Solution Brief
Intel® Omni-Path Architecture
High Performance Computing (HPC)

HIGHER PERFORMANCE AT LOWER COST FOR HPC FABRICS

A fast, scalable fabric is essential to maximize cluster performance, yet traditional InfiniBand* technologies cannot keep pace with growing needs. As a result, organizations are now spending as much as 20 to 40 percent of their infrastructure budget on fabric elements[1]—money that could potentially be used to scale compute and storage resources.

Intel® Omni-Path Architecture (Intel® OPA), an element of Intel® Scalable System Framework, rises to meet this challenge by transforming the performance, scalability, and cost models of HPC fabrics. This end-to-end fabric solution delivers 100 Gbps port bandwidth, while providing low latency that stays low even at extreme scale. Optimized packet protocols, dynamic traffic shaping, and advanced quality of service (QoS) deliver efficient support for diverse traffic types, with high throughput, high packet integrity, and low, deterministic latency for critical MPI messaging.

Reduce Infrastructure Requirements By Up to 50 Percent[2]

Intel developed a new 48-port switch silicon for Intel OPA, which can reduce the number of switches by as much as 50 percent in a typical fat tree configuration.[2] This new silicon powers a family of 24- and 48-port edge switches and 192- and 768-port director-class switches that can be used to support the full range of HPC requirements. With the Intel OPA 48-port leaf cards, a single 5-hop fabric can support up to 27,648 nodes, which is 2.3 times more than can be supported using current InfiniBand switch designs (an InfiniBand fabric would require 7-hops to go beyond 11,664 nodes).[3]

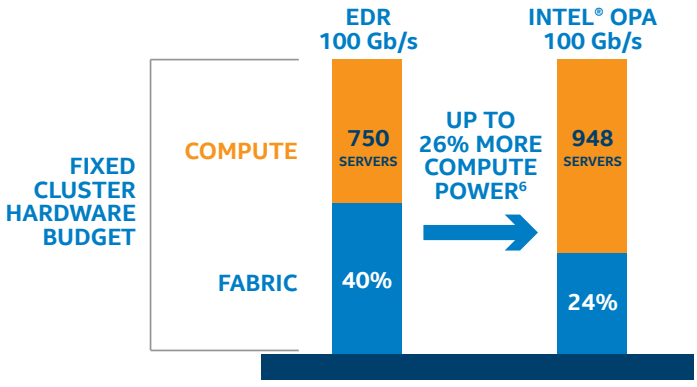
Get More Compute Power Within the Same HPC Budget

An Intel OPA fabric not only scales more easily than current InfiniBand solutions, but also helps to reduce cost, complexity, and power consumption. An Intel OPA fabric for a mainstream HPC cluster with 700 to 750 nodes would require up to 79 percent less rack space than a comparable fabric based on InfiniBand, with two fewer cabinets, which would save 12 square feet of floor space.[4] The Intel OPA fabric would also reduce power and cooling costs by as much as 60 percent over a 3-year period.[5]

Even more importantly, the InfiniBand fabric would soak up as much as 40 percent of the total cluster hardware budget, while an Intel OPA fabric would account for only 21 percent.[1] The money saved could be used to purchase up to 26 percent more servers, along with the fabric elements and cables required to connect the new servers to the cluster.[6] Researchers and engineers would have substantially more compute power to run increasingly complex studies and obtain faster results.

Intel® Omni-Path Architecture versus InfiniBand* EDR

Up to 26 percent more compute power within the same cluster hardware budget[6]



Same Hardware Budget, High Cluster Performance.

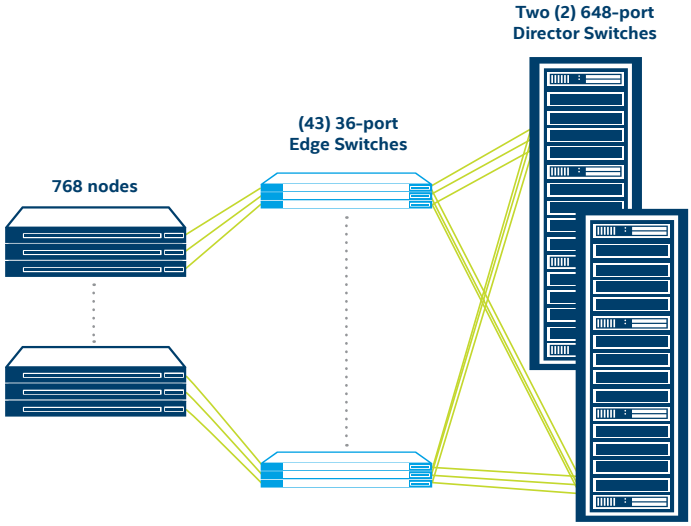
With Intel® Omni-Path Architecture, organizations can spend more on compute resources and less on fabric elements, resulting in higher overall cluster performance within the same budget.[6]

Intel® Omni-Path Fabric's 48 Radix Chip

It's more than just a 33% increase in port count over a 36 Radix chip

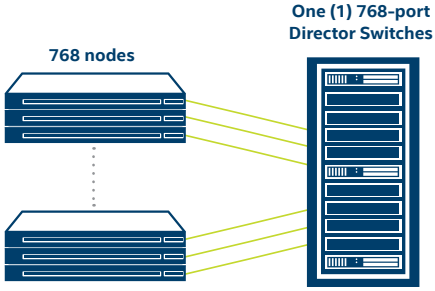
InfiniBand* EDR (36-port Switch Chip)

FIVE-hop Fat Tree



Intel® Omni-Path Architecture (48-port Switch Chip)

THREE-hop Fat Tree



			% Reduction
(43) 36-port	**Edge Switches**	None required	**100%**
1,542	**Cables**	768	**50%**
99u (2+ racks)	**Rack Space[7]**	20u (<1/2 rack)	**79%**
~680ns (5 hops)	**Switch Latency[8]**	300 – 330ns[9]	**51–55%**

Learn More

Intel OPA can help you achieve higher throughput and lower latency for your HPC fabric, while enabling higher overall cluster performance within the same budget.[6] For more information, read the Intel white paper: "Transforming the Economics of HPC Fabrics with Intel® Omni-Path Architecture," and visit **www.intel.com/hpcfabrics**

[1] 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 November 3, 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 over-subscribed Fabric. Intel® OPA pricing based on estimated reseller pricing using Intel MSRP pricing on ark.intel.com as of November 2015.

[2] Reduction in up to ½ fewer switches claim based on a 1024-node full bisectional bandwidth (FBB) Fat-Tree configuration, using a 48-port switch for Intel Omni-Path cluster and 36-port switch ASIC for either Mellanox or Intel® True Scale clusters.

[3] Based on 27,648 nodes for a cluster configured with the Intel Omni-Path Architecture using 48-port switch ASICs, as compared with a 36-port switch chip that can support up to 11,664 nodes.

[4] All Mellanox power and rackspace information is based on Mellanox installation documentation on www.mellanox.com as of November 3, 2015. Cabinet floor space assumes a cabinet that is 24" wide by 37" deep.

[5] Assumes 750-node cluster, and number of switch chips required is based on a full bisectional bandwidth (FBB) Fat-Tree configuration. Intel® OPA uses one fully-populated 768-port director switch, and Mellanox EDR solution uses a combination of director switches and edge switches. Mellanox power data based on Mellanox CS7500 Director Switch, Mellanox SB7700/SB7790 Edge switch, and Mellanox ConnectX-4 VPI adapter card installation documentation posted on www.mellanox.com as of November 1, 2015. Intel OPA power data based on product briefs posted on www.intel.com as of November 16, 2015. All Mellanox power and rackspace information is based on Mellanox installation documentation on www.mellanox.com as of November 3, 2015. Cabinet floor space assumes a cabinet that is 24" wide by 37" deep.

[6] Assumes a 750-node cluster, and number of switch chips required is based on a full bisectional bandwidth (FBB) Fat-Tree configuration. Intel® OPA uses one fully-populated 768-port director switch, and Mellanox EDR solution uses a combination of 648-port director switches and 36-port edge switches. Mellanox component pricing from www.kernelsoftware.com, with prices as of November 3, 2015. Compute node pricing based on Dell PowerEdge R730 server from www.dell.com, with prices as of May 26, 2015. Intel® OPA pricing based on estimated reseller pricing based on projected Intel MSRP pricing at time of launch. All amounts in US dollars.

[7] Rack space based on Mellanox CS7500 Director Switch (28U, 648 ports) and Mellanox SB7700/SB7790 Edge switch (1U, 36 ports) product briefs posted on www.mellanox.com as of November 1, 2015. Intel® OPA Edge Switch 100 Series and Intel® OPA Director Class Switch 100 Series product briefs posted on www.intel.com, November 2015.

[8] Latency data based on Mellanox CS7500 Director Switch (sub-500ns) and Mellanox SB7700/SB7790 Edge switch (90ns) product briefs posted on www.mellanox.com as of November 1, 2015. Intel® OPA Edge Switch 100 Series and Intel® OPA Director Class Switch 100 Series product briefs posted on www.intel.com, November 2015.

[9] Intel measured data that was calculated from difference between back to back osu_latency test and osu_latency test through one switch hop. All tests performed using Intel ® Xeon® E5-2697v3. Pre-production Intel Corporation Device 24f0 – Series 100 HFI ASIC, Series 100 Edge Switch – 48 port.

Printed in USA 1115/EG/HB/PDF 332939-001US