Storage Performance on RHEL with Cisco UCS 1240 & 1280 Virtual Interface Card (VIC)

Storage Performance on Red Hat Enterprise Linux (RHEL) Version 6.2

Contents

Key Findings

This paper presents the storage networking performance characteristics of the Cisco Unified Computing System™ (Cisco UCS®) Virtual Interface Card (VIC) 1240 combined with an I/O Port Expander card in a Cisco UCS blade server. The following observations are presented:

- The Cisco UCS VIC 1240 on Cisco UCS B200 Blade Servers can deliver up to 2.3 GBps of 4K and 3.1 GBps of 8K read-only performance.
- The Cisco UCS VIC 1240 can deliver up to 4.15 GBps of peak read-only performance.
- The Cisco UCS VIC 1240 can deliver up to 2.04 GBps of 4K and 3.3 GBps of 8K write-only performance.
- The Cisco UCS VIC 1240 can deliver up to 3.77 GBps of peak write-only performance.
- The Cisco UCS VIC 1240 can deliver up to 7.5 GBps of peak read and write mixed performance.
- The Cisco UCS VIC 1240 with I/O Expander can deliver up to 5.85 GBps of peak read-only performance.
- The Cisco UCS VIC 1240 with I/O Expander can deliver up to 5.5 GBps of peak write-only performance.
- The Cisco UCS VIC 1240 with I/O Expander can deliver up to 10.2 GBps of peak read and write mixed performance.

Introduction

The Cisco UCS VIC is a virtualization-optimized converged network adapter (CNA) mezzanine card designed for use with Cisco UCS B-Series Blade Servers. The VIC card supports up to 256 Peripheral Component Interconnect Express (PCIe) standards-compliant virtual interfaces (128 for the first generation). These PCIe interfaces can be dynamically configured so that both their interface type ¬(whether network interface card [NIC] or host bus adapter [HBA]) and identity (MAC address and worldwide name [WWN]) are established using just-in-time provisioning. Complete network separation is guaranteed between the PCIe devices using network interface virtualization (NIV) technology.

Configurations for Cisco UCS Virtual Interface Card 1240

Based on the second generation Cisco VIC technology, the VIC 1240 is a modular LAN on motherboard (LOM) that is designed specifically for the M3 generation of Cisco UCS B-Series Servers and offers industry leading performance, flexibility, and manageability. The VIC 1240 is capable of aggregate 8 x 10 Gbps to the half-width blade slot when used with the Port Expander Card for the VIC 1240. Without the Port Expander Card, the Vic 1240 enables 4 ports of 10 Gbps network I/O to each half-width blade server (Figures 1 and 2).
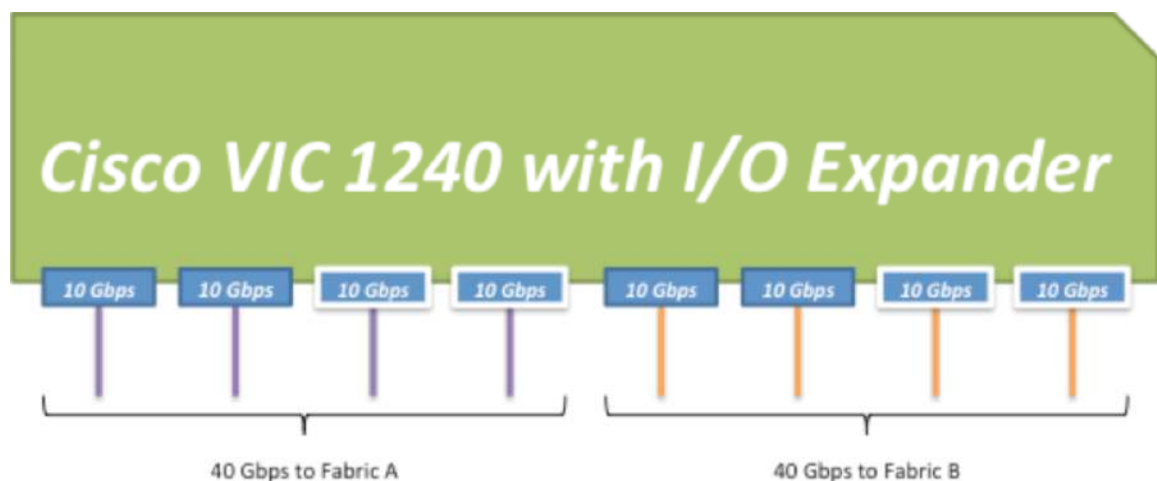
Figure 1. Cisco UCS VIC 1240
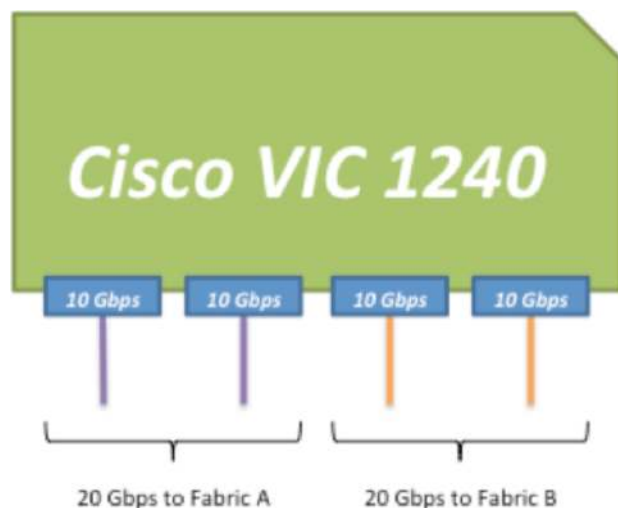
The Cisco UCS VIC 1240 is available in either 4 x 10 Gbps hardware configuration (Figure 1) or 8 x 10 Gbps hardware configuration (Figure 2). In either configuration, half the number of ports is connected to Fabric A, and the other half to Fabric B. A vNIC instantiated on the VIC can be explicitly pinned to either Fabric A or B. The vNIC may also be configured for failover, in which case one fabric-connected path can be designated as the primary path and the other fabric- connected path as secondary or failover path. Third party failover software is required to handle the actual data path failover in case of a path failure event. Please note, this statement is true only when fabric failover is off.

In Figure 2, the boxes outlined in white are additional ports that are enabled through the use of the I/O Port Expander.

Figure 2. Cisco VIC 1240 with I/O Port Expander

Cisco UCS VIC 1280

Cisco UCS VIC 1280 (see Figure 3) is identical in capability to VIC 1240 with the I/O Port Expander. The difference is that the VIC 1280 comes with default 8 x 10 Gbps hardware configuration. Therefore, for all practical purposes, the VIC 1280 has performance characteristics that are identical to the VIC 1240 with I/O Expander.

Figure 3.   Cisco UCS VIC 1280
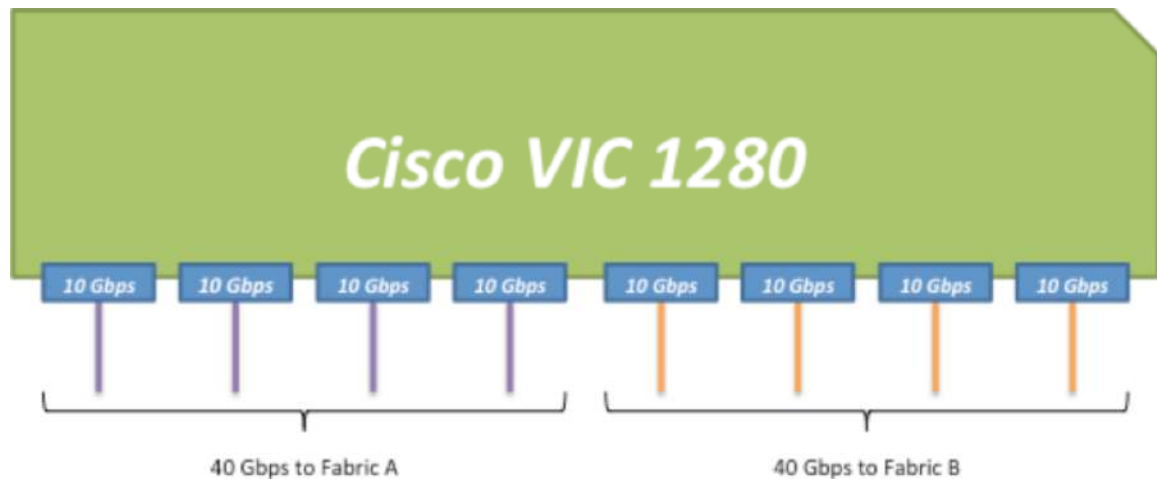
The Cisco VIC is dependent on the chassis I/O module for external connectivity and bandwidth capabilities. The chassis I/O module may be configured to operate in either port-channel mode or non-port-channel mode. Configuring the I/O module in port-channel mode enables up to 80 Gbps of fabric bandwidth per I/O module or up to 160 Gbps of fabric bandwidth across both I/O modules.

The connectivity between the Cisco VIC and the I/O module is always available in port-channel mode. When the I/O module is configured in non-port-channel mode, the bandwidth per half-width slot per I/O module is limited to 10 Gbps. When configured in port-channel mode, the bandwidth per half-width slot per I/O module is limited up to 20 Gbps with the VIC 1240 and up to 40 Gbps with Cisco VIC 1240 with I/O Expander card or Cisco VIC 1280.

System Configuration

Fabric Topology

Figure 4 shows the network topology used for testing.

Figure 4.   Network Topology
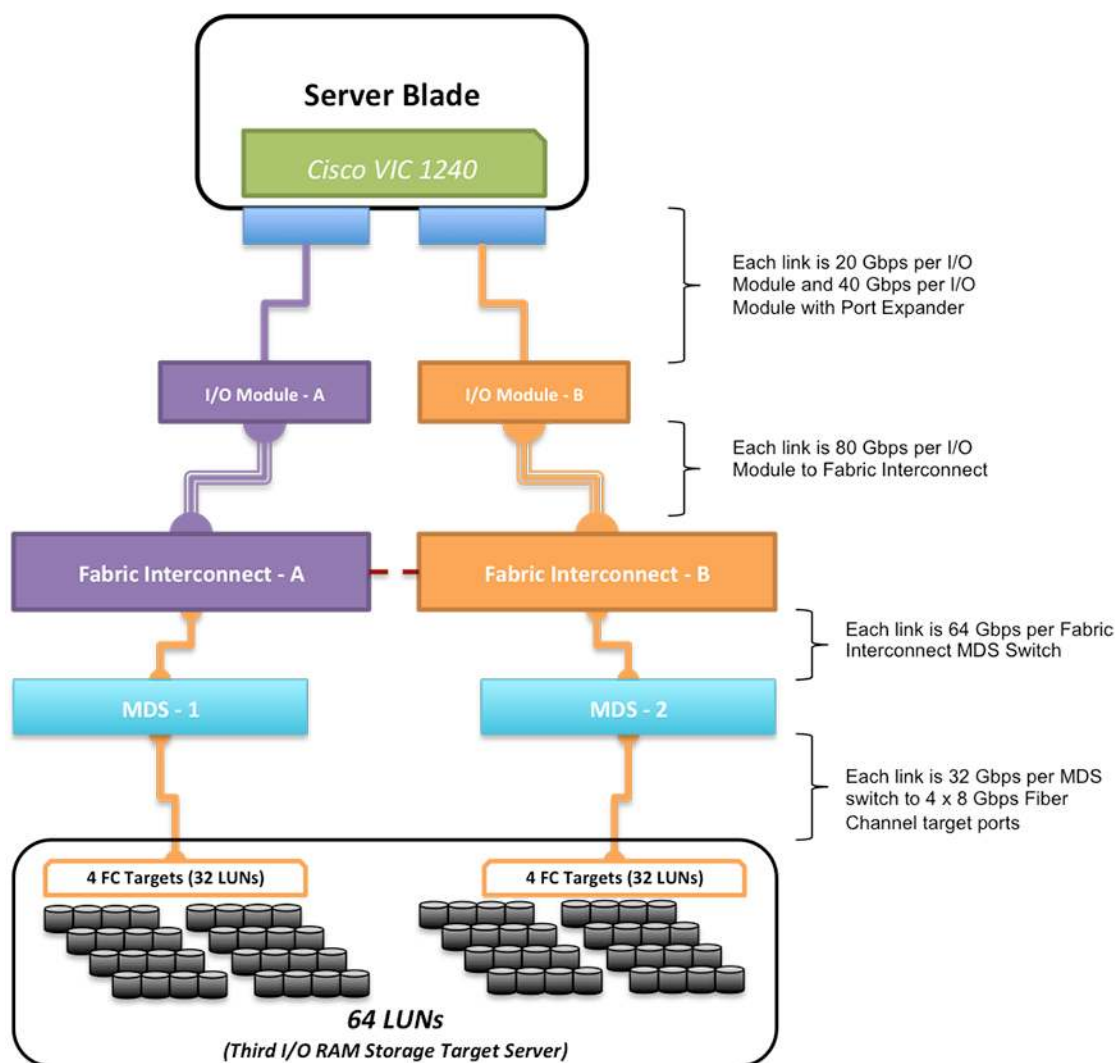
A single server blade is connected to 64 LUNs spread across 8 Fibre Channel targets. Since the virtual host bus adapter (vHBA) configuration was limited to eight vHBAs and each vHBA was pinned to a specific Fibre Channel target, the number of Fibre Channel targets was limited to eight.

With VIC 1240 (Figure 5), eight vHBAs were spread across 4 x 10 Gbps with four vHBAs pinned to Fabric A and four pinned to Fabric B.

Figure 5. vHBA Connectivity on VIC 1240
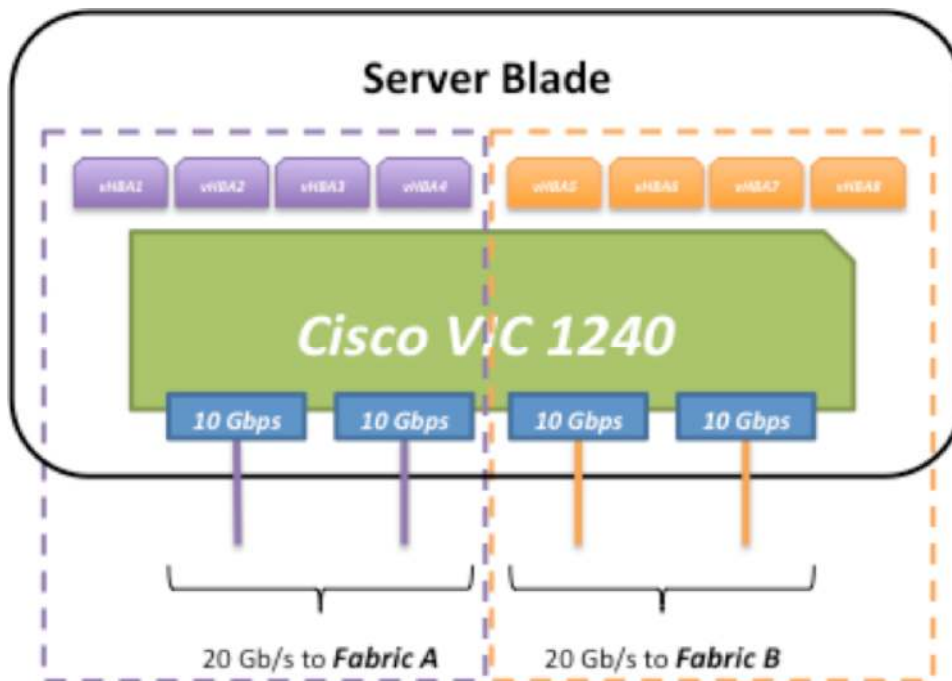
With the VIC 1240 with the I/O Expander, a similar vHBA configuration was set up with 8 x 10-Gbps host interface ports connecting to both fabrics, as illustrated in Figure 6.

Figure 6. vHBA Connectivity on VIC 1240 with I/O Expander
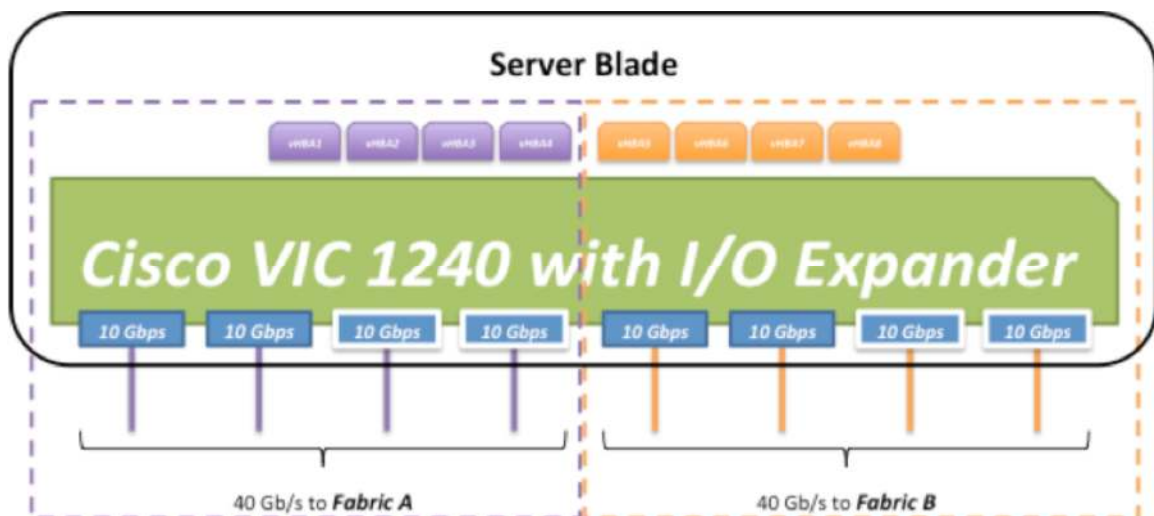
Hardware Configuration

Table 1 shows the basic hardware configuration used for benchmark testing.

Table 1. Hardware Configuration for Benchmark Testing

Hardware
Cisco UCS B200 M3 Blade Server • Two 8-Core Intel Xeon E5-2690 processors at 2.90 GHz • 128 GB DDR3 1600-MHz RAM • Cisco UCS VIC 1240 with I/O Expander (PCIe 2.0 16x)
Cisco UCS • Cisco UCS 5108 Chassis • Cisco UCS 2208XP Fabric Extender • Cisco UCS 6248UP Fabric Interconnects
Target Storage Server Third I/O Iris Kernel Version 5.4 X4 RAM backed storage software running on Cisco UCS C240 M3 Rack Server

The Third I/O solution was chosen for storage targets. Third I/O is a data storage platform solution that uses physical RAM for storage media. This approach allows for very a high level of performance while avoiding the latencies and delays related to conventional, mechanical disk-based storage systems.

Table 2 shows the firmware configuration.

Table 2. Firmware Configuration

Firmware
Cisco UCS Software Release 2.0(4a) was used • Cisco UCS VIC 1240 • Cisco UCS B200 M3 Server Blade BIOS • Cisco UCS 2208XP I/O Modules • Cisco UCS 6248UP Fabric Interconnects

System BIOS Settings

Table 3 lists the system BIOS settings used.

Table 3. System BIOS Settings

CPU Configuration	
Hyper Threading	Disabled
Turbo Mode	Enabled
Intel SpeedStep	Enabled
Direct Cache Access	Enabled
Processor C State	Disabled
Processor C1E	Disabled
Processor C3 Report	Disabled
Processor C6 Report	Disabled
Processor C7 Report	Disabled
CPU Performance	hpc
Package C State Limit	No limit
Memory Configuration	
Memory RAS Config	Maximum performance
NUMA	Enabled
LV DDR Mode	Performance-mode

Adapter Settings

The default adapter settings were used.

Host-Side Settings

I/O Scheduler

The system I/O scheduler was set to **noop** using the **tuned-adm.profile raw** command. You can also set the scheduler on a per LUN basis using the **sysfs** interface.

The **noop** I/O scheduler avoids buffer caching and therefore allows for raw access to the device. Because of caching effects, other I/O schedulers may skew the performance measurements and were therefore avoided.

CPU Settings

The **cpuspeed** service was explicitly configured for **performance** mode. The default setting is **ondemand**. With performance mode, the CPU is always maintained in its highest P-state. This ensures full CPU availability and avoids state transition delay, which may affect performance measurements.

CPU Affinitization

CPU affinitization helps reduce and/or avoid cache misses. Typically, this is handled by **irqbalancer** service on RHEL. However, for the purpose of this test, **irqbalancer** was turned off. We explicitly configured **app** —> **cpu**—> **hardware affinity** using a combination of the **/proc/irq** interface and the **taskset** command.

In Figure 7, each blue circle represents eight **vwrio** processes. Each **vwrio** process is bound to a single LUN. Each **vwrio** process also spawns eight threads against each CPU/Storage Target controller/LUN. Therefore, across the system, for 64 LUNs, a total of 64 **vwrio** processes were launched with a total of 512 threads.

Figure 7. CPU Affinitization
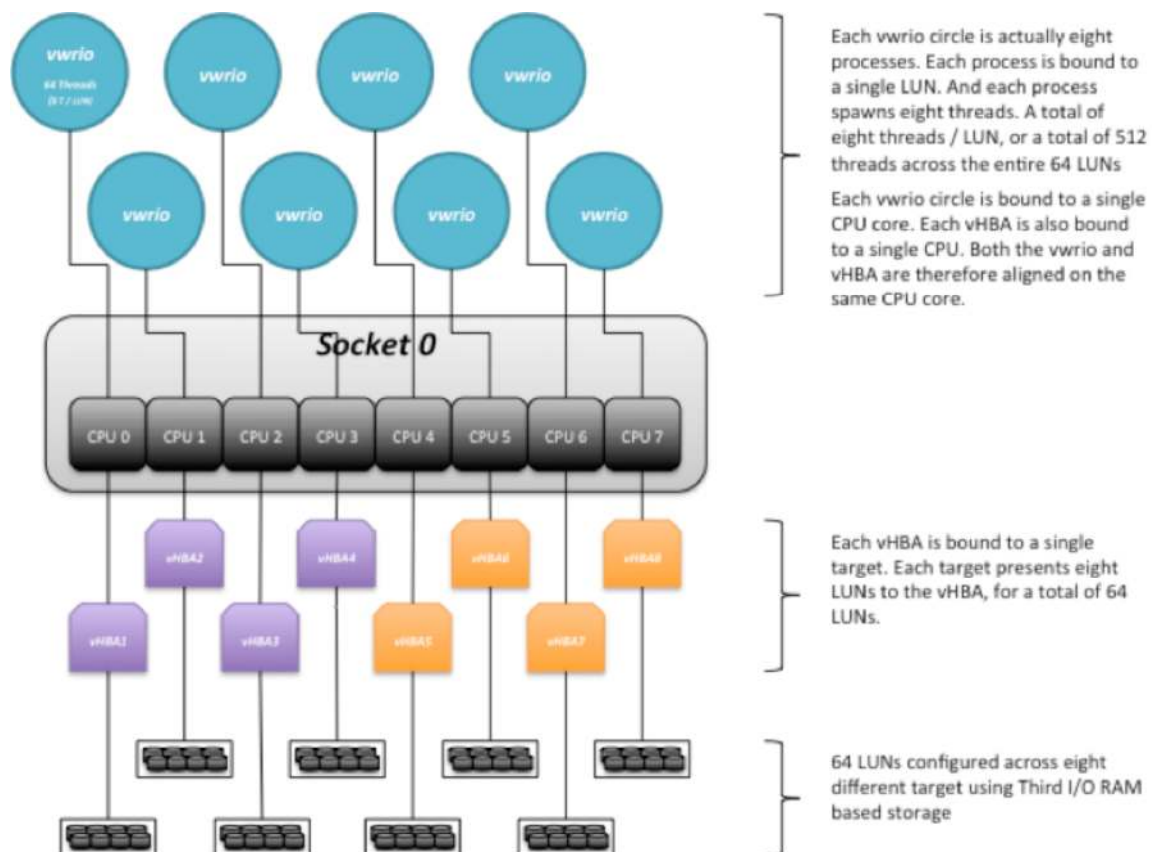
Through CPU affinitization, pinning the vHBA interrupt and the corresponding **vwrio** process on the same CPU core, we can expect efficient performance while relieving the operating system of the task of scheduling the large number of threads efficiently.

Performance Evaluation Tools

The **vwrio** performance tool, Version 1.0, was used to drive the data. **vwrio** is an extremely lightweight and efficient approach to drive raw I/O. It uses the O_DIRECT Linux interface to achieve this. Astute users may already know that the standard **dd** utility also provides O_DIRECT interface. However, **dd** uses a single thread per process and would require multiple processes to drive a large amount of I/O. Even with a sufficiently large number of processes, **dd** might not be as performance efficient.

vwrio provides the same functionality without the overhead of dealing with large number of processes and with the efficiency of Linux **pthreads**.

The following **vwrio** command template was used for running the test:

./vwrio –d /dev/${sd_device} -n ${num_io_threads} –i ${num_iterations} –s ${size_of_sd_device} –b ${block_size} –m ${mode}

For example, for read I/O:

./vwrio –d /dev/sdc –n 8 –i 128 –s 1024 –b 4096 –m R

For mixed I/O:

./vwrio –d /dev/sdc –n 8 –i 128 –s 1024 –b 4096 –m M –M 50

The **-m M** switch specifies mixed mode I/O and **-M 50** switch specifies 50 percent read and 50 percent write. This translates to four I/O threads performing read I/O and the other four performing write I/O.

The results were captured using the Linux **dstat** utility. **dstat** was configured to capture the results in total bytes of throughput per second. Using the following formula, the I/O per second number was derived:

(bytes per second)/block size = I/O per second

Performance Results

Performance results are presented in both graph and tabular formats.

Where the data is presented in graph format, the X-axis corresponds to I/O block size. The primary Y-axis corresponds to MB per second (throughput) and the secondary Y-axis corresponds to I/O per second. This axis format is consistent for all the graphs presented in this paper.

Where the data is presented in tabular format, the results for 4K and 8K block sizes are **bolded**. The numbers **bolded** correspond to peak performance.

A special emphasis was placed on 8K and 4K block sizes since they are the two most commonly used block sizes in enterprise applications.

Read Performance

As Figure 8 shows, the read-only performance for both VIC 1240 and VIC 1240 with I/O Expander, for 4K I/O per second, peaks at around 625K. This is the limitation of the target storage rather than the VIC itself. The almost flat line I/O per second performance for lower block sizes also illustrates this.

Figure 8.   Read Performance
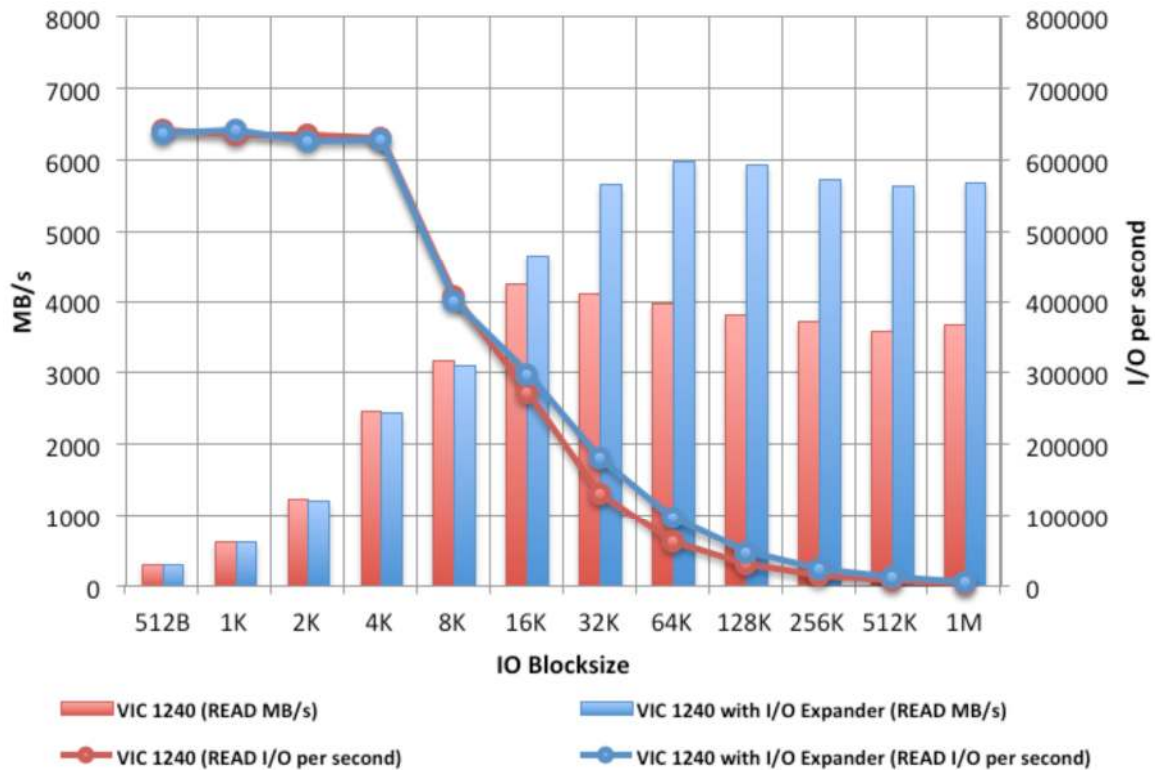
The I/O per second performance for 8K read operations is around 400K, as Table 4 shows.

Table 4.   Read Performance

READ Performance				
	VIC 1240		**VIC 1240 with I/O Expander**	
Blocksize	**I/O per second**	**MB/s**	**I/O per second**	**MB/s**
512B	**642723.90**	313.83	637448.60	311.25
1K	635465.17	620.57	642123.03	627.07
2K	634903.03	1240.04	625926.37	1222.51
4K	**631686.93**	**2467.53**	**628832.67**	**2456.38**
8K	**406618.57**	**3176.71**	**399747.87**	**3123.03**
16K	272152.60	**4252.38**	298658.23	4666.53
32K	131947.83	4123.37	181397.30	5668.67
64K	63558.87	3972.43	95871.73	**5991.98**
128K	30630.77	3828.85	47469.17	5933.65
256K	14881.90	3720.48	22970.43	5742.61
512K	7175.47	3587.73	11301.70	5650.85
1M	3687.93	3687.93	5689.48	5689.48

The decreasing read MB per second trend for the VIC 1240 for block sizes larger than 16K can be attributed to cache miss on the target storage controller.

Write Performance

Figure 9 shows write-only performance results.

Figure 9. Write Performance Results
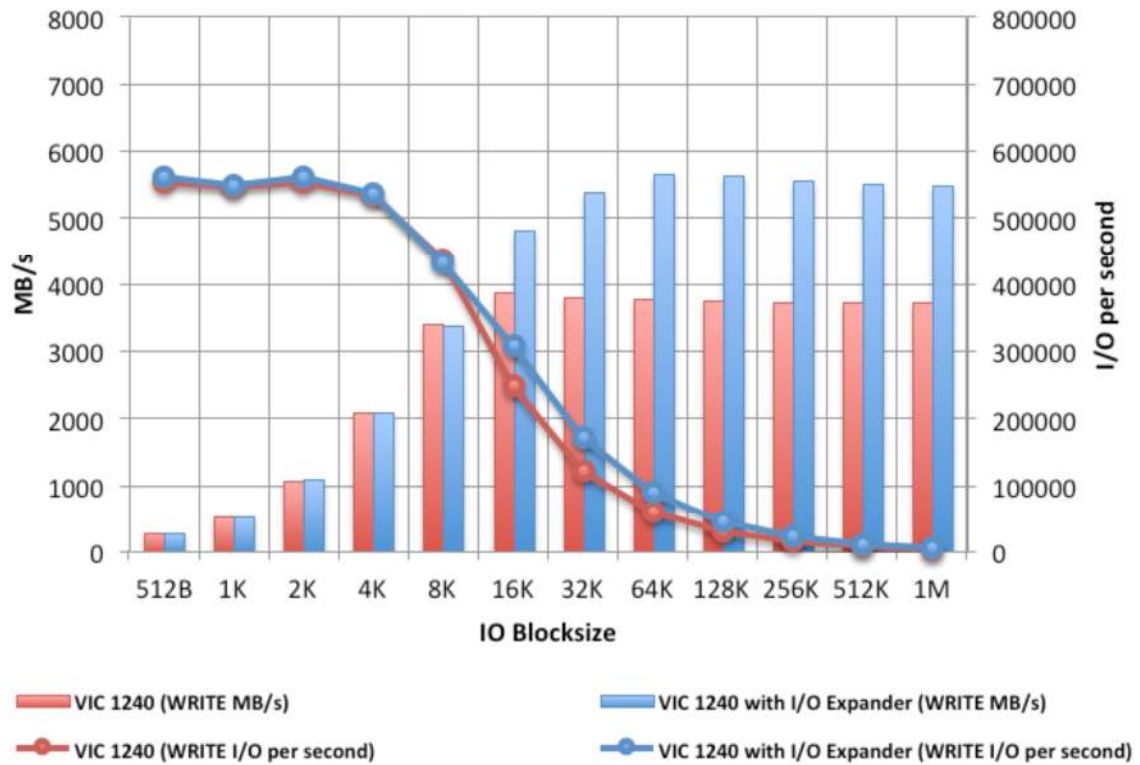
As Table 5 shows, the throughput performance peaks at 3866 MB per second at 16K block size for the VIC 1240.
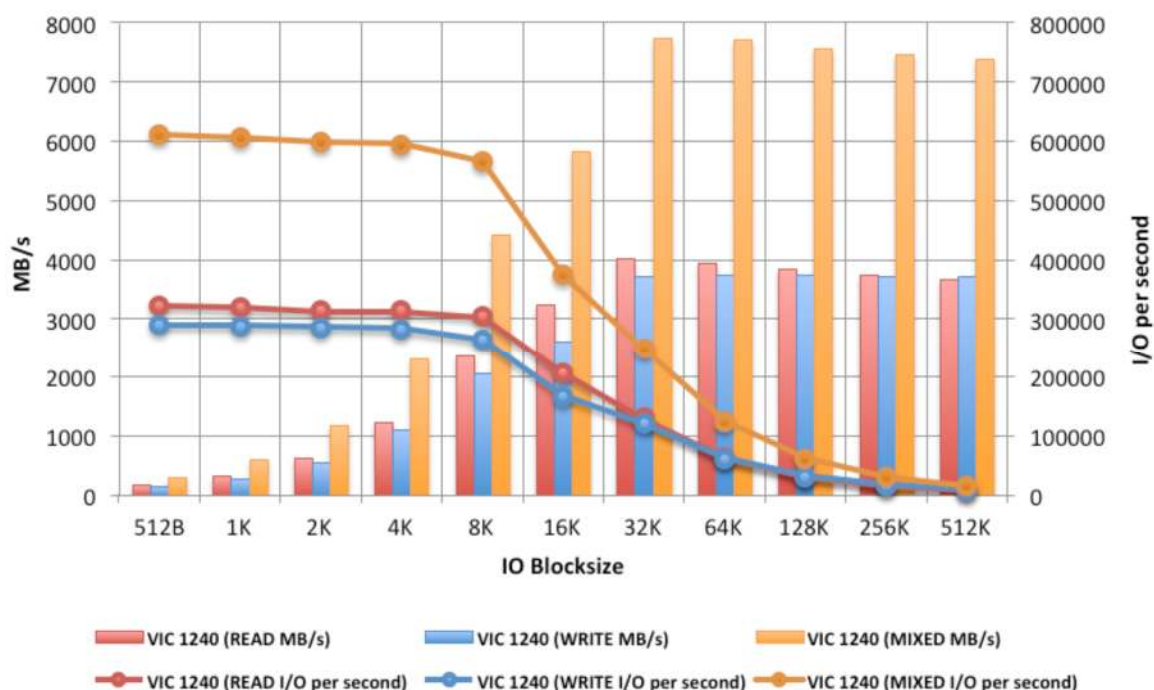
Table 5. Write Performance Results

WRITEPerformance				
	VIC 1240		VIC 1240 with I/O Expander	
Blocksize	I/O per second	MB/s	I/O per second	MB/s
512B	**552537.40**	269.79	561160.07	274.00
1K	545333.23	532.55	548135.93	535.29
2K	553315.33	1080.69	560255.33	1094.25
4K	**533375.10**	**2083.50**	**537354.37**	**2099.04**
8K	**436641.45**	**3411.26**	**432757.67**	**3380.92**
16K	247471.80	**3866.75**	308085.98	4813.84
32K	121816.77	3806.77	172639.21	5394.98
64K	60226.21	3764.14	90534.19	**5658.39**
128K	29921.66	3740.21	45064.86	5633.11
256K	14928.31	3732.08	22239.35	5559.84
512K	7458.77	3729.39	11031.51	5515.75
1M	3721.32	3721.32	5493.60	5493.60

Mixed Read and Write Performance

Figure 10 and Table 6 show results for mixed read and write performance for the Cisco UCS VIC 1240 without I/O Expander.

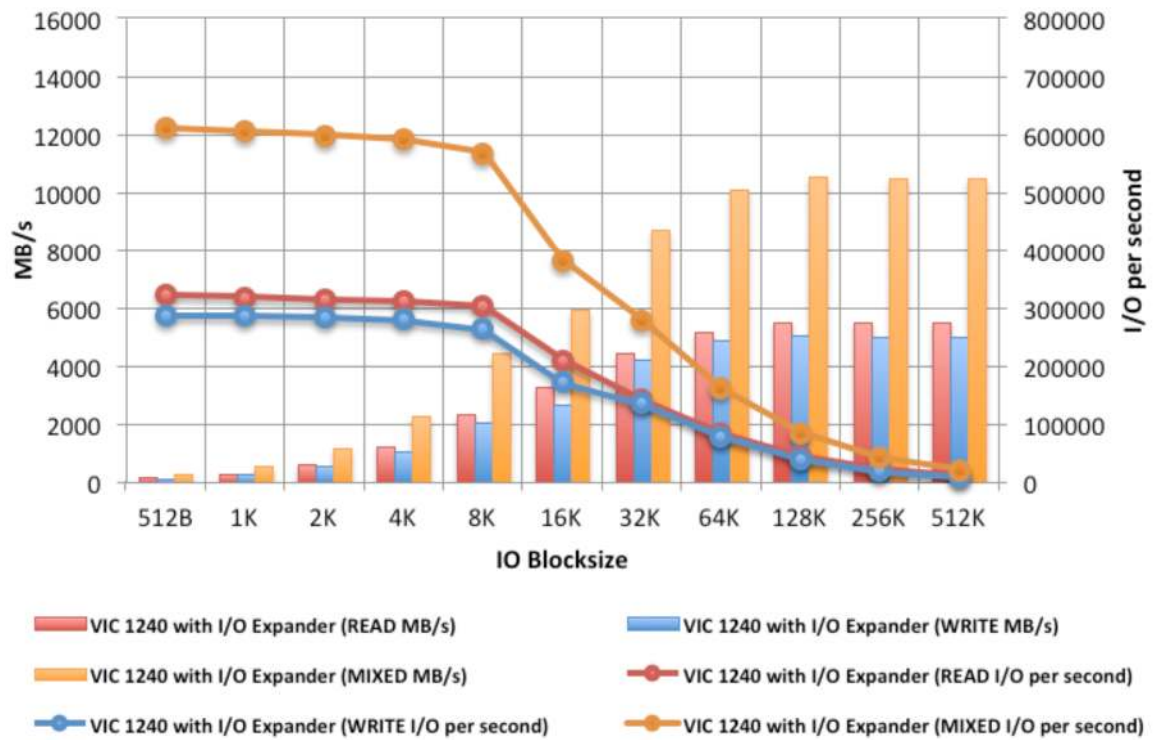**Figure 10.** Mixed Read and Write Performance Results

Table 6. Mixed Read and Write Performance Results

MIXED Performance (VIC 1240)						
	READ		**WRITE**		**Total**	
Blocksize	**I/O per second**	**MB/s**	**I/O per second**	**MB/s**	**I/O per second**	**MB/s**
512B	**322465.67**	157.45	**287331.57**	140.30	**609797.23**	297.75
1K	318840.40	311.37	286778.67	280.06	605619.07	591.42
2K	313189.53	611.70	284583.20	555.83	597772.73	1167.52
4K	**312459.17**	**1220.54**	**281265.30**	**1098.69**	**593724.47**	**2319.24**
8K	**302696.93**	**2364.82**	**262620.93**	**2051.73**	**565317.87**	**4416.55**
16K	206558.67	3227.48	166421.41	2600.33	372980.08	5827.81
32K	128894.30	**4027.95**	118937.80	3716.81	372980.08	**7744.75**
64K	63205.33	3950.33	60052.11	**3753.26**	123257.44	7703.59
128K	30706.53	3838.32	29901.69	3737.71	60608.23	7576.03
256K	14957.13	3739.28	14907.61	3726.90	29864.75	7466.19
512K	7350.30	3675.15	7440.81	3720.40	14791.11	7395.55
1M	3684.15	3684.15	3720.62	3720.62	7404.77	7404.77

Figure 11 and Table 7 show results for the Vic 1240 with the I/O Expander.

Figure 11. Mixed Read and Write Performance Results with I/O Expander

Legend:
- VIC 1240 with I/O Expander (READ MB/s)
- VIC 1240 with I/O Expander (WRITE MB/s)
- VIC 1240 with I/O Expander (MIXED MB/s)
- VIC 1240 with I/O Expander (READ I/O per second)
- VIC 1240 with I/O Expander (WRITE I/O per second)
- VIC 1240 with I/O Expander (MIXED I/O per second)

Table 7. Mixed Read and Write Performance Results with I/O Expander

MIXED Performance (VIC 1240 with I/O Expander)						
	READ		WRITE		Total	
Blocksize	I/O per second	MB/s	I/O per second	MB/s	I/O per second	MB/s
512B	**321934.77**	157.19	**287541.73**	140.40	609476.50	297.60
1K	318150.40	310.69	286678.43	279.96	604828.83	590.65
2K	313147.47	611.62	284644.87	555.95	597792.33	1167.56
4K	**310677.30**	**1213.58**	**280163.80**	**1094.39**	**590841.10**	**2307.97**
8K	**302921.17**	**2366.57**	**263663.82**	**2059.87**	**566584.98**	**4426.45**
16K	209836.37	3278.69	171390.83	2677.98	381227.19	5956.67
32K	142615.80	4456.74	136155.45	4254.86	278771.25	8711.60
64K	82802.35	5175.15	78449.52	4903.09	161251.87	10078.24
128K	44156.40	**5519.55**	40379.13	**5047.39**	84535.53	**10566.94**
256K	22071.51	5517.88	19993.28	4998.32	42064.79	10516.20
512K	10952.70	5476.35	9996.10	4998.05	20948.80	10474.40
1M	5504.90	5504.90	4947.77	4947.77	10452.67	10452.67

Conclusion

The results presented in this paper illustrate that the Cisco UCS VIC 1240 with the I/O Port Expander card running on RHEL Version 6.2 is the best-in-class I/O adapter solution for data center applications.

The Cisco UCS VIC 1240 is a fully standards-compliant, converged network adapter that delivers industry-leading storage-throughput performance. The Cisco UCS VIC 1240 with the I/O Port Expander card is an excellent choice for I/O-intensive server workloads.

Cisco UCS Virtual Interface Card 1240: http://www.cisco.com/en/US/products/ps12377/index.html

Printed in USA

C11-721280-03   02/13