Amanda and Lustre

Backup and Recovery of Lustre

Amanda

Amanda is the world's most popular Open Source Backup and Archiving software. Amanda allows System Administrators to set up a single backup server to back up multiple hosts to a tape- or disk-based storage system. Amanda uses native archival tools and can back up a large number of workstations and servers running various versions of Linux, Unix or Microsoft Windows operating systems.

Lustre

Lustre is a scalable, secure, robust, and highly-available cluster file system. It is designed, developed and maintained by Sun Microsystems, Inc. It includes three functional components, as described below.

Meta Data Servers/MDT

Meta Data Servers store filesystem meta data such as file names, directories, and permissions. Availability of the Meta Data Server is critical for filesystem data. In a typical configuration, there are two MDSs configured for high-availability failover. Since an MDS stores only meta data, the storage (MDT) attached to the filesystem need only store hundreds of gigabytes for a multi-terabyte filesystem.

Object Storage Servers/OST

Object Storage Servers provide access to filesystem data stored on shared or local disks. The files can be striped over multiple OSSs. A Lustre configure has at least 2-10 OSTs. In a large installation, there will be 100s to 1000+ OSTs, in order to have multi-petabyte filesystems.

Clients

The Lustre filesystem is accessed from Lustre clients. All applications that use the Lustre filesystem run on clients. Typical configurations have hundreds of clients. Larger configurations can have thousands of clients. Lustre clients do not store any filesystem data or meta-data.

465 S. Mathilda Ave., Suite 300 • Sunnyvale, CA 94086 • t: (888) 496-2632 • zsales@zmanda.com • www.zmanda.com

Backup/Recovery requirements

Some key requirements for backing up Lustre are:

- Individual files as well as entire filesystem should be recoverable.
- MDS and OSS should be reconstructed quickly, without requiring the entire filesystem to be rebuilt.
- Parallel backup is desirable, as this allows backup of multi-terabyte filesystems within practical backup windows.

Solution

Filesystem backup using multiple Lustre clients

An Amanda client can be used to back up a Lustre filesystem from Lustre clients. The filesystem is statically partitioned into multiple Amanda Disk List Entries (DLEs are unit of Amanda backup configuration) and is backed up from multiple Lustre clients. An Amanda DLE can specify a directory or individual files for backup. Amanda DLE creation and configuration can be managed using the Zmanda Management Console (ZMC). Adding more DLEs and/or Amanda clients allows you to maintain practical backup windows regardless of filesystem size.

Configuring multiple Amanda backup (Amanda “dumper”) threads and adjusting the Amanda client backup threads ("maxdumps" Amanda parameter) can increase the backup parallelism for a DLE.

Amanda intelligently schedules various levels of backups for DLEs within a backup set. Its internal algorithm balances the amount of data backed up across backup runs. This reduces the amount of data backed up in each backup run and keeps the backup window balanced across runs.

Amanda can leverage the Schily tar (**star**) command for backup. The modified **star** command available in Lustre 1.6 can be used with Amanda . The **star** command allows Lustre filesystems to be backed up along with the extended attributes (such as file stripe information).

Amanda can take advantage of the lightweight file scanning tool to generate file lists that can be passed to the **star** program. To take advantage of this tool, an Amanda application wrapper can be developed. Amanda's application plugin interface allows administrators to modify the backup processes as needed.

Backup of Meta data servers/Object Store servers

Amanda can back up meta data servers/OSS using the **ext3fs** dump program. Filesystem dumps can be used as the backup program. The **dump** command can be executed while the data is online. This requires an Amanda client installed on the Lustre MDS/OSS.

Amanda backup server deployment

One or more Amanda backup servers can be used to perform backup of filesystems as well as MDS/OSTs. The number of Amanda backup servers needed depends on the network bandwidth available for backup between backup server and clients, amount of data backed up each backup run as well as number of clients. Backup servers can also be configured to produce a Highly Available (HA) backup catalog.

Alternatively, an Amanda backup server can be installed each Lustre client. This allows direct backup of filesystem information to the media, without using network bandwidth.

Recovery scenarios

The following sections outline the processes necessary for recovering Lustre filesystems from Amanda backups.

Restoration of files/directories in Lustre filesystem

Amanda maintains a catalog of files backed up, and the web-based Zmanda Management Console (ZMC) makes it easy to restore files, directories or entire filesystems.

Restoration of Meta Data Servers/OSS

Meta Data Servers/OSS must be restored when the hardware is being replaced or when there are multiple disk failures. The meta data/object store must be recovered from the Amanda MDS/OST backups. MDS/OST recovery can be managed through the Zmanda Management Console. It is necessary to run a check on the filesystem after recovery from Amanda backup images. The filesystem check should be done as follows:

- Run **e2fsck** on the recovered MDS/OSS with Lustre stopped.
- Do not restore the file **/lov_objids** from the backup image. It tells the MDS which objects are valid on the OSTs (remove the file if it was restored). This file is recreated when Lustre remounts and the MDS contacts the OSTs. This will avoid deleting files on the OSTs which may recovered as part of the distributed file system check.
- Bring the **filesystem** up to reduce the impact on users.
- Run a full **e2fsck** of the MDS to create the MDS database for the distributed file system check on the MDS.
- Run **e2fsck** on the OSS to create the OST database. This step can be performed on all OSS nodes at the same time.
- Run a distributed filesystem check (**lfsck**) from a Lustre client using the MDS db and OST database. This will report inconsistencies. Some of the inconsistencies can be resolved by using **-l** option and examining the files in the **lost+found directory**. If the files are not useful, the **-d** option can be used to delete them. In some cases, files can be restored from the filesystem

backup as needed (as in the case of dangling inodes).

- Fix **lfsck** to "reassemble" objects in the **lost+found** based on attribute data.

Running **e2fsck** on the recovered MDS/OST can be automated as part of Amanda post-restore operation to reduce the outage. The distributed filesystem check (**lfsck**) must be done manually by the system administrator.

References

- Amanda home: http://amanda.zmanda.com/
- Amanda documentation : http://wiki.zmanda.com/
- Lustre documentation: http://wiki.lustre.org/