

QoS for Virtual Machines

In a 100GbE NIC and Open Switch Environment

High VM-Density Driving an Aggregation of I/O

Cloud service providers and large enterprises operate server farms with ultra-high VM-density. According to a CSP survey by Infonetics, the average number of VMs per server was 42 in 2015, growing to 98 in 2017. New, powerful server processors from Intel support even more VMs and IO.

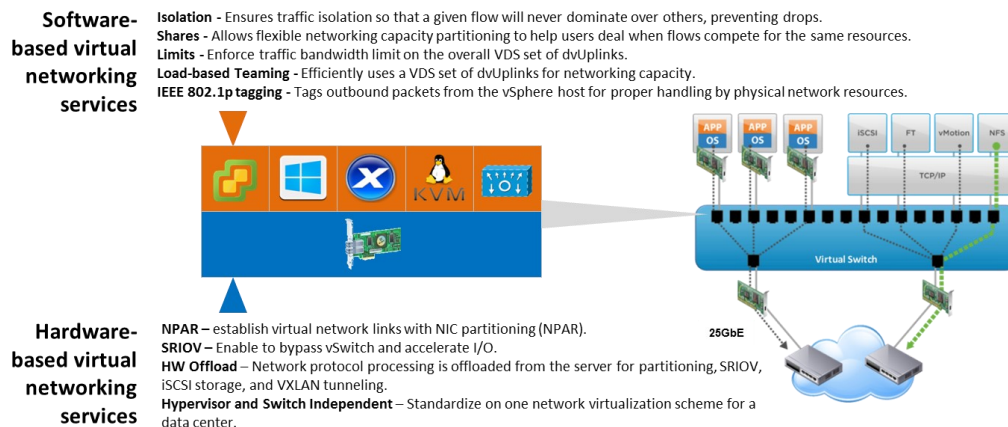
QoS a Must with 100GbE Connectivity to a Virtualized Server

CSPs and enterprise server architects discovered high VM-density leads to an aggregation of I/O, I/O bottlenecks, and that non-critical apps can be noisy neighbors which disrupt traffic to critical apps. Deploying 25, 50 or 100GbE NICs provides the bandwidth needed to eliminate an overall I/O bottleneck, but a QoS policy is necessary to partition the bandwidth to different VMs to ensure performance is guaranteed to specific business critical VMS and applications.

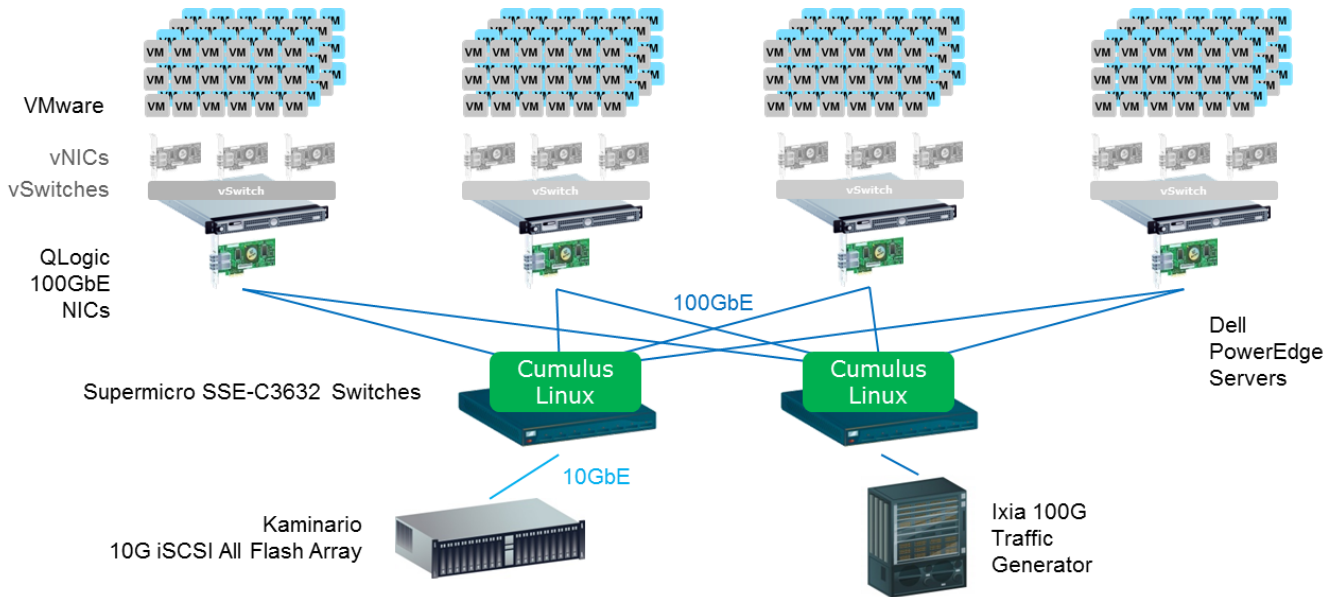
Deploying QoS in a 100GbE Environment

The new generation of high-bandwidth 100GbE NICs with extensive HW offloads has led to a new best practice of deploying NIC-based partitions to enable QoS, then shaping virtual networks by making use of granular controls for ports and resource pools can further increase application performance and availability. The result is maximum use of HW offloads for more efficient use of server CPU, and guaranteed bandwidth to critical applications.

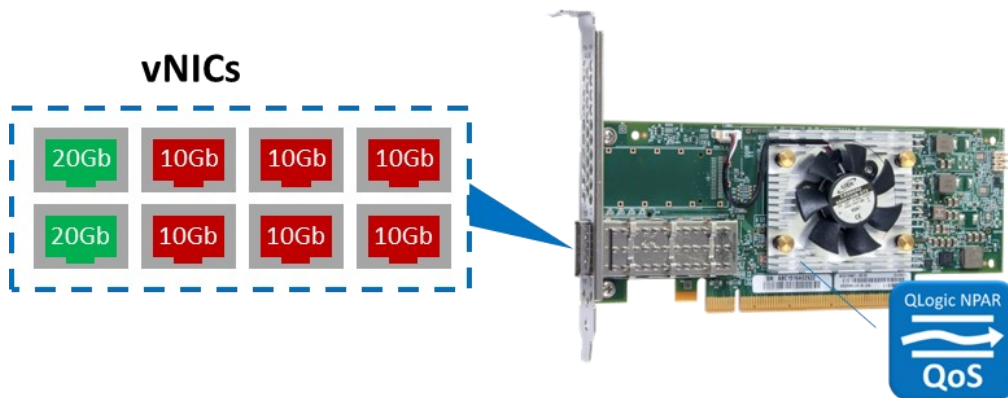
Architecture for Layering SW-based Services on HW-based Virtual Networking Services



QoS in a 100GbE Environment Reference Architecture



The 100GbE NICS in each physical server were partitioned using QLogic NPAR hardware partitioning. Each physical NIC was partitioned into eight vNICs. Two VMs which represented critical business apps were each allocated their own 20GbE port, while the remaining groups of VMs shared 10GbE vNICs. Load balancing and teaming was layered on the hardware-based partitions using VMware ESX Network I/O Control (NIOC). Listed below are the components used in the test configuration.



HW-based partitioning allows OS-independent and switch-independent isolation (per partition), sharing of the same physical port over separate partitions, and limits guaranteed minimum and maximum bandwidth QoS at the adapter hardware level.

Products Used

The products listed below were used in the reference architecture configured in a 2550100 Solutions Lab.

Qty	Product	Model	Description
4	Operating System	Windows Server 2012	Windows Server 2012 R2 is an operating platform which can run the largest workloads with support for up to 64 processors VHDX virtual hard disks up to 64 terabytes .
4	Hypervisor	VMware vSphere 6.0	VMware vSphere v6 features support for per-VM Distributed vSwitch bandwidth reservations to guarantee isolation and enforce limits on bandwidth. V6 also allows vMotion traffic a dedicated networking stack.
4	Servers	Dell PowerEdge 630	PowerEdge R630 support the latest Intel Xeon E5-2600 v4 processors; up to 22 cores; up to 24 DIMMs high-capacity DDR4 memory; Up to 24 1.8” SSDs (23TB); up to 3 PCIe 3.0 expansion slots; and up to 4 Express Flash NVMe PCIe SSDs.
1	Storage	Kaminario All Flash Array	The K2 array is comprised of K-blocks — building blocks that include Active-Active controllers, one or more drive shelves, and connectivity for scaling out. The K2 platform scales-out to two, three, and four K-block configurations. Even when scaling out the cluster remains fully N-ways Active-Active.
2	Switches	Supermicro SSE-C3632	Offering thirty-two QSFP28 Ethernet ports, SSE-C3632S switch enables robust layer-3 IP fabric for flexible layer-2 overlay in Ethernet fabric architecture.
2	Switch OS	Cumulus Linux	Cumulus Linux is a networking-focused Linux distribution. Switches running Cumulus Linux provide standard networking functions such as bridging, routing, VLANs, MLAGs, IPv4/IPv6, OSPF/BGP, access control, VRF, and VxLAN ...
4	Adapters	QLogic QL45000 Series	The QL45611 Series adapters supports a speed of 100Gbps. FastLinQ 45000 Series Controllers enable SR-IOV, RoCE, iSCSI, FCoE, and DCB. They also support PCIe Gen 3.0, along with embedded virtual bridging and other switching technologies for high-performance DMA and VM-to-VM switching.
20	Optical Transceivers	Finisar 100GBASE-SR4 QSFP28	Finisar’s FTLC9551REPM 100G QSFP28 transceiver modules are designed for use in 100 Gigabit Ethernet links over multimode fiber.
1	Traffic Generator	Xcellon-Multis QSFP28 Enhanced Load Module	Xcellon-Multis provides the world’s first 100/50/25GbE multi-rate test system to satisfy equipment maker test needs ranging from basic interoperability and functional test, to high-port count performance tests. As organizations implement this same high-density, high bandwidth networking and network computing equipment in their own networks, they need this same test solution to verify performance and functionality prior to deployment.



FINISAR ixia

kaminario.



QLOGIC

TELEDYNE LECROY
Everywhereyoulook