EMULEX
An Avago Technologies Company

SOLUTIONS BRIEF

Microsoft Partner
Gold OEM

Lenovo Accelerates File Storage I/O by up to 82 Percent

Increase application performance with the Emulex Virtual Fabric Adapter 5 (VFA5) with support for SMB Direct and RDMA over Converged Ethernet (RoCE)

At a glance

Enterprises using Windows Server 2012 R2 can increase their company's competitiveness by leveraging Microsoft SMB Direct with an adapter that supports RoCE, also known as an R-NIC (Network Interface Card). By offloading RDMA processing from the CPU, the R-NIC makes valuable CPU resources available to improve the performance of transaction intensive and performance hungry applications. In addition, by leveraging the low latency capability and the reliable lossless 10Gb Ethernet (10GbE) infrastructure of the R-NIC, IT organizations can provide optimal application responsiveness for increased business revenue. The Emulex Virtual Fabric Adapter 5 (VFA5), based on the OneConnect® OCe14000 Ethernet Network Adapters with RoCE, enable server-to-server data movement directly between application memory without any CPU involvement, providing high throughput and data acceleration on a standard Ethernet fabric without the need for any specialized infrastructure or management in a Microsoft SMB Direct environment.

Key benefits

- Accelerate application file storage I/O by up to 82 percent with RoCE[1]
- Reduce I/O response time by up to 70 percent with SMB Direct[1]
- Increase server electrical power efficiency by up to 80 percent with RoCE[1]
- Up to 750 additional IOPS per server watt with SMB Direct and RoCE[1]

Products

Emulex Virtual Fabric Adapter 5 (VFA5) for System x

Background

The benefits of the Emulex Virtual Fabric Adapter 5 (VFA5) that supports RoCE are easy to understand, but the pieces that come together to provide those benefits can get confusing. Here is some brief background and information on RDMA, SMB Direct and RoCE.

RDMA

RDMA is the remote memory management capability that is fundamentally an accelerated I/O delivery mechanism. It introduces the concept of "zero-copy" data placement, which allows specially designed R-NICs on both ends of a transaction to transfer data directly from the user memory of the source server to the user memory of the destination server, bypassing the operating system (OS) kernel (see Figure 1).

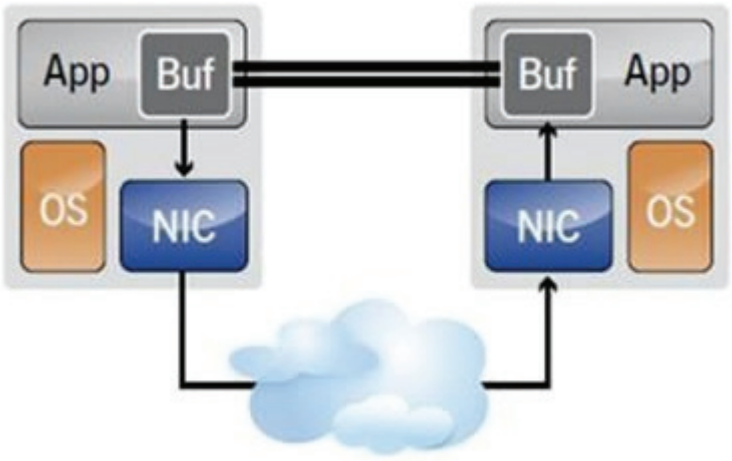


Figure 1. RDMA bypasses the OS kernel.

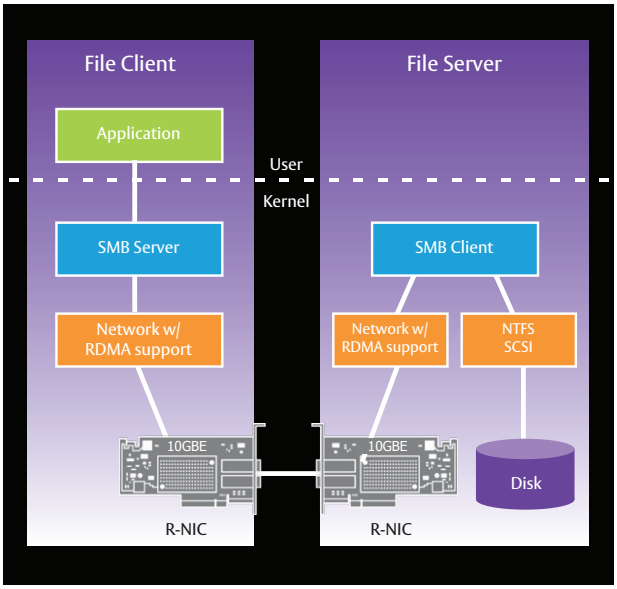
Network adapters that have RDMA can function at full speed with very low latency, while using very little CPU. Since the RDMA data transfer is performed by the Direct Memory Access (DMA) engine on the R-NIC, the CPU is not used for the memory movement, freeing it to perform other tasks, such as hosting more virtual workloads.

[1] IT Brand Pulse, "Blade Server I/O and Workloads of the Future," November 2014

Lenovo Accelerates File Storage I/O by up to 82 Percent

SMB Direct (also called SMB over RDMA)

Microsoft has added a feature to Windows Server 2012 R2 called SMB Direct which allows data transfer from one NIC to the other without the intervention of the CPU on each side, provided the NIC has RDMA support (see Figure 2). With a 10 Gb per second (Gbps) NIC, a lot of CPU resources are used to break all the data down to transfer it over the wire. As shown in Figure 1, RDMA allows the NIC to handle most of that work.



Source—Microsoft blog, Hyper-V over SMB 3.0 in Windows Server 2012, Dec. 2012

Figure 2. Basic SMB Direct architecture.

SMB Direct, when coupled with network adapters that have RDMA capability, delivers increased throughput, lower latency (faster file storage I/O) and lower CPU utilization (faster application performance).

RDMA over Converged Ethernet (RoCE)

RoCE is a networking protocol and standard that enables RDMA to transfer data efficiently with very low latencies on a lossless Ethernet network. It is basically RMDA technology running in a Converged Ethernet environment, but it also directly addresses two key limitations of current compute and networking architectures:

1. Overhead created by data copies between user (application) and kernel memories
2. Latencies introduced by the Transmission Control Protocol/ Internet Protocol (TCP/IP)

The first RoCE specification released in 2010 by the InfiniBand Trade Association (IBTA) included the following benefits:

- Low latency and CPU overhead (eliminated the multiple data copies inside the server)
- High network utilization
- Support for message passing, sockets and storage protocols
- Supported by all major operating systems

IBTA recently released RoCEv2 which enables routing across Layer 3 networks, providing better isolation and enabling very large data center deployments to transmit more data in less time. RoCEv2 addresses the needs of today's evolving data centers which require more efficient data movement over a variety of network topologies.

Lenovo Accelerates File Storage I/O by up to 82 Percent

SMB Direct and RoCE by Lenovo

The VFA5 with RoCE supports SMB Direct, delivering greater application performance and power savings for enterprises using Windows Server 2012 R2. Isolated network testing by IT Brand Pulse reveals that Emulex VFA5 with RoCE excel over previous generations of adapters without it, as well as competitor adapters, as shown in Table 1.

Performance	vs. previous Emulex VFA5 10GbE adapters	vs. 10Gbps Intel X520
Application file storage I/O	Up to 82 percent acceleration	55% greater 4K IOPS
I/O response time	Up to 70 percent reduction	2x read/write data
Server electrical power efficiency	Up to 80 percent increase	N/A
Additional IOPS per server watt	Up to 750 additional IOPS	Up to 750 additional IOPS

Table 1. Performance increases due to RoCE support.

The VFA5 with RoCE also deliver advantages only available in a Converged Ethernet environment, giving Microsoft enterprises even more return on investment, without requiring extra management or special infrastructure. RoCE uses Data Center Bridging (DCB) enhancements to Ethernet which results in a lossless traffic class for RoCE:

- **Priority Flow Control (PFC) for lossless transmission**—Heavily utilized networks normally drop packets which have a "snowball" effect that degrades performance due to TCP retransmissions. RoCE uses PFC which guarantees packet delivery and is not affected by external network congestion.
- **Enhanced Transmission Selection for class of services**—RoCE traffic can be assigned a minimum bandwidth level which can be adjusted. Quality of service (QoS) parameters can be assigned to the RoCE traffic class at the Link Layer (Layer 2) which allows for the precise tuning of network utilization.
- **802.1Q Congestion Notification (QCN) for congestion avoidance**—SMB Direct bypasses potential congestion points in host driver/memory in both initiators and target ports.

Conclusion

Enterprises with Microsoft Windows Server 2012 R2 can extend the benefits of SMB Direct by using I/O connectivity that supports RoCE. The Emulex VFA5 with RoCE includes SMB Direct support and use advances in lossless Ethernet (DCB) to accelerate file storage I/O for faster application performance, higher QoS and greater server power efficiency. With this low cost, low power solution, transaction intensive and performance hungry data center applications, such as financial, content delivery, Web 2.0 and data mining, are accelerated for higher business effectiveness that can translate to increased customer satisfaction and revenue.

EMULEX®
An Avago Technologies Company

For product information, please visit our website at www.emulex.com/lenovo