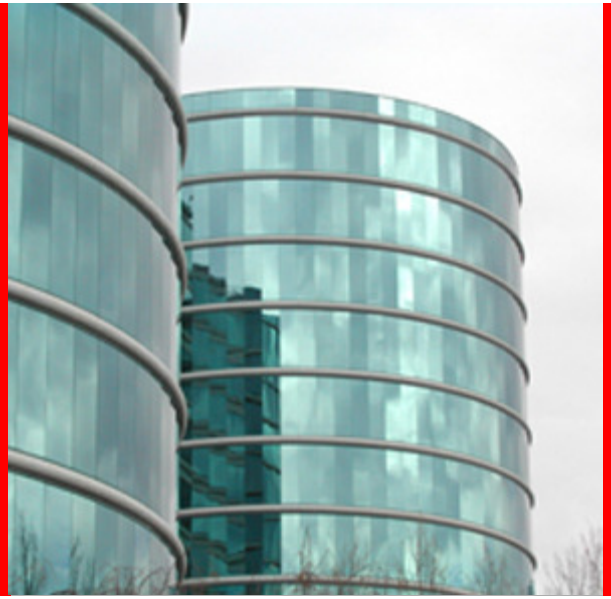


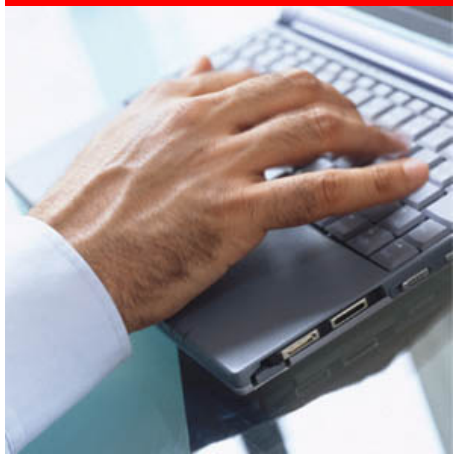
ORACLE

ORACLE THIRTY YEARS 30

Lustre 2.1 “Yangtze I” Update

Nirant Puntambekar, Vitaly Fertman
Steering Committee, October 2010

Agenda



- Since last review
- Scope of Work
- Project Health
- Top Issues and Risks
- Schedule
- Next Steps
- Change Control Request
- In scope features – Deep Dive
- Conclusions/Approvals

Since last review

- Roadmap for Lustre releases formalized
- 2.1 C-team continues to meet daily
- Vitaly is the tech lead and the gatekeeper
- Daily YALA runs continued, HEAD stability being monitored daily.
- Bi-weekly build schedule implemented. QE, Scale and performance testing being conducted

2.1 (Yangtze I) Release Scope

Tier 1

- Asynch Journal Commits
- SMP Scaling
- IO Performance
- Enable ext4
- RHEL6 server support
- Important Bugs

Tier 2

- 16TB LUNs support
- Security (auth &encrypt) Preview only
- Wide Striping
- * Params_tree
- * OI Scrub (Backup/Restore)
- * HSM v0.5

Tier 3

- IB Bonding
-

Project Health – Bug Metrics

- Current Bugs count
  - P2: 37 P3: 216 P4: 102 P5: 28
- Landing Rate
  - 10 patches/week (3 big, 5-10 small)
- Historical Bug Fix Rate (for 2.0)
  - Dedicated team : 0.66 bugs/week/engineer
  - Other teams : 0.23 bugs/week/engineer
- Current team
  - Dedicated : 4 (Vitaly, Vladimir, Zam, Rahul)
  - Other engineers (part time) : 14
- Projection : 5.8 bugs/week fix rate

Project Health – Stability

- Daily YALA runs with both ext3 & ext4. Not much variation observed between the two.
- Ext4 turned on by default with build 5 (Oct 11)
- C-team monitoring results daily and backing out any offending patches after landing.

Stability - HEAD Daily Test Results

Test	August	Sept
obdfilter-survey	61%	75%
ior	93%	89%
metabench	100%	100%
simul	100%	100%
racer	97%	97%
sanity	96%	96%
sanity_benchmark	90%	88%
lustre-rsync-test	93%	93%
sanityn	86%	95%
liblustre	96%	97%
replay_single	95%	79%
ost_pools	66%	79%
recovery_small	95%	97%
replay_dual	90%	97%
sanity_quota	97%	97%
insanity	98%	97%
sanity_sec	87%	100%
performance_sanity	97%	95%
replay_vbr	93%	93%
replay_ost_single	98%	98%
conf_sanity	69%	76%
metadata_updates	100%	100%
parallel_scale	81%	79%
parallel_scale_nfsv3	64%	65%
parallel_scale_nfsv4	92%	85%
lnet_selftest	92%	96%
recovery-mds-scale	86%	65%
recovery-double-scale	76%	59%
recovery-random-scale	90%	67%

Project Health – Performance

- Tracking Bug 22697
- Biweekly tests comparing 1.8.x and 2.1 being run on the Burlington cluster
- Current results in the next few slides
- Next Steps
  - Continue monitoring the runs and watch for IO improvements due to ongoing CLIO work.
  - Observe the effect of SMP patches on MD performance as they land.

Project Health – Performance

Lustre 2.x Performance

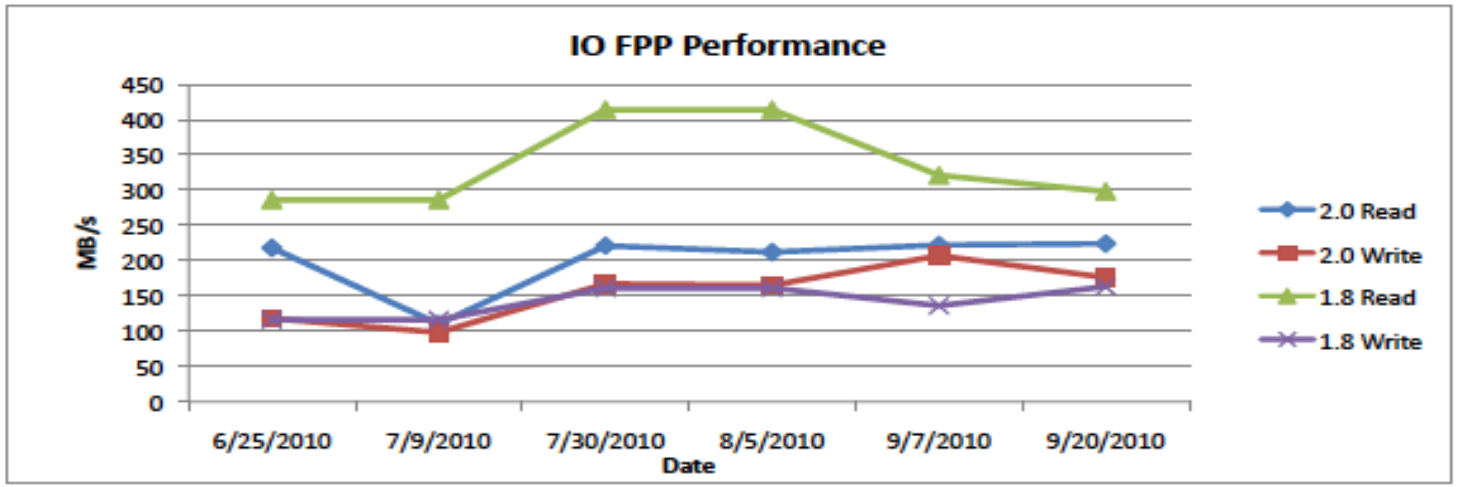
Goal	System	Lustre 2.x	Lustre 1.8.4	% perf of 1.8.x
		MB/s	**MB/s**	
IO FPP Read	Burlington	227	386	58.81%
IO FPP Write	Burlington	206	162	127.16%
IO SSF Read	Burlington	223	335	66.57%
IO SSF Write	Burlington	199	166	119.88%
		ops/s	**ops/s**	
MD Unique Dirs - File Create	Burlington	5574	8745	63.74%
MD Unique Dirs - Dir Create	Burlington	7509	9734	77.14%
MD Unique Dirs - File Stat	Burlington	9929	11807	84.09%
MD Shared Dirs - File Create	Burlington	6091	3569	170.66%
MD Shared Dirs - Dir Create	Burlington	7092	3333	212.78%
MD Shared Dirs - File Stat	Burlington	9868	12042	81.95%

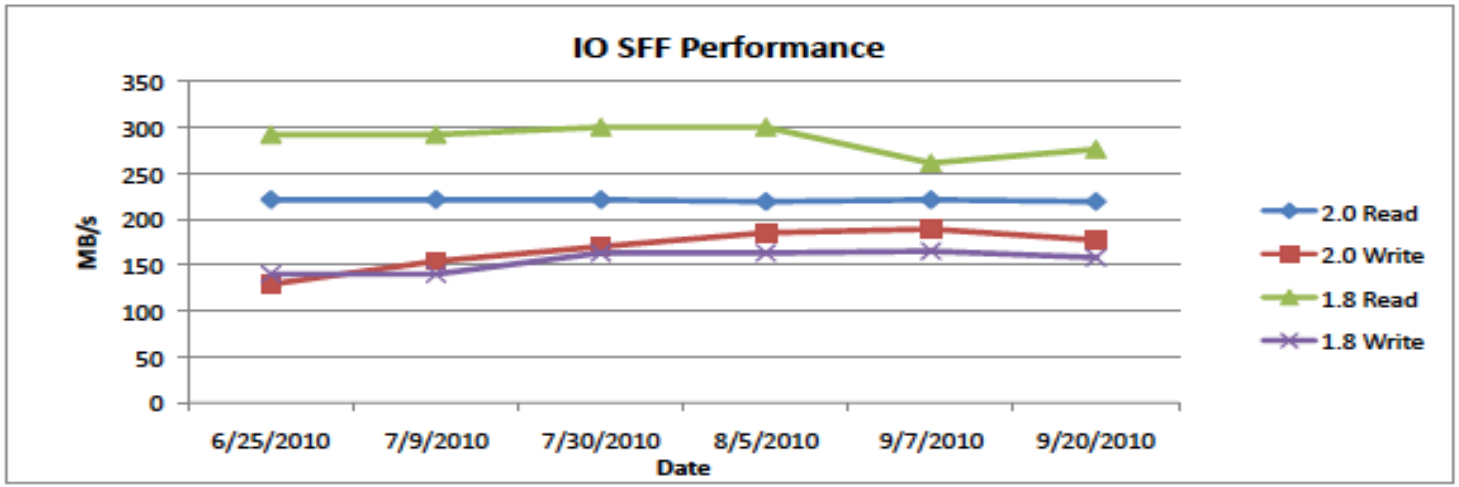
Note : Above data from 09/30 run. History available in Eng metrics report

% Field - Key
Less than 60%
60-90%
90% and above

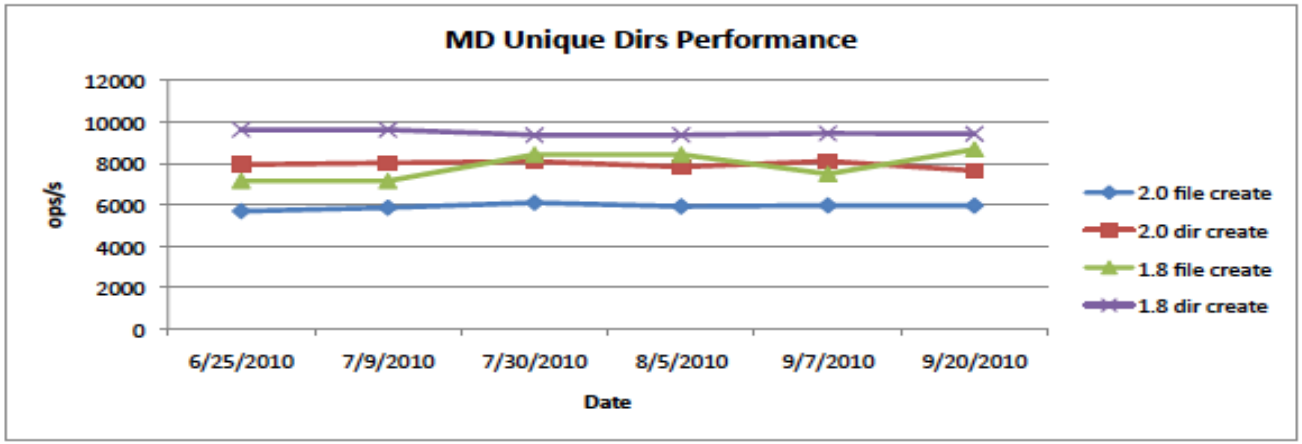
Proiect Health – Performance - IO

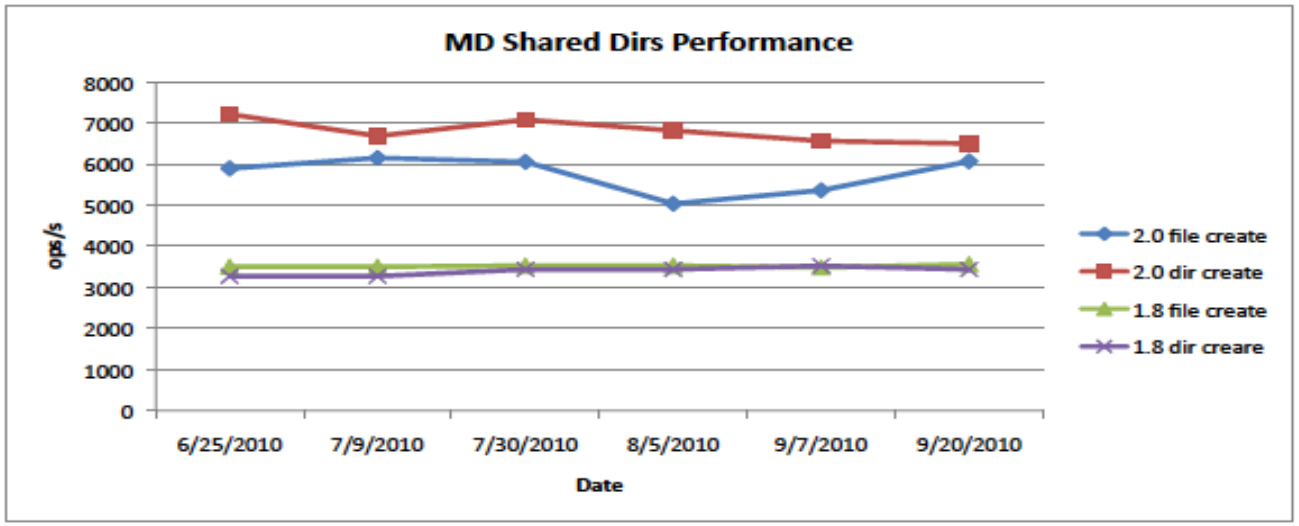
Lustre Performance Metrics



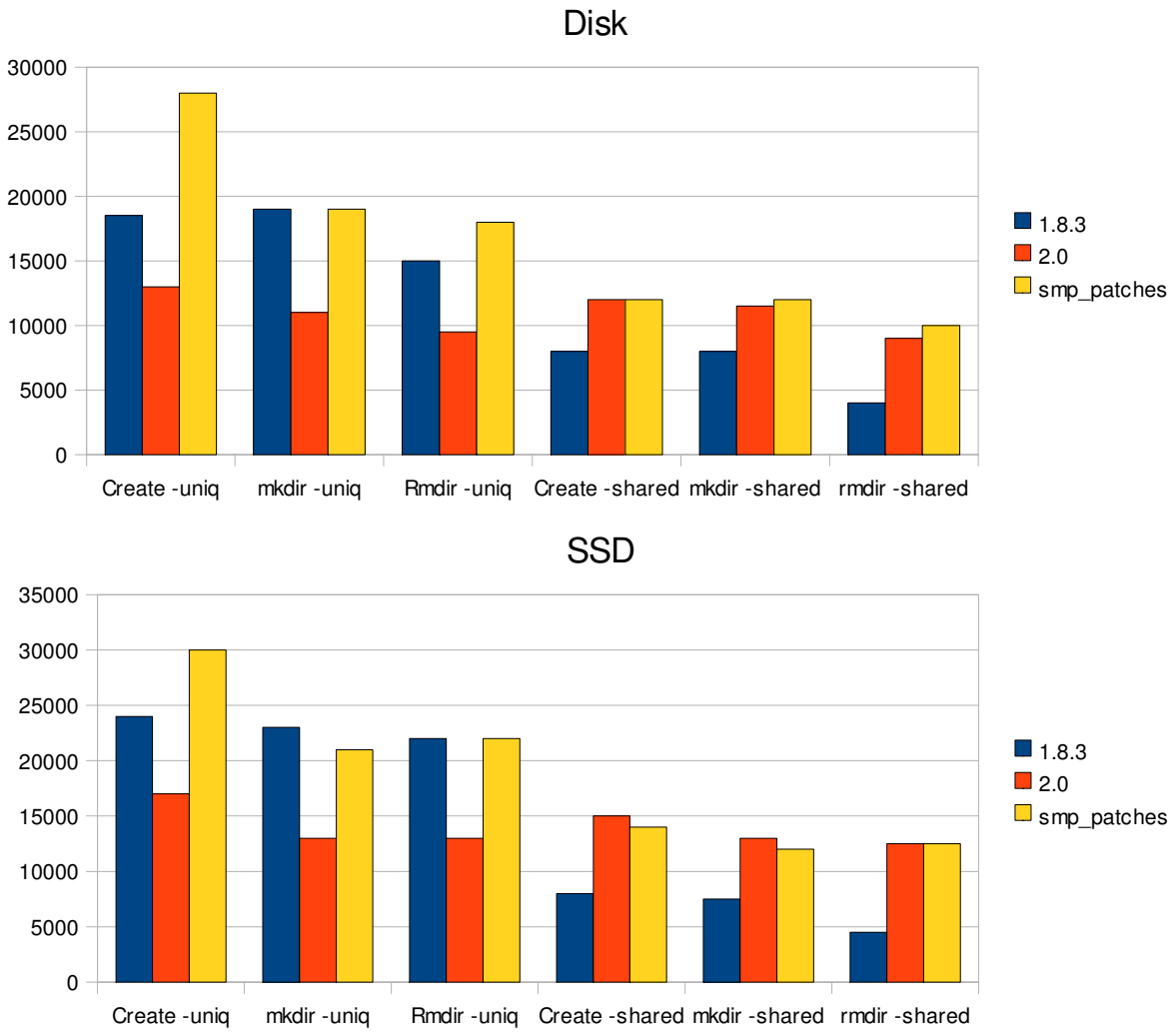


Project Health – Performance - MD





Yangtze MD Performance - Hyperion Testing



Project Macro Schedule

Project Milestones	Schedule	Status	Comments
Bug fixing, Coding, inspections	July – Oct	Y	
Code Freeze (General landings)	Oct 15, 10	Y	
Code Freeze (blocker features only)	Nov 15, 10	Y	
Stabilization (final blocker bugs only)	Nov – Dec 15	Y	
Release	Dec 15, 2010	Y	
Yangtze II	Spring 2011	Y	

Individual Sub-projects : Status

Sub Project	Target	Status	Comments
Asynch Journal Commits	July	B	Completed
SMP Scaling	Mid Nov	Y	Large number of patches
IO Performance	Mid Nov	Y	Investigations in progress
ext4	Mid Oct	G	Close, needs final hyperion retest
RHEL6 servers	Mid Nov	G	Patches close to be posted for inspection
16TB LUNs	Mid Nov	G	Testing in progress on Hillsboro thumper
Security (auth and encryption)	Mid Nov	Y	Preview in 2.0 -, will remain a preview in 2.x, bug fixes only
Params_tree	Mid Nov	G	All patches posted for inspections
Wide Striping	Mid Nov	Y	Patches being tested at ORNL
OI Scrub(Backup/Restore)	Mid Nov	Y	Most of code ready, need DDN CA for final landing
HSM 0.5	Mid Nov	Y	Land whatever we can for this release
IB Bonding	Mid Nov	Y	Work in progress, needs testing in lab

Project Risks & Issues

ID	Category	Description	Mitigation Plan
1	Scope	Large feature set, not all will be complete by code freeze	Track Tier 1 projects closely, others can be deferred to the Spring 2011 release or classified as preview.
2	Schedule	Code stabilization takes longer after code freeze	Controlled landings, daily monitoring of HEAD stability. Assign last month for stabilization only
3	Quality	Performance targets may not be met	Monitor closely the CLIO performance work, if required defer some performance targets to Spring 2011 rel.
4	Resources	Attrition has affected key members of team	Defer the release till mid December to allow for more time for code landings.
5	Testing	Hyperion IB network might not be available post September	Follow up procurement of QDR IB cards. If not test with 10GigE only

Change Control Requests

- None

Change Request History

- Defer release to Dec 15, 2010 (by 1.5 months) to account for attrition, delays and ongoing investigations – Approved Sept SC Meeting

Sub-Projects – Deep Dive

SMP Scaling Optimizations - Yellow

Engineer : Liang Zhen

- Updates
  - Liang to continue work on posting patches till Sept 15th under a CA
  - About 35 patches expected for all the code, 10 posted already
- Risks
  - Not sure if Liang will continue SMP work at WhamCloud
  - Large number of inspections and subsequent landings lined up.
  - Lot of code changes in various subsystems, potential to destabilize HEAD
- Next Steps
  - Track with Liang to make sure all patches are posted
  - Performance testing to validate MD performance

IO Performance - Yellow

Engineer : Oleg Drokin

- Updates
  - Oleg working on removing bloats in the CLIO code, seeing some improvements.
  - EricMei working on CLIO hash optimizations (22683)
  - Other areas identified – parallelize per cpu ptlrpcd, 4M IO, better handling of small chunk IO as in1.8.x
- Risks
  - Some areas identified are still in design phase.
  - Ongoing investigations
- Next Steps
  - Track with Oleg and monitor the biweekly performance runs

Enable ext4 by default - Green

Engineer : Vitaly/Oleg

- Updates
  - Vitaly has landed bugs 21137 & 23368 which might fix the blocking bug 22033
  - On limited Hyperion testing, not able to reproduce the bug after the fixes
  - Daily YALA runs with ext4 have started, also being tested on Hyperion
  - ORNL have tested ext4 with 1.8.4 successfully
- Risks
  - Hyperion full cluster unavailable till end Sept, only 130 nodes currently available
- Next Steps
  - Retest the problem on larger scale on Hyperion once available and switch as default.

RHEL 6 Servers - Green

Engineer : Kalpak/Rahul

- Updates
  - Kalpak has finished about 90% of the work
  - Have recently received RHEL6 rpms from Bull, Kalpak in process of ironing out the final bugs.
- Risks
  - Dependant on external engineer (Kalpak) to complete the work.
- Next Steps
  - Rahul to coordinate the landing and inspections of the patches once posted by Kalpak.

16TB LUNs - Yellow

Engineer : Girish/YuJian

- Updates
  - ORNL have run 16TB LUN tests with 1.8.4, functionally looks good, yet to test performance.
- Risks
  - No real hardware available to test within Oracle.
- Next Steps
  - QE to try to pick up Walter's tests at ORNL and simulate on available hw

Security (Auth & Encryption) - Yellow

Engineer : EricMei/Nico

- Updates
  - Not tested for a while now, some code might need updating
  - Eric Mei feels that a code review with Nico will help.
  - PSC did some 2.0 alpha testing and filed a few bugs
- Risks
  - Eric Mei currently focused on CLIO performance
- Next Steps
  - YuJian to commence testing with help from Eric Mei
  - Create a whitepaper highlighting current state of things

Parameters Tree - Green

Engineer : Emoly

- Updates
  - All patches have been posted for inspections
  - Inspection feedback being worked , patches being updated
- Risks
  - Delays in inspections completion due to other priorities
- Next Steps
  - Track inspection progress and ensure timely landings.

Wide Striping - Red

Engineer : Oleg

- Updates
  - Patches need to be updated for the 2.0 client
  - Need to submit the EA patch upstream for easier maintenance down the line
- Risks
  - Oleg busy with IO performance issues, can only take this on in the background.
  -
- Next Steps
  - Depending on Oleg's availability schedule the work.
  - Only ORNL seems to be wanting this feature, so need to prioritize this accordingly

Backup/Restore - Green

Engineer : Pravin

- Updates
  - One pager completed and inspected
  - Pravin has posted the prototype code for inspection, passed one inspection
- Risks
  - None
- Next Steps
  - Monitor and track the inspections and landings

HSM 0.5 - Yellow

Engineer : TBD/CEA

- Updates
  - Plan is to land as much code as we can for the fall release and only officially support HSM in the Spring 2011 release
- Risks
  - Nathan departure, need to assign a backup
- Next Steps
  - Prepare action plan with Hua on what can be landed

IB Bonding - Red

Engineer : Isaac

- Updates
  - A lot of the initial work for BOM was workarounds for limitations in the IB stack. With 1.5.2 IB stack support the code needs to change
- Risks
  - Isaac has no time to work on this, currently tied up with Cray production problems
  - Limited test hardware in our lab. Only a couple of machines set up with dual port HCA and not true multi rails.
- Next Steps
  - This is next on Isaac's priority list

Appendix

References

2.x.x content aggregator page :

- https://wikis.lustre.org/intra/index.php/Lustre2.x.x

Test Clusters Info

- https://wikis.lustre.org/intra/index.php/TestClusters

HEAD Daily Test Results – 2.0 Release

	Sept	Oct	Nov	Dec	Jan	Feb	Mar	Apr	May	Jun (1-10)	Total Runs
ior	80%	100%	98%	96%	98%	100%	100%	98%	98%	100%	464
simul	95%	72%	94%	96%	98%	100%	98%	98%	98%	100%	453
racer	14%	45%	35%	67%	75%	96%	93%	95%	97%	100%	417
sanity	4%	2%	16%	66%	71%	90%	87%	76%	90%	79%	458
sanity_benchmark	45%	75%	88%	80%	82%	94%	90%	78%	92%	100%	459
lustre-rsync_test	88%	85%	94%	0%	5%	85%	87%	84%	100%	100%	405
sanityn	33%	87%	57%	86%	87%	98%	96%	93%	95%	100%	453
liblustre	94%	98%	80%	98%	88%	98%	95%	51%	100%	100%	461
replay_single	8%	15%	14%	86%	76%	98%	93%	90%	87%	79%	459
metabench	NT	100%	NT	98%	100%	100%	100%	98%	87%	93%	306
recovery_small	50%	82%	98%	90%	83%	90%	93%	94%	98%	100%	457
replay_dual	98%	79%	78%	94%	71%	92%	95%	99%	92%	100%	462
insanity	100%	95%	100%	96%	94%	98%	100%	100%	100%	100%	458
sanity_quota	33%	82%	92%	88%	85%	96%	91%	96%	92%	100%	468
sanity_sec	100%	100%	100%	96%	98%	100%	100%	100%	100%	100%	464
performance_sanity	8%	75%	67%	85%	75%	80%	82%	85%	83%	57%	463
replay_vbr	98%	71%	67%	94%	83%	96%	93%	96%	77%	93%	460
replay_ost_single	10%	91%	92%	96%	88%	100%	91%	96%	95%	100%	462
conf_sanity	79%	84%	73%	67%	60%	90%	95%	91%	93%	69%	459
parallel_scale	0%	18%	55%	27%	57%	38%	63%	60%	33%	50%	87
lnet_selftest	100%	100%	100%	100%	100%	88%	100%	91%	100%	100%	80
recovery-mds-scale	NT	70%	91%	100%	100%	63%	100%	64%	100%	100%	78
recovery-double-scale	NT	44%	91%	91%	71%	50%	63%	73%	91%	100%	77
recovery-random-scale	NT	78%	100%	100%	71%	75%	50%	91%	100%	100%	77
ost_pools	NT	20%	33%	45%	29%	25%	50%	91%	93%	93%	141
parallel_scale_nfsv4	NT	0%	0%	0%	NT	50%	0%	50%	15%	100%	64
parallel_scale_nfsv3	NT	0%	0%	0%	NT	13%	88%	70%	86%	75%	67
metadata_updates	NT	0%	0%	0%	0%	0%	0%	55%	93%	100%	73

The preceding is intended to outline our general product direction. It is intended for information purposes only, and may not be incorporated into any contract. It is not a commitment to deliver any material, code, or functionality, and should not be relied upon in making purchasing decisions.
The development, release, and timing of any features or functionality described for Oracle's products remains at the sole discretion of Oracle.


ORACLE IS THE INFORMATION COMPANY