Mellanox®
TECHNOLOGIES

SX1012: High Performance Small Scale Top-of-Rack Switch

Introduction

As new data center applications are used, new solutions are required to meet the demand for higher throughput without increasing the power consumption or cost.

A common way to connect servers or storage devices to the data center infrastructure is to deploy Top-of-Rack (ToR) switches. Most ToR switches available on the market today are 1RU in size and provide 48 to 64 10GbE ports. In some cases, the utilization level of these switches is lower than their maximum capacity for various reasons. This low utilization of switch ports equates to both high capital costs and higher operational costs (larger switches consume more power). Mellanox's SX1012 switch solution is designed to occupy half the width of a 19" rack in order to solve this issue while maintaining the high standards of performance switching that are found in other SwitchX-based systems while allowing for growth of the network.

Smaller Footprint Equals Cost Savings

Most high performance Ethernet switches available on today's data center market are 1 RU in size and the chassis are fully occupied with ports. For various reasons, data center racks are frequently not fully occupied with servers and storage systems. Running standard switches with less than half of their ports in use results in very inefficient use of expensive hardware, translating to high outlay costs and high power consumption.

Mellanox's SX1012 provides a smaller footprint and a more flexible configuration that improves the utilization of the hardware, lowers power consumption, and therefore reduces the capital and operational costs significantly.

Pay As You Grow Strategy

Because the SX1012 switch only uses half of a rack unit width, it is always possible to add a new SX1012 side-by-side with the existing one in order to add nodes for additional servers or storage. In this manner, switch capacity is expanded dynamically without any disruption to the existing switch and without having to use more rack space. SX1012 therefore offers the advantage of a switching solution that grows with your business.
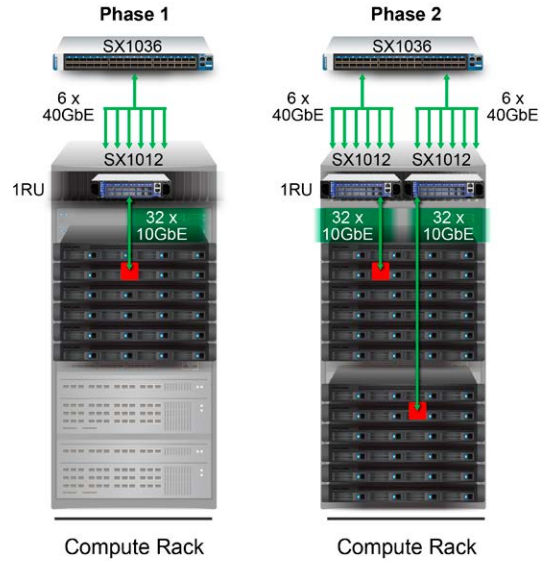
Figure 1. *Pay As You Grow Strategy*

Optimal ToR for Small-Scale Deployments

SX1012 is the optimal ToR switch for small-scale high-performance computing, storage, and database deployments, providing 12 QSFP interfaces that can operate at speeds of 1GbE, 10GbE, 40GbE, or 56GbE. Flexibility is built-in, as each QSFP interface can be used with a passive breakout cable to create four discrete SFP+ interfaces to be used as 10GbE ports, for a maximum of 48 10GbE ports on the smallest form factor.
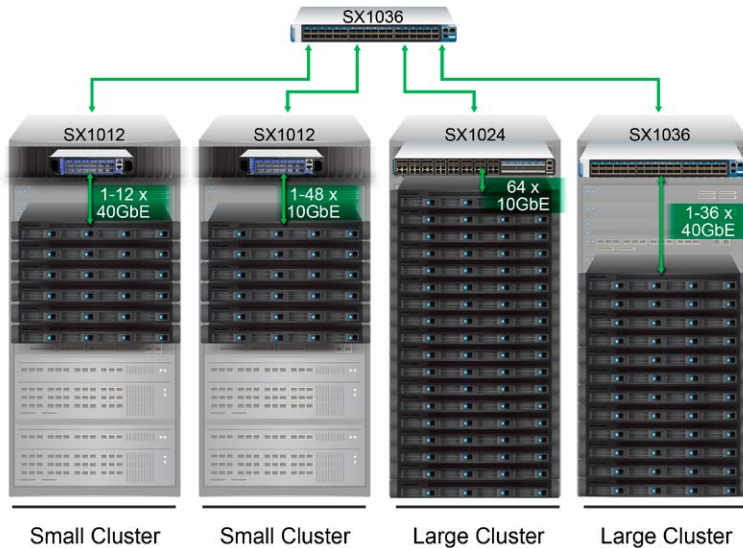

Figure 2. *SX1012 Optimal Switch for Small-to-Medium Clusters*

Flexible Port Assignment

Any port or group of ports can be dynamically assigned as access ports or uplink ports to aggregation switches. The flexibility of the port configuration allows for dynamic design of clusters, enabling a converged design of compute and storage servers in the same enclosure.

40 GbE Ports	12	10	8	6	4	2	0
10 GbE Ports	0	8	16	24	32	40	48

Table 1. *SX1024 Flexible Port Configurations*

Since the SX1012 switch supports data transport at various speeds (1GbE, 10GbE, 40GbE, or 56GbE) simultaneously, it can operate in a rack that includes servers and storage nodes from different generations, making server upgrade possible without the need to change the ToR switch.

Enabler of Converged Infrastructure

One example of a small converged cluster is a rack containing 8 servers and 4 storage systems that require a non-blocking connection to the aggregation network. With SX1012, this requirement can easily be fulfilled by assigning eight 10GbE ports to the servers, four 40GbE ports to the storage, and six 40GbE ports to the uplinks to the aggregation switch.
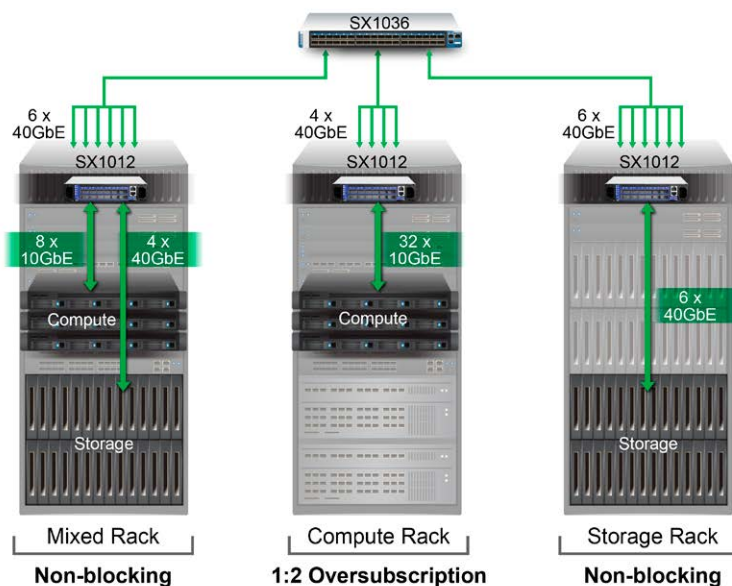

Figure 3. *SX1012 Flexibility and Low Cost Mixed Racks of Servers and Storage*

56GbE Uplink Ports – 40% Higher Link Capacity

Another possible cluster configuration using SX1012 is a rack consisting of 32 10GbE servers that need to be connected to the aggregation network. Opting for an oversubscription ratio of 2:1, four 40GbE ports can be assigned as aggregation ports.

Or there could be a cluster in which a storage rack with six 40GbE storage units is connected to the network. Since it is a storage network, non-blocking connectivity is the best option, which mandates a minimum of 240GbE uplink capacity. SX1012 can easily satisfy this requirement by assigning six 40GbE ports as downlink and six 40GbE ports as uplinks.

These examples demonstrate the flexibility of SX1012 and its cost effectiveness as an interconnect switch solution for small clusters.

The SX1012 switch platform has thus far been considered with its uplink ports running 40GbE. There is, however, an advantage in utilizing the 56GbE capability of its ports as the uplink speed to the data center aggregation. When operating at 56GbE, the 40% higher bandwidth can be used to further reduce oversubscription or to free more ports for server connectivity. Activation of the 56GbE mode is enabled via a license key, with no change of equipment or of the software running on it.

If in the previous example of four 40GbE uplink ports the SX1012 is configured to operate instead at 56GbE, the oversubscription ratio for connecting the 32 10GbE ports (servers) is easily reduced from a 2:1 ratio to 1.42:1.

High Availability Solution in a Standard 1RU

The SX1012 encompasses a unique design to accommodate high availability cluster solutions. HA clusters require redundant links from each server and storage unit connected to resilient switches. A common way to provide server resilience is to deploy servers with two Network Interface Cards (NICs) or with a NIC with two interfaces, and to connect them to two different ToR switches. This requirement usually requires two full-size 1RU switches, occupying 2RU in a standard 19" rack. However, the small form factor of SX1012 allows for the placement of two SX1012 switches side-by-side that occupy 1RU height, which makes SX1012 the perfect choice for high availability solutions.

A resilient rack design with up to 32 servers can easily be fulfilled by SX1012, by mounting two SX1012 switches side-by-side in the same rack unit, connecting each of the two 10GbE server NICs to switches. The remaining four 40GbE ports of the SX1012 can be used as uplinks to the aggregation switches, providing an oversubscription factor of 2:1.
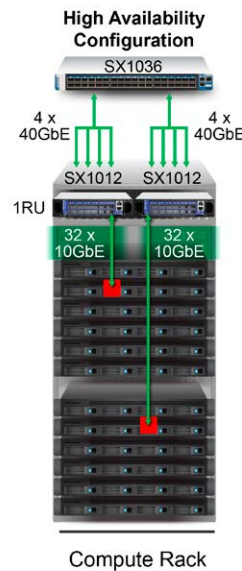

Figure 4. SX1012 as High Availability Rack Switch

Operational Benefits from Full Front Access

Apart from its smaller form factor, SX1012 comes fully equipped with the hardware and software capabilities that exist in other Mellanox SX10xx series switches, running MLNX-OS™ management software that provides full L2 and L3 functionality; this enables full flexibility when designing the network infrastructure.

To ease operation and maintenance, the switch comes with full front access of both network and management interfaces, as well as redundant power outlets and redundant fan units that give yet another level of high availability within the switch.

Actual Deployments Using SX1012

Database solutions need to deliver a high level of availability and scalability, giving the user the ability to scale out the database over a number of servers in active-active configuration.

Applications such as DB2 pureScale and Oracle RAC require high bandwidth and low latency to and from the caching facility, the disk storage system, and all members of the DB2/RAC. They also require that the members/nodes be connected to the application servers using it.

Improving throughput and resilience in the event of Ethernet network adapter failures is handled through Ethernet bonding. Ethernet bonding (or channel bonding) is a setup in which two or more network interfaces are combined. Similarly, redundant SAN interface cards can be used. The SX1012 complements database and storage solutions by providing the required port count of both 10GbE and 40GbE ports and fitting into the stringent requirement of rack space while still providing the highest level of resiliency.

Summary – Selecting the Right ToR Switch

Several ToR switch properties must be considered when installing a new rack. Newer servers offer faster connectivity for higher bandwidth requirements, so the ToR switch should be able to handle both 10GbE and 40GbE (and higher) bandwidth. It should furthermore be able to connect to additional servers over time, future-proofing the rack by encouraging scalability. In many cases, resiliency is also an important property for the rack's servers.

Mellanox's SX1012 Ethernet switch is the ideal ToR switch for small clusters by offering flexibility, scalability, and future-proofing in the smallest form factor in the industry. SX1012 is based on the Mellanox SwitchX®-2 IC, which provides market-leading 230-nanosecond latency and unmatchable power efficiency.

Through a simple change of license, the SX1012 switch system is ready for adding InfiniBand functionality, and enables a Virtual Protocol Interconnect (VPI) gateway.

The SX1012 switch system is also ready for Mellanox Open Ethernet™, an initiative that enables customers to terminate the dependency on traditional closed-code Ethernet switches. With Open Ethernet, customers gain control over the switch and turn it into a flexible and customizable platform, allowing for optimized utilization, efficiency and higher overall return on investment. This customization can allow, for example, improved behavior of specific scenarios or shorter turnaround time to implement new specifications.

Mellanox 12-Port 40/56Gb Ethernet Switches – Solution Components

Ordering Part Number	Description
Switch Systems	
MSX1012B-1BFS	SwitchX®-2 based 12-port QSFP+ 40GbE, 1U Ethernet switch, 1 power supply, short depth, PSU-side to connector-side airflow, RoHS-6 Rail Kit must be purchased separately
MSX1012B-2BFS	SwitchX®-2 based 12-port QSFP+ 40GbE, 1U Ethernet switch, 2 power supplies, short depth, PSU-side to connector-side airflow, RoHS-6 Rail Kit must be purchased separately
MSX1012B-2BRS	SwitchX®-2 based 12-port QSFP+ 40GbE, 1U Ethernet switch, 2 power supplies, short depth, connector-side to PSU-side airflow, ROHS-6 Rail Kit must be purchased separately
MSX60-DKIT	Rack installation kit for SX6005/SX6012 and SX1012 series short depth 1U switches, allows installation of one or two switches side-by-side into standard depth racks
MC2609130-003	Mellanox passive copper hybrid cable, ETH 40GbE to 4x10GbE, QSFP to 4xSFP+, 3m
MC2210128-003	Mellanox passive copper cable, ETH 40GbE, 40Gb/s QSFP, 3m

Mellanox Ethernet Switch Portfolio

Switch	Application
SX1036	Ideal ToR/Core switch providing 36 40/56GbE or 64 10GbE from server to aggregation
SX1024	Ideal non-blocking ToR switch providing 48 10GbE from server and 12 40/56GbE to aggregation layer
SX1012	Ideal ToR/pod switch for small deployments providing 12 40/56GbE or 48 10GbE from server to aggregation