and that the output of the following command matches what was provided in the Intel® EE for Lustre* software download:
   rpm - qa|grep e2fsprog

5. Install the Lustre server packages on all Lustre servers.

   a. Log onto a Lustre server as the root user and change directory to the directory containing the Lustre RPMs.

   b. Use the yum command to install the packages:
      # yum localinstall ./lustre- *.rpm

      Note: If no yum command is available (in SLES for example) the rpm command may be used:
      # rpm – ivh ./lustre- *.rpm

   c. Verify the packages are installed correctly and that the output of the following command matches what was provided in the Intel® EE for Lustre* software download:
      rpm - qa|egrep "lustre"

   d. Reboot the Lustre server.

   e. Repeat steps 5a – 5d on each Lustre server.

6. Install the Lustre client packages on each of the Lustre clients to be upgraded.

   Note: The kernel should be upgraded when completing the steps below. If the kernel is not upgraded, the version of the kernel running on a Lustre client must be the same as the version of the lustre-client-modules-ver package being installed. If not, a compatible kernel must be installed on the client before the Lustre client packages are installed. Depending on the exact distribution and what version you are upgrading from, there may be more extensive updates or upgrades needed than just installing the correct kernel.

Note: For SLES, an article about how to upgrade to SLES11 SP3 may be found at: http://www.novell.com/support/kb/doc.php?id=7012368

a. Log onto a Lustre client as the root user.

b. Use the yum command to install the packages:
```
# yum localinstall lustre-                client*
```

Note: If no yum command is available (in SLES for example) the rpm command may be used:
```
# rpm –   ivh lustre-        client*
```

c. Verify that the packages were installed correctly:
```
# rpm -    qa|egrep "lustre"
```

d. Repeat steps 6a - 6c on each Lustre client.

7. (Optional) To enable wide striping on an existing MDT, first make sure the MDT is not mounted, and then run the following command on the MDT:
```
mdt# tune2fs -        O large_xattr device
```

Note: For more information about wide striping, see the Lustre Operations Manual.

8. (Optional) To format an additional MDT, complete these steps:

a. Determine the index used for the first MDT (each MDT must have unique index). Enter:
```
client$ lctl dl | grep mdc
36 UP mdc lustre-        MDT0000-mdc- ffff88004edf3c00           \
                4c8be054-   144f-   9359-  b063-  8477566eb84e 5
```

In the above example, because the index of the existing MDT is 0, from MDT-0000, the next available index is 1.

b. Add the new block device as a new MDT at the next available index by entering (on one line):
```
mds# mkfs.lustre --reformat              --fsname=filesystem_name          --mdt    \
   --mgsnode=mgsnode         --index 1 /dev/mdt1_device
```

9. (Optional) If you are upgrading from Lustre software release 2.2 or earlier and want to enable the quota feature, complete these steps:

a. Before mounting the server devices, enter the following command on the MDS and OSSs:
   tunefs.lustre –quota device

b. After the server devices are mounted, enter the following command on the MDS:
   conf_param $FSNAME.quota.mdt=$QUOTA_TYPE
   conf_param $FSNAME.quota.ost=$QUOTA_TYPE

10. (Optional) If you are upgrading from Lustre software release 1.8, you must manually enable the FID-in-dirent feature. Warning: This step is not reversible. Do not complete this step until you are sure you will not be downgrading the Lustre software.

    On the MDS, enter the following command:
    tune2fs –O dirdata /dev/mdtdev

    This step only enables FID-in-dirent for newly created files. You can use LFSCK 1.5 to enable FID-in-dirent for existing files. For more information about FID-in-dirent and related functionalities in LFSCK 1.5, see the Lustre Operations Manual.

11. Start the Lustre file system by starting the components in the order shown in the following steps:

    a. If you have a separate MGT, mount the MGT. On the MGS, run:
       mgs# mount - a - t lustre

    b. Mount the MDT(s). On each MDS node, run:
       mds# mount - a - t lustre

    c. Mount all the OSTs. On each OSS node, run:
       oss# mount - a - t lustre

    d. Mount the file system on the clients. On each client node, run:
       client# mount - a - t lustre

Note : The above mount commands assume that all the MGT, MDT(s), and OSTs are listed in the /etc/fstab file. Targets that are not listed in the /etc/fstab file must be mounted individually by running the mount command:

mount - t lustre /dev/block_device /mount_point

The mounting order described in the steps above must be followed for the initial mount and registration of a Lustre file system after an upgrade. For a normal start of a Lustre file system, the mounting order is MGT, OSTs, MDT(s), clients.

Additional Instructions

For more information about installing and configuring Lustre, see the Lustre Operations Manual available at https://build.hpdd.intel.com/job/lustre-manual/lastSuccessfulBuild/artifact/lustre_manual.xhtml.

Enabling Wide Striping

In Lustre software release 2.2 and later, a feature has been added that allows striping across up to 2,000 OSTs. By default, this "wide striping" feature is disabled. It is activated by setting the large_xattr or ea_inode option on the MDT using either mkfs.lustre or tune2fs.

For example, after upgrading an existing file system earlier that Lustre software release 2.2 to Intel® EE for Lustre* software, wide striping can be enabled by running the following command on the MDT device before mounting it:

```
tune2fs - O large_xattr
```

Once the wide striping feature is enabled and in use on the MDT, it is not possible to directly downgrade the MDT file system to an earlier version of the Lustre software that does not support wide striping.

Support for Multiple MDTs

In Lustre software release 2.4 and later, a new feature allows using multiple MDTs , which can each serve one or more remote sub-directories in the file system. The root directory is always located on MDT0.

Note that clients running a release prior to the Lustre software release 2.4 can only see the namespace hosted by MDT0 and will return an IO error if an attempt is made to access a directory on another MDT.

Troubleshooting

You may encounter difficulty while installing Lustre on a client. Errors may pertain to a Lustre-client test dependency. Existing Lustre clients may have the lustre-client-tests package installed. Before upgrading Lustre on the clients, remove the lustre-client-tests package:

```
# yum remove lustre- client- tests- <ver>
```

If no yum command is available (in SLES for example) the rpm command may be used instead:

```
# rpm –   - erase lustre-        client-      tests-    <ver>
```

Appendix A - Becoming an Intel® Solutions for Lustre* Software Reseller, accessing downloads and documentation

1. Go to this page: https://www-ssl.intel.com/content/www/us/en/software/intel-solutions-for-lustre-software-become-a-reseller.html
2. Click on the 'Become a Reseller' link.
3. Complete the registration form.
4. Accept the 'Click to Accept' Intel® EE for Lustre* software Reseller Partner Agreement
5. You will receive an email with a download link within one or two days.

Note: If you are already registered as an Intel® Solutions for Lustre reseller, you can obtain the Intel® Enterprise Edition for Lustre software distribution from here, by entering your registration details:

https://registrationcenter.intel.com