October 15-17, 2013

COMPUTERWORLD SNIA

SNW

Intelligent Architecture for the Data-Driven Business

I/O Virtualization
The Next Virtualization Frontier

Dennis Martin, President, Demartek

Demartek®

This presentation is available at http://www.demartek.com/Demartek_Presenting_SNWUSA_2013-10.html

Agenda

- Demartek – About Us
- Why Do We Virtualize?
- I/O Virtualization – What Is It?
- Virtualizing the PCIe Bus
- SR-IOV, MR-IOV and Hybrid
- NPIV and Virtual Fibre Channel
- SR-IOV Performance Results
- References

[View the Demartek Video Library]

Why Do We Virtualize?

- De-couple the logical from the physical
  - Hardware can be split into smaller logical units
  - Hardware can be represented as multiple units
  - Hardware can be combined into larger units
- Want to use computing resources more effectively, especially the under-utilized assets
- Improves deployment time
- Allows expensive resources to be shared, or shared more widely

Examples Available Today

- NIC Teaming or Bonding
  - Logically join two or more NIC ports and present them as one NIC port
  - Can be used for failover or load balancing
  - Often, but not always, specific to one brand of NIC
- NPIV
  - FC HBA that can present multiple virtual ports using the same physical port
  - Can be used in VM environments

I/O Virtualization – What Is It?

- Virtualizing the *I/O path* between a server and an external device
- Can apply to anything that uses an adapter in a server, such as:
  - Ethernet Network Interface Cards (NICs)
  - Disk Controllers (including RAID controllers)
  - Fibre Channel Host Bus Adapters (HBAs)
  - Graphics/Video cards or co-processors
  - SSDs mounted on internal cards

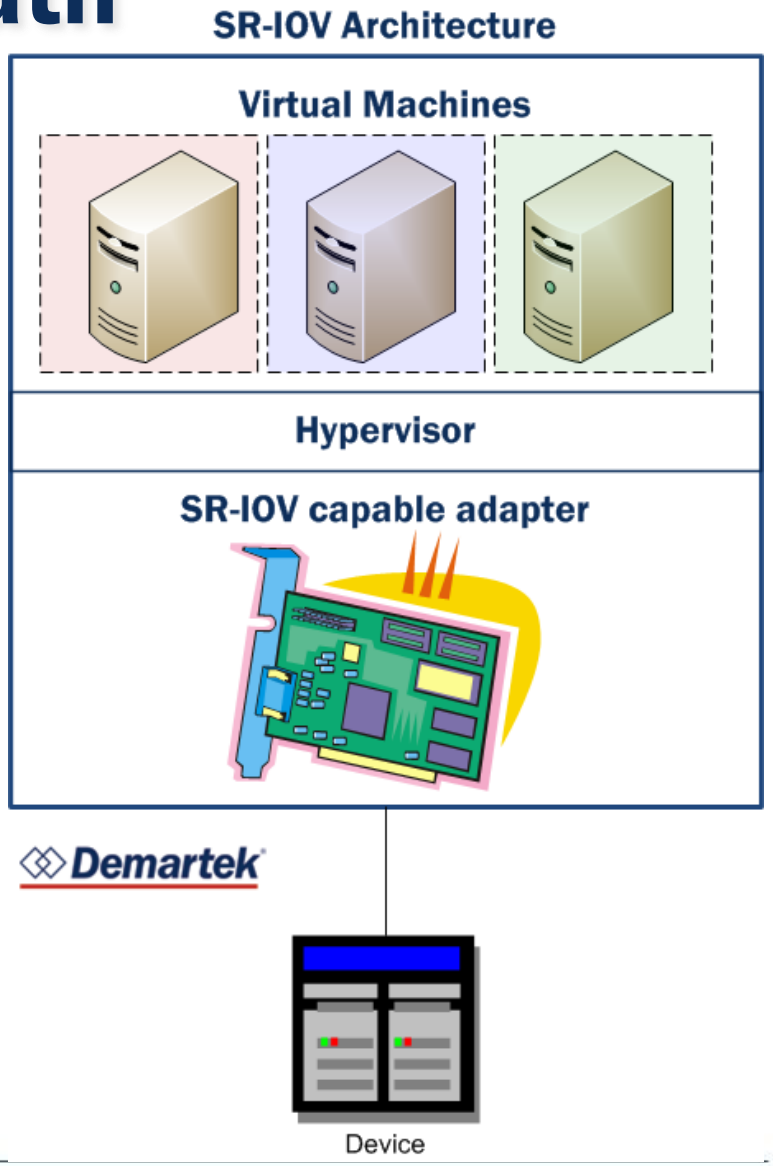
Virtualizing the PCIe Bus

- In June 2008, the PCI-SIG®, the Special Interest Group responsible for PCI Express® (PCIe®) industry-standard I/O technology, announced the completion of the PCI-SIG I/O Virtualization (IOV) suite of specifications
- Works with system virtualization technologies
- Allows multiple operating systems to natively share PCI-Express devices

Virtualizing the I/O Path

Single-Root (SR-IOV)

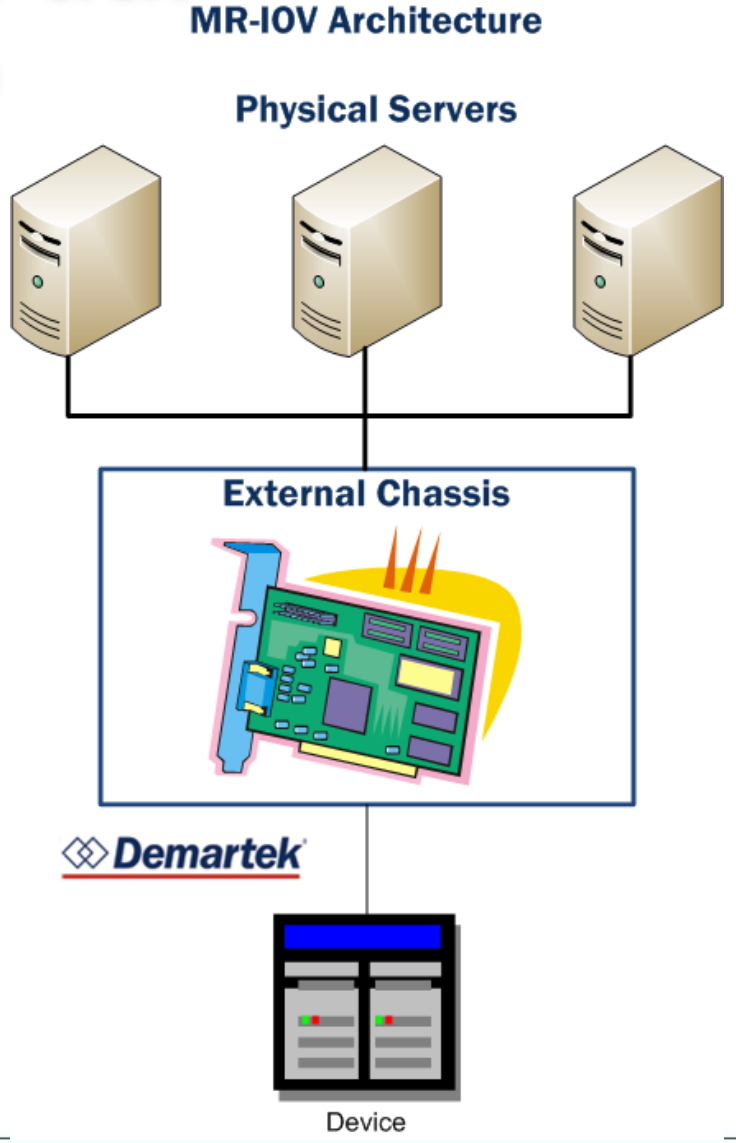
- Multiple VMs sharing one I/O adapter
- Bandwidth of the I/O adapter is shared among the VMs
- Virtual adapters created and managed by SR-IOV adapter (not hypervisor)
- Improved performance for VMs and their apps (near-native)



Virtualizing the I/O Path

Multi-Root (MR-IOV)

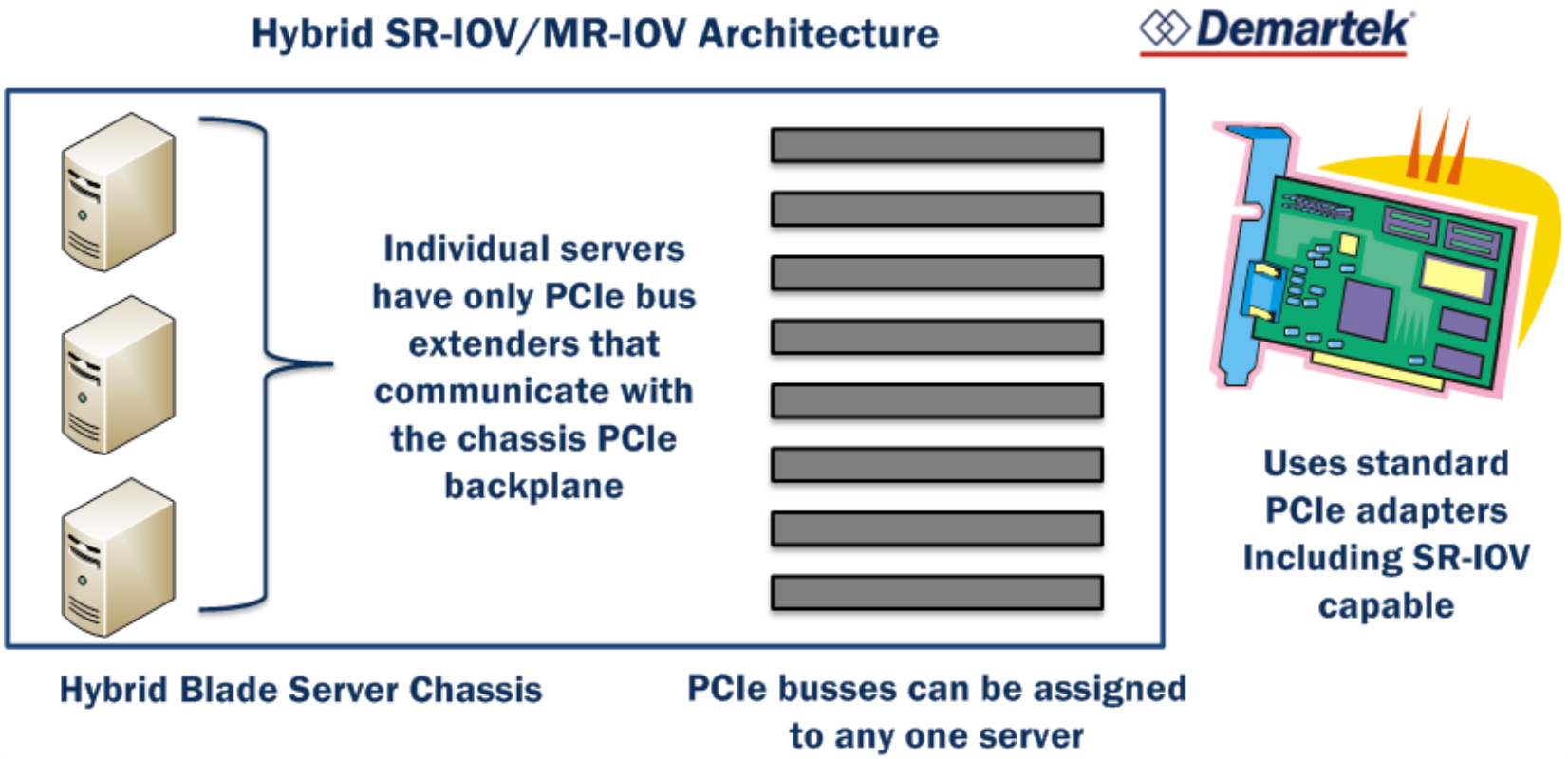
- Multiple servers & VMs sharing one I/O adapter
- Bandwidth of the I/O adapter is shared among the servers
- The I/O adapter is placed into a separate chassis
- Bus extender cards are placed into the servers



Virtualizing the I/O Path

Hybrid SR-IOV/MR-IOV

- **Common PCIe backplane for several servers in new hybrid blade/rack server**



Hairpin Turns

- In an IOV-capable environment, traffic can be sent out of one virtual adapter and received into another virtual adapter
- These two virtual adapters could reside on the same physical adapter, resulting in a “hairpin turn”
- The IOV adapters or IOV switches could act as LAN or SAN switches within the PCIe fabric (at lower cost)

Virtual Functions (VF)

- Virtual functions are the way that the adapter makes multiple versions of itself visible to the Hypervisor
- The Hypervisor assigns a VF to a guest
  - Many VFs can be created per physical port
  - Hypervisor has no visibility into the VF
- Guest sees a new adapter that it can use for anything that adapter can do

SR-IOV Today

- SR-IOV adapter availability
  - Ethernet NICs (1Gb and 10Gb)
  - FC HBAs
  - Some RAID controller prototypes
- Hypervisor support (Ethernet only)
- Citrix XenServer
  - RHEL 6 KVM
  - VMware 5.1 and higher
    - vMotion is not supported with SR-IOV
  - Windows Server 2012 (Hyper-V and guest)
    - Live Migration is supported with SR-IOV

SR-IOV Hardware Dependencies

- Hardware Checklist
  - Processor (CPU) support
  - Motherboard support (chipset, etc.)
  - BIOS support
  - SR-IOV capable NIC
- Many, but not all PCIe 2.0 and 3.0 servers will meet all of these criteria

SR-IOV Software Dependencies

- Hypervisor IOV-enabled virtual switch
- VF driver for the guest O.S.
- Windows registry: IovEnableOverride
- Set in the parent (Hyper-V) partition
- Windows is more strict about making sure that the hardware checklist is satisfied

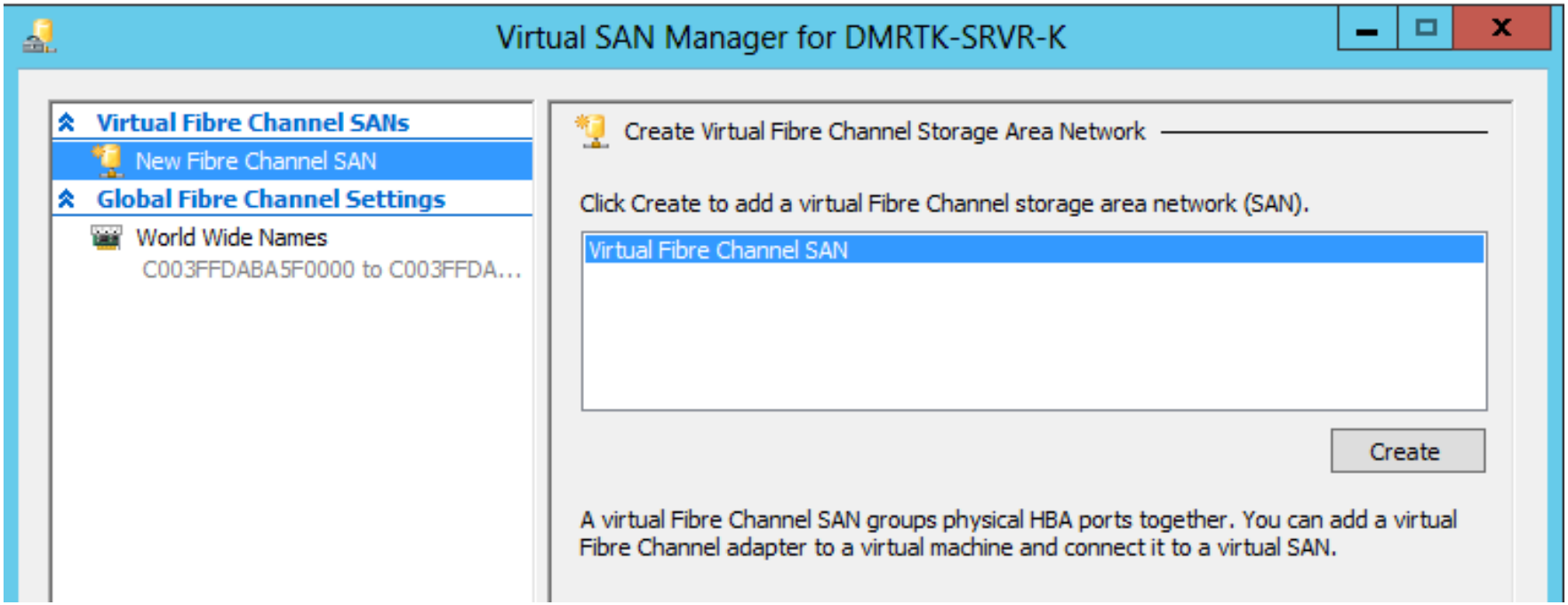
SR-IOV Bottom Line

- High traffic VMs can be assigned directly to a PCI Express Virtual Function provided by SR-IOV support in the OS, Server Platform and Network Controller by-passing the Hypervisor and Virtual Switch
- Advantage: Guest sees higher performance than it would without SR-IOV
- Disadvantage: Hypervisor has no visibility

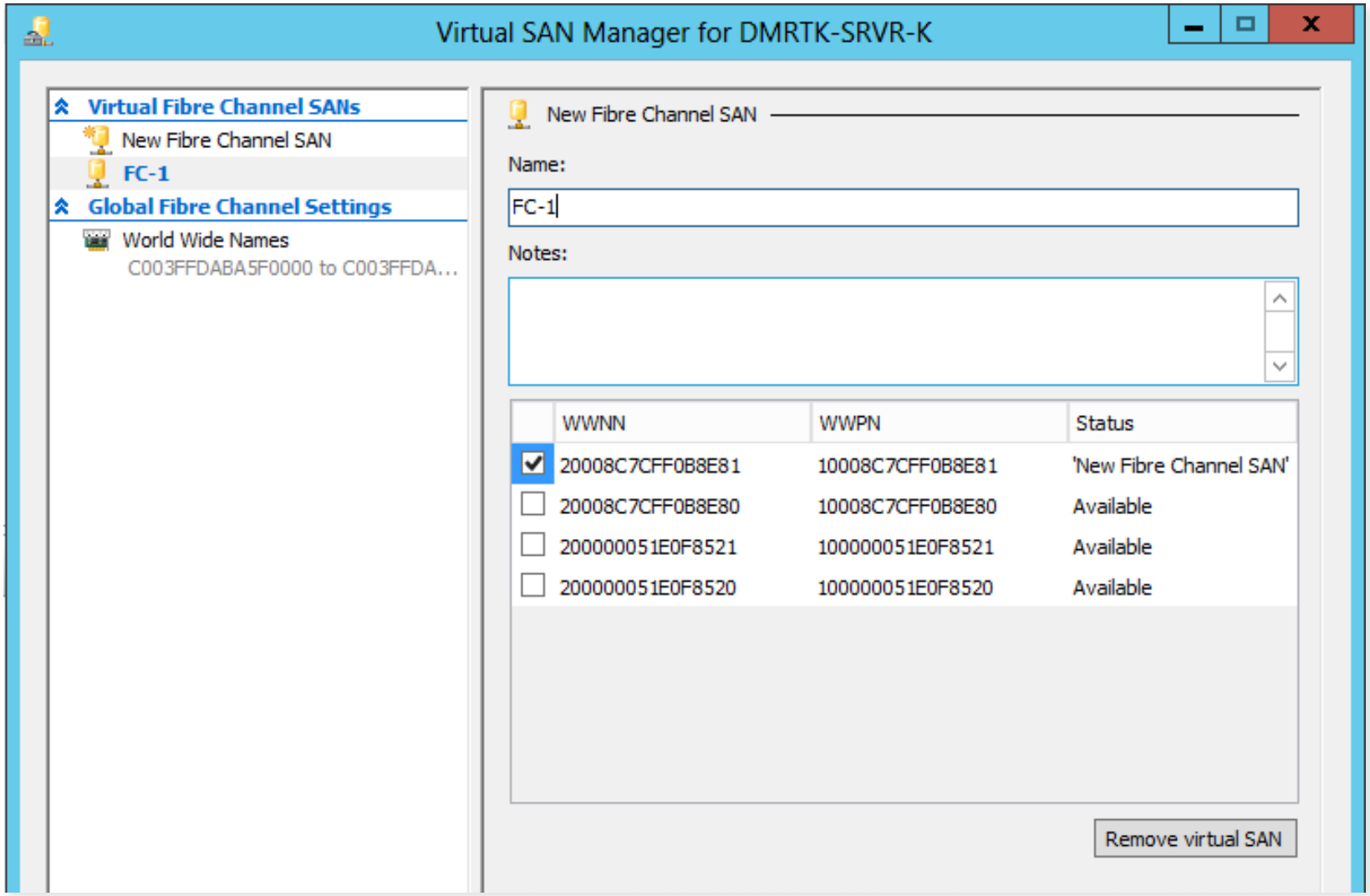
NPIV & Virtual Fibre Channel

- N-Port ID Virtualization (NPIV) for Fibre Channel behaves in a similar manner to SR-IOV for Ethernet
- When enabled, the FC HBA has VFs that can be assigned to guests
- Guest VM sees a virtual FC HBA and has access to same FC infrastructure as a physical server

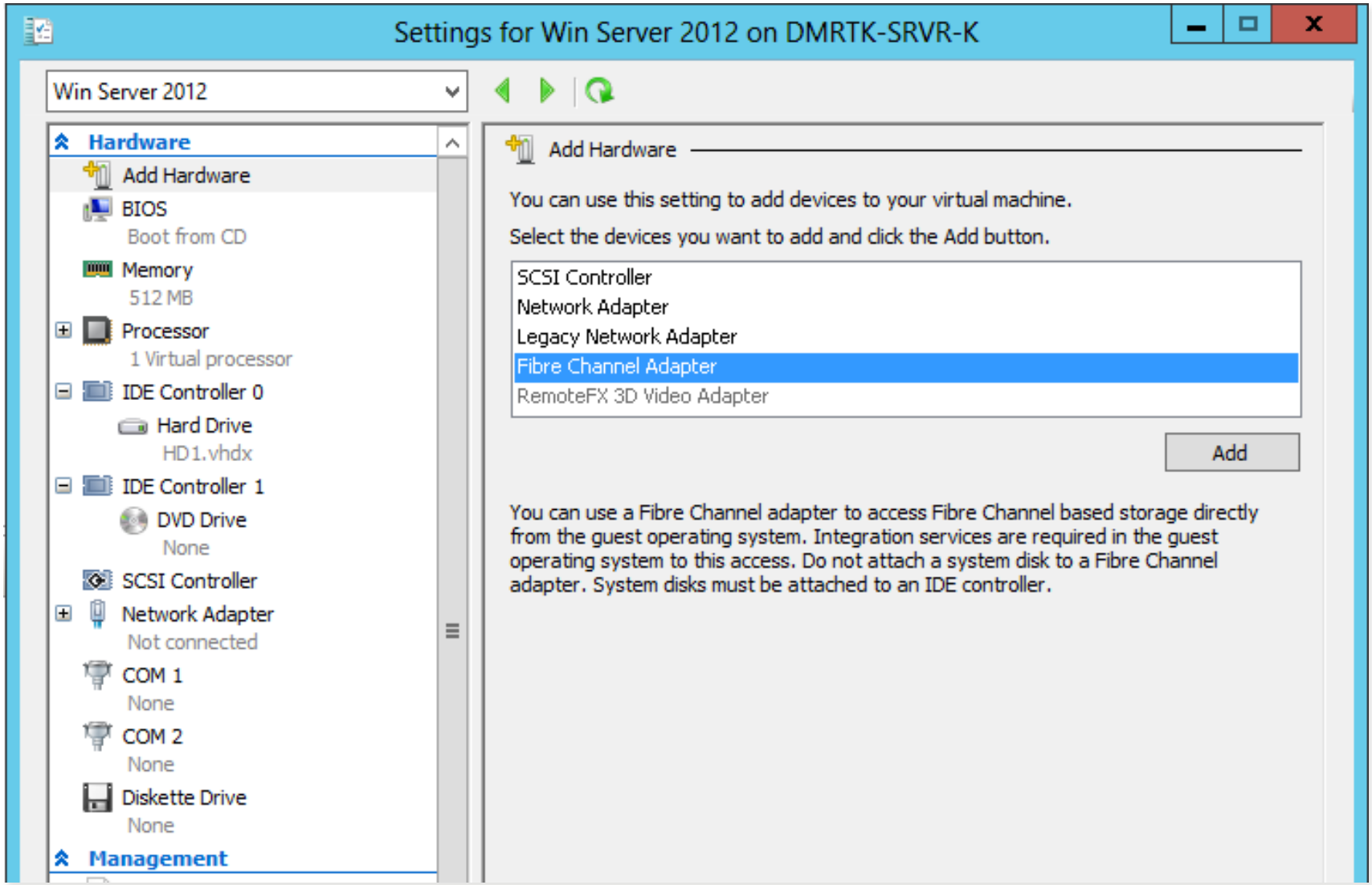
Virtual Fibre Channel – Hyper-V



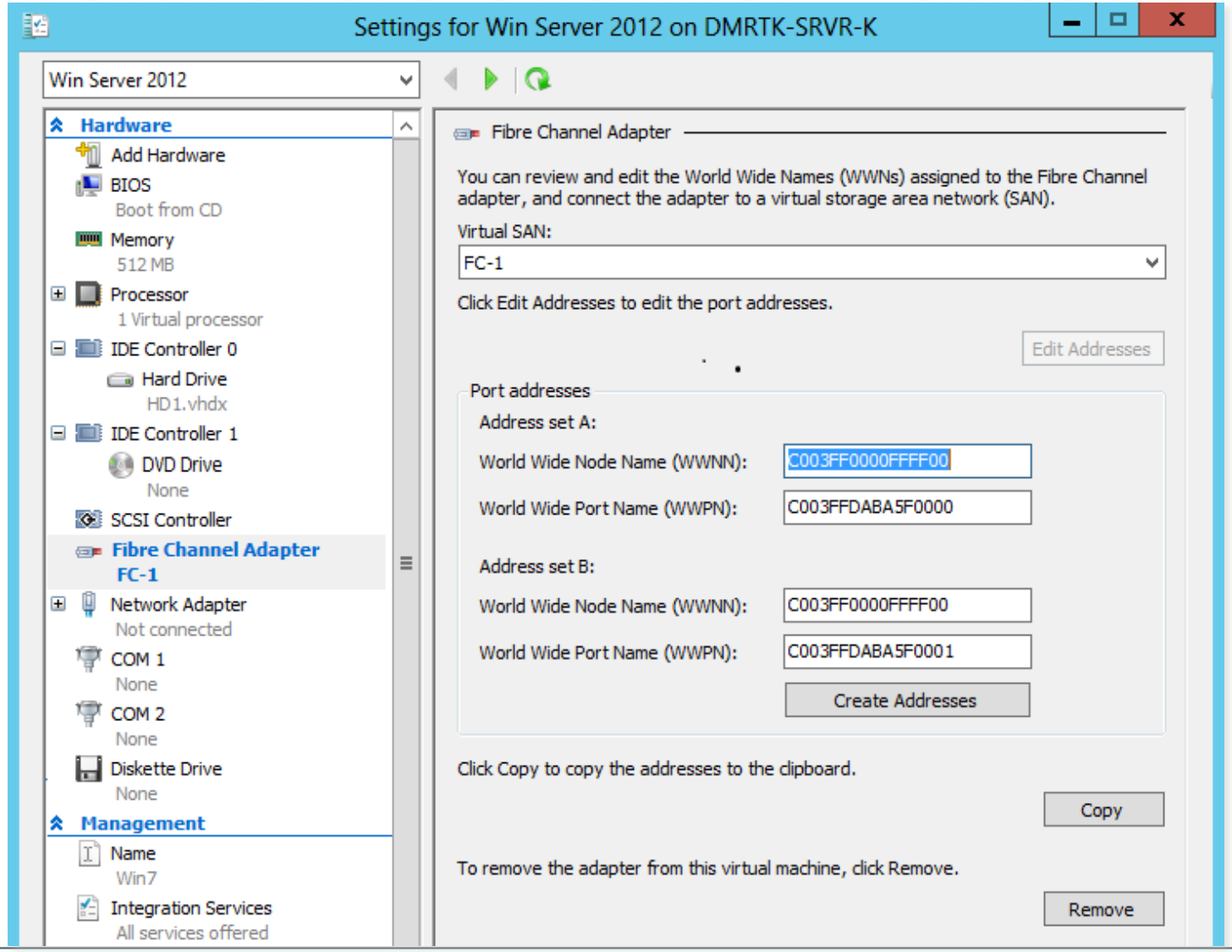
Virtual Fibre Channel – Hyper-V



Virtual Fibre Channel – Guest VM



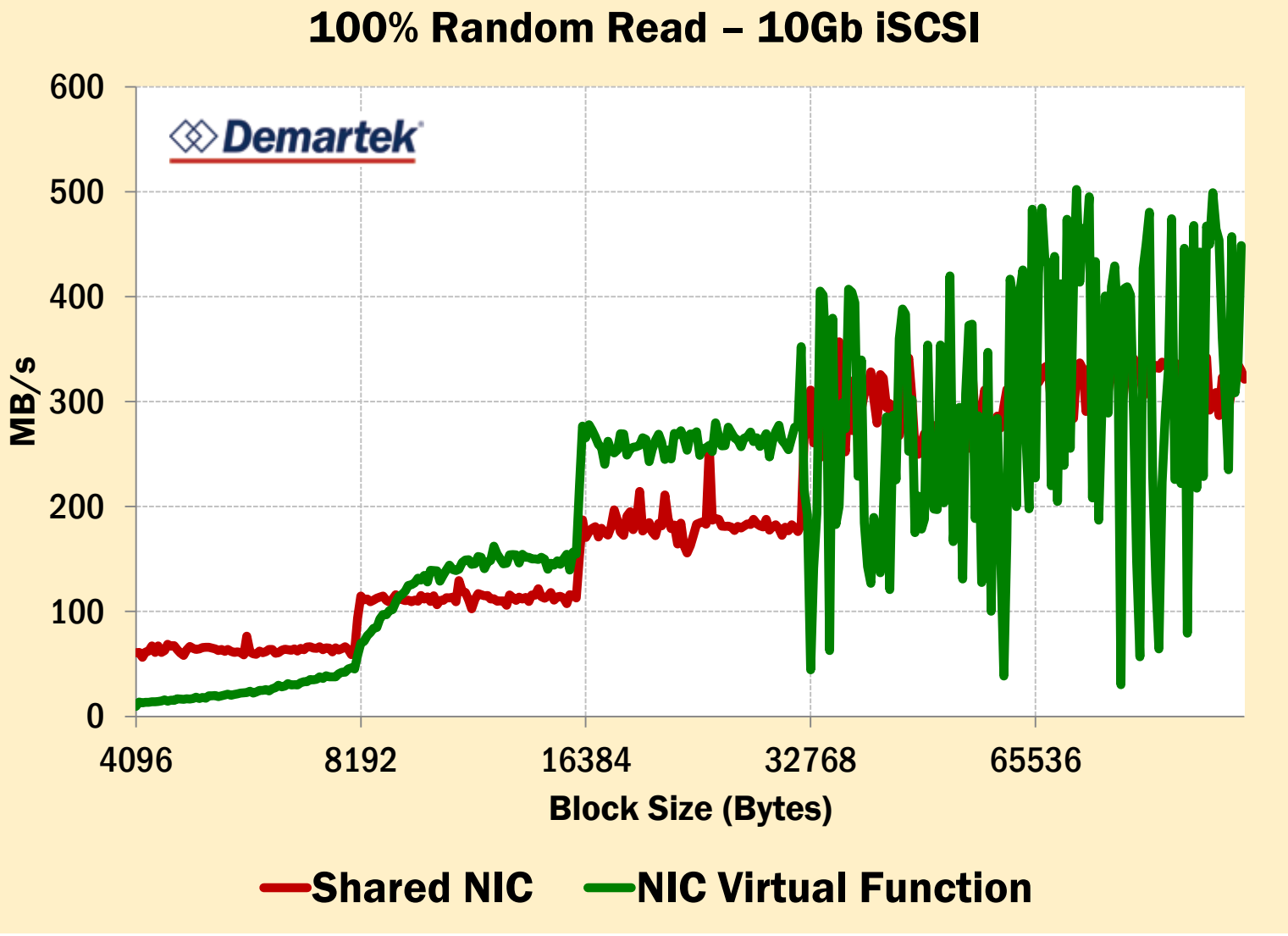
Virtual Fibre Channel – Guest VM

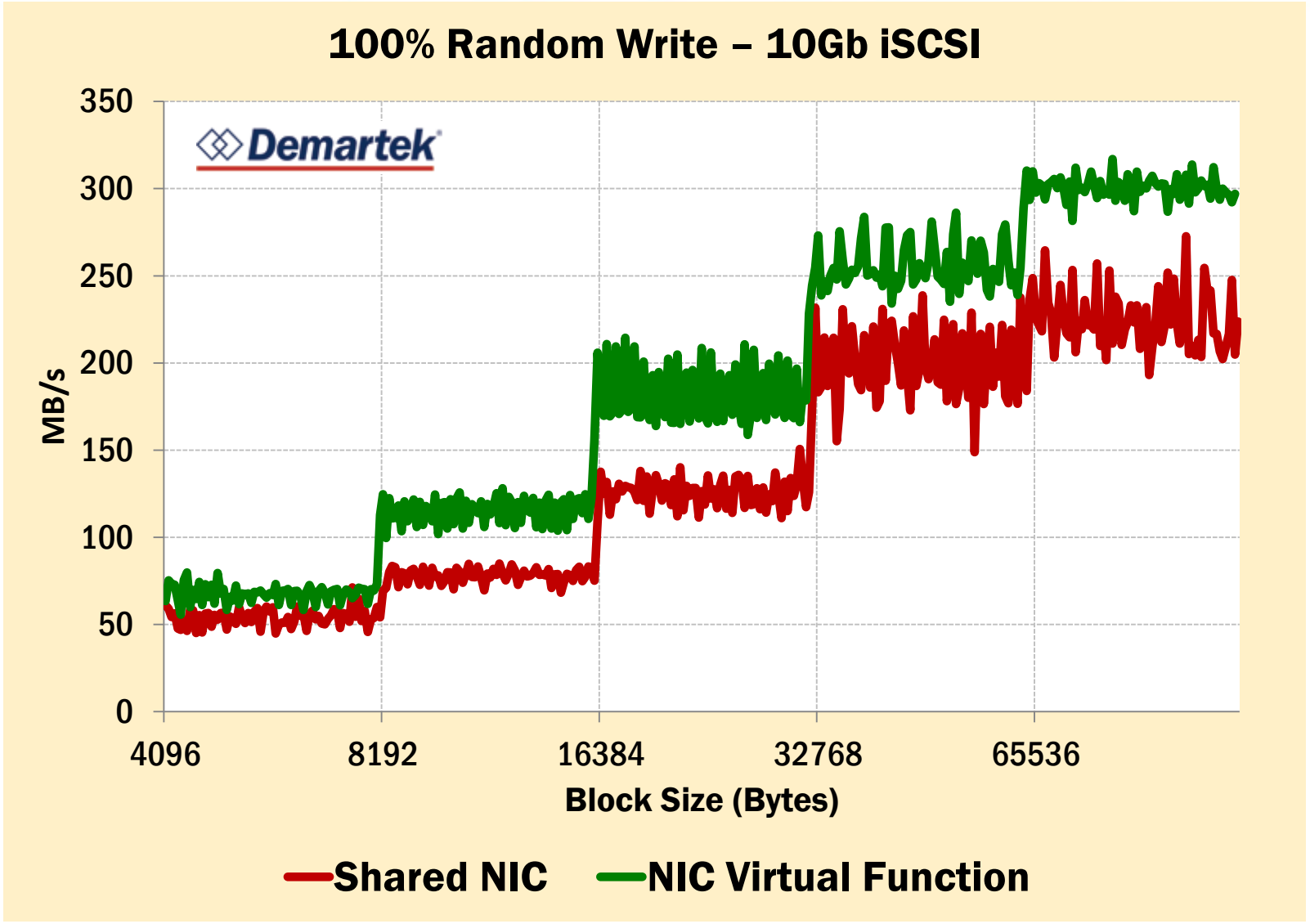


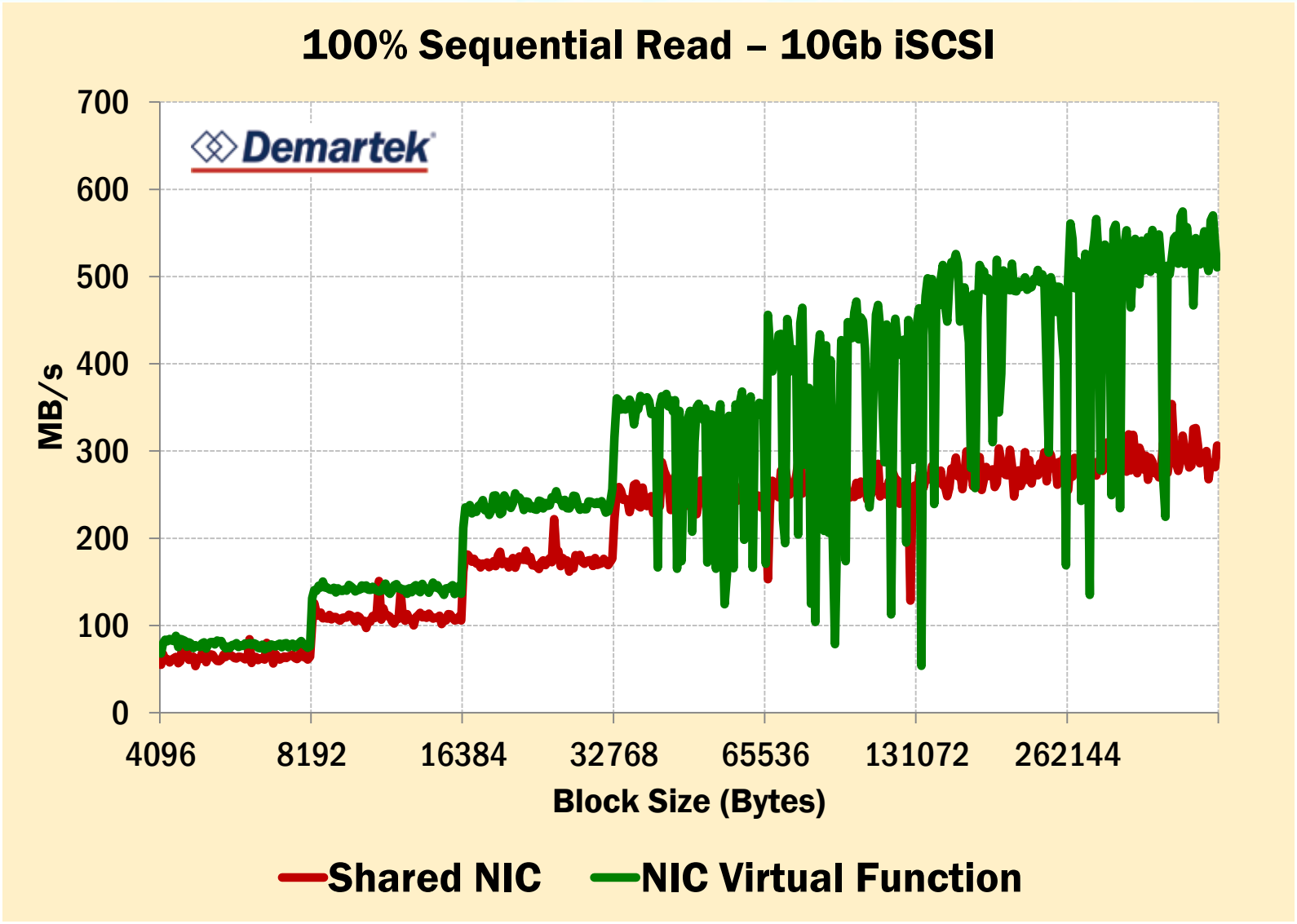
Performance Examples

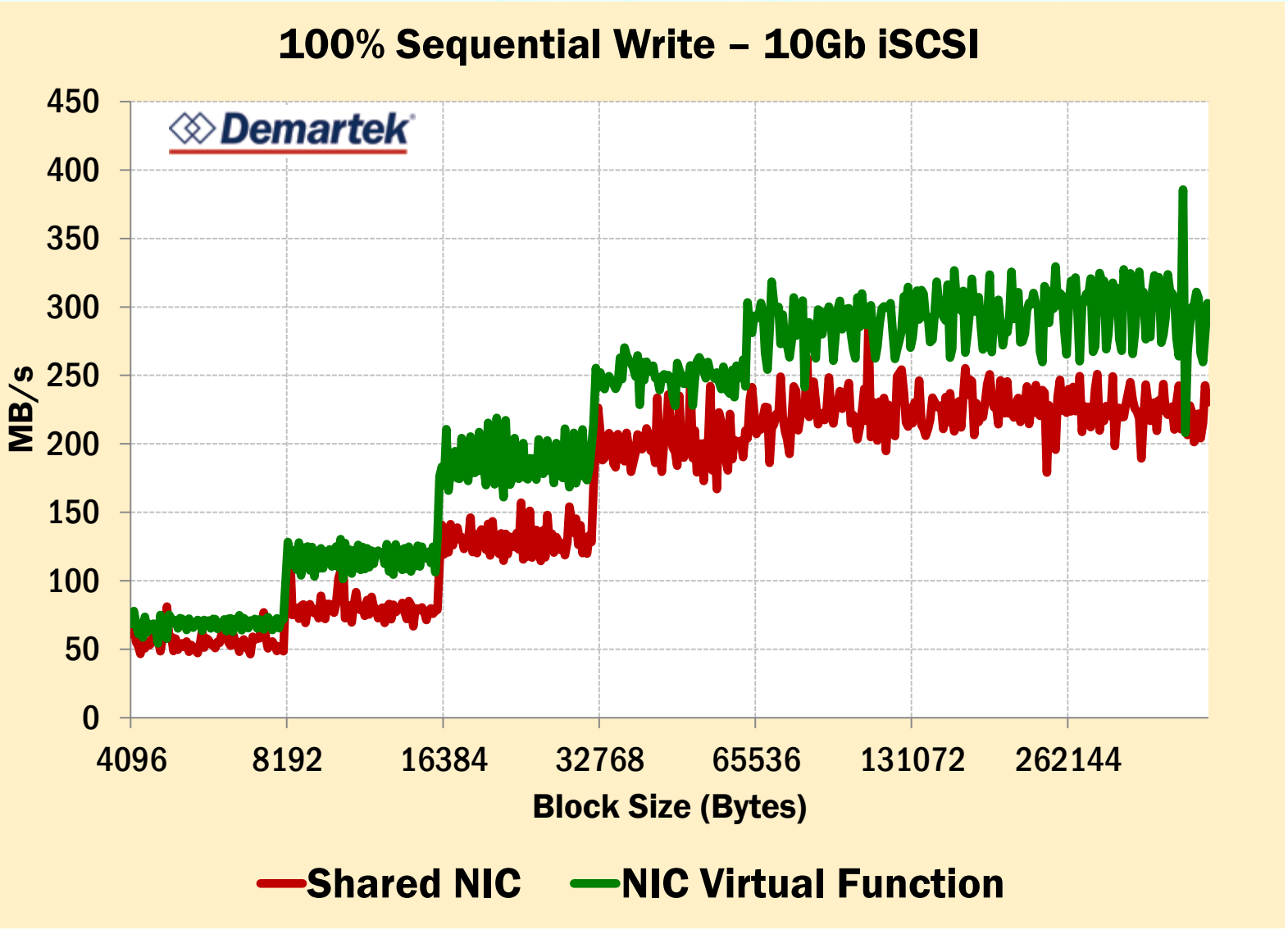
SR-IOV Performance Test

- We compared the performance of an SR-IOV capable 10GbE NIC connected via iSCSI to an all-flash storage target
  - SR-IOV enabled (VF)
  - SR-IOV disabled (shared NIC)
- Used vdbench to drive the workload
- Tested with one guest
- Demartek lab server
  - 2x Intel Xeon X5690, 3.46 GHz, 144 GB RAM with Intel X520 10GbE Server Adapter


100% Random Read – 10Gb iSCSI
Demartek
600
500
400
300
200
100
0
MB/s
4096
8192
16384
32768
65536
Block Size (Bytes)
Shared NIC
NIC Virtual Function


100% Random Write – 10Gb iSCSI
Demartek
350
300
250
200
150
100
50
0
MB/s
4096
8192
16384
32768
65536
Block Size (Bytes)
Shared NIC
NIC Virtual Function


100% Sequential Read – 10Gb iSCSI
Demartek
MB/s
700
600
500
400
300
200
100
0
4096
8192
16384
32768
65536
131072
262144
Block Size (Bytes)
Shared NIC
NIC Virtual Function


100% Sequential Write – 10Gb iSCSI
Demartek
MB/s
450
400
350
300
250
200
150
100
50
0
4096
8192
16384
32768
65536
131072
262144
Block Size (Bytes)
Shared NIC
NIC Virtual Function

Demartek References

- **Demartek Storage Interface Comparison**
  - www.demartek.com/Demartek_Interface_Comparison.html
- **Demartek SSD Zone**
  - www.demartek.com/SSD
- **Demartek SSD Deployment Guide**
  - www.demartek.com/Demartek_SSD_Deployment_Guide.html
- **Demartek Commentary – Horses, Buggies & SSDs**
  - www.demartek.com/Demartek_Horses_Buggies_SSDs_Commentary.html
- **Demartek Free Monthly Newsletter**
  - www.demartek.com/Newsletter

Thank You!

Dennis Martin, President
dennis@demartek.com
www.linkedin.com/in/dennismartin

Demartek

(303) 940-7575
www.demartek.com
http://twitter.com/Demartek
www.youtube.com/Demartek
Skype: Demartek

To learn more about Demartek:

- Download the Aurasma App (Android/iPhone)
- Search and follow "Demartek"
- View image below with viewfinder.

*also on the back of Dennis' business card

Powered by: AURASMA