Lustre Releases

Peter Jones
Lustre Support and Releases Manager
September 24th 2012
Community Lustre Roadmap

- **Maintenance Releases on an ad-hoc basis**
  - 1.8.7-wc1
  - 1.8.8-wc1

- **Maintenance Releases every quarter**
  - 2.1.1
  - 2.1.2
  - 2.1.3
  - 2.1.4
  - 2.1.5
  - 2.1.6
  - 2.4.1

- **Feature Releases**
  - 2.2
  - 2.3
  - 2.4
  - 2.5

**Sponsor for Intel Development and Releases:**
- ORNL
- OpenSFS
- LLNL
- Intel

**Third Party Development:**
- CEA
- Xyratex

- Imperative Recovery
- Dirop SMP Scaling
- Wide Striping
- Statahead Speedup
- Server Stack SMP Scaling
- Online check/scrub
- Job Stats
- OSD restructuring
- DNE Phase 1
- HSM
- Network Request Scheduler
- DNE Phase 2
- LFSCK MDT-OST Consistency

Q4 2011

Q1 2012

Q2 2013

Q3 2013

Q4 2013
Lustre 1.8.x

Majority of Lustre production sites still running this codeline
  • Many new deployments still occurring using 1.8.x releases
  • OpenSFS survey showed 67% using WC 1.8.x release

Support life lengthened
  • Lustre 1.8.8-wc1 was released in May; 1.8.9-wc1 likely at some point
  • 1.8.x client interop still planned for 2.4 (but deprecated in 2.5)

Oracle still making 1.8.x releases
  • 2.x interop not tested or supported
Lustre 2.1.x

Increasing number of sites in production on 2.1.x

- LLNL, NASA and CEA all running in production
- OpenSFS survey showed 21% running 2.1.x releases

Will remain maintenance release stream until 2.4 GA

- 2.1.3 released last month
- Quarterly bugfix releases scheduled until Q2 2013

RHEL6 servers and larger LUN size main attraction
Lustre 2.2.x

Feature release only; no maintenance releases scheduled

• Went GA March 30\textsuperscript{th}
• OpenSFS survey showed 8% running 2.2
• Cray plan to baseline next release on 2.2; Sanger running stably

A number of new features:

Asynchronous Glimpse Lock/Statahead (LU-925/LU-389)

• Improved performance for \texttt{ls \textendash l/find} and accessing object attributes (file sizes/xtime etc)
• Development funded by ORNL
Client Parallel Checksums (LU-884)
  - Improved support for mmap and better performance using checksums
  - Development funded by ORNL

Imperative Recovery (LU-580)
  - Faster recovery
  - Development funded by ORNL

Large Xattrs (aka Wide Striping) (LU-80)
  - Maximum stripe size raised from 160 to 2000; max file size increased from 320 TB to 64PB
  - Completing work by Sun funded by ORNL
  - ORNL/Xyratex helped with this initiative
Lustre 2.2.x (cont2)

MDS-survey (LU-593/LU-633)
• Tool for MDS performance benchmarking
• Development funded by LLNL

Parallel Directory Operations (LU-50)
• Improved performance when multiple processes access/change the same directory in parallel
• Development funded by OpenSFS
Lustre 2.3.x

Feature release only; no maintenance releases scheduled

• Scheduled for GA September 30th; in final release testing
• Large number of contributing organizations (Bull, CEA, Cray, DDN, EMC, Fujitsu, Intel, LLNL, ORNL, TACC, Ultrascale, UVT, Xyratex)

A number of new features

CRC (LU-1201/LU-1339)
• Improved version of CRC-32
• Development by Xyratex
Lustre 2.3.x (cont)

Flock Improvements (LU-1156/LU-1157)
• Improved flock handling on the servers
• Development by Xyratex

IO Engine Rewrite (LU-1030)
• Performance improvement for CLIO RPC formation
• Development by Intel

Job Stats (LU-694)
• Send client-side job scheduler ID to server for IO/metadata statistics
• Development by Intel
Lustre 2.3.x (cont2)

OI Scrub(LU-957)
  • Inode iterator and OI scrub (lfsck replacement phase 1)
  • Development funded by OpenSFS

Server Stack SMP (LU-56/LU-1315/LU-1316)
  • Performance improvements for running with larger numbers of cores
  • Development funded by OpenSFS
Lustre 2.4.x

Will become maintenance release stream for 18 months following GA

- Scheduled for GA April 30\textsuperscript{th} 2013
- Will drop RHEL5 and SLES11 SP1 clients; add SLES11 SP2 clients

A number of features candidates for inclusion

DNE Phase 1 (LU-1187)

- Multiple MDS/MDTs in a single filesystem, partitioned by subdir
- Development funded by OpenSFS
Lustre 2.4.x (cont)

HSM (LU-169/LU-827/LU-941/LU-1333/LU-1338)
• Enables data to be transferred between different storage types
• Development by CEA

Network Request Scheduler (LU-398)
• Mechanism to apply policies to how RPC requests are handled
• Development by Xyratex/Intel

ZFS Support (LU-1305)
• Will provide Lustre servers without kernel patches
• Online check/scrub/repairs (no more e2fsck!)
• Development funded by LLNL
Lustre 2.5.x and Beyond

Content of future releases will become clearer soon

- Next OpenSFS Development RFP due out shortly
- Roadmap updates discussed during OpenSGS CDWG
- Overview of development in progress can be found at http://wiki.whamcloud.com/display/PUB/Lustre+Community+Development+in+Progress
Thank You