



Moving Lustre* Forward What We've Learned and What's Coming

Brent Gorda | General Manager | High Performance Data Division

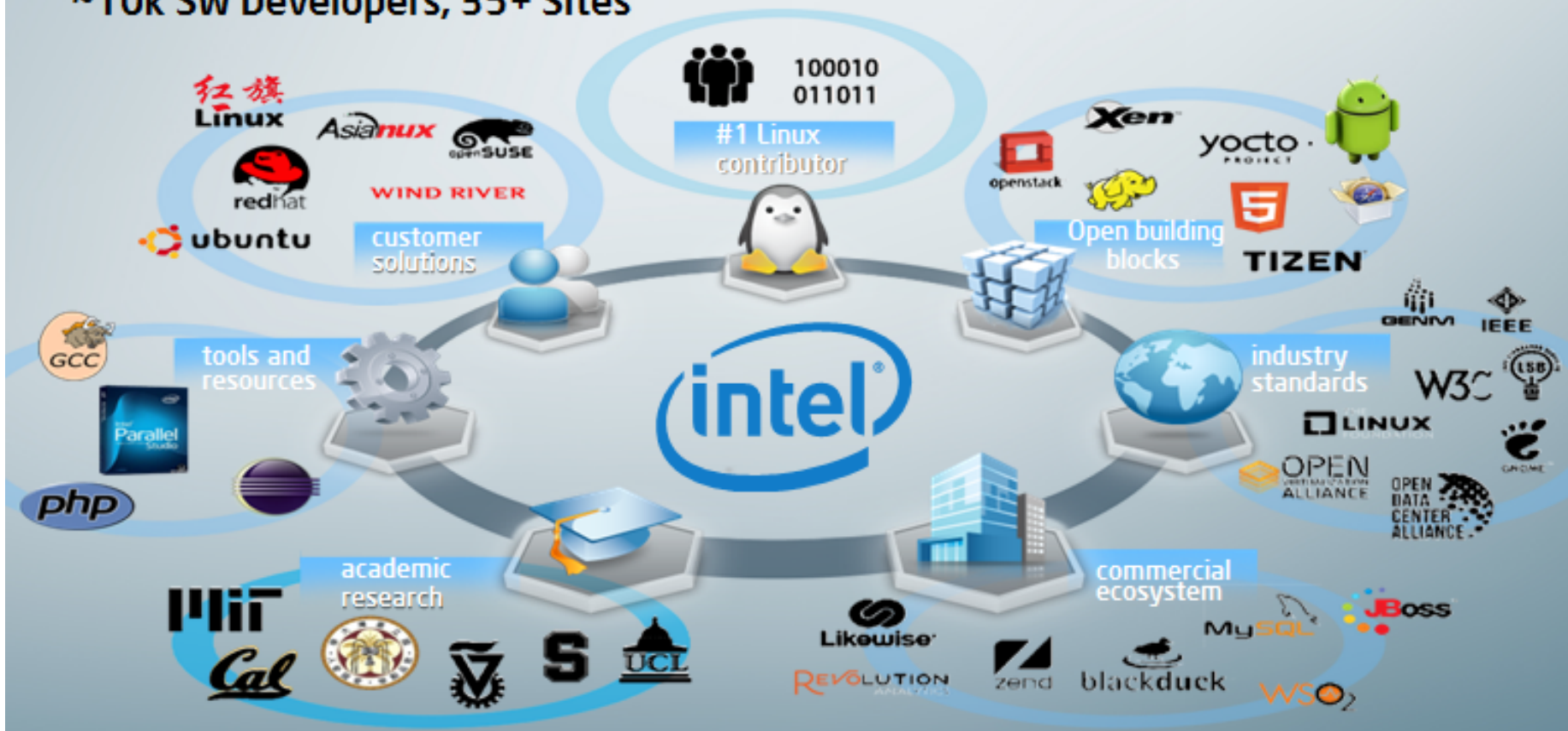


Intel is Firmly Committed to Open Source

Intel the Open Source Software Company



~10k SW Developers, 35+ Sites



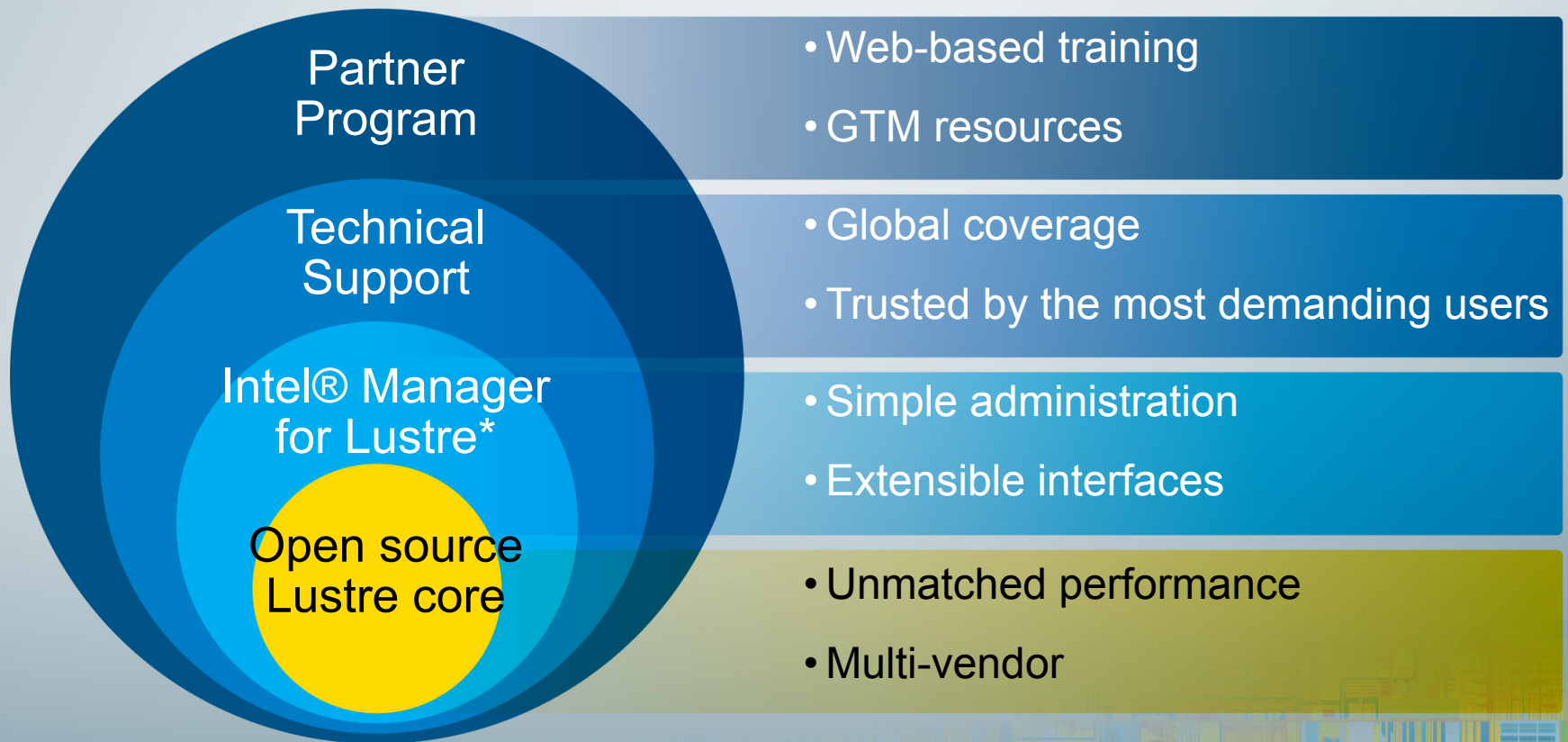
Our activity with Open Tree

- Open Source Tree Stewardship
 - Feature Releases in association with funding from OpenSFS
 - Maintenance Releases
 - Development with funding from OpenSFS and others
 - Gate Keeping in association with the tree contract from OpenSFS
- Hosting the Open Lustre* Assets
 - Code repository, bug database, documentation
- We are software focused: we do not sell Lustre storage

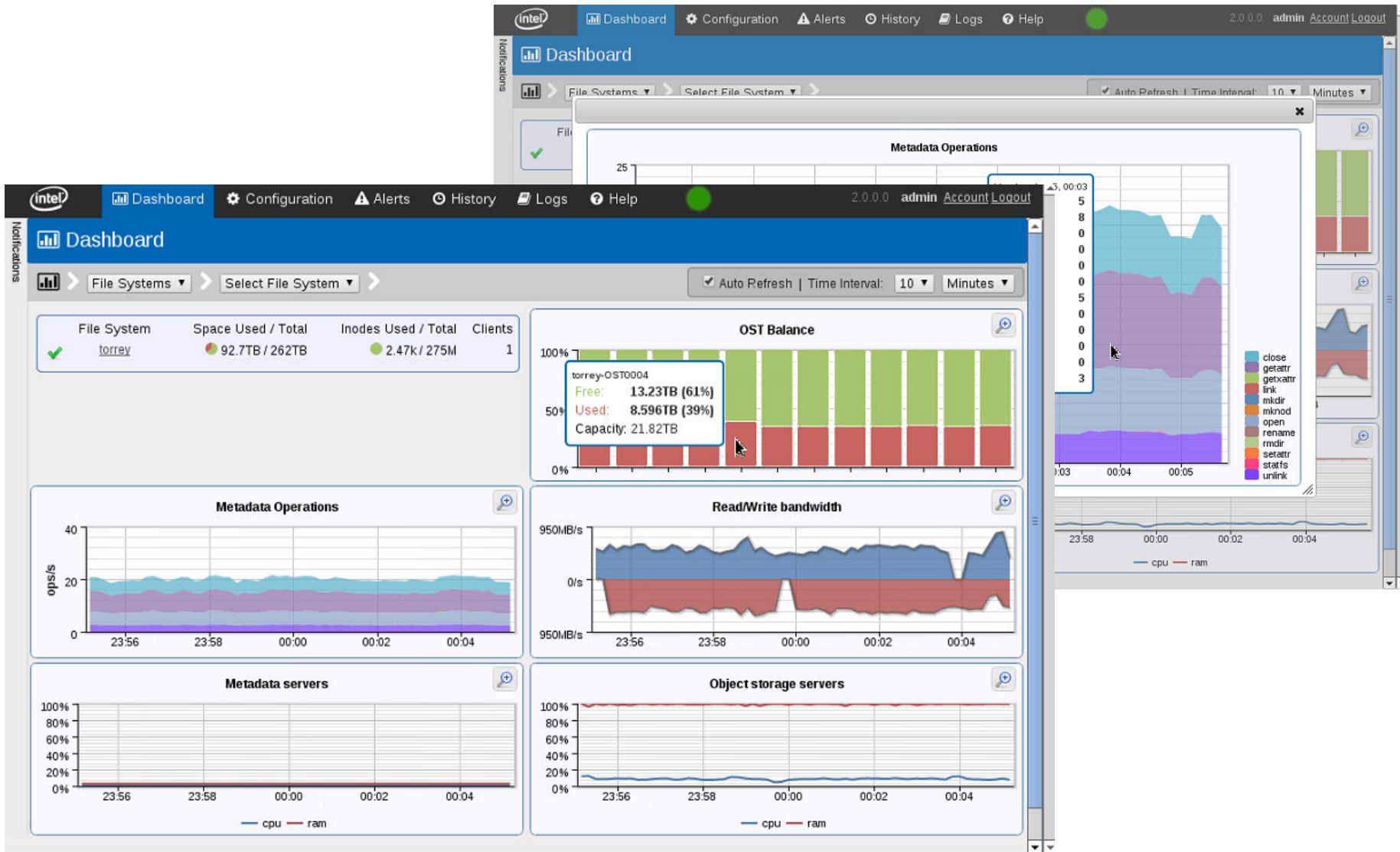
What we've learned in the past year

- Lustre* is safe, but needs a strong community to remain so
 - EOFS and OpenSFS are the core of the Open Source Community
 - Intel is a strong supporter and major contributor
- We learned the market wants more than just an open release
 - Many want the best support available for a technology they rely on
 - Customers asking for a branded offering, backed by Intel
- “Beyond HPC” opportunities have started to appear
 - Customers with “Big Data” problems gained sufficient confidence to talk
 - IDC has a new phrase for this: HPDA High Performance Data Analytics

Intel Enterprise Edition for Lustre*



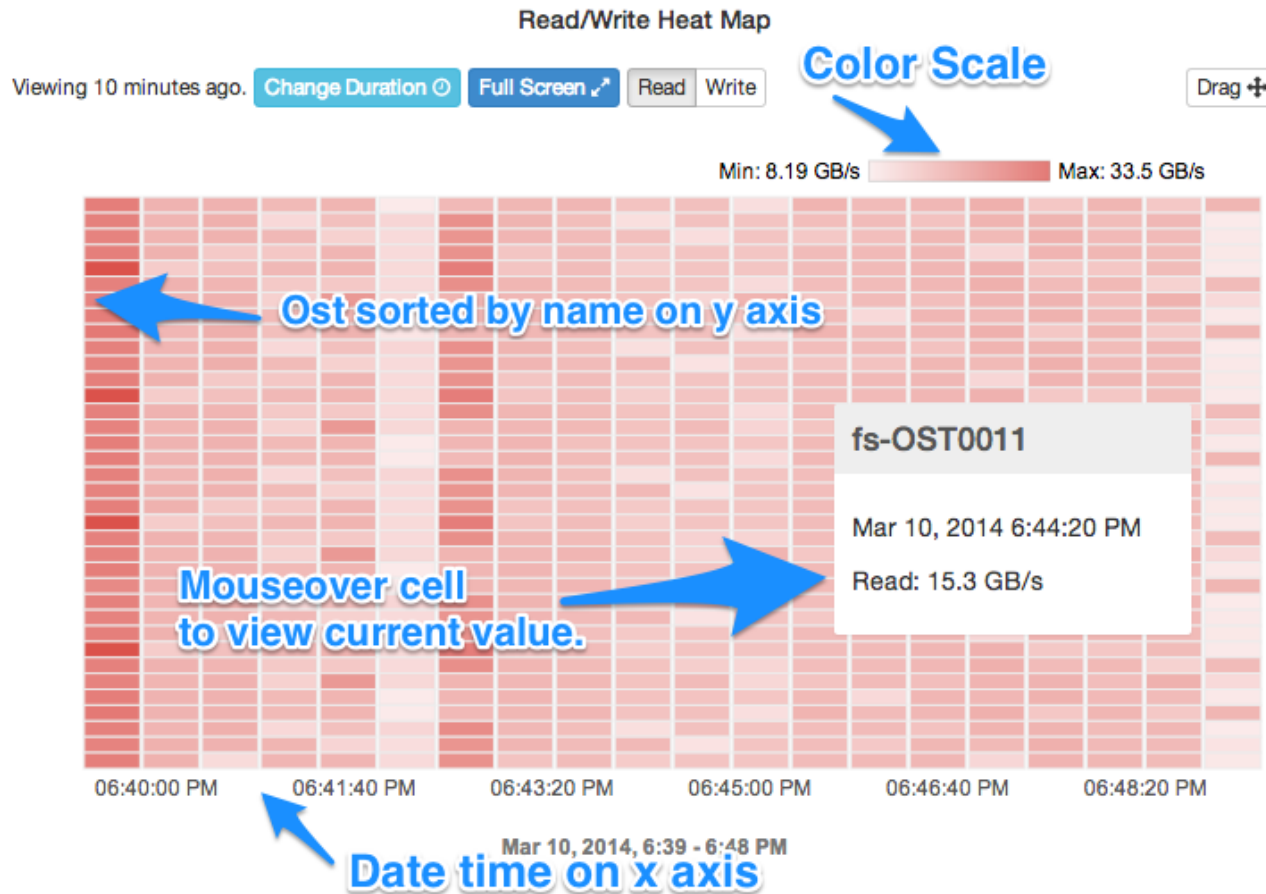
Intel Manager for Lustre*



* Some names and brands may be claimed as the property of others.



IML getting "heat" maps



High Performance Data Analysis (HPDA)

HPC workloads create and keep LOTS of data...

Hadoop uses local, direct-attached storage

But, HPC nodes are diskless

- PAIN POINT → storage efficiency and management complexity

Recent IDC research uncovered:

- ~67% of HPC sites are running Hadoop workloads on their HPC systems
- Hadoop workloads consume about 30% of their computing cycles

The image shows the cover page of an IDC report. At the top left is the IDC logo with the tagline 'Analyze the Future'. The title of the report is 'IDC HPC End-User Special Study of High-Performance Data Analysis (HPDA): Where Big Data Meets HPC'. Below the title, the authors are listed as Steve Conway and Earl C. Joseph, Ph.D., and Chirag Dekate, Ph.D. The report is categorized as a 'SPECIAL STUDY' and an 'IDC OPINION'. A vertical sidebar on the left contains the text 'MA 0101 USA P 2010021000 F 0010004015 WWW.IDC.COM'. The main body of text begins with 'This study is part of the third edition of IDC's end-user special study of the worldwide high-performance computing (HPC) market. IDC coined the term high-performance data analysis (HPDA) to refer to data-intensive ("Big Data") workloads that require or benefit greatly from HPC resources, even though not all HPDA beneficiaries consider themselves HPC users. These workloads include established data-intensive simulations and newer advanced analytics problems. The common denominator for HPDA problems is a degree of algorithmic complexity that is atypical for operational business workloads. Findings include:'. A bullet point follows: 'Especially during the rapid rise of clusters since 2002, HPC vendors have aggressively advanced the processor peak performance of their systems while paying less attention to data-intensive applications, HPC storage, and I/O capabilities. The areal densities of magnetic disks have increased dramatically, but improvements to disk I/O performance and access density have greatly lagged behind advances in disk capacity and processor speeds. And as HPC users have deployed ever-larger parallel servers, the fundamental imbalance — the gap between the server and storage sides of HPC — has grown worse. The road maps of HPC cluster vendors indicate that in relation to the continuing data explosion, tomorrow's HPC systems will be even more unbalanced ("flap sided") than today's. This imbalance heavily affects operations in the emerging HPDA market.'



+18% CAGR for HPDA storage, twice HPC storage growth

Lustre* + Hadoop: Open Platform for High Performance Data Analytics

Value Prop: Features, Functions, and Benefits



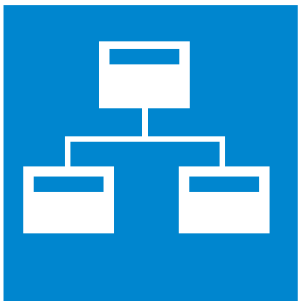
Performance

- **Bring compute to the data: Run MapReduce* on Lustre without code changes**
- **Run MapReduce faster: Avoid the intermediate file shuffle with shared storage**



Efficiency

- **Avoid Hadoop* islands in the sea of HPC systems**
- **Run MapReduce jobs alongside HPC workloads with full access to the cluster resources**



Manageability

- **Use the seamless integration to manage one common platform for Hadoop and HPC**
- **Develop with multiple programming models and deploy on shared storage**

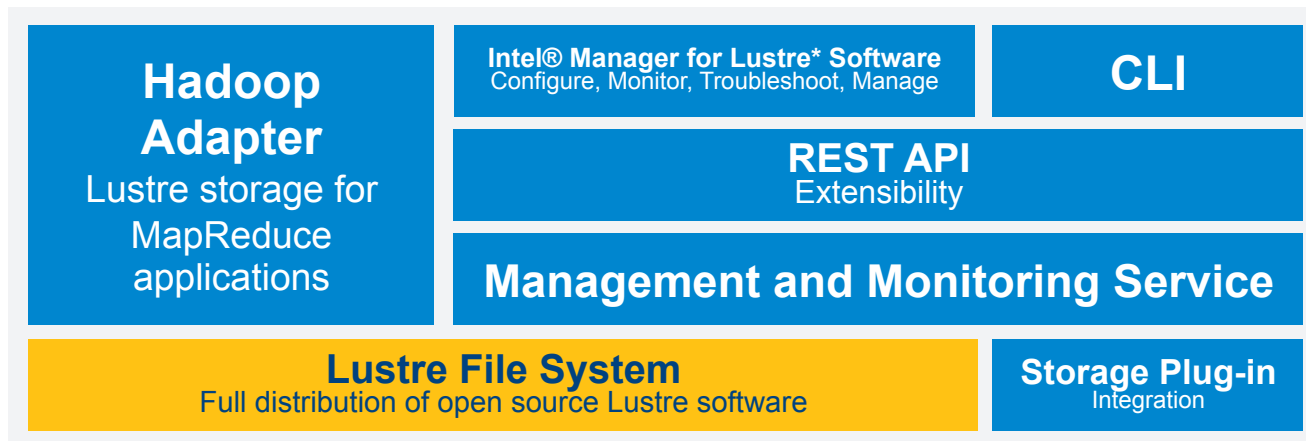
* Some names and brands may be claimed as the property of others

Solution: Intel® Enterprise Edition for Lustre* Software

Integration and support of Lustre* out of the box for Cloudera Hadoop

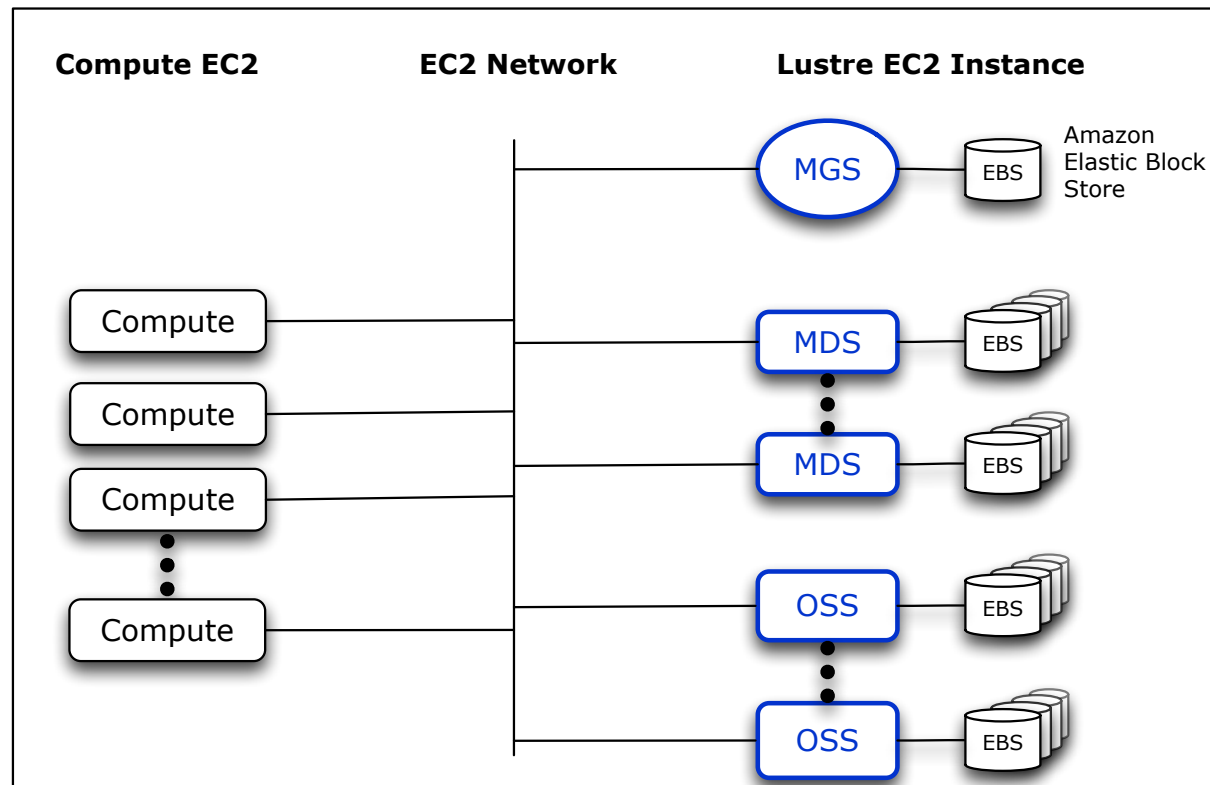
Intel® Enterprise Edition for Lustre* Software

- Full open source core
- Simple GUI for install and management with central data collection
- Direct integration with storage HW and applications
- Global support
- Storage plug-in; deep vendor integration
- REST API—extensibility
- Hadoop* Adapter for shared simplified storage for Hadoop



Intel Cloud Edition for Lustre* Software

- Available now in the AWS Marketplace
- More exciting news soon...



What to expect in the next year

- Continued significant contributions to open activities
 - Partnered with OpenSFS for feature releases
 - New features added as you'll learn about this week
 - A review and focus on development process and stability
- Intel Enterprise Edition of Lustre*
 - Focus on channel partner priorities, releases and support
 - Enabling technologies such as the Hadoop adaptor

