IDC HPC User Forum

Intel's vision: the Scalable system framework

Bret Costelow

September 2015

Legal Disclaimers

Intel technologies features and benefits depend on system configuration and may require enabled hardware, software or service activation. Performance varies depending on system configuration. No computer system can be absolutely secure. Check with your system manufacturer or retailer or learn more at [intel.com].

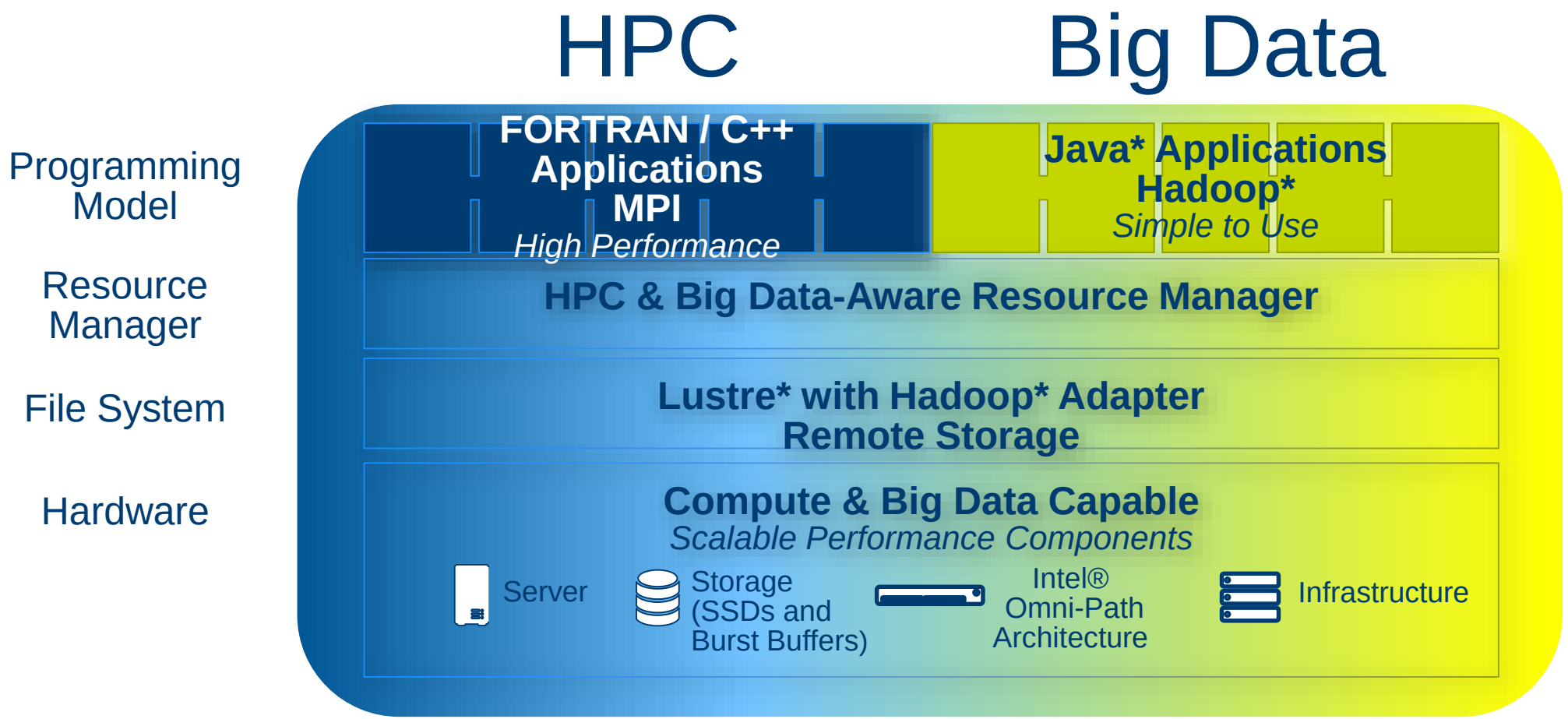
Software and workloads used in performance tests may have been optimized for performance only on Intel microprocessors. Performance tests, such as SYSmark and MobileMark, are measured using specific computer systems, components, software, operations and functions. Any change to any of those factors may cause the results to vary. You should consult other information and performance tests to assist you in fully evaluating your contemplated purchases, including the performance of that product when combined with other products.

All information provided here is subject to change without notice. Contact your Intel representative to obtain the latest Intel product specifications and roadmaps.

Results have been estimated or simulated using internal Intel analysis or architecture simulation or modeling, and provided to you for informational purposes. Any differences in your system hardware, software or configuration may affect your actual performance.

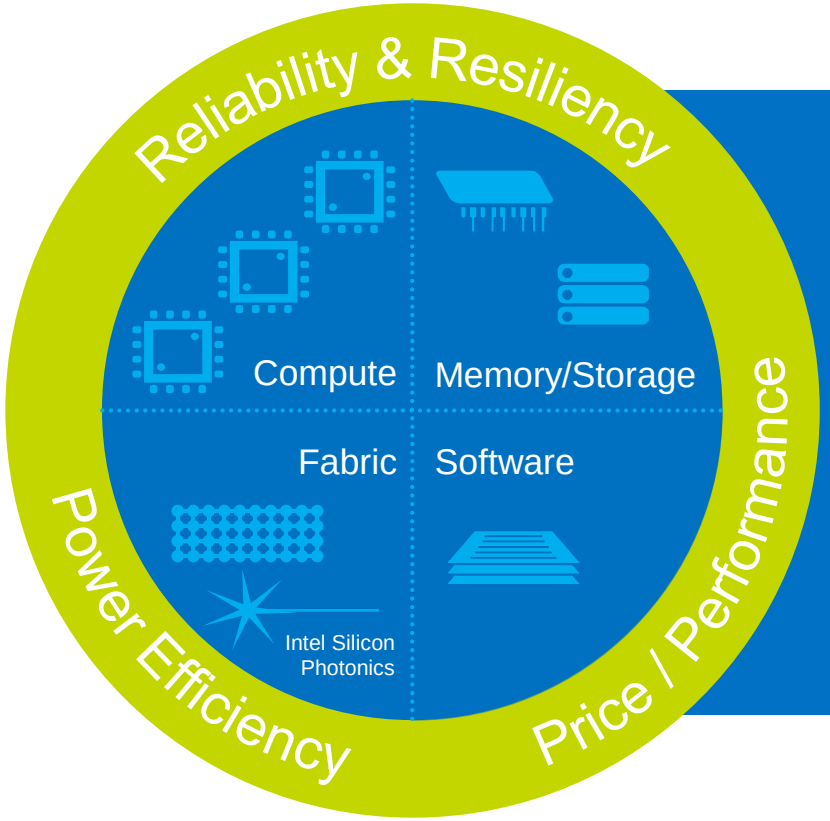
Intel technologies' features and benefits depend on system configuration and may require enabled hardware, software or service activation. Performance varies depending on system configuration. No computer system can be absolutely secure. Check with your system manufacturer or retailer or learn more at https://www-ssl.intel.com/content/www/us/en/high-performance-computing/path-to-aurora.html.

Converged Architecture for HPC and Big Data



Intel's Scalable System Framework

A Configurable Design Path Customizable for a Wide Range of HPC & Big Data Workloads



Small Clusters Through Supercomputers

Compute and Data-Centric Computing

Standards-Based Programmability

On-Premise and Cloud-Based

Intel® Xeon® Processors

Intel® Xeon Phi™ Coprocessors

Intel® Xeon Phi™ Processors

Intel® True Scale Fabric

Intel® Omni-Path Architecture

Intel® Ethernet

Intel® SSDs

Intel® Lustre-based Solutions

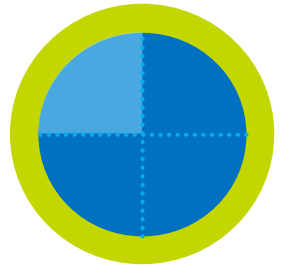
Intel® Silicon Photonics Technology

Intel® Software Tools

HPC Scalable Software Stack

Intel® Cluster Ready Program

Another huge leap in CPU Performance



Intel's HPC Scalable System Framework

Extreme
Parallel
performance

Parallel performance

- **72** cores; 2 VPU/core; 6 DDR4 channels with 384GB capacity
- **>3** Teraflop/s per socket[1]

Integrated memory

- **16GB; 5X** bandwidth vs DDR[2]
- **3 configurable** modes (memory, cache, hybrid)

Integrated fabric

- **2** Intel Omni-Path Fabric Ports (more configuration options)

Market adoption

- **>50** systems providers expected[3]
- **>100** PFLOPS customer system compute commits to-date[3]
- **>**Software development kits shipping Q4'15
- **>Standard IA Software Programming Model and Tools**

[1] Source: Intel internal information. [2]Projected result based on internal Intel analysis of STREAM benchmark using a Knights Landing processor with 16GB of ultra high-bandwidth versus DDR4 memory only with all channels populated. 2Over 3 Teraflops of peak theoretical double-precision performance is preliminary and based on current expectations of cores, clock frequency and floating point operations per cycle. FLOPS = cores x clock frequency x floating-point operations per second per cycle.

Intel's HPC Scalable System Framework

Parallel compute technology

Intel Xeon

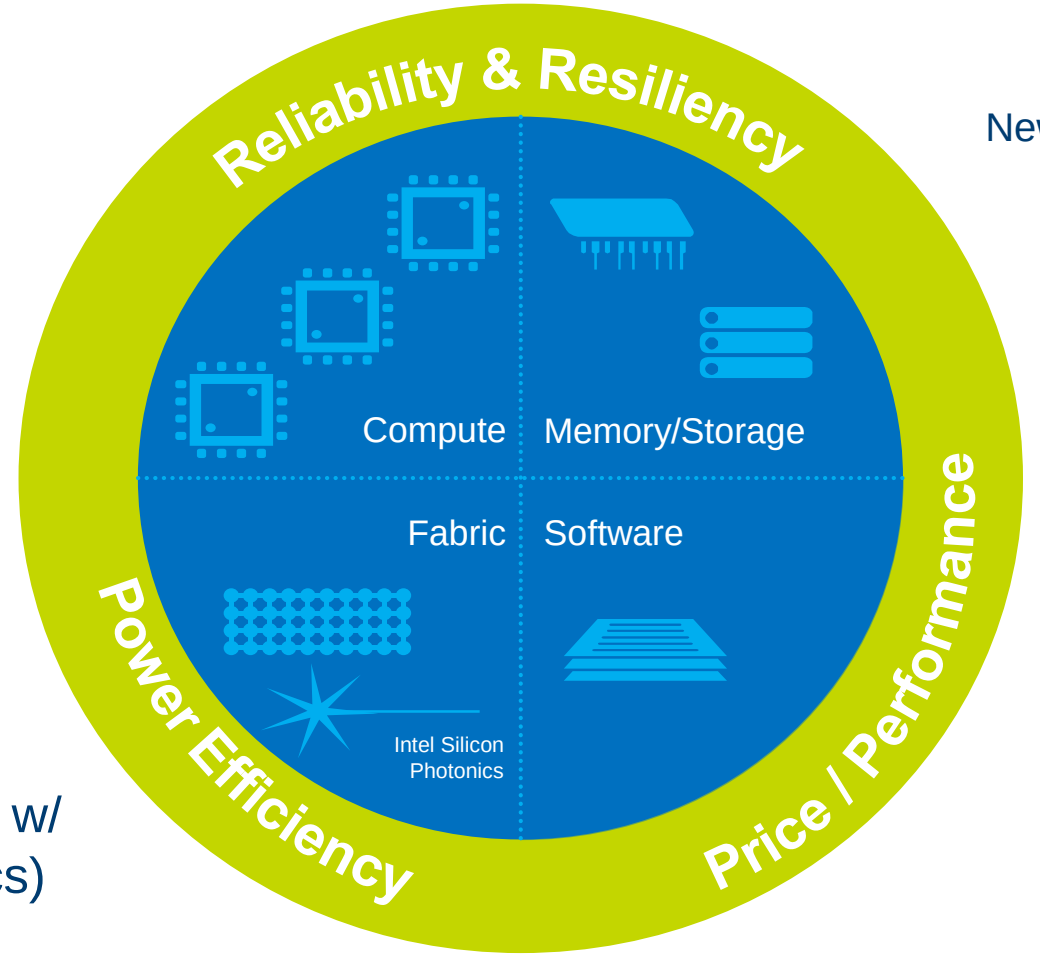
Intel Xeon Phi (KNL)

(eliminate offloading)

New fabric technology

Omni-Path Arch

(KNL integration and w/ Intel Silicon Photonics)



New memory/storage technologies

3D XPoint Technology

Next Gen NVRAM

Powerful Parallel File System

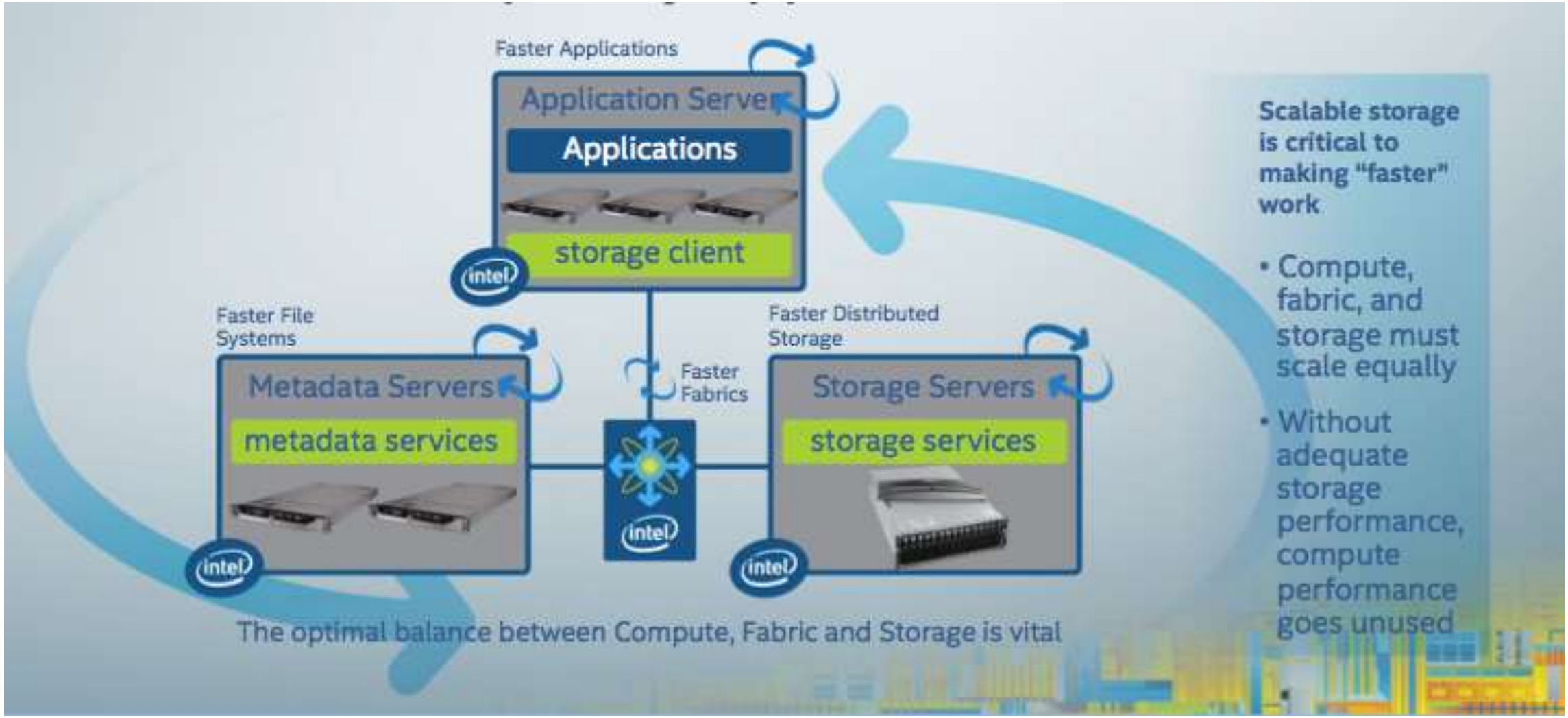
Intel Enterprise Lustre*

Phi Support today

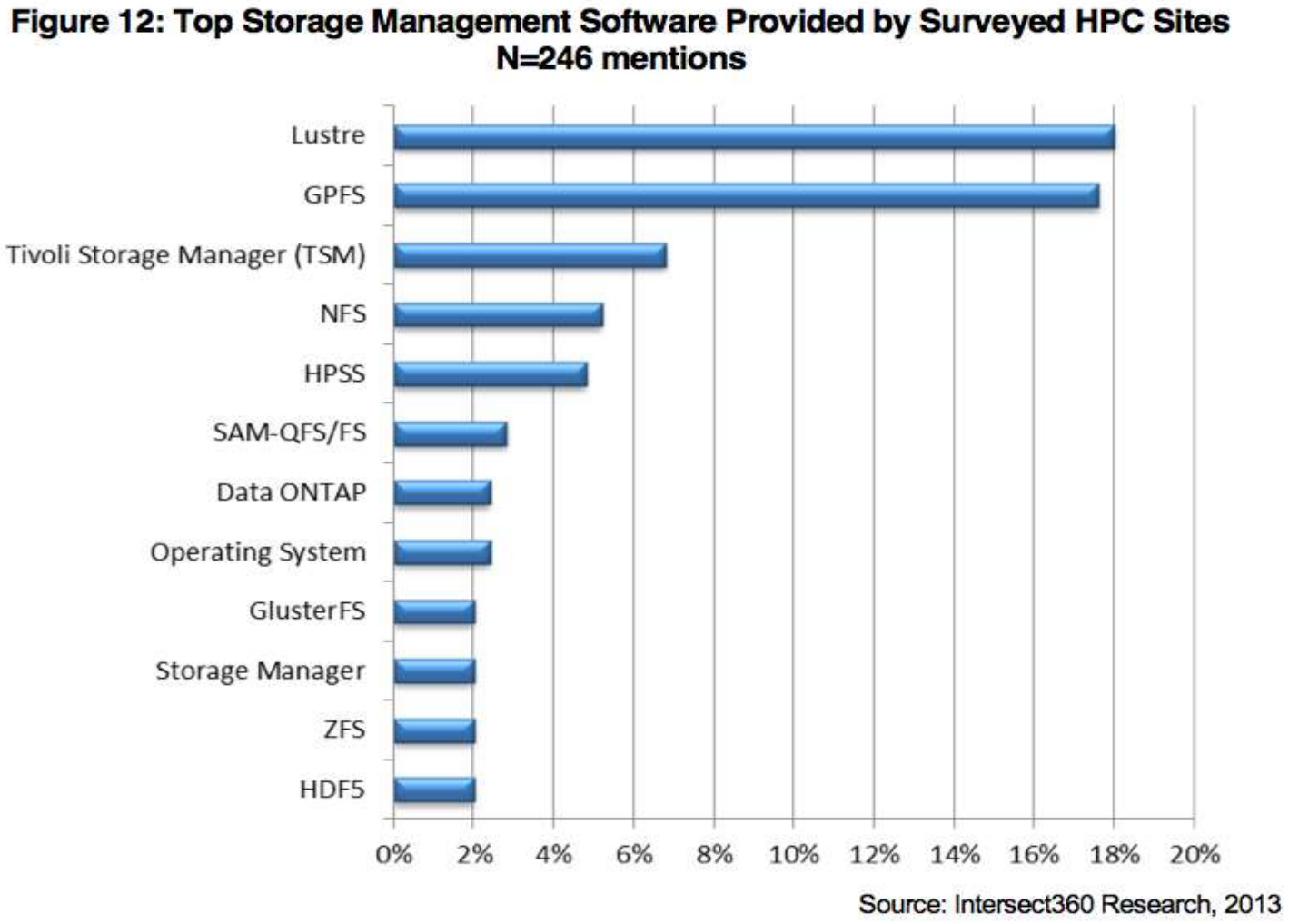
Hadoop Adapters

Intel Manager for Lustre*

Multidisciplinary Approach to HPC Performance



Intel moves Lustre forward

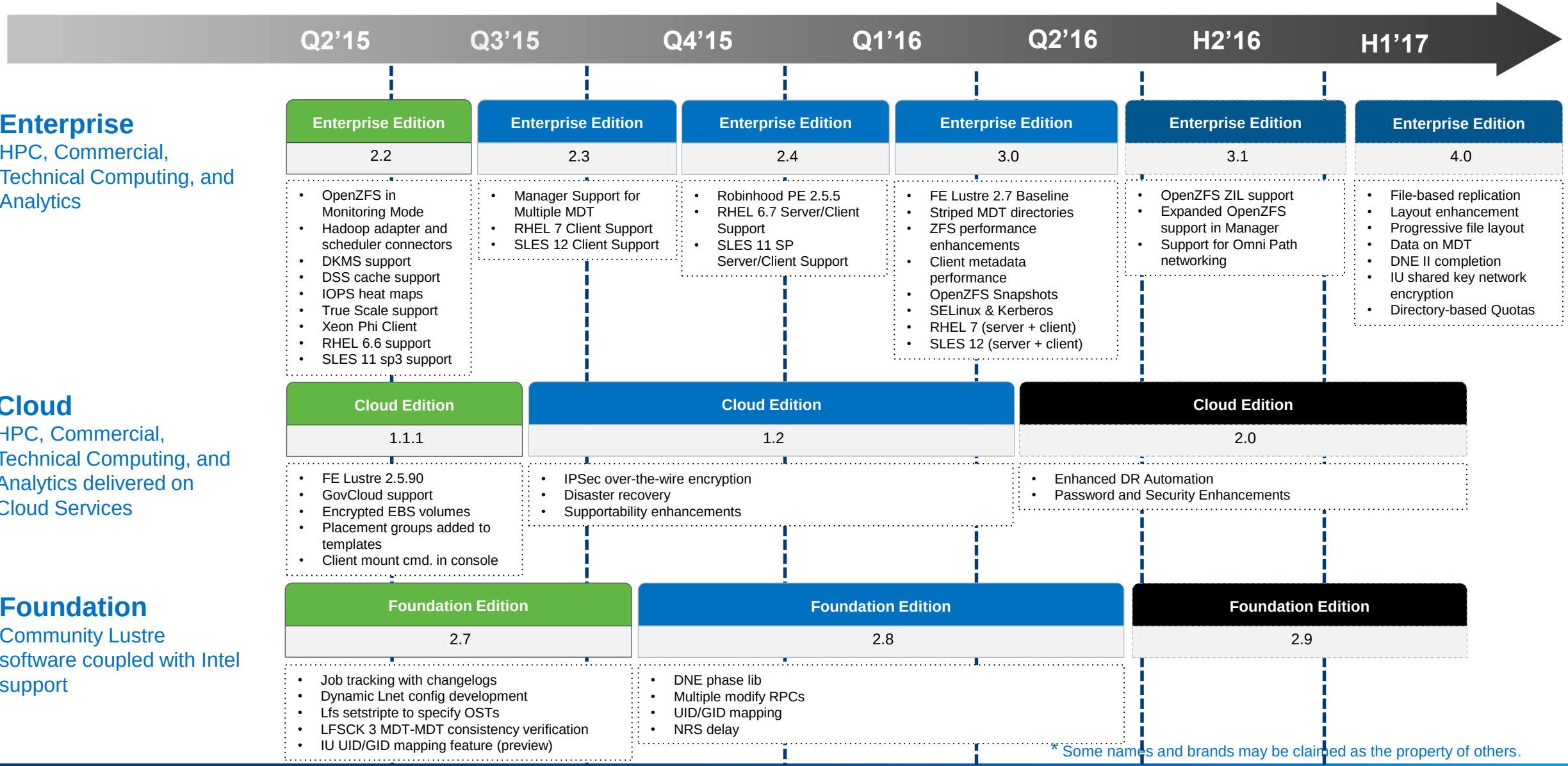


- 8 of Top10 Sites
- Most Adopted
- Most Scalable
- Open Source
- Commercial Packaging
- 4th largest SW company
- More than 80% of code
- Vibrant Community

HPC User Site Census: Middleware – April 2014. Lustre is a registered trademark of Seagate

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

Intel® Solutions for Lustre* Roadmap



Product placement not representative of final launch date within the specified quarter. For more details refer to ILU and 5 Quarter Roadmap.

Intel's HPC Scalable System Framework

Parallel compute technology

Intel Xeon

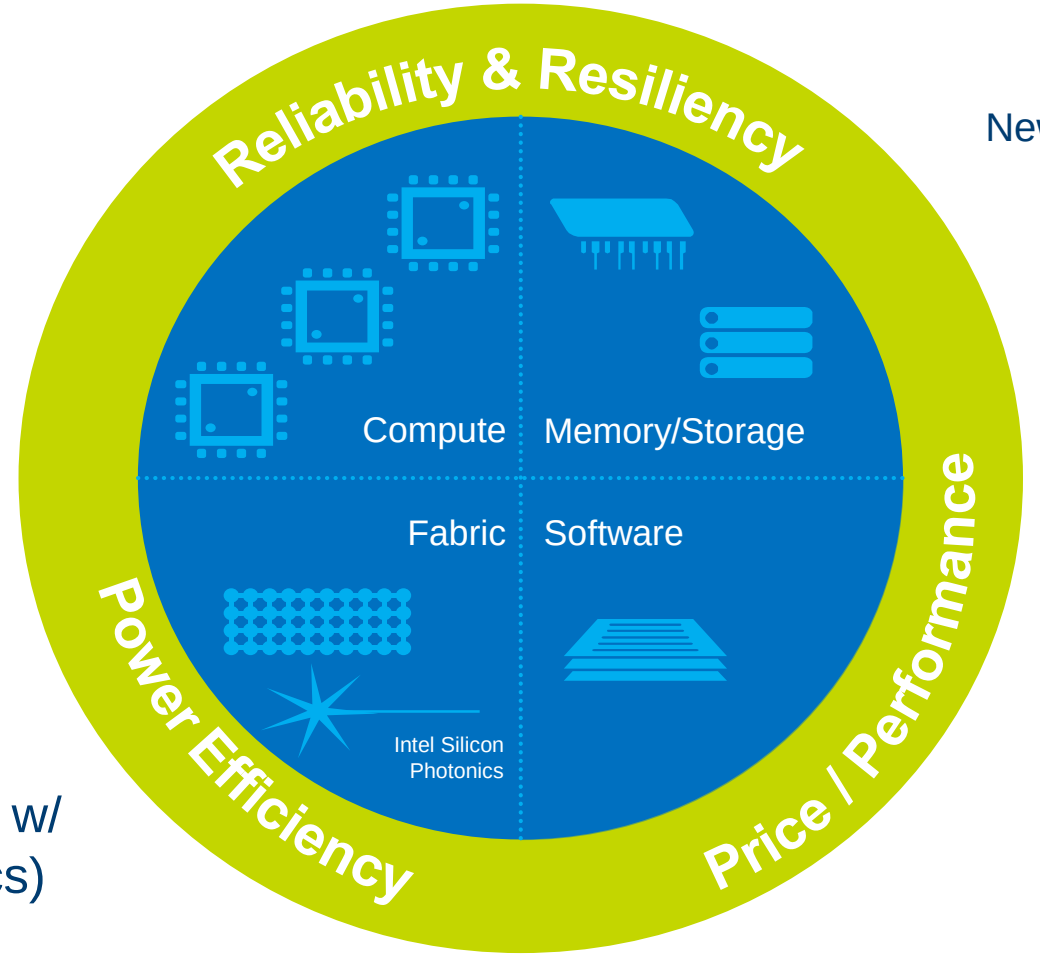
Intel Xeon Phi (KNL)

(eliminate offloading)

New fabric technology

Omni-Path Arch

(KNL integration and w/ Intel Silicon Photonics)



New memory/storage technologies

3D XPoint Technology

Next Gen NVRAM

Powerful Parallel File System

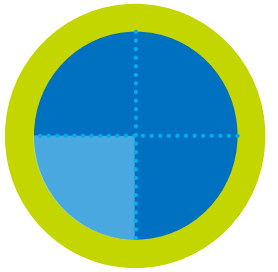
Intel Enterprise Lustre*

Phi Support today

Hadoop Adapters

Intel Manager for Lustre*

A High Performance Fabric



Intel's HPC Scalable System Framework

Intel® Omni-Path Fabric

HPC's Next Generation Fabric

Better scaling vs InfiniBand*

- **48** port switch chip – lower switch costs
- **73%** higher switch MPI message rate[2]
- **33%** lower switch fabric latency[3]

Configurable / Resilient / Flexible

- **Job prioritization** (Traffic Flow Optimization)
- **No-compromise resiliency** (Packet Integrity Protection, Dynamic Lane Scaling)
- **Open source fabric management suite**, OFA-compliant software

Market adoption

- **>100** OEM designs[1]
- >**100ku** nodes under contract/bid[1]

[1] Source: Intel internal information. Design win count based on OEM and HPC storage vendors who are planning to offer either Intel-branded or custom switch products, along with the total number of OEM platforms that are currently planned to support custom and/or standard Intel® OPA adapters. Design win count as of July 1, 2015 and subject to change without notice based on vendor product plans. [2] Based on Prairie River switch silicon maximum MPI messaging rate (48-port chip), compared to Mellanox CS7500 Director Switch and Mellanox SB7700/SB7790 Edge switch product briefs (36-port chip) posted on www.mellanox.com as of July 1, 2015. [3] Latency reductions based on Mellanox CS7500 Director Switch and Mellanox SB7700/SB7790 Edge switch product briefs posted on www.Mellanox.com as of July 1, 2015, compared to Intel® OPA switch port-to-port latency of 100-110ns that was measured data that was calculated from difference between back to back osu_latency test and osu_latency test through one switch hop. 10ns variation due to "near" and "far" ports on an Eldorado Forest switch. All tests performed using Intel ® Xeon® E5-2697v3, Turbo Mode enabled. Up to 60% latency reduction is based on a 1024-node cluster in a full bisectional bandwidth (FBB) Fat-Tree configuration (3-tier, 5 total switch hops), using a 48-port switch for Intel Omni-Path cluster and 36-port switch ASIC for either Mellanox or Intel® True Scale clusters. Results have been estimated or simulated using internal Intel analysis or architecture simulation or modeling, and provided to you for informational purposes. Any differences in your system hardware, software or configuration may affect your actual performance

Intel® Omni-Path Architecture:
Changing the Fabric Landscape

CPU-Fabric Integration

Optimizing

- **Performance**
- **Density**
- **Power**
- **Cost**

Next Generation

Next Intel® Xeon Phi™ processo (Knights Hill)

Next Intel® Xeon® processor

Intel® Xeon Phi™ processor (Knights Landing)
Multi-chip package integration

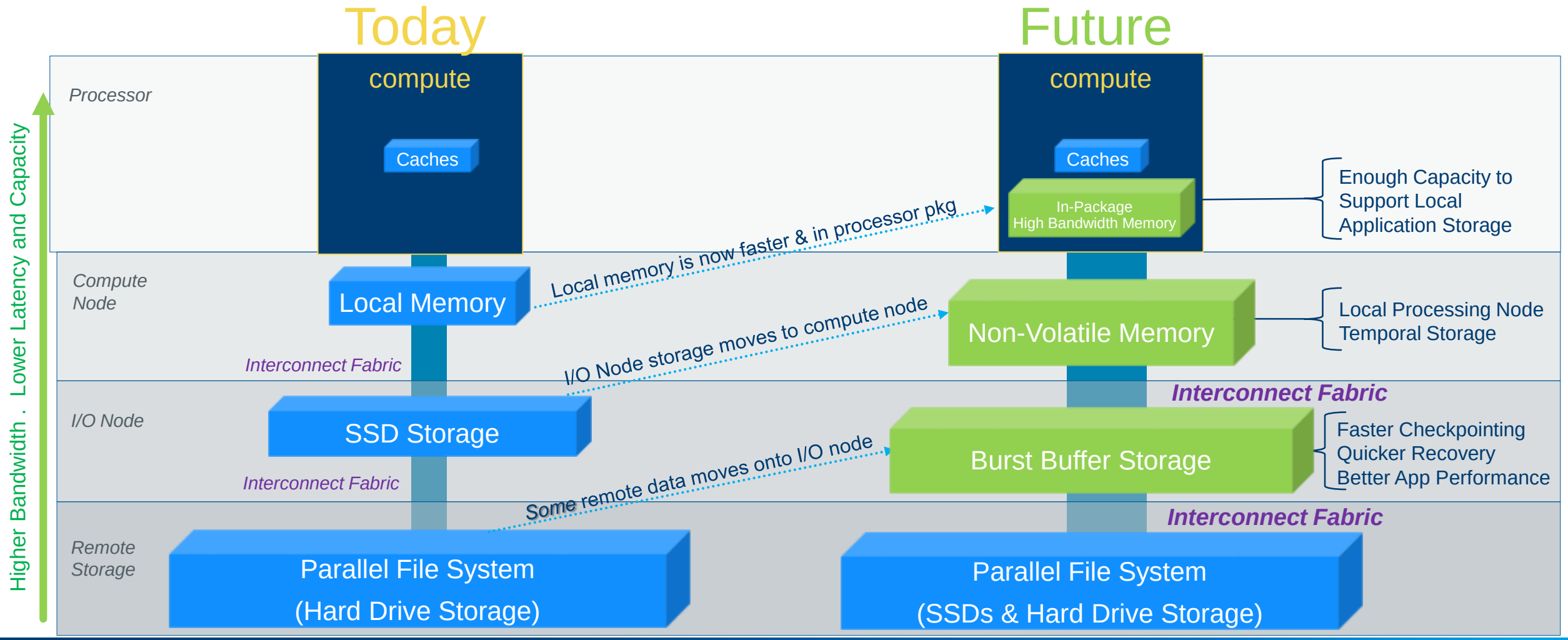
Next Intel® Xeon® processor
Discrete PCIe HFI

Intel® Xeon® processor E5-2600 v3
Discrete PCIe HFI

Time

Intel® Omni-Path: Key to Intel's Next Gen Architecture

Combines with advances in memory hierarchy to keep data closer to compute
→ better data-intensive app performance and energy efficiency



Aurora | Built on a Powerful Foundation

Breakthrough technologies that deliver massive benefits

Compute

>17X performance†

FLOPS per node

>12X memory bandwidth†

>30PB/s aggregate
in-package memory bandwidth

Integrated Intel® Omni-Path Architecture

Processor code name: Knights Hill

Interconnect

>20X faster†

>500 TB/s bi-section bandwidth

>2.5 PB/s aggregate node link bandwidth

File System

>3X faster†

>1 TB/s file system throughput

>5X capacity†

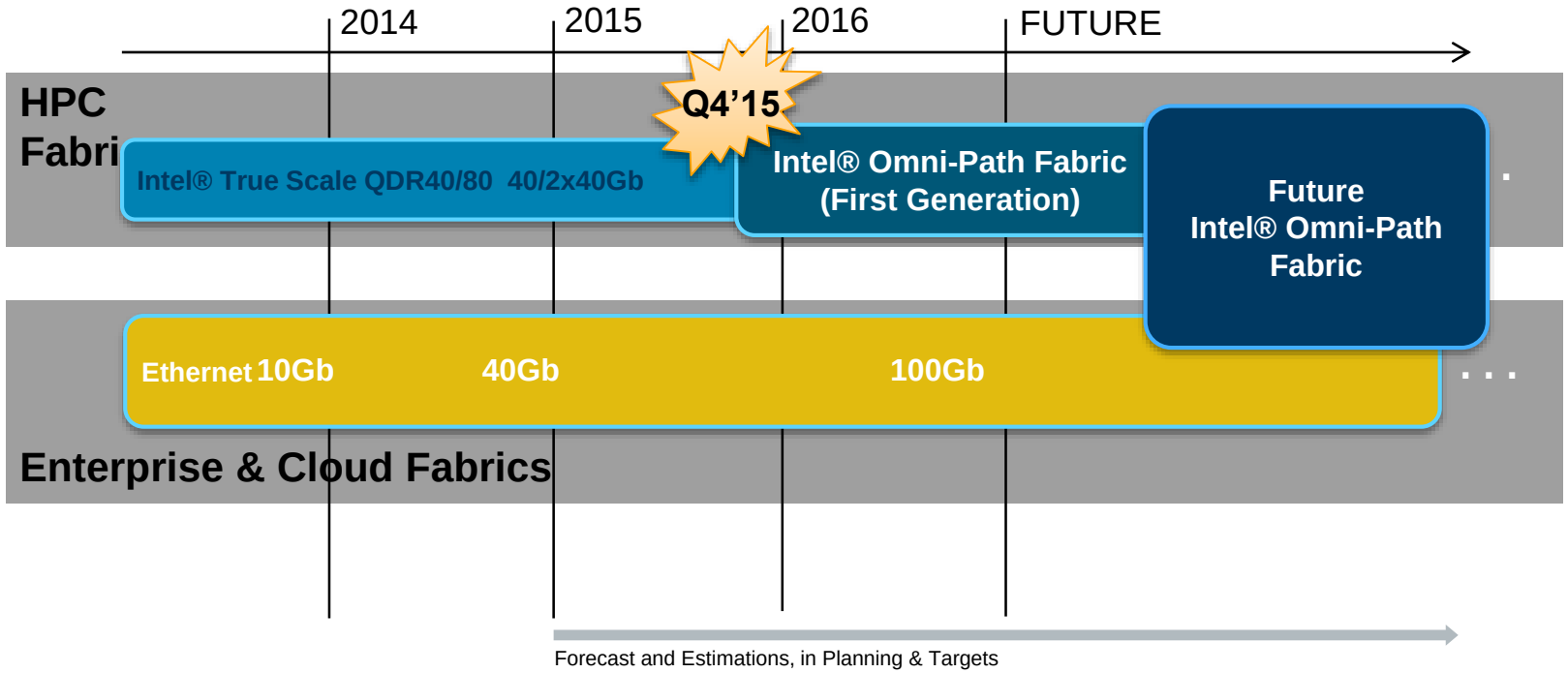
>150PB file system capacity

Source: Argonne National Laboratory and Intel.
† Comparison to ANL's largest current system, MIRA.

Backup

Tying it all together:
The Intel® Fabric Product Roadmap Vision

Intel's HPC Scalable System Framework



The FABRIC is fundamental to meeting the growing needs of the datacenter

Intel is extending fabric capabilities, performance, scalability

Potential future options, subject to change without notice.
All timeframes, features, products and dates are preliminary forecasts and subject to change without further notification.

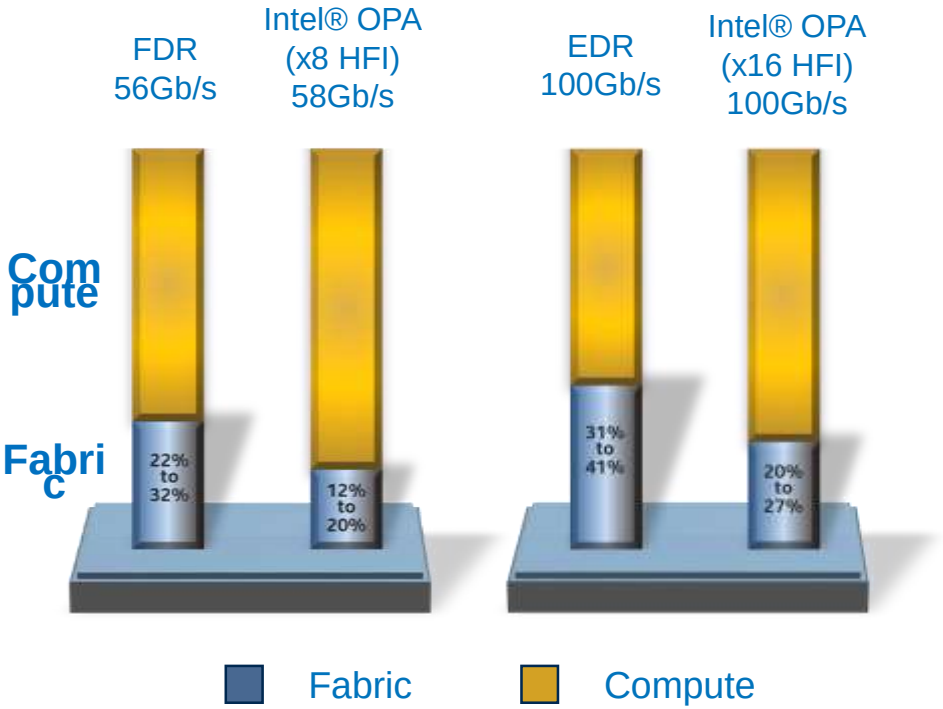
The Interconnect landscape

Compute / interconnect cost ratio has changed

- Compute price/performance improvements continue unabated
- Current corresponding fabric metrics unable to keep pace as a percentage of total cluster costs which includes compute and storage

Challenge: Keeping fabric costs in check to free up cluster $$$ for increased compute and storage capability

Hardware Cost Estimate[1]



More Compute = More FLOPS

1 Source: Internal analysis based on a 256-node to 2048-node clusters configured with Mellanox FDR and EDR InfiniBand products. Mellanox component pricing from www.kernelsoftware.com Prices as of August 22, 2015. Compute node pricing based on Dell PowerEdge R730 server from www.dell.com . Prices as of May 26, 2015. Intel® OPA (x8) Utilizes a 2-1 Oversubscribed Fabric. Analysis Limited Fabric Sizes up to 2K nodes

Evolutionary Approach, Revolutionary Features, End-to-End Solution

High performance, low latency fabric designed to scale from entry to the largest supercomputers

Builds on proven industry technologies

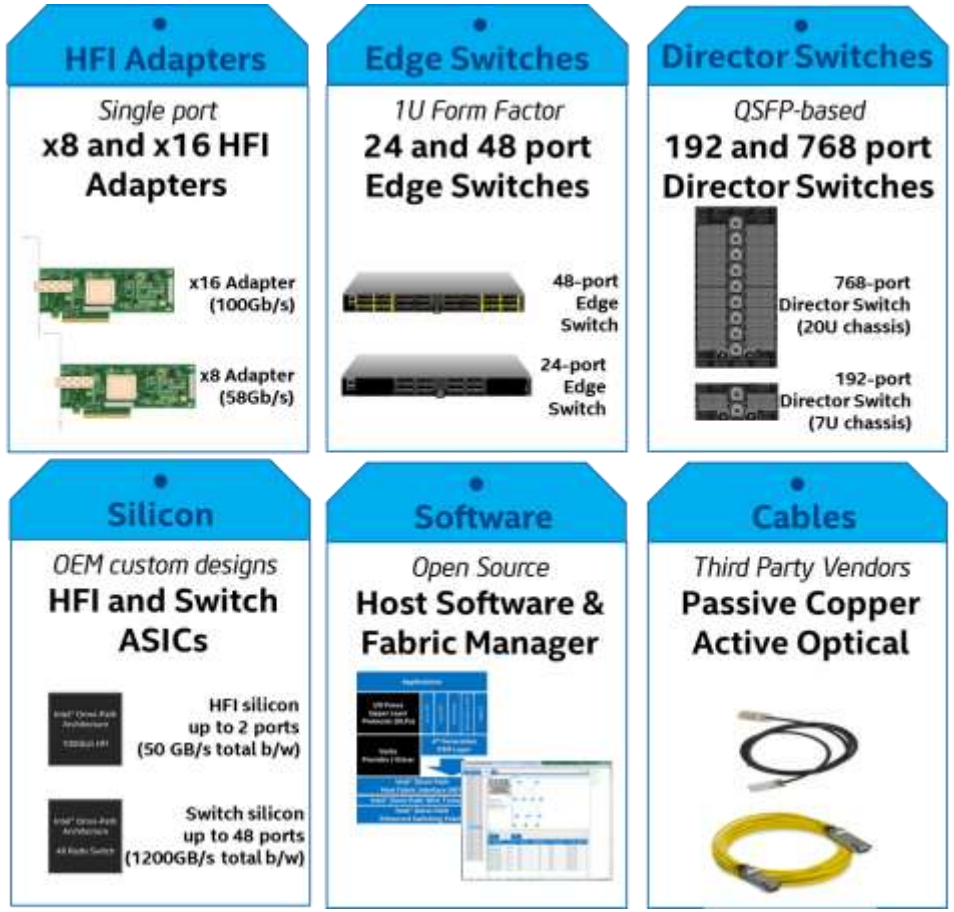
- Innovative new features and capabilities to improve performance, reliability and QoS

Highly leverage existing Aries and True Scale fabric

- Re-use of existing OpenFabrics Alliance* software

Early market adoption

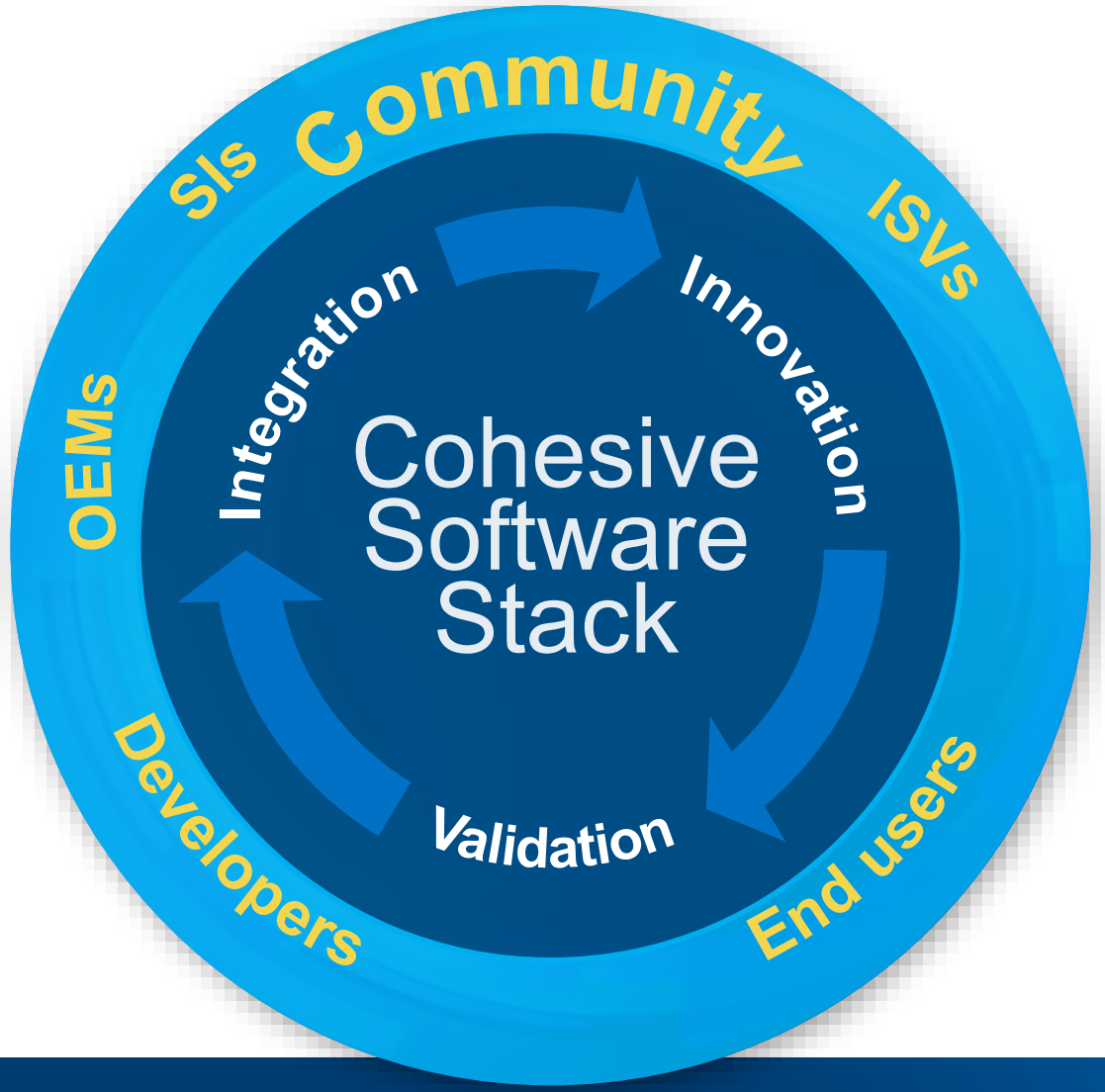
- >100 OEM designs[1]
- >100ku nodes under contract/bid[1]



[1] Source: Intel internal information. Design win count based on OEM and HPC storage vendors who are planning to offer either Intel-branded or custom switch products, along with the total number of OEM platforms that are currently planned to support custom and/or standard Intel® OPA adapters. Design win count as of July 1, 2015 and subject to change without notice based on vendor product plans. [2] Based on Prairie River switch silicon maximum MPI messaging rate (48-port chip), compared to Mellanox Switch-IB* product (36-port chip) posted on www.mellanox.com as of August 13, 2015. [3] Latency reductions based on Mellanox SB7700/SB7790 Edge switch product briefs posted on www.Mellanox.com as of August 13, 2015 with a stated latency of 90ns, compared to Intel® OPA switch port-to-port latency of 100-110ns that was measured data that was calculated from difference between back to back osu_latency test and osu_latency test through one switch hop. 10ns variation due to "near" and "far" ports on an Eldorado Forest switch. All tests performed using Intel ® Xeon® E5-2697v3, Turbo Mode enabled. Up to 33% latency reduction is based on a 1024-node cluster in a full bisectional bandwidth (FBB) Fat-Tree configuration, using a 48-port switch for Intel Omni-Path cluster (3 switch hops) and 36-port switch ASIC for either Mellanox or Intel® True Scale clusters (5 switch hops). Results have been estimated or simulated using internal Intel analysis or architecture simulation or modeling, and provided to you for informational purposes. Any differences in your system hardware, software or configuration may affect your actual performance

New Highly Scalable Open Software Stack

Intel-led industry collaboration underway to create a complete and cohesive HPC system stack



Intel's HPC Scalable System Framework

An open community effort

- Broad range of ecosystem partners
- Designing, developing and testing
- Open source availability enables differentiation

Benefits the entire HPC ecosystem

- Innovative system hardware and software
- Simplify configuration, management and use
- Accelerate application development
- Turnkey to customizable Intel-supported versions

Lustre.intel.com