

Double Your Storage System Efficiency

Enable Higher IOPS while Maximizing CPU Utilization



Executive Summary

As virtualization is now a standard technology in the modern data center, IT managers are now seeking ways to increase efficiency by adopting new architectures and technologies that enable faster data processing and execute more jobs over the same infrastructure, thereby lowering the cost per job. Since CPUs and the storage systems are the two main contributors to infrastructure cost, using fewer CPU cycles and accelerating access to storage are keys toward achieving higher efficiency.

Deploying Microsoft's Storage Spaces with SMB Direct (SMB 3.0 over RDMA) over Dell PowerEdge servers, Dell Networking switch, and Mellanox 10GbE with RoCE (RDMA over Converged Ethernet) enabled NICs, boosts applications' access to storage, and cuts CPU cycles by 50%, which enable more jobs to run over the same equipment, enhancing the overall user experience.

The solution has been developed and tested at Dell's Solutions Center.

Overview

With the ongoing demand to support mobility and real-time analytics of constantly increasing amounts of data, new architectures and technologies must be used, specifically those with smarter usage of expensive CPU cycles and as a replacement of old storage systems that were very efficient in the past, but that have become hard to manage and extremely expensive to scale in modern virtualized environments.

With an average cost of low thousands per CPU, about 50% of compute server cost is due to the CPUs. The I/O controllers, on the other hand, cost is in the lower hundreds. Thus, offloading tasks from the CPU to the I/O controller frees expensive CPU cycles, increasing the overall server efficiency. Other expensive components, such as SSD, will therefore not need to wait the

extra cycles for the CPU. This means that using advanced I/O controllers with offload engines results in a much more balanced system that increases the overall infrastructure efficiency.

RDMA (Remote Direct Memory Access) is one of the most effective offload engines to be integrated into Mellanox's ConnectX[®] I/O controller family. This engine enables significant I/O acceleration by allowing application software to bypass most layers of software and communicate directly with the hardware. Plus, it enables servers to directly place information into the memory of another computer, thereby reducing the application latency and minimizing the CPU overhead.

As for the storage system, the exponential growth of data has forced the industry to replace traditional SAN with higher capacity, easier-to-scale, and lower cost scale-out systems. This new architecture requires lower latency, which can easily be achieved when RDMA is used to access a distributed cache storage (mostly SSD-based). These new scale-out systems, such as Storage Spaces over SMB Direct, provide the same RAID capabilities as traditional SAN¹. Moreover, by using I/O adapters with offload engines such as erasure coding, which is often used instead of traditional RAID because of its ability to reduce the time and overhead required to reconstruct data, these scale-out systems can further increase the performance and lower the cost.

Solution

In order to meet customer needs, Dell and Mellanox worked together on a joint solution that increases CPU utilization and accelerates access to storage. The solution includes standard off-the-shelf components that reduce the Total Cost of Ownership. The solution components are listed in Table 1.

Function	Type
Client Server	2 x Dell PowerEdge R630
Switch	Dell S6000 Networking 10/40GbE Switch
File Servers	2 x Dell PowerEdge R730
NIC	Mellanox ConnectX-3 10GbE with RoCE
Cables	Mellanox LinkX™ SFP+ Copper Cables

Table 1. Solution components

Figure 1 displays the solution which includes two clients using Dell’s PowerEdge R630, each running two VMs, each with 16 vCPU and 16GBytes of RAM. The client servers are connected to two file servers using Dell’s PowerEdge R730, via a Dell Networking S6000 10/40GbE switch, and connected over 10GbE to Dell’s PowerVault MD1200 storage system with 24 SSDs (Solid State Drives). The entire solution uses Mellanox’s ConnectX-3 10GbE with RoCE NICs and Mellanox LinkX SFP+ copper cables.

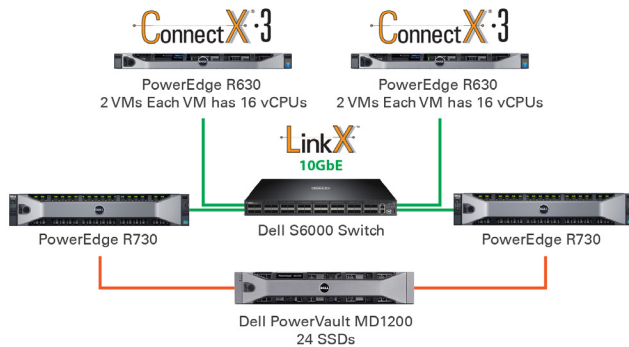


Figure 1. Solution topology

To compare the efficiency boost that RoCE enables, we compared the IOPS when running different workloads over 4KByte and 8Kbyte block sizes. Analysis of the results show that in all workloads, the number of IOPS when running Storage Spaces over SMB Direct was higher than when running over TCP/IP. See Figure 2 for further details.

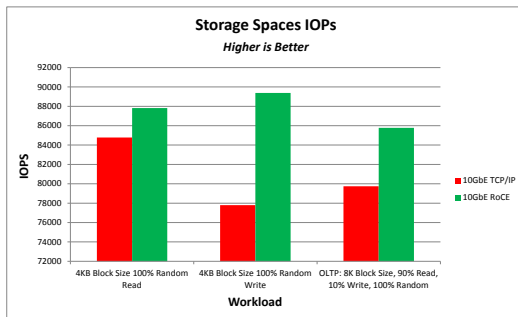


Figure 2. Storage Spaces IOPS over 10GbE TCP/IP vs. 10GbE RoCE

To compare the efficiency in CPU utilization, we ran different workloads using a 512Kbyte block size. The results, displayed in figure 3, show that when running over SMB Direct, Storage Spaces uses less than 50% of the CPU it uses when running over TCP/IP.

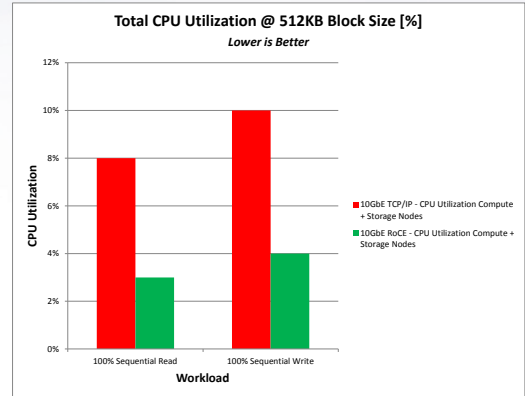


Figure 3. Storage Spaces CPU utilization over 10GbE TCP/IP vs. 10GbE RoCE

Summary

Faster access to storage with lower CPU usage enables more jobs per second to run over the same infrastructure, thereby maximizing the ROI.

Running Microsoft’s Storage Spaces over Dell PowerEdge servers and Mellanox 10GbE RoCE interconnect provides a much better alternative to TCP/IP-based networking by:

- Delivering 15% more IOPS
- Increased CPU efficiency by over 50%
- Reducing power consumption by achieving more jobs per second over the same infrastructure

¹<http://www.microsoft.com/en-us/download/details.aspx?id=42960>



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