Milestone Completion for the LFSCK 1.5 Subproject 3.1.5 on the Lustre FSCK Project of the SFS-DEV-001 contract.

Revision History

Date	Revision	Author
2013-01-02	Original	R. Henwood

Contents

Introduction

The following milestone completion document applies to Subproject 3.1.5 – LFSCK 1.5: FID-in-Dirent and LinkEA of the Lustre FSCK within Amendment No. 1 on the OpenSFS Lustre Development contract SFS-DEV-001 agreed October 10, 2012.

Subproject Description

Per the contract, Implementation milestone is described as follows:

> *LFSCK 1.5 will implement the functionality to verify and rebuild FID-in-Dirent and linkEA entries. To achieve this, the functionality of the inode iterator (see Subproject 3.1) will be enhanced. This enhancement will ensure the FID-in-Dirent name entry is consistent with the FID in the inode LMA on Lustre 2.x file systems. If the name is inconsistent it will be repaired or rebuilt as necessary.*
> *LFSCK 1.5 will also verify the name entry for normal file (non-directory) is correctly referenced by the inode linkEA and the inode linkEA points to the valid name entry. An unmatched or redundant inode linkEA will be removed on discovery and a correct or missed inode linkEA will be added.*
> *Finally, the FID-in-Dirent process will add IGIF FIDs to the dirent for upgraded 1.8 file systems. This additional step will avoid the need to look up the inode when doing readdir.*

Milestone Completion Criteria

Per the contract, Implementation milestone is described as follows:

> *Contractor shall complete implementation and unit testing for the approved solution. Contractor shall regularly report feature development progress including progress metrics at project meetings and engineers shall share interim unit testing results as they are available. OpenSFS at its discretion may request a code review. Completion of the implementation phase shall occur when the agreed to solution has been completed up to and including unit testing and this functionality can be demonstrated on a test cluster. Code Reviews shall include:*
>
> a. *Discussion led by Contractor engineer providing an overview of Lustre source code changes*
> b. *Review of any new unit test cases that were developed to test changes*

Location of Completed Solution

The agreed solution has been completed and is recorded in the following patches:

Code Review	Description
3268	LU-1866 lfsck: user space interfaces for lfsck_namespace
4807	LU-1866 lfsck: FID-in-dirent and linkEA consistency

New functional tests

New functional tests to automatically verify the acceptance criteria agreed in the LFSCK1.5 Solution Architecture are available in the patch-set:

http://review.whamcloud.com/4807

The tests are contained in modifications to the files:

lustre/tests/sanity-lfsck.sh
lustre/tests/sanity-scrub.sh
lustre/tests/sanity.sh
lustre/tests/test-framework.sh

Demonstration of LFSCK 1.5 functionality.

Functional testing was completed on 2012-12-23. The detailed results are recorded in Appendix A. Section 5 of the LFSCK1.5 Solution Architecture describes the acceptance tests including:

Acceptance test	**Corresponding code test from Appendix A**
5.1 Start/stop FID-in-dirent and linkEA consistency check/repair through userspace commands	`== sanity-lfsck test 0: Control LFSCK manually ==`
5.2 Monitor FID-in-dirent and linkEA consistency check/repair	Present in all tests.
5.3 The FID-in-dirent can be rebuilt after the MDT is restored from file-level backup	`== sanity-lfsck test 4: FID-in-dirent can be rebuilt after MDT file-level backup/restore ==`
5.4 Build FID-in-dirent and linkEA for the MDT upgraded from 1.8-based device	`== sanity-lfsck test 5: LFSCK can handle IFIG object upgrading ==`
5.5 Verify files with multiple hard links	`== sanity-lfsck test 2c: LFSCK can find out and recover missed name entry ==` `== sanity-lfsck test 3a: LFSCK can find and repair incorrect object nlink (1) ==`

	== sanity-lfsck test 3b: LFSCK can find and repair incorrect object nlink (2) ==
5.6 Resume FID-in-dirent and linkEA consistency check/repair from the latest checkpoint	== sanity-lfsck test 6a: LFSCK resumes from last checkpoint (1) == == sanity-lfsck test 6b: LFSCK resumes from last checkpoint (2) ==
5.7 Rate control for FID-in-dirent and linkEA consistency check/repair	== sanity-lfsck test 9a: LFSCK speed control (1) == == sanity-lfsck test 9b: LFSCK speed control (2) ==
5.8 The Lustre system is available during LFSCK for FID-in-dirent and linkEA consistency check/repair	== sanity-lfsck test 10: System is available during LFSCK scanning ==
Core functionality tests:	== sanity-lfsck test 1a: LFSCK can find and repair crashed FID-in-dirent == == sanity-lfsck test 1b: LFSCK can find and repair missed FID-in-LMA == == sanity-lfsck test 1c: LFSCK can find out and repair missed object == == sanity-lfsck test 2a: LFSCK can find and repair crashed linkEA == == sanity-lfsck test 2b: LFSCK can find out and remove invalid linkEA entry == == sanity-lfsck test 7a: non-stopped LFSCK should auto restarts after MDS remount (1) == == sanity-lfsck test 7b: non-stopped LFSCK should auto restarts after MDS remount (2) == == sanity-lfsck test 8: LFSCK state machine ==

The successful test is recorded in Maloo at:
https://maloo.whamcloud.com/test_sessions/d9c100a0-4df6-11e2-9dc7-52540035b04c

Conclusion

Implementation has been completed according to the agreed criteria.

Appendix A: functional test results from 2012-12-29

```
Logging to shared log directory: /tmp/test_logs/1355693517
Checking servers environments
Checking clients RHEL6 environments
Loading modules from /root/Work/Lustre/L11/lustre-release/lustre
detected 2 online CPUs by sysfs
Force libcfs to create 2 CPU partitions
../libcfs/libcfs/libcfs options: 'cpu_npartitions=2'
debug=vfstrace rpctrace dlmtrace neterror ha config ioctl super
subsystem_debug=all -lnet -lnd -pinger
../lnet/lnet/lnet options: 'accept=all'
gss/krb5 is not supported
quota/lquota options: 'hash_lqs_cur_bits=3'
Setup mgs, mdt, osts
Starting mds1:   -o loop /tmp/lustre-mdt1 /mnt/mds1
Started lustre-MDT0000
Starting ost1:   -o loop /tmp/lustre-ost1 /mnt/ost1
Started lustre-OST0000
Starting ost2:   -o loop /tmp/lustre-ost2 /mnt/ost2
Started lustre-OST0001
Starting client: RHEL6: -o user_xattr,flock RHEL6@tcp:/lustre /mnt/lustre
Starting client RHEL6: -o user_xattr,flock RHEL6@tcp:/lustre /mnt/lustre
Started clients RHEL6:
RHEL6@tcp:/lustre on /mnt/lustre type lustre (rw,user_xattr,flock)
Using TIMEOUT=20
enable jobstats, set job scheduler as procname_uid
disable quota as required
excepting tests:

== sanity-lfsck test 0: Control LFSCK manually == 05:32:23 (1355693543)
formatall
setupall
preparing... 10 * 10 files will be created.
prepared.
stop mds1
start mds1
fail_val=3
fail_loc=0x1600
Started LFSCK on the MDT device lustre-MDT0000: namespace.
name: lfsck_namespace
magic: 0xa0629d03
version: 2
status: scanning-phase1
flags:
param:
time_since_last_completed: N/A
time_since_latest_start: 0 seconds
time_since_last_checkpoint: 0 seconds
latest_start_position: 115, N/A, N/A
last_checkpoint_position: 115, N/A, N/A
first_failure_position: N/A, N/A, N/A
checked_phase1: 0
```

```
checked_phase2: 0
updated_phase1: 0
updated_phase2: 0
failed_phase1: 0
failed_phase2: 0
dirs: 0
M-linked: 0
nlinks_repaired: 0
ents_added: 0
success_count: 0
run_time_phase1: 0 seconds
run_time_phase2: 0 seconds
average_speed_phase1: 0 items/sec
average_speed_phase2: N/A
real-time_speed_phase1: 0 items/sec
real-time_speed_phase2: N/A
current_position: 116, N/A, N/A
Stopped LFSCK on the MDT device lustre-MDT0000.
Started LFSCK on the MDT device lustre-MDT0000: namespace.
fail_loc=0
fail_val=0
Resetting fail_loc on all nodes...done.
PASS 0 (39s)

== sanity-lfsck test 1a: LFSCK can find out and repair crashed FID-in-dirent ==
05:33:02 (1355693582)
formatall
setupall
preparing... 1 * 1 files will be created.
prepared.
stop mds1
start mds1
Starting client: RHEL6: -o user_xattr,flock RHEL6@tcp:/lustre /mnt/lustre
fail_loc=0x1501
fail_loc=0
10.211.55.5@tcp:/lustre /mnt/lustre lustre rw,flock,user_xattr 0 0
Stopping client RHEL6 /mnt/lustre (opts:)
Started LFSCK on the MDT device lustre-MDT0000: namespace.
Starting client: RHEL6: -o user_xattr,flock RHEL6@tcp:/lustre /mnt/lustre
fail_loc=0x1505
fail_loc=0
Resetting fail_loc on all nodes...done.
PASS 1a (44s)

== sanity-lfsck test 1b: LFSCK can find out and repair missed FID-in-LMA ==
05:33:46 (1355693626)
formatall
setupall
preparing... 1 * 1 files will be created.
prepared.
stop mds1
start mds1
Starting client: RHEL6: -o user_xattr,flock RHEL6@tcp:/lustre /mnt/lustre
fail_loc=0x1502
fail_loc=0
```

```
10.211.55.5@tcp:/lustre /mnt/lustre lustre rw,flock,user_xattr 0 0
Stopping client RHEL6 /mnt/lustre (opts:)
fail_loc=0x1506
Started LFSCK on the MDT device lustre-MDT0000: namespace.
fail_loc=0
Starting client: RHEL6: -o user_xattr,flock RHEL6@tcp:/lustre /mnt/lustre
fail_loc=0x1505
fail_loc=0
Resetting fail_loc on all nodes...done.
PASS 1b (38s)

== sanity-lfsck test 1c: LFSCK can find out and repair missed object == 05:34:24
(1355693664)
formatall
setupall
preparing... 1 * 1 files will be created.
prepared.
stop mds1
start mds1
Starting client: RHEL6: -o user_xattr,flock RHEL6@tcp:/lustre /mnt/lustre
fail_loc=0x1503
10.211.55.5@tcp:/lustre /mnt/lustre lustre rw,flock,user_xattr 0 0
Stopping client RHEL6 /mnt/lustre (opts:)
Started LFSCK on the MDT device lustre-MDT0000: namespace.
Starting client: RHEL6: -o user_xattr,flock RHEL6@tcp:/lustre /mnt/lustre
fail_loc=0
Resetting fail_loc on all nodes...done.
PASS 1c (41s)

== sanity-lfsck test 2a: LFSCK can find out and repair crashed linkEA entry ==
05:35:05 (1355693705)
formatall
setupall
preparing... 1 * 1 files will be created.
prepared.
stop mds1
start mds1
Starting client: RHEL6: -o user_xattr,flock RHEL6@tcp:/lustre /mnt/lustre
fail_loc=0x1603
fail_loc=0
10.211.55.5@tcp:/lustre /mnt/lustre lustre rw,flock,user_xattr 0 0
Stopping client RHEL6 /mnt/lustre (opts:)
Started LFSCK on the MDT device lustre-MDT0000: namespace.
Starting client: RHEL6: -o user_xattr,flock RHEL6@tcp:/lustre /mnt/lustre
Resetting fail_loc on all nodes...done.
PASS 2a (40s)

== sanity-lfsck test 2b: LFSCK can find out and remove invalid linkEA entry ==
05:35:45 (1355693745)
formatall
setupall
preparing... 1 * 1 files will be created.
prepared.
stop mds1
start mds1
```

```
Starting client: RHEL6: -o user_xattr,flock RHEL6@tcp:/lustre /mnt/lustre
fail_loc=0x1604
fail_loc=0
10.211.55.5@tcp:/lustre /mnt/lustre lustre rw,flock,user_xattr 0 0
Stopping client RHEL6 /mnt/lustre (opts:)
Started LFSCK on the MDT device lustre-MDT0000: namespace.
Starting client: RHEL6: -o user_xattr,flock RHEL6@tcp:/lustre /mnt/lustre
Resetting fail_loc on all nodes...done.
PASS 2b (44s)

== sanity-lfsck test 2c: LFSCK can find out and recover missed name entry ==
05:36:29 (1355693789)
formatall
setupall
preparing... 1 * 1 files will be created.
prepared.
stop mds1
start mds1
Starting client: RHEL6: -o user_xattr,flock RHEL6@tcp:/lustre /mnt/lustre
fail_loc=0x1605
fail_loc=0
10.211.55.5@tcp:/lustre /mnt/lustre lustre rw,flock,user_xattr 0 0
Stopping client RHEL6 /mnt/lustre (opts:)
Started LFSCK on the MDT device lustre-MDT0000: namespace.
Starting client: RHEL6: -o user_xattr,flock RHEL6@tcp:/lustre /mnt/lustre
Resetting fail_loc on all nodes...done.
PASS 2c (43s)

== sanity-lfsck test 3a: LFSCK can find out and repair incorrect object nlink (1)
== 05:37:12 (1355693832)
formatall
setupall
preparing... 1 * 1 files will be created.
prepared.
stop mds1
start mds1
Starting client: RHEL6: -o user_xattr,flock RHEL6@tcp:/lustre /mnt/lustre
fail_loc=0x1606
fail_loc=0
10.211.55.5@tcp:/lustre /mnt/lustre lustre rw,flock,user_xattr 0 0
Stopping client RHEL6 /mnt/lustre (opts:)
Started LFSCK on the MDT device lustre-MDT0000: namespace.
Starting client: RHEL6: -o user_xattr,flock RHEL6@tcp:/lustre /mnt/lustre
Resetting fail_loc on all nodes...done.
PASS 3a (42s)

== sanity-lfsck test 3b: LFSCK can find out and repair incorrect object nlink (2)
== 05:37:54 (1355693874)
formatall
setupall
preparing... 1 * 1 files will be created.
prepared.
stop mds1
start mds1
Starting client: RHEL6: -o user_xattr,flock RHEL6@tcp:/lustre /mnt/lustre
```

```
fail_loc=0x1607
fail_loc=0
10.211.55.5@tcp:/lustre /mnt/lustre lustre rw,flock,user_xattr 0 0
Stopping client RHEL6 /mnt/lustre (opts:)
Started LFSCK on the MDT device lustre-MDT0000: namespace.
Starting client: RHEL6: -o user_xattr,flock RHEL6@tcp:/lustre /mnt/lustre
Resetting fail_loc on all nodes...done.
PASS 3b (43s)
```

== sanity-lfsck test 4: FID-in-dirent can be rebuilt after MDT file-level backup/restore == 05:38:37 (1355693917)

```
formatall
setupall
preparing... 3 * 3 files will be created.
prepared.
stop mds1
file-level backup/restore on mds1:/tmp/lustre-mdt1
backup EA
/root/Work/Lustre/L11/lustre-release/lustre/tests
backup data
reformat new device
restore data
restore EA
/root/Work/Lustre/L11/lustre-release/lustre/tests
remove recovery logs
removed `/mnt/brpt/CATALOGS'
start mds1 with disabling OI scrub
fail_val=1
fail_loc=0x1601
Started LFSCK on the MDT device lustre-MDT0000: namespace.
fail_loc=0
fail_val=0
Starting client: RHEL6: -o user_xattr,flock RHEL6@tcp:/lustre /mnt/lustre
fail_loc=0x1505
fail_loc=0
Resetting fail_loc on all nodes...done.
PASS 4 (46s)
```

== sanity-lfsck test 5: LFSCK can handle IFIG object upgrading == 05:39:23 (1355693963)

```
formatall
fail_loc=0x1504
setupall
preparing... 1 * 1 files will be created.
fail_loc=0
prepared.
stop mds1
file-level backup/restore on mds1:/tmp/lustre-mdt1
backup EA
/root/Work/Lustre/L11/lustre-release/lustre/tests
backup data
reformat new device
restore data
restore EA
/root/Work/Lustre/L11/lustre-release/lustre/tests
```

```
remove recovery logs
removed `/mnt/brpt/CATALOGS'
start mds1 with disabling OI scrub
fail_val=1
fail_loc=0x1601
Started LFSCK on the MDT device lustre-MDT0000: namespace.
fail_loc=0
fail_val=0
Starting client: RHEL6: -o user_xattr,flock RHEL6@tcp:/lustre /mnt/lustre
fail_loc=0x1505
fail_loc=0
Resetting fail_loc on all nodes...done.
PASS 5 (45s)

== sanity-lfsck test 6a: LFSCK resumes from last checkpoint (1) == 05:40:08
(1355694008)
formatall
setupall
preparing... 10 * 10 files will be created.
prepared.
stop mds1
start mds1
fail_val=1
fail_loc=0x1600
Started LFSCK on the MDT device lustre-MDT0000: namespace.
fail_loc=0x80001608
fail_val=1
fail_loc=0x1600
Started LFSCK on the MDT device lustre-MDT0000: namespace.
fail_loc=0
fail_val=0
Resetting fail_loc on all nodes...done.
PASS 6a (50s)

== sanity-lfsck test 6b: LFSCK resumes from last checkpoint (2) == 05:40:58
(1355694058)
formatall
setupall
preparing... 10 * 10 files will be created.
prepared.
stop mds1
start mds1
fail_val=1
fail_loc=0x1601
Started LFSCK on the MDT device lustre-MDT0000: namespace.
fail_loc=0x80001609
fail_val=1
fail_loc=0x1601
Started LFSCK on the MDT device lustre-MDT0000: namespace.
sanity-lfsck.sh: line 622: [: N/A: integer expression expected
fail_loc=0
fail_val=0
Resetting fail_loc on all nodes...done.
PASS 6b (42s)
```

== sanity-lfsck test 7a: non-stopped LFSCK should auto restarts after MDS remount (1) == 05:41:40 (1355694100)

```
formatall
setupall
preparing... 10 * 10 files will be created.
prepared.
stop mds1
start mds1
fail_val=1
fail_loc=0x1601
Started LFSCK on the MDT device lustre-MDT0000: namespace.
stop mds1
start mds1
fail_loc=0
fail_val=0
Resetting fail_loc on all nodes...done.
PASS 7a (43s)
```

== sanity-lfsck test 7b: non-stopped LFSCK should auto restarts after MDS remount (2) == 05:42:23 (1355694143)

```
formatall
setupall
preparing... 2 * 2 files will be created.
prepared.
stop mds1
start mds1
Starting client: RHEL6: -o user_xattr,flock RHEL6@tcp:/lustre /mnt/lustre
fail_loc=0x1604
fail_val=3
fail_loc=0x1602
Started LFSCK on the MDT device lustre-MDT0000: namespace.
stop mds1
start mds1
fail_loc=0
fail_val=0
Resetting fail_loc on all nodes...done.
PASS 7b (43s)
```

== sanity-lfsck test 8: LFSCK state machine == 05:43:06 (1355694186)

```
formatall
setupall
preparing... 20 * 20 files will be created.
prepared.
stop mds1
start mds1
Starting client: RHEL6: -o user_xattr,flock RHEL6@tcp:/lustre /mnt/lustre
fail_loc=0x1603
fail_loc=0x1604
fail_val=2
fail_loc=0x1601
Started LFSCK on the MDT device lustre-MDT0000: namespace.
Stopped LFSCK on the MDT device lustre-MDT0000.
Started LFSCK on the MDT device lustre-MDT0000: namespace.
fail_loc=0x80001609
fail_loc=0x1600
```

```
Started LFSCK on the MDT device lustre-MDT0000: namespace.
fail_loc=0x160a
stop mds1
fail_loc=0x160b
start mds1
fail_loc=0x1601
Started LFSCK on the MDT device lustre-MDT0000: namespace.
stop mds1
fail_loc=0x160b
start mds1
fail_val=2
fail_loc=0x1602
Started LFSCK on the MDT device lustre-MDT0000: namespace.
fail_loc=0
fail_val=0
Resetting fail_loc on all nodes...done.
PASS 8 (55s)

== sanity-lfsck test 9a: LFSCK speed control (1) == 05:44:01 (1355694241)
formatall
setupall
preparing... 70 * 70 files will be created.
prepared.
stop mds1
start mds1
Started LFSCK on the MDT device lustre-MDT0000: namespace.
Resetting fail_loc on all nodes...done.
PASS 9a (73s)

== sanity-lfsck test 9b: LFSCK speed control (2) == 05:45:14 (1355694314)
formatall
setupall
preparing... 0 * 0 files will be created.
prepared.
stop mds1
start mds1
Starting client: RHEL6: -o user_xattr,flock RHEL6@tcp:/lustre /mnt/lustre
Another preparing... 50 * 50 files (with error) will be created.
fail_loc=0x1604
fail_loc=0x160c
Started LFSCK on the MDT device lustre-MDT0000: namespace.
fail_loc=0
Started LFSCK on the MDT device lustre-MDT0000: namespace.
Resetting fail_loc on all nodes...done.
PASS 9b (80s)

== sanity-lfsck test 10: System is available during LFSCK scanning == 05:46:34
(1355694394)
formatall
setupall
preparing... 1 * 1 files will be created.
prepared.
stop mds1
start mds1
Starting client: RHEL6: -o user_xattr,flock RHEL6@tcp:/lustre /mnt/lustre
```

```
fail_loc=0x1603
fail_loc=0x1604
fail_loc=0x1605
fail_loc=0
10.211.55.5@tcp:/lustre /mnt/lustre lustre rw,flock,user_xattr 0 0
Stopping client RHEL6 /mnt/lustre (opts:)
Starting client: RHEL6: -o user_xattr,flock RHEL6@tcp:/lustre /mnt/lustre
Started LFSCK on the MDT device lustre-MDT0000: namespace.
10.211.55.5@tcp:/lustre /mnt/lustre lustre rw,flock,user_xattr 0 0
Stopping client RHEL6 /mnt/lustre (opts:)
Starting client: RHEL6: -o user_xattr,flock RHEL6@tcp:/lustre /mnt/lustre
Resetting fail_loc on all nodes...done.
PASS 10 (73s)
Stopping clients: RHEL6 /mnt/lustre (opts:)
Stopping client RHEL6 /mnt/lustre opts:
Stopping clients: RHEL6 /mnt/lustre2 (opts:)
Stopping /mnt/mds1 (opts:-f) on RHEL6
Stopping /mnt/ost1 (opts:-f) on RHEL6
Stopping /mnt/ost2 (opts:-f) on RHEL6
Loading modules from /root/Work/Lustre/L11/lustre-release/lustre
detected 2 online CPUs by sysfs
Force libcfs to create 2 CPU partitions
debug=vfstrace rpctrace dlmtrace neterror ha config ioctl super
subsystem_debug=all -lnet -lnd -pinger
gss/krb5 is not supported
Formatting mgs, mds, osts
Format mds1: /tmp/lustre-mdt1
Format ost1: /tmp/lustre-ost1
Format ost2: /tmp/lustre-ost2
== sanity-lfsck test complete, duration 969 sec == 05:48:06 (1355694486)
```