



Lustre Development Update

- Dan Ferber
Whamcloud, Inc.
dferber@whamcloud.com

Agenda

- Lustre current status
 - Community & Roadmap
 - Releases
 - Current development processes
- Looking forward
 - Chroma™
 - Exascale



Whamcloud Today

- ~55 people worldwide
 - Unique advantage:
 - Critical mass for Lustre technology
 - Time to delivery or resolution (engineers, project managers)
- ~175 supported sites worldwide
- Our offerings:
 - Worldwide Lustre support
 - Lustre development
 - Training
 - Community releases (OpenSFS)
 - Chroma



Whamcloud is widely recognized as the source for Lustre
We have the only HW vendor-neutral offering

Broad Community Success

Lustre is stable today

- Technically: v1.8.x is very solid, 2.x features
- Politically: the community has stepped up
 - The tree is safe, stable, and reliable
 - Development expected to be done in the open for review

The Ecosystem is growing:

- EOFS + OpenSFS + Whamcloud
- Single tree from which to pull
- Active dev community: Whamcloud + others
- More storage vendors shipping Lustre today

Lustre Community - Resources

- Whamcloud community membership
<http://www.opensfs.org/release-planning-group/technical-working-group>



- Whamcloud maintains the community assets
 - Wiki + roadmap: <http://wiki.whamcloud.com>
 - All Lustre releases: <http://www.whamcloud.com/downloads>
 - Jira bug tracker: <http://bugs.{OpenSFS.org,whamcloud.com}>
 - Git repositories: <http://git.whamcloud.com>
 - Gerrit code review: <http://review.{OpenSFS.org,whamcloud.com}>
 - Build: <http://build.whamcloud.com>
- No copyright assignment on source contributions
 - Ensures no single entity can own whole copyright on Lustre
 - Has support of OpenSFS and EOFS

LUG 2012 in Austin, Texas

<http://insidehpc.com/category/events/lug-2012/>



Whamcloud Community Lustre Releases

- Single community-wide source tree
 - Hosted at Whamcloud, tested and available via RPM
 - Formally recognized by community
- Whamcloud defines two release streams
 - The designated maintenance release stream is targeted for conservative users wishing to use a well-proven release
 - The feature release stream is targeted for those requiring first access to new features
- Bugfix releases every quarter for maintenance release stream
 - Currently 2.1.x is the designated maintenance release stream
 - Use combination of support metrics and customer demand to determine when to switch
- Feature releases every six months
 - Not all feature releases will have maintenance releases

Lustre 1.8.x

- Whamcloud continues to support Lustre 1.8.x
 - We will offer Lustre 1.8.x support as long as there is sufficient demand
- Whamcloud to release 1.8.8-wc1
 - A supplemental release to provide RHEL6.2 support

Lustre 2.1.x

- Lustre 2.1.0 released Q3 2011
- Lustre 2.1.1 released in Q1
 - Provided bugfixes and RHEL6.2 server and client support
- Lustre 2.1.2 scheduled for release in Q2
 - Will incorporate bugfixes from large sites running 2.1.x in production
- Intention is to match Lustre 1.8.x for stability
 - Still early days in this process

Lustre 2.2

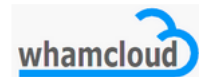
- Released March 30th as per the schedule
- Changelog is available at:
<http://wiki.whamcloud.com/display/PUB/Changelog+2.2>
- *First release to benefit from OpenSFS funding*
 - Covers some of the costs of maintaining the community tree
 - The remaining costs are covered by Whamcloud support contracts
- Regular updates throughout release cycle
 - JIRA filters provide dynamic list of blockers
 - Open access to issues (JIRA); patches (git/gerrit) and test results(maloo)
 - Biweekly email update to wc-discuss and CDWG mailing lists
 - Quarterly reports posted on OpenSFS CDWG wiki

Lustre 2.2 Changelog

Dashboard > Whamcloud Community Space > Wiki Front Page > Changelog 2.2

Browse ▾ Dan Ferber ▾ Search

-
- ☐ Changelog 1.8
 - ☐ Changelog 2.1
 - ☒ **Changelog 2.2**
 - ☒ Community Job Board
 - ☐ Community Lustre Roadmap
 - ☐ Community Resources and Mail Lists
 - ☐ Documentation
 - ☒ Getting started with Lustre
 - ☐ Lustre 2.2
 - ☐ Lustre Community Development in Progress
 - ☒ Lustre Development
 - ☒ Lustre Presentations and Meetings
 - ☐ Lustre Releases
 - ☐ Lustre Support Matrix
 - ☒ Lustre Tools
 - ☒ Lustre Training and Community News
 - ☐ Partner News -- Content Store
 - ☐ Why Use Lustre



Changelog 2.2

Added by Peter Jones, last edited by Peter Jones on Mar 30, 2012 (view change)

 Edit  Add ▾  Tools ▾

version 2.2.0

Support for networks:

o2ibld - OFED 1.5.4

Server support for kernels:

2.6.32-220.4.2.el6 (RHEL6)

Client support for unpatched kernels:

2.6.18-274.18.1.el5 (RHEL5)
2.6.32-220.4.2.el6 (RHEL6)
2.6.32.36-0.5 (SLES11)

Recommended e2fsprogs version:

1.41.90.wc4

Known Issues in 2.2.0:

LU-1188 Panics can occur running sanity test while running with kernel debug options turned on
LU-1191 Due to a known NFS-related bug in the kernel - https://bugzilla.kernel.org/show_bug.cgi?id=38572 - users wanting to re-export Lustre via NFS should apply the kernel patch for this
Powered by a free **Atlassian Confluence Open Source Project License** granted to Lustre. Evaluate Confluence today.

Powered by Atlassian Confluence 3.3, the Enterprise Wiki | Report a bug | Atlassian News

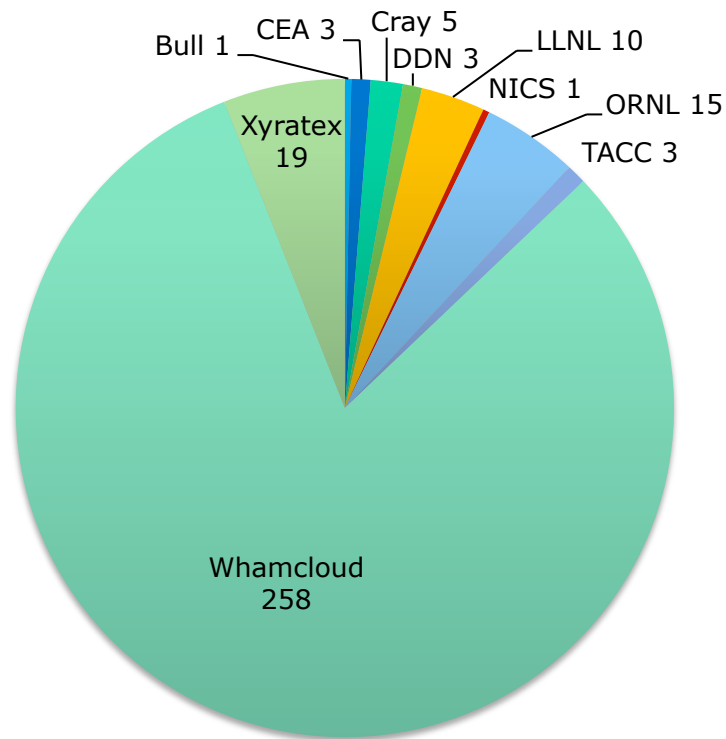
Lustre 2.2 – Accomplishments

- Amount of test automation greatly increased
 - Reduces the overhead of making releases and increases the likelihood of catching regression earlier in the release cycle
- Test reports facility added
 - <https://maloo.whamcloud.com/reports>
 - Highlights areas where tests need improving
- Testing resources at IU and FZJ utilized during test cycle
- System wide Hyperion testing conducted
 - 1100+ clients
- Community Development wiki established
 - <http://wiki.whamcloud.com/display/PUB/Lustre+Community+Development+in+Progress>
 - Aim is to foster greater collaboration and prevent duplication of effort with development across organizations

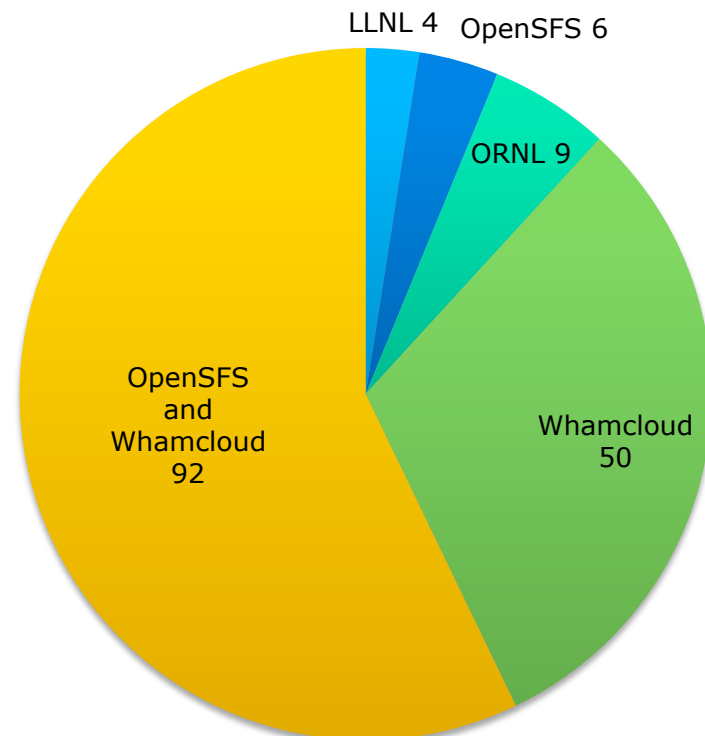
Lustre 2.2 – Features

- **Asynchronous Glimpse Lock/Statahead (LU925/LU389)**
 - Improved performance for ls -l/find and accessing object attributes (file sizes/ctime etc)
 - Development funded by ORNL
- **Client Parallel Checksums (LU884)**
 - Improved support for mmap and better performance using checksums
 - Development funded by ORNL
- **Imperative Recovery (LU580)**
 - Faster recovery
 - Development funded by ORNL
- **Large Xattrs (aka Wide Striping) (LU80)**
 - 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
- **Mds-survey (LU593/LU633)**
 - Tool for MDS performance benchmarking
 - Development funded by LLNL
- **Parallel Directory Operations (LU50)**
 - Improved performance when multiple processes access the same directory in parallel
 - Development funded by OpenSFS

Lustre Community Work



Lustre 2.2 Landings
by Company



Lustre Landings
by Contract

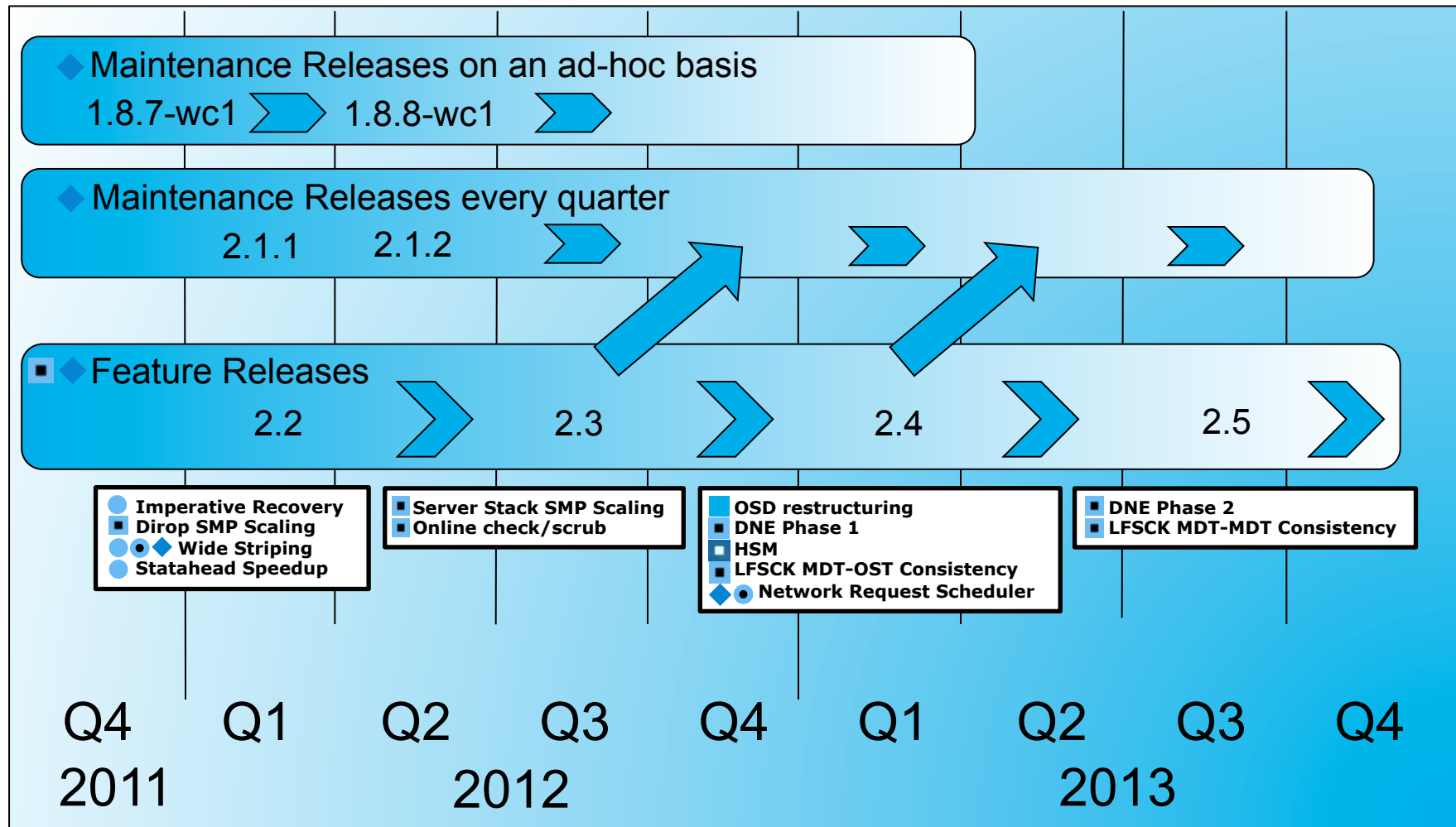
Lustre 2.3 and Beyond

- Lustre 2.3 (September 2012)
 - Server Stack SMP Scaling (LU-56)
 - LFSCCK Online OSD check/ OI scrub (Ph.1, LU-957)
 - OSD restructuring (ZFS OST capability, LU-1305)
- Lustre 2.4 (March 2013)
 - OSD Restructuring (ZFS on MDTs, LU-1305)
 - Distributed Namespace (Ph.1 Remote Directories, LU-1187)
 - HSM (CEA implementation, LU-941,169,827,1338,1333)
 - LFSCCK Online check/scrub – Distributed Repair (Ph.2, LU-957)
- Lustre 2.5 (September 2013)
 - LFSCCK MDT-MDT Consistency (Ph. 3, LU-957)
 - Distributed Namespace (Ph.2 Distributed Directories, LU-1187)

Other Areas of Interest

- Networking
 - LNET Dynamic Configurations,
 - Channel Bonding,
 - Health Networks,
 - IPv6
- Storage Management
 - Tiered Storage: Policy-driven object storage placement
 - Migration: OST rebalancing, Async mirroring
 - Small File Performance: Unified Targets
- Other
 - Administrative Shutdown
 - Test frameworks
 - JobStats
 - Hadoop Integration

Community Lustre Roadmap



Sponsor for Whamcloud Development and Releases: ● ORNL ■ OpenSFS ■ LLNL ◆ Whamcloud

Third Party Development: ■ CEA ● Xyratex

Lustre Feature Lookup

- Go to:
<http://wiki.whamcloud.com/display/PUB/Lustre+Community+Development+in+Progress>
- Find your feature of interest
- Find the feature on the roadmap (schedule)
- Look at the bug in Jira (details on feature)

The Tools We Use Today

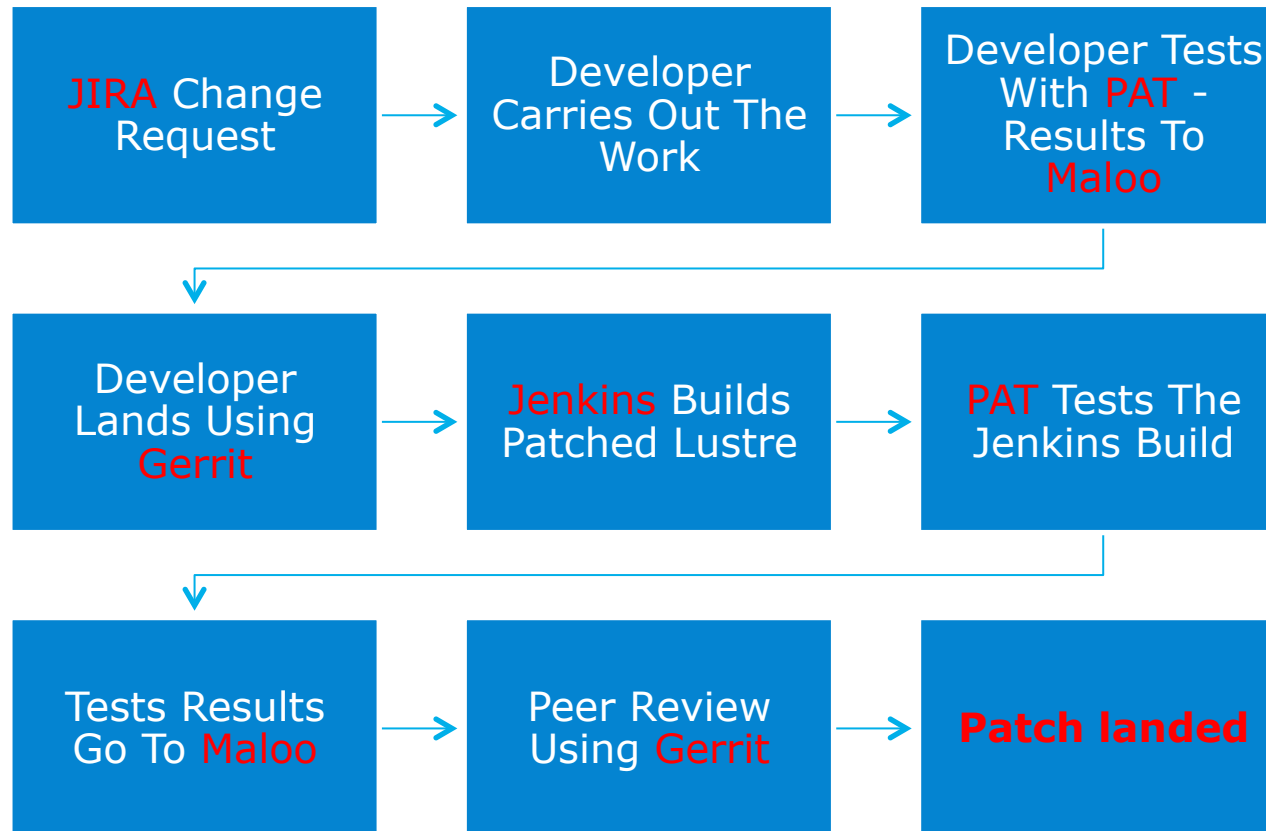
Jira, Jenkins
Git and Gerrit

Tools Live Today

jira.whamcloud.com
build.whamcloud.com
review.whamcloud.com

- **JIRA** is Whamcloud's Issue and Agile management tool
- **JENKINS** is the build tool that continuously builds mainstream branches and all patches submitted by the community
- **GIT** is source code tool used for managing the Lustre canonical tree
- **GERRIT** is code review tool that allows the whole community to be part of the code review process

Work Flow



<http://wiki.whamcloud.com/display/PUB/Submitting+Changes>

Gerrit


- When one developer writes code, another developer is asked to review that code
- A careful line-by-line critique
- Happens in a non-threatening context
- Goal is cooperation, not fault-finding
- An integral part of the Lustre coding process

Maloo

- Maloo is the authoritative test results database
 - Autotest **and** Developer results are stored in Maloo
- Testing results from development
 - Results from development provide landing collateral
 - Failures are as important as passes
 - Good to see the transition from failure to pass
- Landing requires passing results in Maloo
 - Maloo / Jenkins / Gerrit work in unison to ensure Reviews, Build and Test have all occurred.

Maloo Screen Shots

latest results | latest sessions | search results | test statistics | upload results

Chris Gearing [settings | logout] 

Test sessions

Sessions for user:

All users

Go

1 2 3 4 5 6 7 8 9 ... 28 29 Next →

Host	Group	User	Run at	Imported at	Sets passed	Links
client-23-ib	review	Whamcloud Autotest	2011-05-20 09:58:44 UTC	2011-05-20 13:52:22 UTC	15/16	gerrit:12f8dcc47
client-20-ib	review	Whamcloud Autotest	2011-05-20 09:17:00 UTC	2011-05-20 09:17:00 UTC	15/16	gerrit:c45e7eacc
client-23-ib	review	Whamcloud Autotest	2011-05-20 03:02:25 UTC	2011-05-20 09:32:47 UTC	16/17	gerrit:33340bf6c
client-23-ib	review	Whamcloud Autotest	2011-05-20 00:43:45 UTC	2011-05-20 02:40:59 UTC	1/2	gerrit:4f1aa57ed
client-20-ib	review	Whamcloud Autotest	2011-05-19 22:40:48 UTC	2011-05-20 05:03:31 UTC	16/17	gerrit:2c57a6a60
zwicky1	acc-sm-zwicky1	Prakash Surya	2011-05-19 20:25:08 UTC	2011-05-19 20:45:41 UTC	1/1	
client-8-ib	development	Chris Gearing	2011-05-19 18:03:08 UTC	2011-05-19 21:45:23 UTC	8/10	
client-23-ib	regression	Whamcloud Autotest	2011-05-19 17:05:32 UTC	2011-05-20 00:20:51 UTC	21/21	
client-20-ib	review	Whamcloud Autotest	2011-05-19 16:38:28 UTC	2011-05-19 22:21:54 UTC	17/17	gerrit:1b4f9f99f
client-23-ib	review	Whamcloud Autotest	2011-05-19 09:51:18 UTC	2011-05-19 15:25:47 UTC	17/17	gerrit:f4cded4b3
client-20-ib	regression	Whamcloud Autotest	2011-05-19 06:28:56 UTC	2011-05-19 14:57:04 UTC	20/21	

This is a link to the test suite detail

Looking Forward

- Chroma™
- Exascale



Chroma™

Chroma Terminology

<http://insidehpc.com/category/events/lug-2012/>

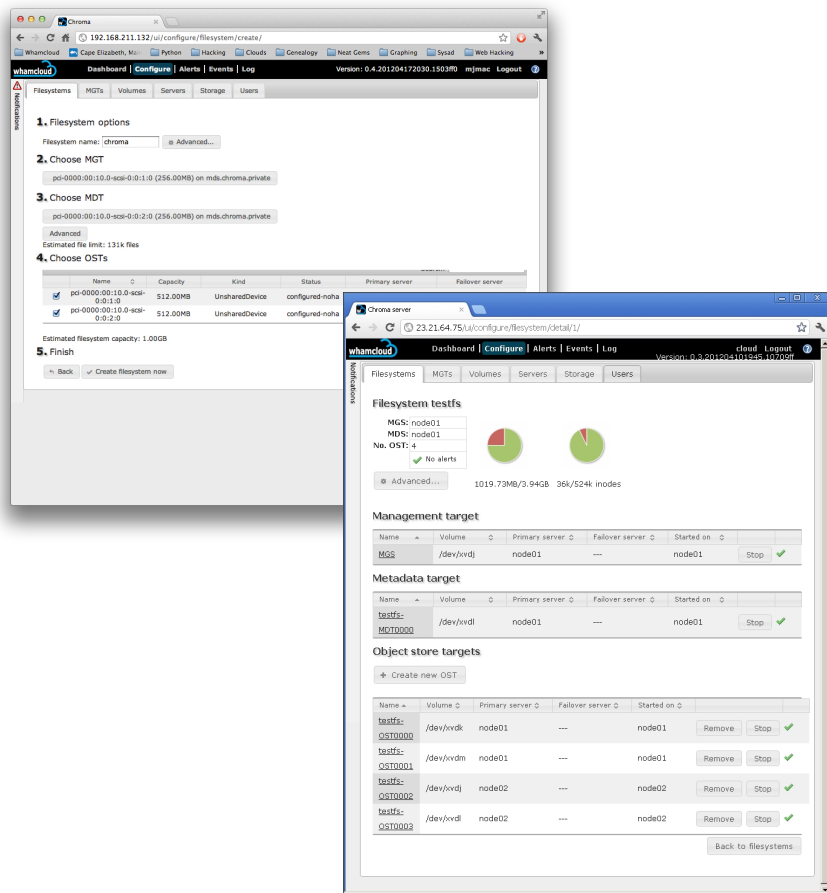
- Chroma Manager
 - The User Interface for FS Management
- Chroma Storage
 - Software appliance (includes Lustre)
 - Installed on bare-metal servers
 - Attached to Chroma-supported storage
- Chroma Enterprise
 - Complete Scalable Storage Solution
 - Consists of Chroma Manager and Chroma Storage



Drive Chroma the Way You Want to

- Management console
 - Simplify provisioning, monitoring and maintenance of Lustre storage
 - Monitor and analyze filesystem performance
 - Deep integration with storage vendor partner products
 - Appliance toolkit
- Scriptable user interface
 - Command line
 - Web service
- REST API for Lustre management
 - Architecture for using standard web protocols
 - Separate front-end and server logic
 - Used internally by GUI and CLI
 - Enables deep integration with 3rd party site/cluster management products

Chroma Manager



The image shows two overlapping screenshots of the Chroma Manager web interface. The top screenshot displays the 'Filesystem options' configuration page, which includes sections for '1. Filesystem options', '2. Choose MGT', '3. Choose MDT', '4. Choose OSTs', and '5. Finish'. The bottom screenshot shows the 'Chroma server' dashboard, which includes a 'Filesystem testfs' section with pie charts for MGS and MDS, a 'Management target' table, a 'Metadata target' table, and an 'Object store targets' table.

Name	Volume	Primary server	Falover server	Started on	Stop
MGS	/dev/vxvdj	node01	---	node01	Stop

Name	Volume	Primary server	Falover server	Started on	Stop
testfs	/dev/vxvdj	node01	---	node01	Stop

Name	Volume	Primary server	Falover server	Started on	Remove	Stop
testfs	/dev/vxvdj	node01	---	node01	Remove	Stop
OST0000	/dev/vxvdj	node01	---	node01	Remove	Stop
testfs	/dev/vxvdj	node02	---	node02	Remove	Stop
OST0002	/dev/vxvdj	node02	---	node02	Remove	Stop
testfs	/dev/vxvdj	node02	---	node02	Remove	Stop
OST0003	/dev/vxvdj	node02	---	node02	Remove	Stop



Chroma Enterprise 1.0

Available from DDN Now

- Chroma Manager
 - Provides graphical and command-line interfaces for filesystem management
 - Central repository of filesystem configuration, state, and measurements
- Chroma Storage
 - Software appliance (includes Lustre)
 - Installed on bare-metal servers
 - Attached to Chroma-supported storage
- Other partners working on Chroma
- Chroma is targeted for partners

Exascale Challenges

<http://insidehpc.com/category/events/lug-2012/>

Application data + metadata

- Explosive growth
 - Large, sophisticated models
 - Uncertainty Qualification
 - Billions – trillions of “Leaf” data objects
 - Complex analysis
- Filesystem namespace pollution
 - Keep filesystem namespace for storage management / administration
 - Separate namespace for application data + metadata
 - Distributed Application Object Storage (DAOS) containers
- Preserve model integrity in the face of all possible failures
 - Very large atomic, durable transactions
 - Integrity APIs at all levels of the I/O stack
- Search / query / analysis
 - Non-resident index maintenance & traversal / non-sequential data traversal
 - Move query processing to global storage
 - Same programming model as apps?

