



High-Performance Computing Dell Intel EE Lustre Storage

Melbourne Big Data User Group
January 2016

Andrew Underwood
Lead HPC Technologist

The future of data is BIG, and HPC is needed to power it

The digital universe is growing 40%* a year into the next decade...

~90%

of the worlds data
has been created
in the last 2 years

44ZB

of data will exist
in the digital universe
by the year 2020

~37%

of the data generated
in 2020 will be used for
analysis and processing

2016

2020



The convergence of HPC and Big Data are driving change

1

Scalable performance and massive capacity

2

Stable, predictable and reliable

3

Balanced configuration, designed for parallel input-output

4

Support compute intensive, and data intensive 'Big Data' workloads with Hadoop

5

Enterprise grade technology with 24/7 access to data



Introducing our latest generation Dell EE Lustre Storage



11GB/s

Peak Read
per building block

7GB/s

Peak Write
per building block

Up to **4PB**
raw capacity
per rack

Up to **44GB/s**
throughput
per rack

Limitless

Endless Scalability

Parallel

For ultimate scale-out

Hadoop

Converged Platform

The **Ultimate** HPC File System



What is Lustre?

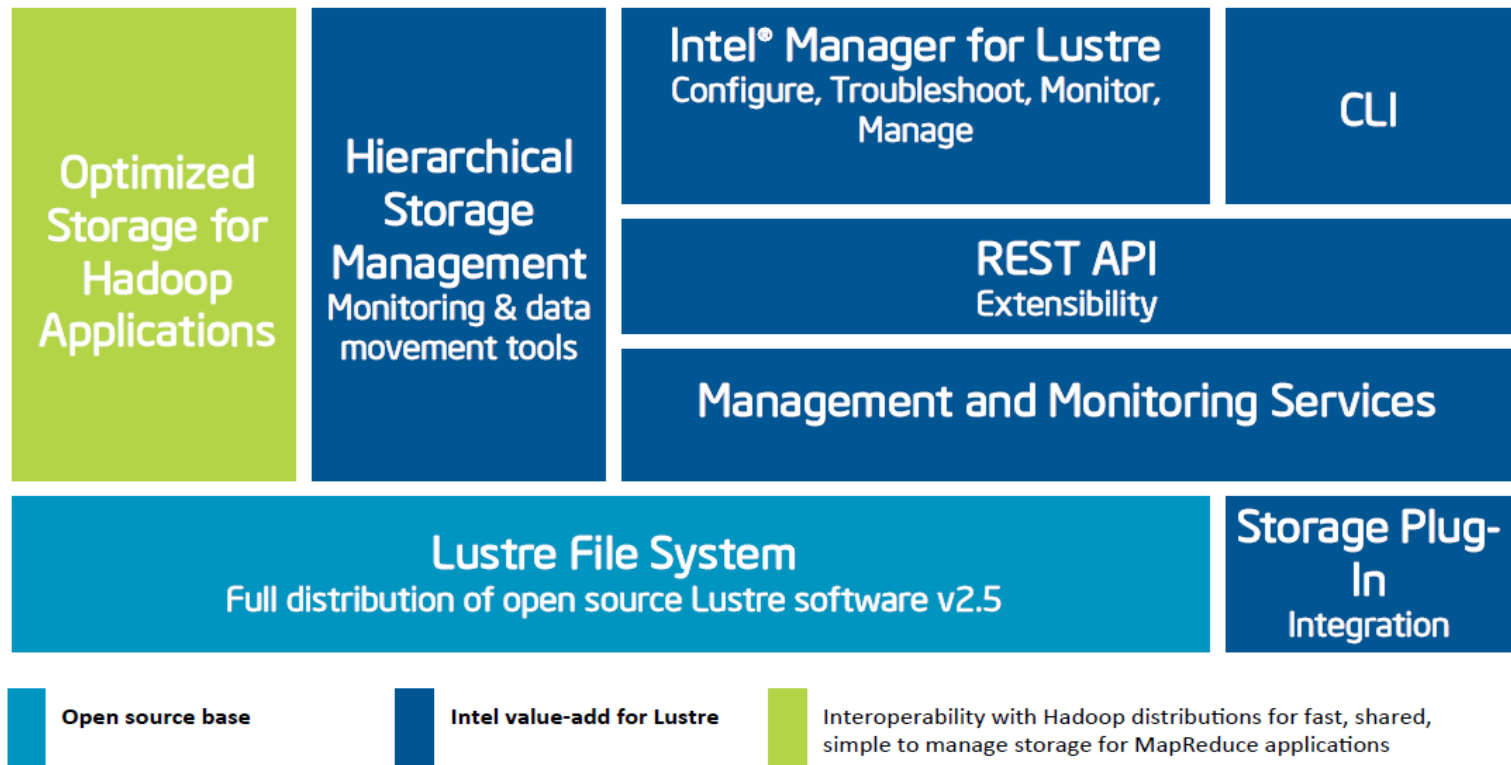


Designed for maximum performance and scalability....

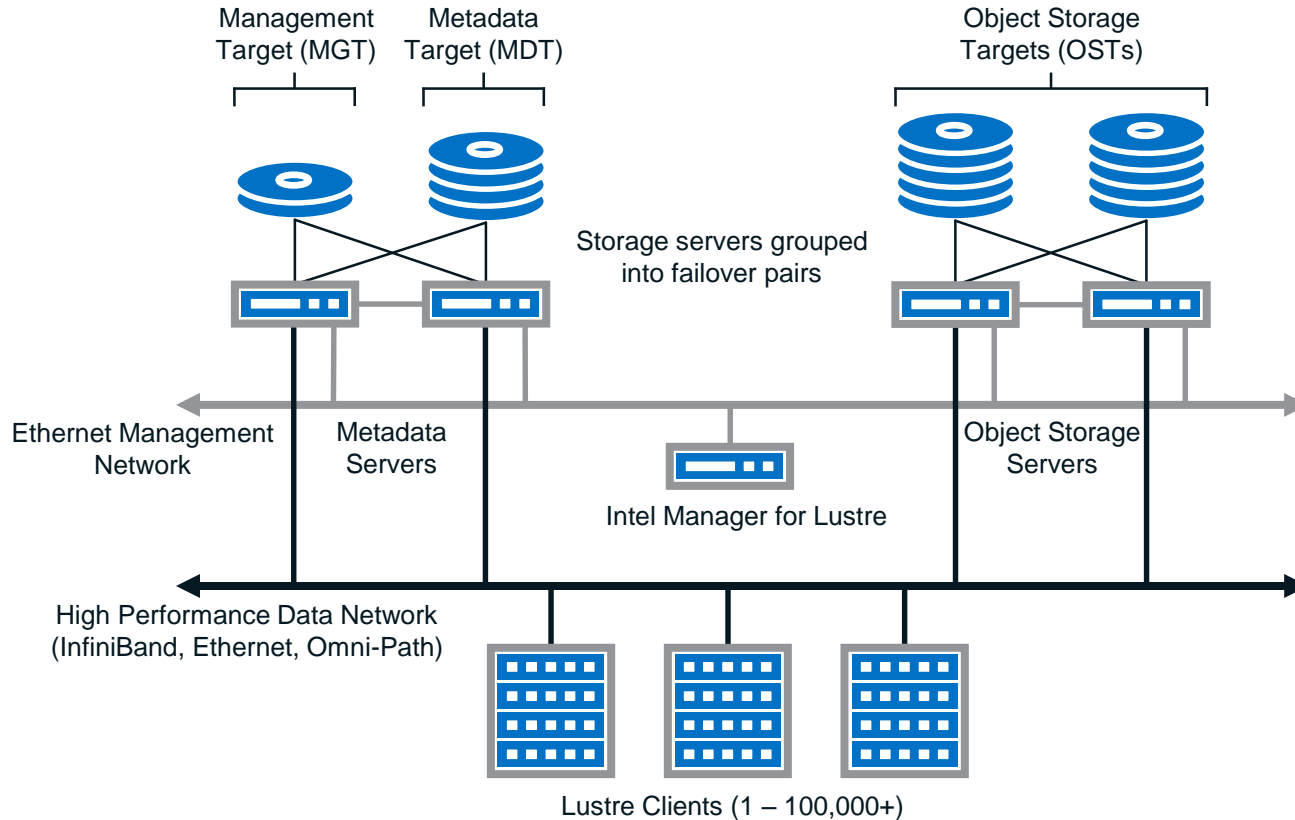
- Open-Source parallel file system built on open standards hardware
- Global, shared name space – accessible by over 25,000+ clients
- Object based file system with distributed file stripe across storage targets
- Highly available design, with no single point of failure
- Scalable beyond an Exabyte in capacity, and over a Petabyte of sustained throughput
- Accessible by clients over network (Ethernet, InfiniBand, Omni-Path Architecture)



The Intel Enterprise Edition of Lustre



Inside the Lustre File System

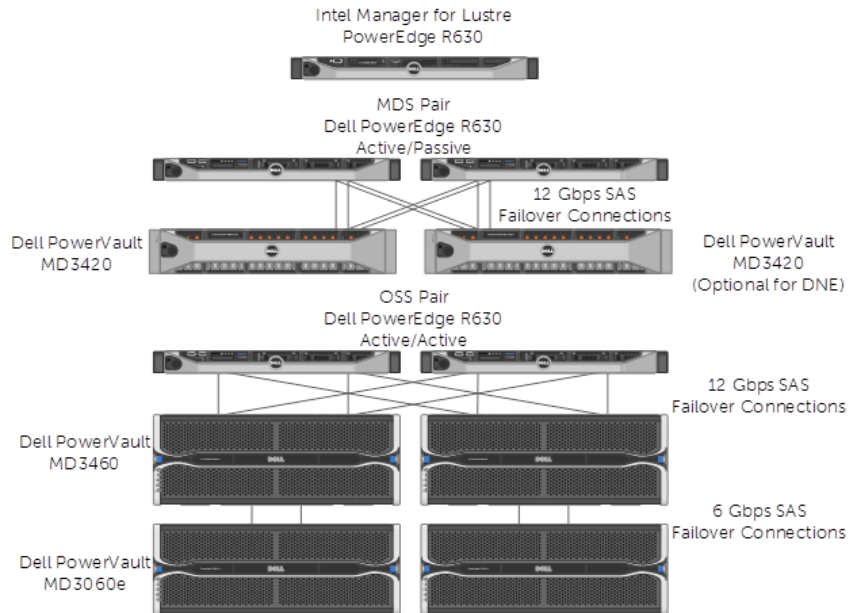


A scalable building block design

Designed to scale to the Exabyte, with a Petabyte of throughput

Solution benefits & Dell differentiation

- File system based on Intel Enterprise Edition for Lustre
- Single file system namespace scalable to high capacities and performance
- Engineered by Dell HPC Engineering to provide optimal performance on Dell hardware platform
- Design providing maximum throughput per building block with on-the-fly storage expansion
- Solution design for Big Data workloads using Intel Hadoop Adapter for Lustre (HAL)
- Share data with other file systems utilizing optional NFS/CIFS gateway
- Dell Networking 10/40GbE, InfiniBand, or Omni-Path



How Lustre Works

Basic File Storage Principles of the Lustre File System

The file system consists of many object storage targets (OST) which are presented as a single unified file system. IO can be increased in most cases by using file striping. File striping divides data into chunks that are distributed across OSTs within the file system.

Lustre Write File Data Flow

- 1) Client requests to write a file to the file system
- 2) Client contacts the MDS with a write request
- 3) MDS checks the user authentication and intended location of the file
- 4) MDS responds to the Client with a list of OSTs that the Client can write the file to
- 5) Client receives the response and writes to the assigned OSTs without further communication with the MDS

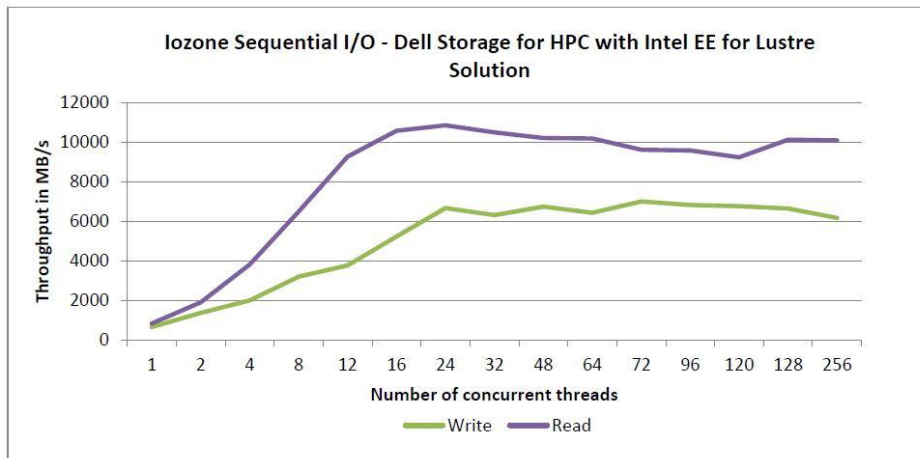
Lustre Read File Data Flow

- 1) Client requests to read a file from the file system
- 2) Client contacts the MDS with a read request
- 3) MDS checks the user authentication and file location
- 4) MDS responds to the Client with a list of OSTs that the stripes of the file are located
- 5) Client receives the response and reads the data from the OSTs without further communication with the MDS



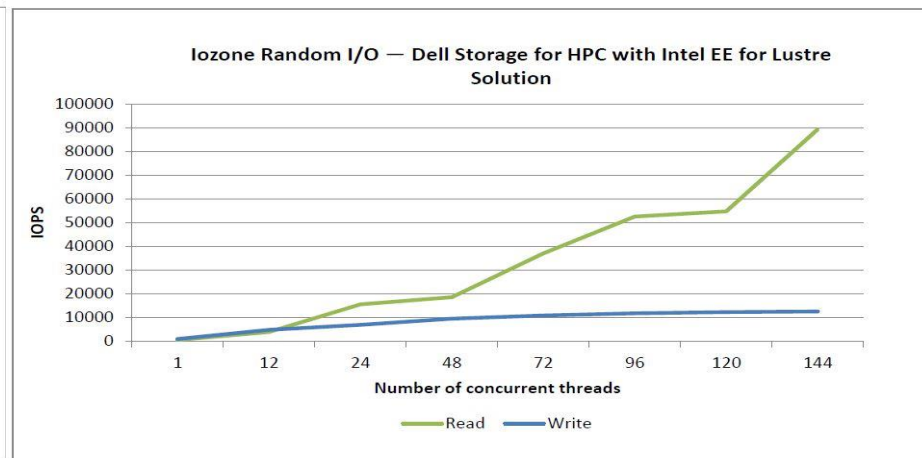
Where performance meets scalability

Sequential IO Performance – 1 MB Stripe



- Peak Write = 7GB/s
- Peak Read = 11GB/s

Random IO Performance – 4 MB Stripe

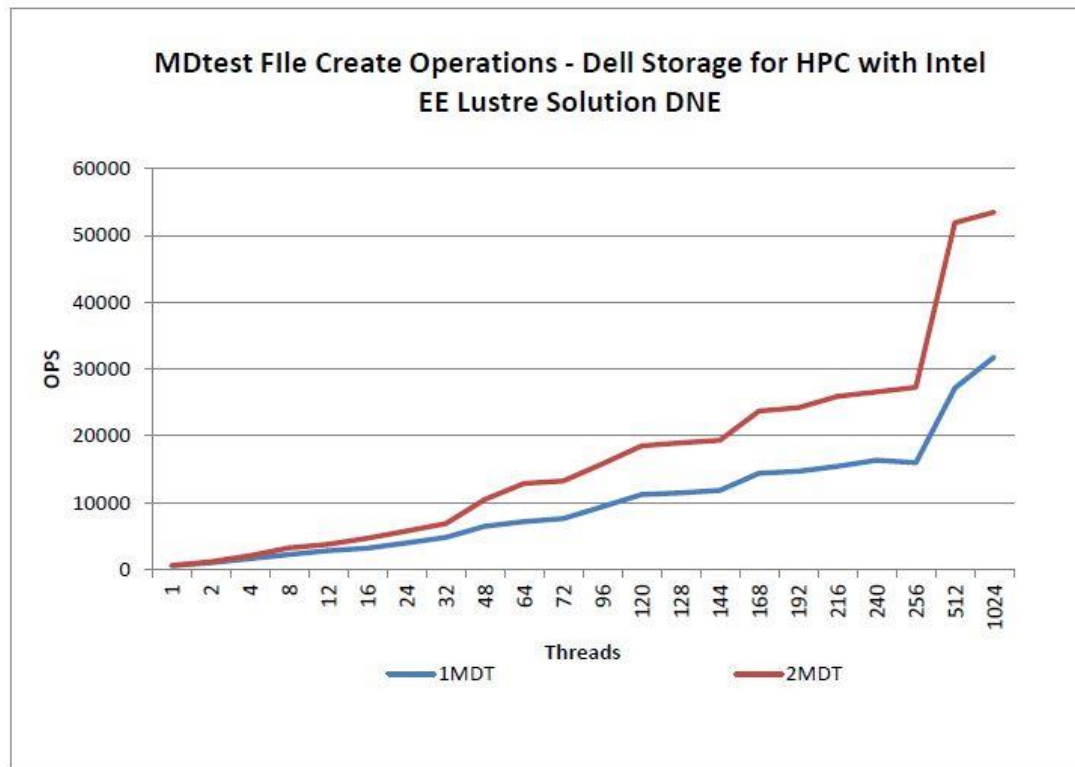


- Peak Write = 12.6K IOPS
- Peak Read = 96K IOPS



Increase metadata performance with Luster Distributed Namespace feature (Lustre DNE)

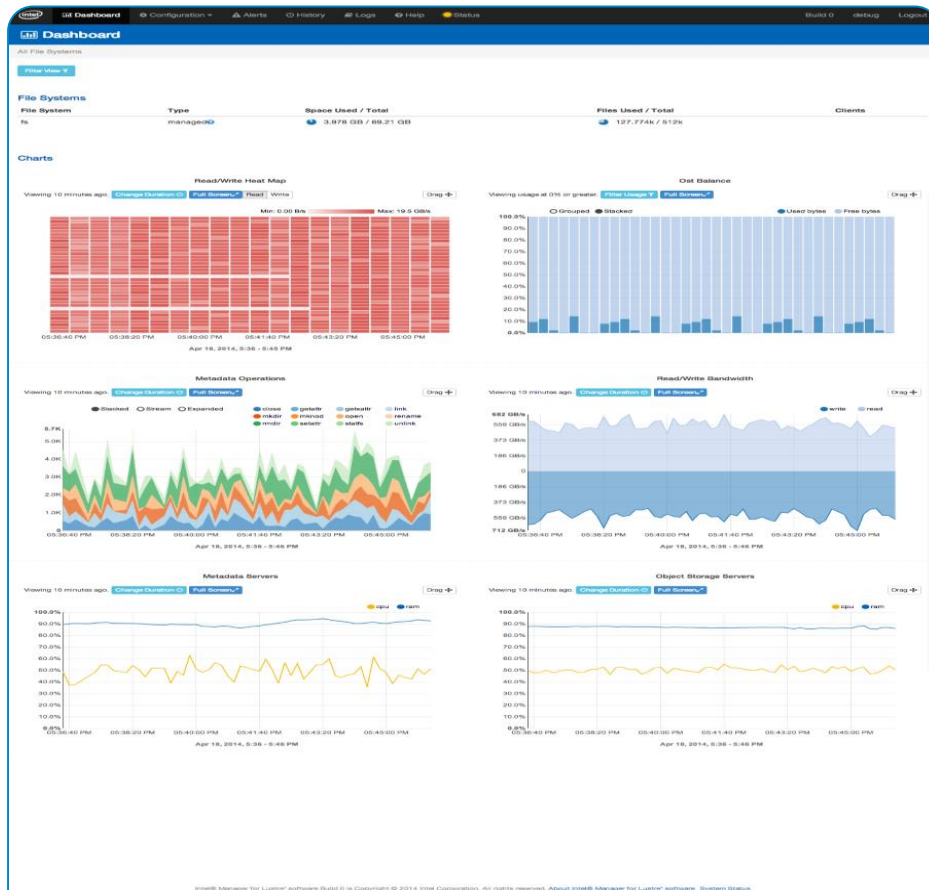
- Lustre DNE Phase 1 is the designation for Luster DNE Remote Directories.
- Lustre sub-directories can now be distributed across multiple MDTs to increase metadata capacity capabilities and performance.



Configure, optimize and manage using Intel Manager for Lustre

IML simplifies your management workflow:

- UI driven configuration, monitoring, and overall management lowers complexity and cost
- Advanced charting options illustrate storage performance in near-real time
- Automated configuration of storage servers pairs for increased high-availability
- Configure and manage power distribution units for automated fail-over
- Smart, intuitive alerts and logs help storage administrators monitor storage performance

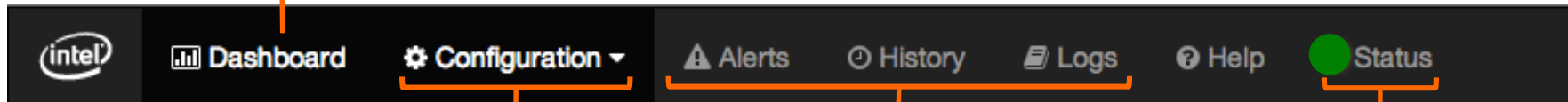


Intel Manager for Lustre (IML) software dashboard



The 'dashboard' canvas displays a variety of dynamic charts illustrating performance levels and resource utilization. Administrators can easily view file systems, check resource consumption for jobs, and monitor performance.

In depth storage hardware reporting is possible when combined with optional hardware vendor provided plug-ins.



Easily configure servers, volumes and power controls. Optionally, enable HSM per file system

Intelligent, intuitive log files – quickly understand how your storage is performing

System status indicator provides the status for all managed file systems. Click to go to detailed information.

Where Big Data meets high-performance computing

As data sets expand, the infrastructure that supports them needs to be faster, bigger, more scalable

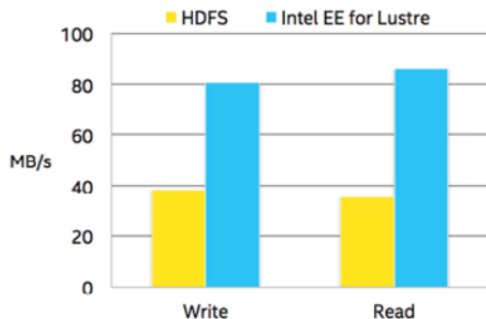
Hadoop over Lustre (HAL)

Intel EE Lustre has plugins for Apache Hadoop and Cloudera Distribution of Hadoop, which require no changes to the Lustre architecture and allows for big data workloads to take advantage of high-performance infrastructure, by replacing the HDFS portion of Hadoop with the Lustre file system.

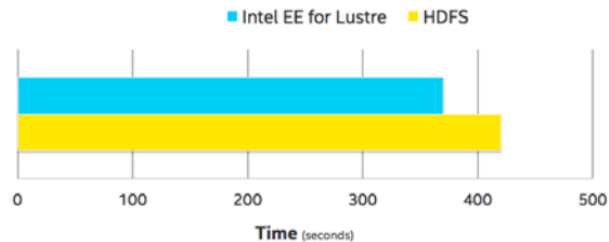
This is driving a new market for High-Performance Data Analytics, where HPC boosts performance of Big Data workloads.

Measuring the performance of Hadoop running on Lustre – [Check out this video of Dell deploying Hadoop on Lustre in production](#)

Results from Intel testing both file systems using TestDFSIO (higher is better)



Results from Intel testing both file systems using TeraSort (shorter is better)



* Source: Tests are from Intel and Tata Consultancy Services Lustre Big Data White Paper ([Click Here](#))



Start by making your data future ready!

Contact your local Dell Account Executive for a free HPC workshop, and we can help identify bottlenecks, opportunities for optimization, and provide a plan for how Lustre can boost your performance, and build a future ready architecture for the convergence of HPC and Big Data





The power to do more