

Real Solutions for the Challenges of the Post-Petascale Era

Introduction	1
Moving More Data in Less Time	1
Reducing Power Consumption: A Key Challenge	2
Accelerating GPU Computing	2
Delivering Advanced Network Capabilities	2
Supporting Next Generation Compute Architectures	3
Mellanox Solutions Today Are Enabled For Tomorrow	3
About Mellanox	3

Introduction

Pushing the frontiers of science and technology will require extreme-scale computing with machines that are 500-to-1,000 times more capable than today's supercomputers. As researchers continuously refine the models and push increased resolutions, the demand for more parallel computation and advanced networking capabilities is paramount. As a result of the ubiquitous data explosion and the ascendance of big data, especially unstructured data, today's systems need to move enormous amounts of data as well as perform more sophisticated analysis; the interconnect truly becomes the critical element of enabling the use of data.

Moving More Data in Less Time

Mellanox's Switch-IB™ is the world's first EDR InfiniBand switch system, capable of delivering 7.2Tb/s of non-blocking bandwidth. The port-to-port latency was reduced compared to previous generations to 130ns, and each of the 36 ports provides 100Gb/s full bidirectional bandwidth. Among the advanced capabilities of this new technology, Mellanox added support for InfiniBand routing, which enables new designs of larger scale systems with virtually no limitations in terms of topologies. Switch-IB also provides advanced adaptive routing capabilities (dynamic network routing paths) for improved application performance at scale, especially those applications that are network latency sensitive.

Mellanox also introduced the LinkX™ line of copper and fiber cables supporting the EDR technology. Bringing EDR switch technology for the data center switch network is a game changer. The EDR network aggregation offers lower cost, less equipment, and improved performance. For instance, a full non-blocking 648-node cluster would take 54 FDR switches, but only 39 switches when using EDR Switch-IB – this translates to fewer switches, fewer cables, and lower latency.

For the server connectivity, both Connect-IB™ and the latest ConnectX®-4 EDR 100Gb/s HCA enable advanced performance and scalability features, such as Dynamically Connected (DC) transport. The DC transport is a new scalable transport, in which resource requirements scale based on node characteristics, with no dependency on system size. The combination of the world's fastest 100Gb/s InfiniBand HCAs from Mellanox and our latest Switch-IB switching capabilities delivers leading performance and scalability, reduces infrastructure costs, and helps address the very real issue of power constraints in today's HPC and enterprise data centers.

As larger Exascale class systems are being designed to solve the even larger scope of problems being addressed by high performance computing, the paramount issue we all are challenged with is reducing the power consumption. This is not just a challenge for the Exascale deployments that are several years away, but even affects the entire HPC community today – government, academia, and commercial institutions alike.

Reducing Power Consumption: A Key Challenge

As advancements are being made in lowering the processor's power consumption, similar efforts are underway in optimizations in other sub-system components including the network elements. Mellanox is working aggressively to address the power challenges of today and tomorrow, including continued work within the ECONET consortium, a project dedicated to advancements of dynamic power scaling and focused on network-specific energy saving capabilities. The goal is to reduce energy requirements of the network by 50 to 80%, while ensuring end-to-end quality of service. Mellanox is becoming an increasingly important industry leader in reducing the overall carbon footprint of next generation high performance computing deployments.

Accelerating GPU Computing

Most can agree that in the past decade one of the most transformative technologies in HPC has been GPGPU computing. Lately, GPUs have become commonplace in the high performance computing space. In November 2013, Mellanox and NVIDIA released a new technology called GPUDirect® RDMA. This technology allows direct peer-to-peer communication between remote GPUs over the Mellanox fabric, completely bypassing the need for CPU and host memory intervention to move data. This capability reduces latency for internode GPU communication by upwards of 70%.

Coupled with the advanced customization abilities of cloud computing, the HPC community is pursuing the use of clouds incorporating GPUs for their computing needs. Enablement of Infrastructure as a Service with GPUs into cloud infrastructures will enable researchers to rapidly adopt and deploy their own private HPC clouds using GPUs more effectively. This again translates to the advanced network capabilities of the Mellanox interconnect, using less hardware more efficiently and further reducing power consumption.

Delivering Advanced Network Capabilities

The Top500 list is a great tool for looking at the technology trends of HPC. One key metric for value of an HPC installation is the efficiency rating. InfiniBand is dominant in the Top500 list today, primarily because InfiniBand is the only interconnect that delivers not only RDMA capabilities, but also advanced network capabilities that improve the efficiency of the overall system. InfiniBand is proven to be capable of delivering up to 99.8% efficiency for HPC systems in the Top500 list today. Additionally, with advanced offloading of collective communication onto the Mellanox fabric, we free up the processors to do meaningful computation instead of spending time on network communication.

We have touched on some of the additional capabilities already available today that are critical elements to reach Exascale class environments, such as robust adaptive routing, support for new topologies, InfiniBand routing, acceleration of GPUs, and offloading the processor from network tasks; but what about the future of microprocessor architectures?

Supporting Next Generation Compute Architectures

In all probability, there will not be any single dominant microprocessor architecture for next generation Exascale class installations, especially as workflows for tomorrow's computational science are driving requirements. Alternative architectures to x86, such as Power and 64-bit ARM are already picking up adoption. The highly advanced, lower-power architectures capable of handling demanding HPC workflows will rely upon a best-in-class interconnect that is scalable, sustainable, and able to exploit application performance at extreme scale.

Mellanox Solutions Today Are Enabled For Tomorrow

Mellanox continues to demonstrate our commitment to the future of high performance computing by delivering a highly optimized network infrastructure today that is already a generation ahead of any other interconnect solution. We believe our goals are aligned with building end-to-end scalable fabric solutions for system software performance. We continue to thrive as the leader in interconnect technology capable of handling communications of massive parallelism, tuned for the underlying infrastructure and optimized for moving enormous volumes of data, yet also entirely energy conscious in regard to how that data is delivered.

About Mellanox

Mellanox Technologies (NASDAQ: MLNX) is a leading supplier of end-to-end InfiniBand and Ethernet interconnect solutions and services for servers and storage. Mellanox interconnect solutions increase data center efficiency by providing the highest throughput and lowest latency, delivering data faster to applications and unlocking system performance capability. Mellanox offers a choice of fast interconnect products: adapters, switches, software and silicon that accelerate application runtime and maximize business results for a wide range of markets including high performance computing, enterprise data centers, Web 2.0, cloud, storage and financial services.

More information is available at www.mellanox.com.



350 Oakmead Parkway, Suite 100, Sunnyvale, CA 94085
Tel: 408-970-3400 • Fax: 408-970-3403
www.mellanox.com