MDT-OST Consistency Scope Statement

> This document has not been approved.

Introduction

The following scope statement applies to the MDT-OST Consistency Scope Statement project within the SFS-DEV-001 contract/SOW dates 08/01/2011.

Problem Statement

MDT-OST consistency will implement functionality for distributed verification and repair of the MDT inode to OST object layout. This will add functionality while the MDT is iterating over the inodes (see Subproject 3.1 Inode Iterator and OI Scrub) to check the file layout (LOV EA) to verify the ostid therein, that the OST objects referenced by the file layout exist, and that each OST object has a back reference to the correct MDT inode. Incorrect or missing back pointers on the OST objects will be corrected, and missing objects will be recreated when detected.

The UID and GID of OST objects will also be verified to match that of the MDT inode to ensure correct quota allocation. After the MDT iteration is complete, any unreferenced OST objects will be linked into the Lustre lost+found directory.

Subproject 3.1, 3.1.5 and this sub-project (3.2) together constitute a complete, scalable replacement of the existing e2fsprogs-based lfsck utility. This will allow the distributed checking and repair of Lustre inter-server state for non-DNE file systems while the file system is on-line.

Project Goals

1. Distributed verification and repair of MDT inode to OST object mappings.
2. Measure the resulting code performance characteristics.

In-Scope

- User-space control code for MDS-OST consistency tool.
- The detected/repaired inconsistencies will be reported to the administrator.
- Administrative documentation in the form of a man page and update to Lustre 2.x manual.
- Documentation for user interface.

Out of Scope

- Design and implementation will strive to be agnostic to the back-end OSD file system. Only ldiskfs OSDs will be tested and landed for this phase, pending the development of a ZFS object iterator.
- Only support verification of currently supported LOV_MAGIC_{V1,V3} RAID-0 layouts. Verifying new layouts should be done by the projects adding the new layouts.
- OST iteration will skip object orphan cleanup if the parent FID is on a MDT that has not been processed.
- Detection and resolution of internal file system inconsistencies is provided by e2fsck and is not within scope of this phase.
- Additional Lustre specific attributes, for example SOM, are not considered for verification.

Project Constraints

- The properly skilled engineer available for this work is a Chinese national.

Project Assumptions

- OpenSFS Functional Test Cluster will be available during the sub-project life time.
- Test Cluster will be available for exclusive use by Intel HPDD foreign nationals of Chinese citizenship.
- Test Cluster will have 8 server nodes and 24 compute nodes.

Key Deliverables

- Signed Milestone documents for project phases:
    - Solution Architecture.
    - Implementation & Test.
    - Acceptance Testing (OpenSFS executed).
- Intel HPDD developed test plan.
- Source code that meets feature requirements and runs with Lustre 2.x Master branch on customer’s site.
- Source code for new test cases.
- MDT-OST consistency code landed in on Lustre 2.x Master branch.
- LU ticket for public visibility of MDT-OST consistency work.

Key Milestones

Milestones and schedule recorded at Project Review

Glossary

Lustre 2.x Master branch - Lustre 2.x community code base maintained by Intel HPDD.