Cray's Lustre Support Model and Roadmap

Cory Spitz

CUG

5/1/2012

Introduction

- Explain the Cray model for integrating Lustre into Cray
- Show how Cray works with the Lustre community
- Introduce an updated roadmap for our Lustre SW releases

Overview

- Lustre model
- OpenSFS & Cray
- Community Lustre development
- Cray Lustre development
- Lustre for External Services
- Direct Attached Lustre
- Sonexion
- Cray lustre-utils tools
- Cray Lustre roadmap summary

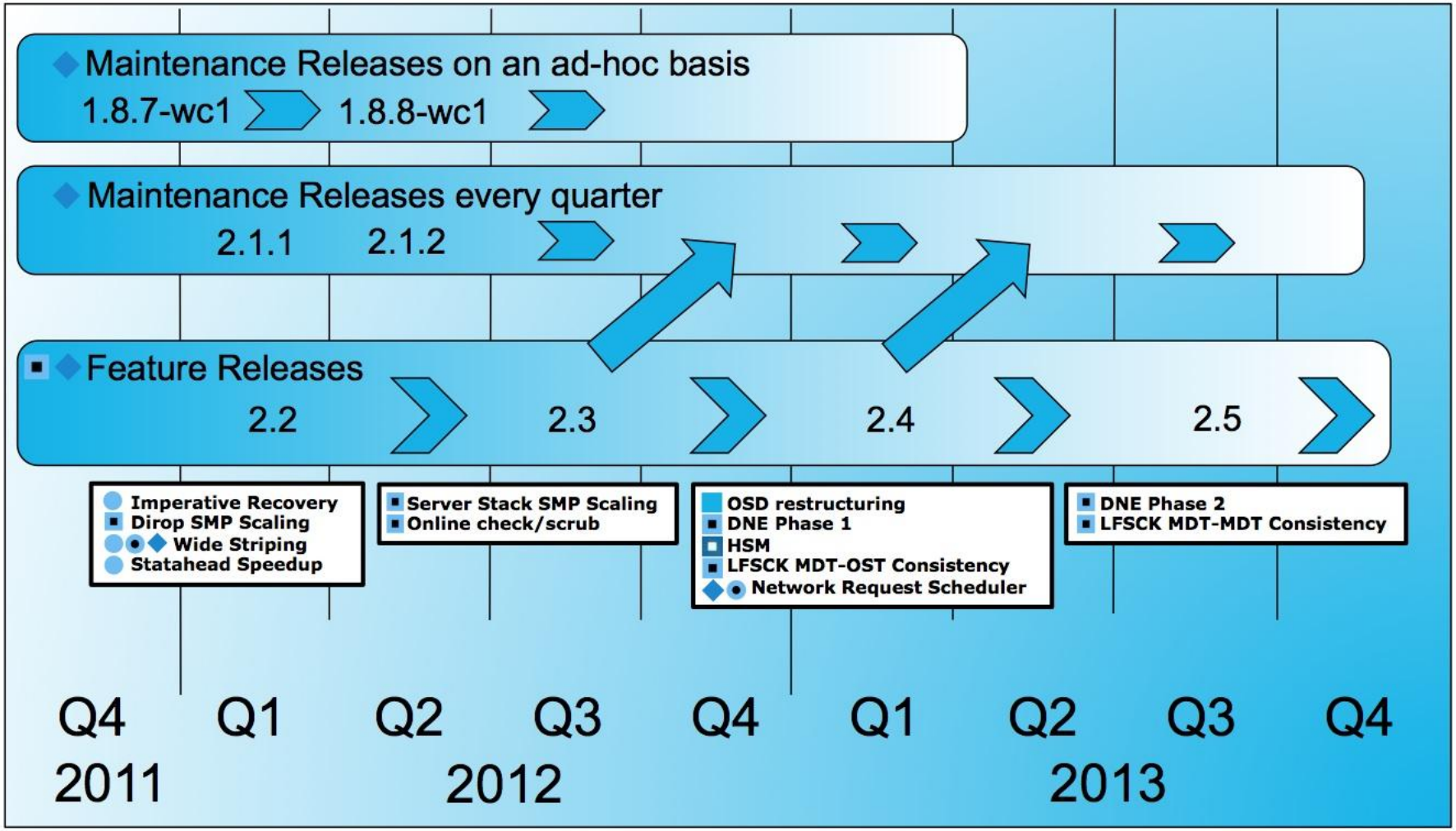
Our Lustre Model Ensures Stability & Productivity

- **Invest in Lustre through OpenSFS**
  - Cray is an OpenSFS founder and promoter with $500K annual dues
- **Acquire Lustre from canonical upstream source**
  - Cray no longer bases on Oracle
  - Regularly rebase on OpenSFS/Whamcloud canonical version
  - Patch CLE with fixes & enhancements
  - Push changes upstream for minimal deviation
  - Stabilize Cray Lustre version and release
- **Monolithic releases**
  - Clients, servers, and routers tested together
  - One Lustre base for internal direct-attached, esLogin, esFS, compute
- **Cray partners with Xyratex for level III Lustre support**
  - Covers direct-attached, esFS, Sonexion, and compute clients
  - Cray reports issues privately to Xyratex
  - Cray & Xyratex work with community on patches
  - Cray doesn't close tickets until patch landed 'upstream'

OpenSFS & Cray – Promoting Lustre

- **Cray is an OpenSFS 'promoter' with $500K annual dues**
  - David Wallace holds a seat (and vote) on the board
- **Cray has taken a leadership role within OpenSFS**
  - John Carrier co-chairs the Technical Working Group (TWG)
  - John is also involved with the Benchmarking Working Group (BWG)
  - Cory Spitz is involved with the Community Development Working Group (CDWG) and the TWG
  - John and Cory are TWG Project Approval Committee (PAC) members
- **OpenSFS funds development of features Cray customers desire**
  - 2011 – MDS single server enhancements, MDS scale out, OI-Scrub
  - 2012 – TBD, Requirements by 5/24, RFP @ ISC '12, SOW by SC '12
    - Current Lustre Requirements: http://goo.gl/63u9Q
- **Please join OpenSFS**
  - Your participation is needed to offset Lustre costs
  - Currently OpenSFS covers roughly half of 'tree maintenance'

Community Lustre Roadmap



Sponsor for Whamcloud Development and Releases: ORNL OpenSFS LLNL Whamcloud

Third Party Development: CEA Xyratex

Cray Lustre Development In a Year

- **Completed development**
  - CLE 4.0 UP00, UP01, and UP02 w/Lustre 1.8.4 released
    - Based on Oracle 1.8.4 & SLES 11 SP1 kernel support from 1.8.5
  - CLE 4.0 UP03 w/Lustre 1.8.6 released
    - Based on Whamcloud 1.8.6-wc1
- **Active development**
  - Cray has added Linux 3.0 support for Lustre clients
    - 'patchless' client for SLES 11 SP2
    - LU-812 (http://jira.whamcloud.com/browse/LU-812)
  - LNET best practices
    - ORNL: "I/O Congestion Avoidance via Routing and Object Placement"
    - Fine Grained Routing (FGR) and tuning
  - Cray lustre-utils tools
  - Lustre for new CLE releases, Sonexion, and External Services products

Productization of Lustre for External Services

- **esFS is 3rd party hardware, Cray software**
  - External Lustre servers
  - Connects to Cray mainframe clients via LNET routers
  - DDN or NetApp storage hardware
- **ESF is the codename for the new esFS SW release**
- **ESL is the codename for the new esLogin SW release**
- **ESL is tied to specific CLE release**
  - Same SLES release as CLE
- **ESF will be paired and tested with specific CLE releases**
  - Uses CentOS
- **ESL & ESF use the same Lustre stack as CLE**
  - One common source tree for three products

ESL & ESF Software Coming Soon

- **Includes esFSmon for esFS failover**
  - Requires esMS Management Server
- **lustre_control for command and control**
  - Same lustre_control as CLE for familiarity and common code base
  - Requires esMS
- **Release roadmap**
  - ESL & ESF Koshi UP01
    - GA in December ’12 w/CLE Koshi UP01
    - Lustre 2.2
  - ESL & ESF Nile UP02
    - GA in March ’13 w/CLE Nile UP02
    - Lustre 2.2
  - ESL & ESF releases are supported for 18 months
- **Upgrade Migrations**
  - Cray will provide a migration plan for upgrades
  - ESF w/Lustre 2.2 must be installed before Lustre 2.x CLE clients

EOL Planning for Direct Attached Lustre

Direct Attached Lustre is the traditional Cray Lustre offering with Lustre servers on mainframe I/O service nodes

Release Roadmap

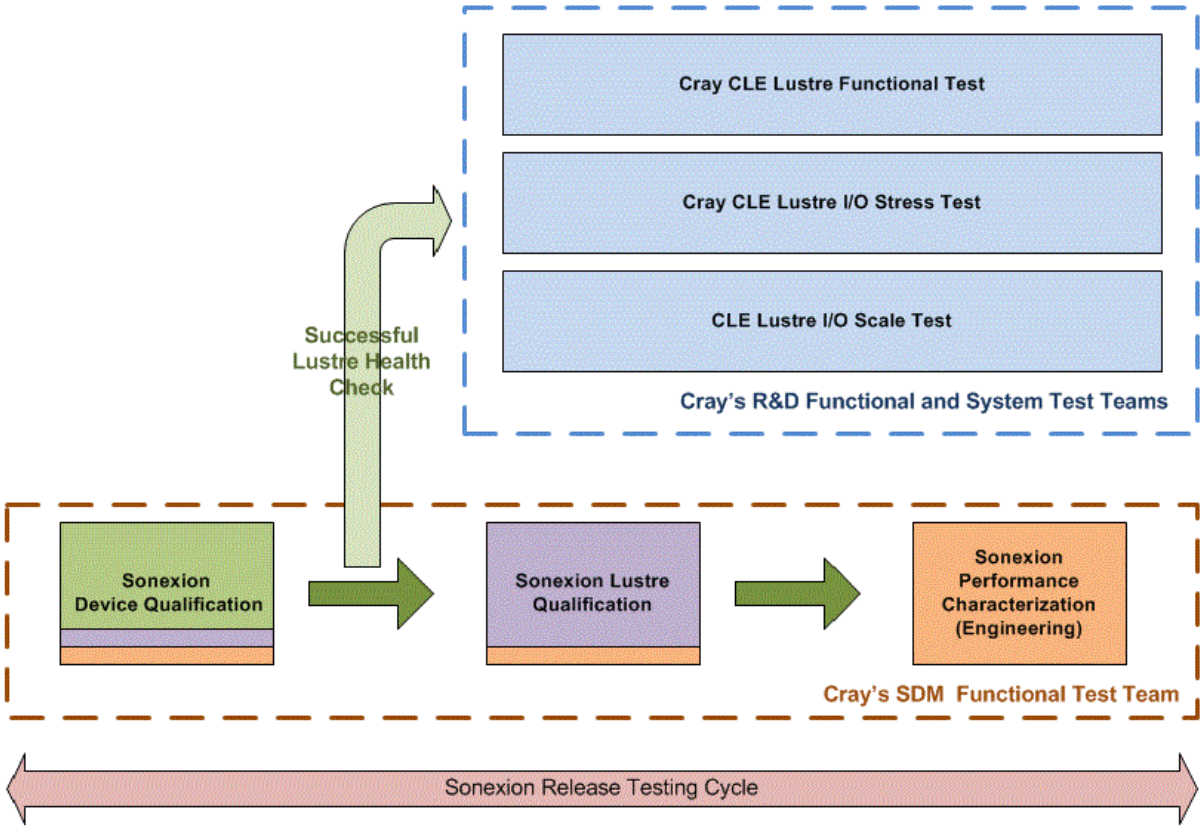
- CLE 4.0 UP03 w/Lustre 1.8.6
  - Patch support available through mid-2013
- CLE Koshi UP01 w/Lustre 1.8.x GA in December
  - Lustre version TBD, likely 1.8.7-wc1 or 1.8.8-wc1 based
  - Patch support available through mid-2014

Lustre 2.x is not available for Direct Attached Lustre

Direct Attached Lustre is not available beyond Koshi

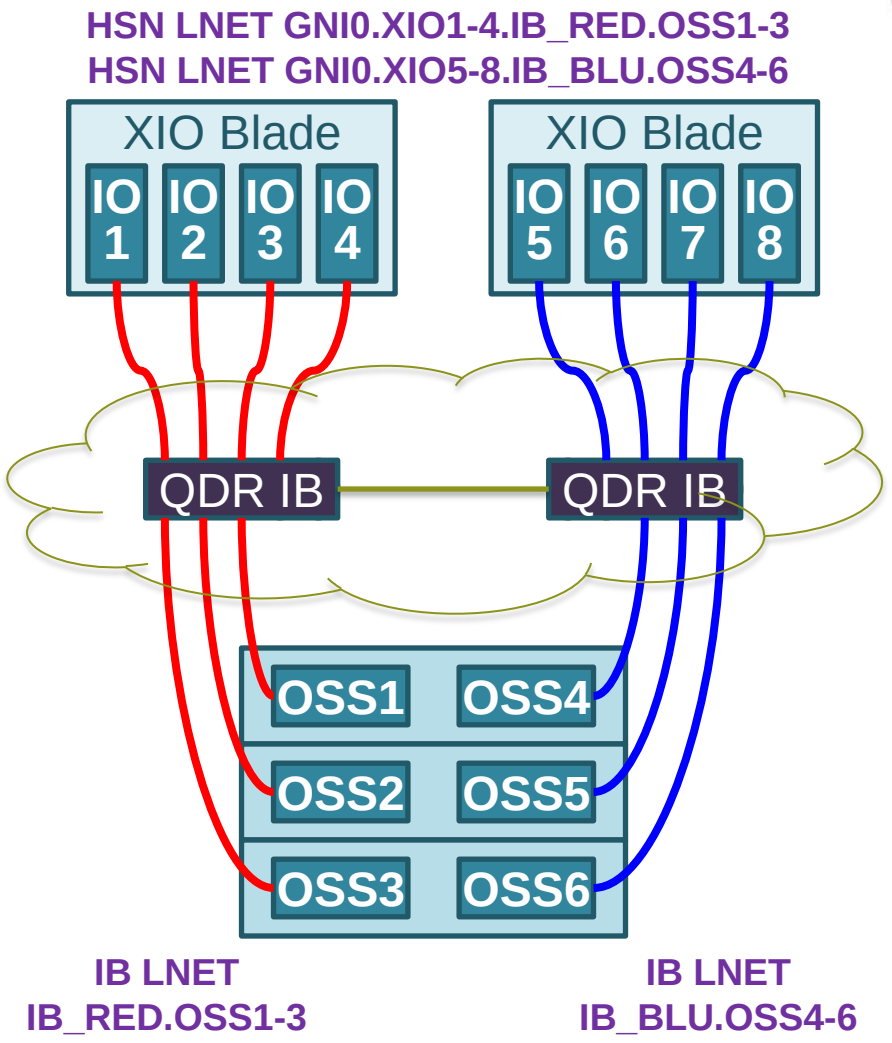
Cray Sonexion and SDM Focused Cray Testing

- Cray branded OEM solution from Xyratex
- Scalable storage units with integrated servers
- Lustre 2.x based servers
- Deployed with LNET FGR routing for extreme scale
- Paired and tested with specific CLE releases
- Close collaboration with Xyratex & Whamcloud on release updates and patches



What is LNET Fine Grained Routing?

- Lustre networking can create logical subnets within a fabric
- Define multiple LNETs to isolate I/O to specific physical paths through the fabric
- LNET Fine Grained Routing groups routers and OSSes together
- Eliminates congestion on both fabrics
- Reduces cost of IB fabric
- Easy to configure with clcvt
- Easy loading with LU-1071



Cray lustre-utils tools for ease of use

- **lustre_control**
  - Redesigned w/esFS support
  - Enhancements:
    - Improved file system definition and configuration
    - Operate on multiple file systems with a single command
      - Automatically updates the SDB if using Direct Attached Lustre w/failover
    - Control of mount/umount of service node clients and compute nodes
    - Failover and failback control including interface with esFSmon for esFS
    - Lustre server status reporting
    - Parallel Lustre target consistency checking with fsck
    - Configuration verification (verifies correct target for correct device)
    - Emplaces Lustre tuneables a la lctl set_param
- **clcvt – Cray LNET configuration and validation tool**
  - Sonexion LNET FGR only initially
  - Automatically generates 'ip2nets' and 'routes' for LNET configuration
  - Generates a cable map to aid install
  - Performs live validation of the cabling
  - Performs live validation of the LNET configuration

Cray Lustre Roadmap Summary

- **Koshi UP01 GA December ’12**
  - Includes new lustre_control and clcvt
  - CentOS 6.2 for ESF
  - SLES 11 SP1 for ESL and CLE
  - Lustre 1.8.x for Direct Attached Lustre in CLE
  - Lustre 2.2 client for ESL & CLE
  - Lustre 2.2 server for ESF
  - Patch support ends 18 months after GA – mid-2014

- **Nile UP02 GA March ’13**
  - CentOS 6.2 for ESF
  - SLES 11 SP2 for ESL and CLE
  - Lustre 2.2 CLE client for ESL & CLE
  - Lustre 2.2 server for ESF
  - Patch support ends 18 months after GA – late-2014

Questions?

Thank You