WHITE PAPER

March 2015

CloudX Hyper-Converged Infrastructure for OpenStack using EMC ScaleIO

Introduction

Mellanox offers a complete 10/25/40/50/56/100 Gb/s InfiniBand and Ethernet adapter, switch, cable product line which together deliver an Efficient Virtual Network (EVN)[1]. The Mellanox CloudX™ OpenStack reference architecture takes advantage of the integration of the key features of the Efficient Virtual Network drivers into the OpenStack cloud management framework. While OpenStack provides a flexible framework for cloud integration and management, there are certain limitations to the classical architecture which can be overcome by using advanced scale-out software defined storage technologies. This whitepaper will demonstrate how the EMC Scale-IO Elastic Converged Storage can be integrated within the OpenStack framework to build both converged and hyper-converged infrastructure. This CloudX reference architecture based on the integration of OpenStack and ScaleIO provides a converged infrastructure platform to build the most efficient and scalable private, public and hybrid clouds possible.

ScaleIO Elastic Converged Storage for Converged Infrastructure

EMC ScaleIO Elastic Converged Storage is a software defined storage solution which allows for the convergence of server, storage, and application. ScaleIO enables scale-out "Server-SAN" storage that creates server based, shared networked storage from local attached storage, and eliminates the requirement for a dedicated Storage Area Network (SAN).

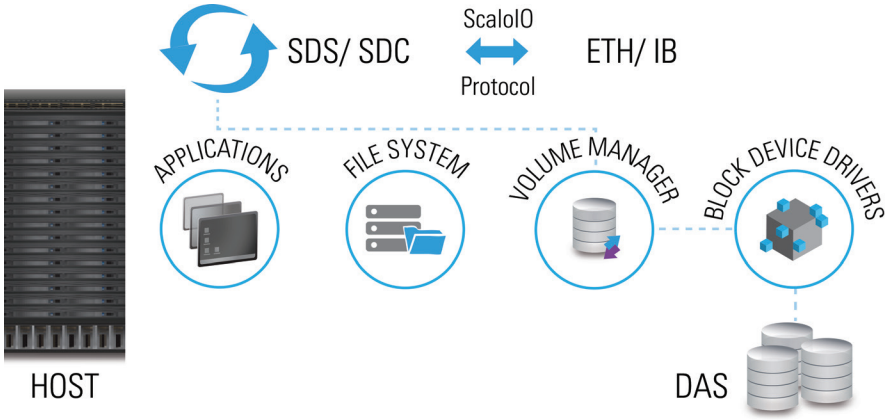

Figure 1. *ScaleIO Architecture*

The Key components that make up a ScaleIO platform include:

- SDC (ScaleIO Data Client)
  - Installed on an application server node
  - Exposes ScaleIO volume to application
- SDS (ScaleIO Data Server)
  - Manages local storage
  - Fulfills IO requests from SDC in a cluster (local or remote)

[1] For more details read the whitepaper: Efficient Virtual Networks: The Key to building an Efficient Cloud

- MDM (ScaleIO Meta Data Manager)
  - Tracks storage and data mapping
  - Monitors capacity, performance and load balancing
  - Makes data migration decisions
- Scale IO Gateway
  - Required to bootstrap SDS installation

Traditional OpenStack Architecture with Cinder Storage

The integration of the Mellanox components into the classical OpenStack architecture is shown in the diagram below. A key component of the integration is the Cinder storage servers which integrates Mellanox drivers to provide iSCSI over RDMA capabilities that improve block storage performance by up to 6X when compared to conventional iSCSI performance.

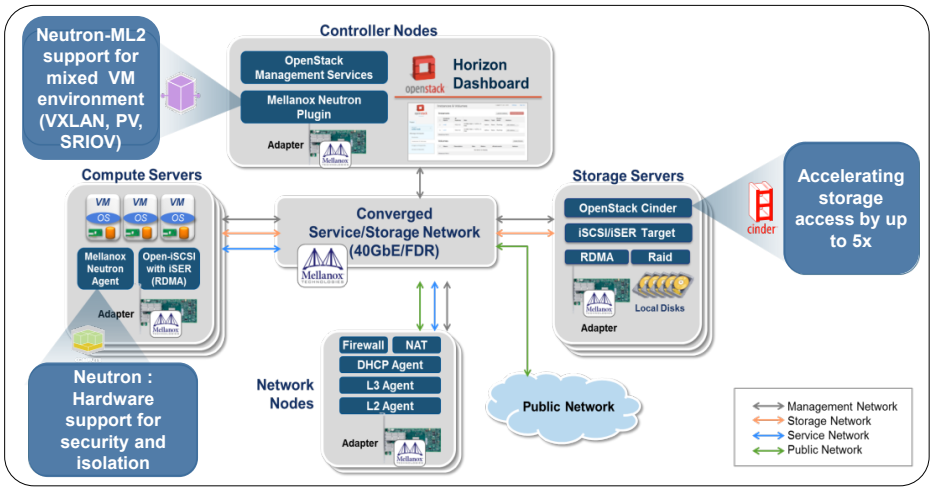

Figure 2. *Classical OpenStack Integration with Mellanox Plugins*

While the performance gains of Mellanox RDMA are impressive, the storage architecture based on Cinder has certain limitations that will be discussed further below.

Converged Infrastructure Integration of OpenStack and ScaleIO

The most basic OpenStack integration simply has the ScaleIO storage nodes operate within the Cinder framework to deliver block level storage. This integration is shown in Figure 2 and has been described in this presentation[2].

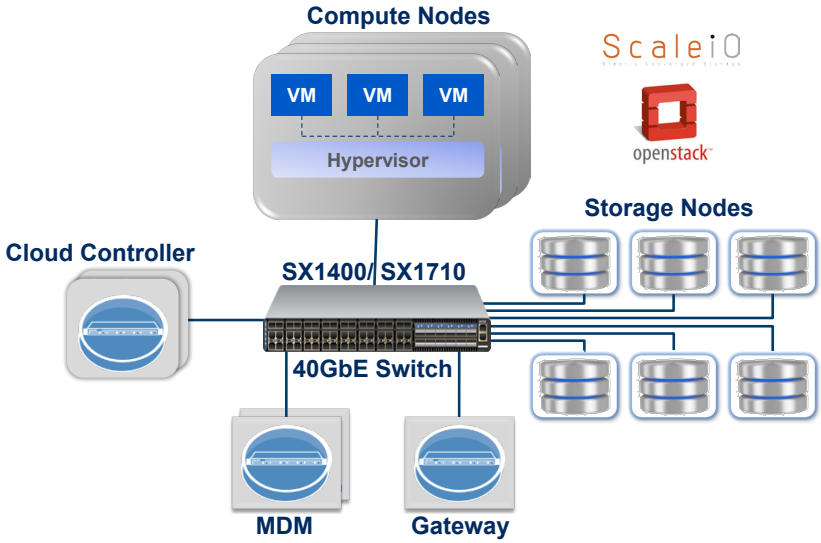

Figure 3. *Basic ScaleIO OpenStack Integration*

[2] https://www.openstack.org/assets/presentation-media/Laying-Cinder-Blocks-v4.ppt

To accomplish this, the integration requires that the SDS be installed on each Cinder node and the SDC on each Nova compute node. In addition separate nodes are required for both the MDM and Gateway nodes. In this way the ScaleIO block storage protocol is incorporated into the Cinder framework.

The principal advantage of this integration is its simplicity and the elasticity and scalability offered by ScaleIO storage. Its simplicity does however have some drawbacks as the direct mapping carries along some unfortunate limitations of the Cinder architecture. These include the requirement to have dedicated storage nodes and 1:1 mapping to Cinder storage controller.

Hyper-Converged Infrastructure Integration of OpenStack and ScaleIO with CloudX

Fortunately a more sophisticated integration is possible that overcomes these limitations and further improves the architecture making it hyper-converge infrastructure. As can be seen with this integration, the separate storage nodes have been eliminated and instead storage has been moved to the compute nodes. This is accomplished by combining the SDS and SDC instances with the Nova compute nodes – thereby converging compute and storage nodes. This hyper-converged infrastructure implements what is frequently called 'Server-SAN' storage and allows the locally attached storage in each server node to be treated as a giant shared storage pool across the entire cluster. Furthermore it is straightforward to instantiate an application such as VDI, MySQL/LAMP stack, or MongoDB on the compute nodes thereby forming a completely hyper-converged platform.

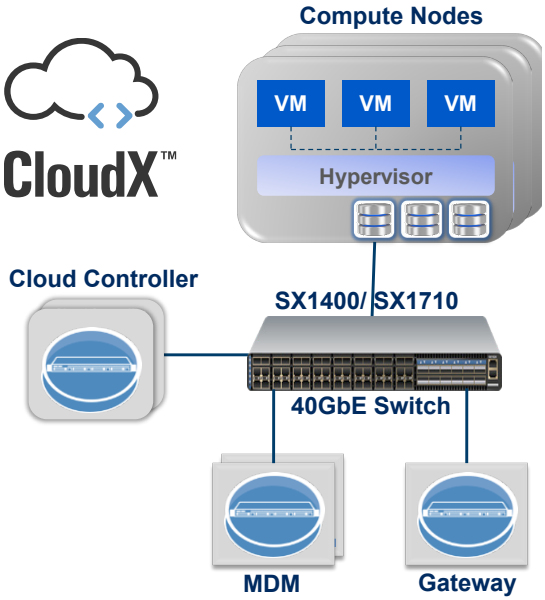

Figure 4. *CloudX Hyper-Converged Infrastructure ScaleIO-OpenStack Integration*

	Configuration
Server Machines	Supermicro SuperServer SYS-F627R2-F73 with: • Ivy Bridge 8-core E5-2640 • 64GB DDR3 RAM • 4x Micron M500DC -800GB
Network Fabric	Mellanox SX1400/ SX1710 Switch
	Mellanox ConnectX-3 Pro VPI Adapter
OpenStack Distribution	Mirantis Fuel 5.1
Storage	EMC ScaleIO 1.31.1

Table 1. *CloudX Hyper-Converged ScaleIO Configuration*

Table 1 describes the configuration of this hyper-converged solution. This hyper-converged platform eliminates the additional cost of separate storage nodes as well as the complexity and cost of a dedicated storage area network. The Mellanox interconnect is the key to making this CloudX hyper-converged architecture feasible as it offers ultra-high performance networks, low-latency connections, and traffic isolation.

This hyper-converged platform eliminates , the additional cost of separate storage nodes as well as the complexity and cost of a dedicated storage area network. The Mellanox interconnect is the key to making this CloudX hyper-converged architecture feasible as it offers ultra-high performance networks, low-latency connections, and traffic isolation.

Streamlining the Architecture with Additional Integration

Streamlining the architecture even further is made possible by the Mellanox SX1400/ SX1710 of high performance Virtualized 10/40/56 Gb/s Ethernet switches. In addition to offering full L2/L3 switching, routing, and data center bridging capabilities; these switches support a fully virtualized KVM control plane that allow VM's and applications to be hosted on the switch.

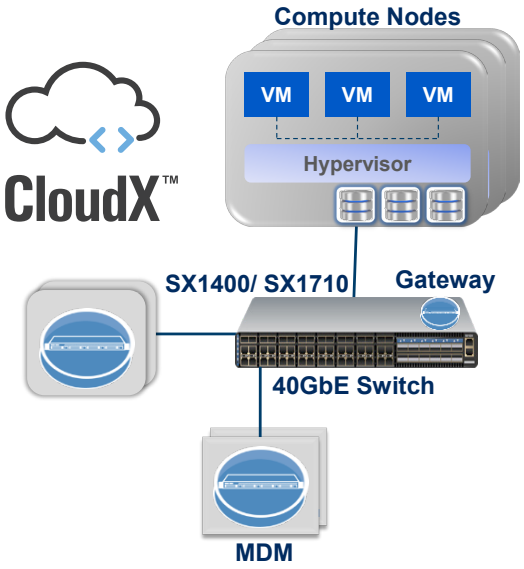

Figure 5. CloudX Integration of ScaleIO Gateway App onto Switch VM

In this case the ScaleIO gateway node is eliminated, and instead pre-installed as an application running in a Windows VM on the switch. This not only simplifies the configuration but as can be seen in Figure 5, the integration also eliminates a costly dedicated server gateway node required for the installation and initialization of the other ScaleIO components.

Summary

In summary, we have shown in this innovative CloudX reference architecture that, with the help of Mellanox's Efficient Virtual Interconnect, we could use the elasticity and efficiency offered by EMC ScaleIO software only storage in OpenStack, as a converged or a hyper-converged infrastructure. Our reference architecture, allows not just building hyper-converged appliance for OpenStack using off the shelf components, but also helps in eliminating some of the limitations, otherwise present in the Cinder architecture. We further streamline this integration, by preinstalling the ScaleIO Gateway on Mellanox SX1400/ SX1710, thus eliminating the need for dedicated servers.

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