Mellanox TECHNOLOGIES

Leading European Web 2.0 Company Offers Advanced Capabilities by Upgrading to Mellanox Interconnect

Background

When one of the leading European Web 2.0 companies decided it needed to upgrade its network to better support its Hadoop application, it turned to Mellanox to create a high performance cluster using RDMA over Converged Ethernet (RoCE).

The Hadoop application provides the company with distributed pattern-based searching and sorting of data. Advanced solutions such as this require reliable, consistent, high-performing interconnect to support both the volume and lightning-fast analysis required to maintain the company's core business.

Prior to working with Mellanox, a 1GbE network was employed that was insufficient to handle the company's interconnect needs, leading to dropped packets and serious performance issues. The switches rarely enabled the company to reach line rate despite a non-blocking architecture. After considering an upgrade of its hardware to 10GbE switches from the same supplier, the company approached Mellanox to discuss alternatives.

Mellanox offered a proof of concept that proved to be so successful that it led to a deal in which Mellanox provides its interconnect solution for the Hadoop application.

The Solution

Mellanox offered a RoCE version 1.5 cluster with an additional IP encapsulation. The additional header allows for routing to be handled via IP without affecting the RoCE structure. This allows the company's routed network to use RoCE traffic.

Although RDMA traffic traditionally require the use of flow control mechanisms, the Mellanox implementation of the network fabric, due to its unique ASIC design, was able to handle the projected traffic pattern without any traffic flow mechanism enabled on the switch

To accomplish this, Mellanox provided SX1024 Top-of-Rack switches, using 10 uplinks of 56GbE from each switch, as well as SX1036 aggregation switches. With 40 servers per rack, each ToR switch used 40 x 10GbE for the downlinks.

RoCE

OVERVIEW

Mellanox provided a RoCE v1.5 cluster that allows for routing via IP without any traffic flow mechanism enabled on the switch, significantly improving network performance and achieving full line rate bandwidth across the board.
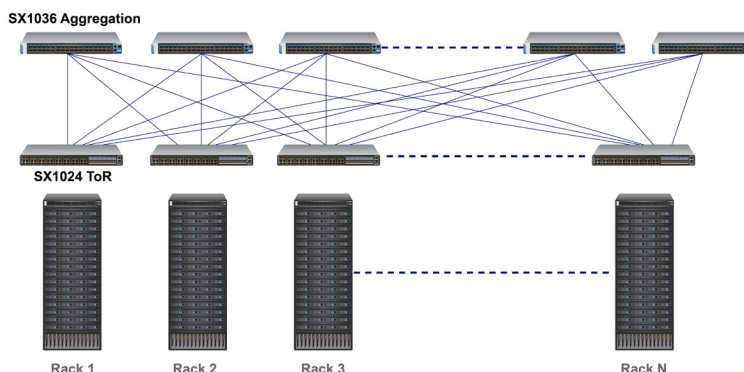

Figure 1. *Mellanox RoCE Cluster*

Mellanox's SX1036, offering the world's highest density by running 36 ports at 56GbE, assisted in reducing the number of switches required to form a non-blocking ~1,000-node fabric, while Mellanox's SX1024 enabled non-blocking connectivity at the Top-of-Rack level.

Conclusion

The Web 2.0 company has seen a significantly better performing network since the Mellanox solution was deployed. The network now provides full line rate bandwidth of 10GbE across the board. Such efficiency could only be achieved using RDMA. With its previous setup, the company could not reach a peak nor a sustainable line rate of over 2G.

With its new Mellanox interconnect, the internet company is now able to promote its high performance services, including a more advanced data analytics capability, thereby differentiating itself from its competitors.

Moreover, Mellanox has deployed an architecture for its customer that is future-proof. Already having improved the network's bandwidth from 1 GbE to 10GbE, there is additional room to increase the network size by 30% (by adding aggregation switches) without changing the topology of the cluster.
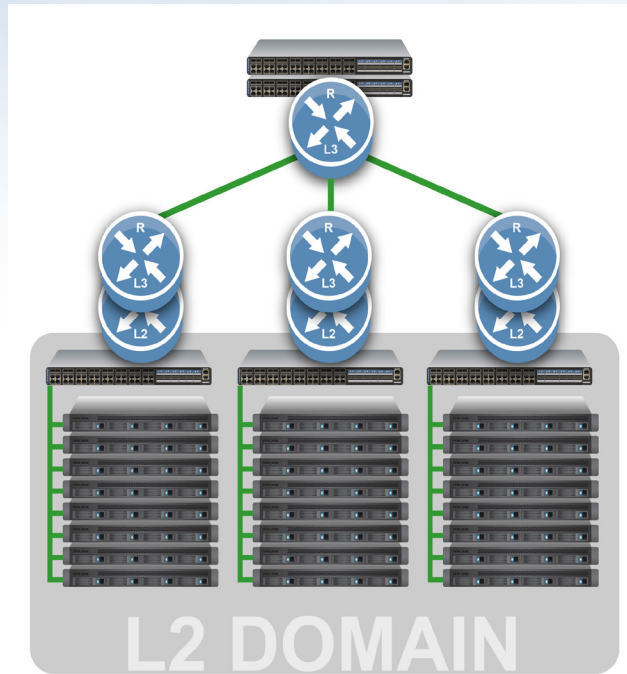

Figure 2. RoCE v1.5

Furthermore, the Hadoop application has become a model for other small clusters within the company, such that Mellanox solutions offer room for growth and continued improvement of the interconnect architecture.

15-3552CS
Rev 1.0