Performance Study of

2nd Generation IBM Power Systems SAS RAID Adapters

Designed for Solid State Storage

Version 2

September 2015

By Lakshmi Subramanian, Clark Anderson, Mugdha Dabeer and Steven Kerchberger
IBM Systems and Technology Group

Table of Contents

Executive overview

In January 2014 IBM announced a new generation of SAS RAID adapters optimized for solid state storage speeds. These include the PCIe3 12GB Cache RAID SAS Adapter Quad-port 6 Gb x8, Custom Card Identification Number (CCIN) 57CE and the PCIe3 RAID SAS Adapter Quad-port 6 Gb x8, CCIN 57B4. Both are constructed with PCIe Gen3 ASICs that evolved from a prior generation FPGA based family of adapters also optimized for solid state storage. The ASIC includes higher internal and external bus bandwidths, more function migrated from software to hardware accelerators, faster microprocessors and new features.

These adapters can control HDDs and SSDs installed in EXP24S SFF Gen2-bay drawers and FC 5803 Expansion units. They perform 2x faster than the previous FPGA based adapters. This study shows the performance comparison between the PCIe Gen2 generation of adapters with the newer PCIe Gen3 ASIC based adapters.

The newer generation adapters studied in this paper bring a new level of high-performance density to the server storage market as well as affordability. The lower cost ASIC also makes affordable imaginative new solutions that analytically crunch through a myriad of data in a short time period.

Abstract

Results from the 1st release of this paper in 2014, studied the throughput of the IBM newer generation 57CE and 57B4 adapters with four-corners workloads and three online transaction processing (OLTP) workloads. The three OLTP applications were also used to show response times provided by the direct attached storage (DAS) subsystems. The current 2nd release of this paper in 2015, showcases the performance of 57CE adapters in POWER8 PCIe Gen3 slots.

The newer generation 57CE and 57B4 SAS adapters can be configured with varying numbers of SSDs , HDDs and protection levels. The results in this paper can be used to help guide system designers to define how many devices are needed to support applications. Additionally, this paper will help readers decide the applications for which these adapters are well suited. Readers can also use the resultant rules of thumb to help properly size and tune storage I/O subsystems with the new SAS adapters for many applications.

Introduction

IBM's next generation SAS RAID adapters, 57CE and 57B4 announced in January 2014, continue the legacy of enterprise class direct attach storage for Power Systems with additional features for optimized Solid State Drive (SSD) performance. The technology used in these adapters is part of IBM's overall Flash investment to bring break-through Flash performance to client's applications. This is IBM's 10th generation of enterprise storage adapters and the 2nd generation designed and optimized for demanding SSD applications. Over 10 patents directly related to hardware accelerators and overall performance are implemented in these adapters, demonstrating IBM's commitment to innovation that matters.

Benefits of 57CE Cache RAID SAS Adapter

The 57CE utilizes IBM's latest RAID on Chip ASIC controller. This new ASIC utilizes various advanced hardware accelerators designed to improve storage performance while maintaining enterprise class requirements for data protection and availability. The ASIC, commonly referred to as a RoC (RAID on Chip), contains an integrated PowerPC microprocessor. As an integrated part of the RoC, the microprocessor runs 2X faster than the separate microprocessor used in the prior generation 57B5 adapter. The 57CE includes non-volatile, fully redundant cache over 6X larger than the cache in the 57B5. The non-volatile nature of the cache is enabled by a patented design to save/restore the cache to on-card flash in the event of a power outage. Similar to the 57B5, there are no batteries used in this design.

Note:

The announcement made in January 2014 was for support of the 57CE in POWER7 systems via 5802/5877 12X I/O drawers even though 57CE was Gen3 capable. As such, the 57CE will only run at Gen1 speeds in those configurations. Hence the tests with the POWER7 System Configuration (Test Configuration #1 –

57CE & 57B4 in PCIe Gen2 Hardware) and its results were performed with 57CE in PCIe Gen1 hardware in 2014 when the original paper was published.

Now that POWER8 systems have been announced which include PCIe Gen3 slots, the 57CE adapter was measured in those slots for this paper. The results are compared to the previously published 57CE adapter results when running in PCIe Gen1 slots in POWER7 systems.

This paper compares the performance of,

- 57CE adapters placed in the PCIe Gen1 hardware on POWER7 systems via 5803/5873 2X I/O drawers (Results from 1st release of this paper in 2014)
- 57CE adapters placed in the PCIe Gen3 hardware on POWER8 system's CEC (Results from current study in 2015)

Note:

It must be noted that, in the system configuration with PCIe Gen1 hardware, the 57CE ran only at Gen1 speeds. The 57CE has four x4 6Gb/s SAS connectors compared to only three, two x4 and one x3, SAS connectors on 57B5. This enables a larger number of devices. With 57CE, up to 48 SSDs or 96 total devices (HDDs and SSDs) may be attached enabling larger storage configurations with fewer adapters. The 57CE also supports 1 or 2 adapter-to-adapter (AA) links providing the potential for better performance in write intensive workloads. Internally, an integrated compression/decompression hardware accelerator effectively doubles the size of the cache, while the cache's controller quadruples in bandwidth, readying the 57CE to unleash PCIe Gen2 and Gen3 throughput rates.

The firmware that runs on the integrated PowerPC microprocessor has been significantly enhanced to provide new functions and deliver improved SSD performance as compared to the prior generation of adapters. RAID 10 is now fully supported with the 57CE on AIX and Linux. Extensive redesign of the firmware is ready to deliver a new level of performance for RAID 5 write operations. Firmware architecture and implementation to take advantage of the new generation RoC delivers RAID 5 small random write operations better than 155,000 IOps, a 2x increase over the 57B5. Additionally, RAID 6 performance results with SSDs are much closer to RAID 5 results than in the past. If using SSDs, the additional protection provided by RAID 6 no longer has the performance penalty typical of RAID 6 configurations.

Benefits of 57B4 RAID SAS Adapter

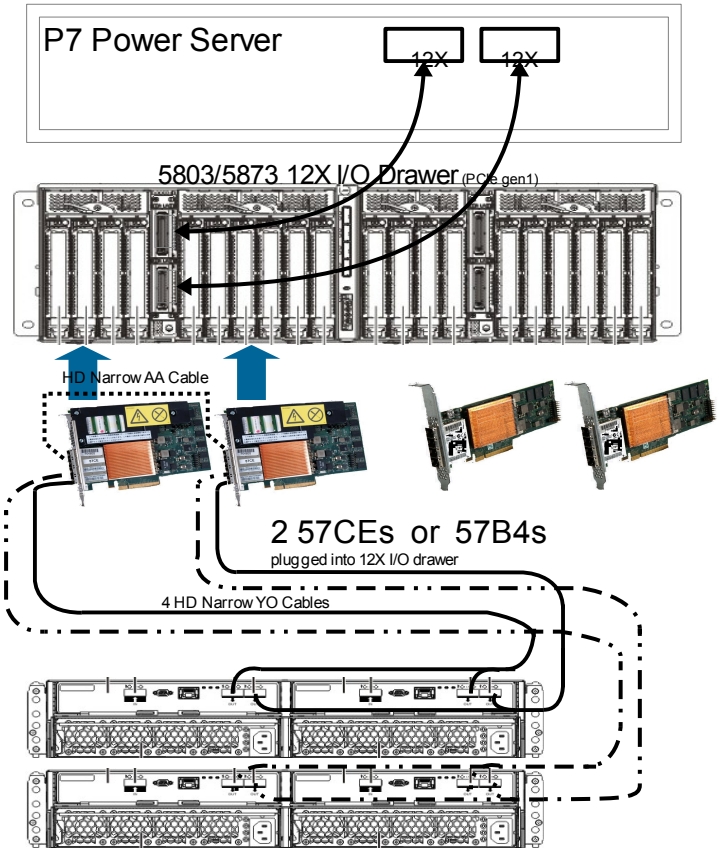
The 57B4 is a derivative of the 57CE adapter. It utilizes the same next generation advanced RoC ASIC and other components. Like the 57CE, the 57B4 is fully PCIe Gen3 capable. And the 57B4 has four 6Gb/s SAS x4 connectors.

The 57B4 is a cacheless RAID adapter. The removal of the cache allows for a more cost effective solution especially for legacy HDD (spinning disk drive) connectivity. Additionally, the 57B4 provides support for LTO5 and LTO-6 SAS tape devices. Even though the 57B4 does not have a dedicated write cache, it still supports up to 8 tape drives per adapter.

Test Setup Overview

Test Configuration #1 – 57CE & 57B4 in PCIe Gen2 Hardware

A POWER7 based Power System was connected to a 5803 I/O drawer via a 12X interface. The 5803 drawer had the newer generation 57CE and 57B4 adapter pairs installed in it.



Drawing 1: Hardware Configuration

Each adapter's top-most ports were connected together with a HD Narrow AA cable.

2 HD Narrow X cables were used for this test to connect the adapters to two EXP24S drawers each containing 24 CCIN 58B9 387GB (gen3 eMLC) SSDs. This is the largest SSD configuration currently supported.

2 X cables were used instead of the 4 HD Narrow YO cables as shown in Drawing 1. Even though YO cables were not used in the tests, the performance of 4 YO cables results in the identical performance as 2 X cables due to the exact same number of SAS lanes used.

YO cables are recommended for concurrent maintenance capability when EXP24S drawers are configured for a single, Mode 1, SAS zone. X cables are recommended when the EXP24S is configured for Mode 2 SAS zoning.

For comparisons with the previous generation PCIe2 RAID SAS adapter Dual-port 6 Gb (CCIN 57C4) and PCIe2 1.8 GB Cache RAID SAS adapter Tri-port 6 Gb (CCIN 57B5) those adapters were placed in PCIe Gen2 slots in a Power server.

Note this difference:

Since the 5803 drawers have PCIe Gen1 slots, the newer generation 57CE and 57B4 ran at PCIe Gen1 speeds which is slower than the PCIe Gen2 speed of older generation adapters. This is important to realize when looking at workloads that are limited by PCIe link speed.

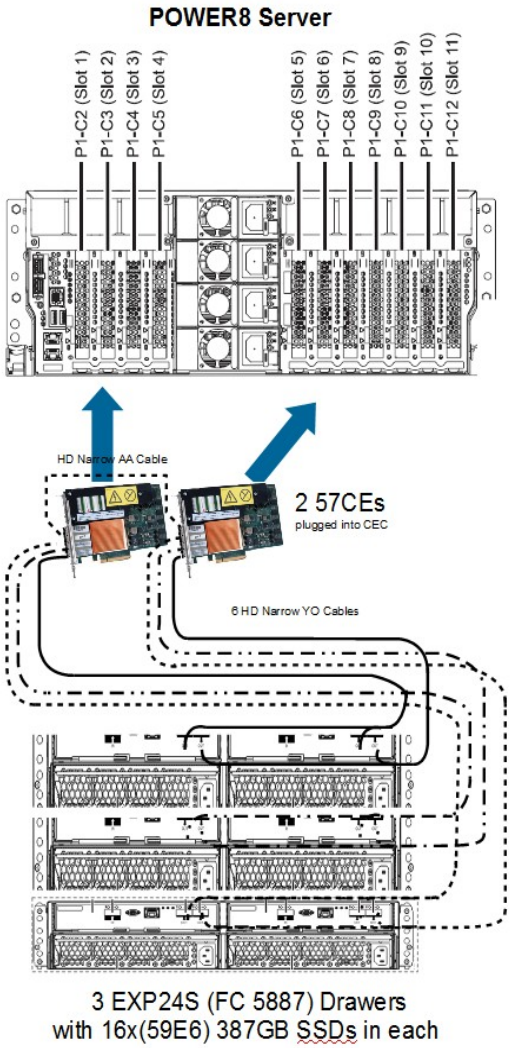
The 12GB caching 57CE adapters can be ordered with Feature Code (FC) EJ0L. The non-caching 57B4 adapters can be ordered with Feature Codes EJ0J or EJ0X, depending upon what enclosure will be used and whether they are intended for DASD (SSD or HDD) use or LTO5/6 tape use.

A single AA cable was used for these tests. When only 1 or 2 EXP24S enclosures are used and high write throughputs are expected, 2 AA cables are recommended.

Test Configuration #2 – 57CE in PCIe Gen3 Hardware

A POWER8 based Power System was connected to the EXP24S I/O drawers via 2x57CE adapters.

Three EXP24S drawers with drives were connected to the adapters as shown. The 57CE adapter pair was installed in the Gen3 Slots of the POWER8 System.



Drawing 2: Hardware Configuration

Each adapter's top-most ports were connected together with a HD Narrow AA cable.

3 HD Narrow X cables were used for this test to connect the adapters to three EXP24S drawers each containing 16 CCIN 58B9 387GB (gen3 eMLC) SSDs.

This is the largest SSD configuration currently supported. The previous test with gen 1 slots only used 2 drawers. This test used three since the higher speed PCIe Gen3 slots are able to take advantage of the increased SAS bandwidth available from a third drawer. When the 57CEs are plugged into gen1 slots, the added SAS bandwidth of a third drawer can be taken advantage of. You can still set up a maximum SSD configuration with 2 drawers, but data throughput bandwidth will be limited by those 2 drawers at approximately 8 GB/s.

IBM AIX Performance

An AIX partition using all system resources was built on the Power Server.

For RAID 0 configurations, each hdisk was built using one drive. Thus 48 hdisks were built. For RAID 5 configurations, 3 drives were used to build one hdisk. Thus 16 hdisks were built. For RAID 6, 4 drives were used to build one hdisk. Thus 12 hdisks were built. For RAID 10, 2 drives were used to build one hdisk. Thus 24 hdisks were built.

A pair of adapters was used in each case for high availability. The adapters were used in active-active configuration to maximize performance. This was done by optimizing an equal number of arrays to each adapter in a pair. Load balancing between the adapters was thus handled by the OS

Performance tuning on the system

The following parameters were changed on the system to maximize performance:

1. Hardware pre-fetch was disabled with the following command:
    - dscrctl -n -b -s 1
2. Schedo (disk proc stats collection) was turned off by:

    schedo -o proc_disk_stats=0
3. atime updating was disabled with the following command:
    - mount -o noatime,remount /
4. RAS was turned off to disable lightweight memory trace with the following command:
    - raso -y -r -o mtrc_enabled=0
5. Queue depth on each hdisk was set to 64 for the Gen1 slots testing and for all the Gen3 slots testing the default queue depth of 16 was set to each disk in an array.

    Experimentation with the 57CEs in Gen3 slots showed that peak throughput could be obtained with shallower queue depths than were used in the Gen 1 testing. The reduced latency of the faster PCIe bus and the fewer chip hop count of the POWER8 test configuration vs the POWER7 I/O topology enabled the lower queue depths to be used. The queue depth was set by the below command.

    - chdev -a queue_depth=64 -l hdisk#

The adapters were tuned to best performance with the following:

1. MSI-X (multiple channnels) on the adapter cards were used by executing the following for each adapter:
    - chdev -l sissas# -a nchan=15 -P
    - bosboot -aD
2. The IO traffic was equally divided between the two cards by optimizing half of the arrays beneath each adapter. Setting the HA access characteristics for a disk array specifies which controller is preferred to be optimized for the disk array and perform direct read and write operations to the physical devices. The HA access characteristics can be viewd by completing the following steps.
    - Navigate to the IBM® SAS Disk Array Manager by using the steps found in Using the Disk Array Manager
    - Select Manage HA Access Characteristics of a SAS Disk Array.
    - Select the IBM SAS RAID controller.
    - HA access characteristics are displayed on the IBM SAS Disk Array Manager screen.

The following access state settings are valid.

Optimized: The selected controller performs direct access for this disk array. This gives I/O operations that are performed on selected controller optimized performance compared to the remote controller.

Non Optimized: The selected controller performs indirect access for this disk array. This gives I/O operations that are performed on selected controller nonoptimized performance compared to the remote controller.

Cleared: Neither an Optimized nor Non Optimized access state has been set for this disk array. By default the disk array is optimized on the primary controller.

The HA access characteristic can be displayed on either the primary or secondary controller. However, as with all other disk array management, the HA access characteristics can only be changed from the primary controller. Setting the preferred HA access characteristic is accomplished by selecting one of the disk arrays. This will bring up the Change/Show HA Access Characteristics of a SAS Disk Array screen. The Preferred Access state can be changed when the disk array is selected from the primary controller. The Preferred Access state cannot be changed if the disk array is selected from the secondary controller. If this is attempted, an error message is displayed. Changing the Preferred Access state from the primary controller stores the settings in the disk array and automatically shows the opposite settings when viewed from the secondary controller.

By default all disk arrays are created with a Preferred Access state of Cleared. To maximize performance, when appropriate, create multiple disk arrays and optimize them equally between the controller pair. This is accomplished by setting the Preferred Access to Optimized for half of the disk arrays and Non Optimized to the other half.

3. Queue depth on the cards was set to maximum with :
    - chdev -l sissas# -a max_cmd_elems=8,984,0 -P

Comparisons

We measured the performance of RAID 5 since it typically provides the lowest purchase price of a subsystem with data redundancy. Testing of RAID 0 assesses performance with mirroring at levels above the IOAs (typically in the OS). Not enabling the mirroring function in the IBM AIX® OS allowed the raw RAID 0 performance to be assessed. Knowing raw RAID 0 speed allows OS-managed mirroring by any supported OS to be estimated. RAID 10 managed by the IOAs was also measured. RAID 6 was also investigated for the caching adapter. The extra device activity resulting from RAID 6 writes is the reason we do not recommend running a cacheless adapter with RAID 6.

With the IOA running under AIX, users have the ability to disable the redundant, super capacitor-backed write caches. The write caches are enabled by default. Unless explicitly mentioned all tests were run with cache enabled. Tests were executed with write caches enabled and disabled to show when it helps or hurts the perceived speed of a particular workload. The IOA write cache is managed with the IBM SAS Disk Array Manager utilities that can be started from System Management Interface Tool (SMIT), command line or Diagnostics.

Four corners description

The four corners of adapter performance are as follows: random read, random write, sequential read and sequential write speed. Random accesses are generally small in size, while sequential accesses tend to be larger.

Our first test writes a 4 KB block of data in a completely random location over all of the drives to simulate the random access at varying number of processes. The results reported are in average IOPS (I/O Operations per second) over the duration of the test. The random read 4 KB tests were performed similarly with a 4 KB block of data read from completely random locations over the array and results were reported in average IOPS.

To measure sequential performance we ran a 1MB sequential test over the entire span of the drive at a queue depth of 1, 24, 48 and 96. The results reported are in average MB/s over the entire test duration.

OLTP Benchmark Descriptions

Three different workloads were run that mix reads and writes, various transfer lengths, spatial localities and request rates. These workloads simulate the I/O operations performed by various OLTP applications.

OLTP1: Read/write ratio is 60/40 with 4 KB op size. I/O operations are random over the full capacity of each device, so there is little opportunity for cache hits in the adapter.

OLTP2: Read/write ratio is 90/10 with 8 KB op size. I/O operations are random over the full capacity of each device, so there are almost no cache hits in the adapter.

OLTP3: Read/write ratio is 70/30 with 4 KB op size. Half of the read operations are random and half are intended to elicit cache hits. 33 percent of write operations are random, 33 percent are to logical block addresses recently read, and 34 percent are intended to elicit write cache hits.

Results

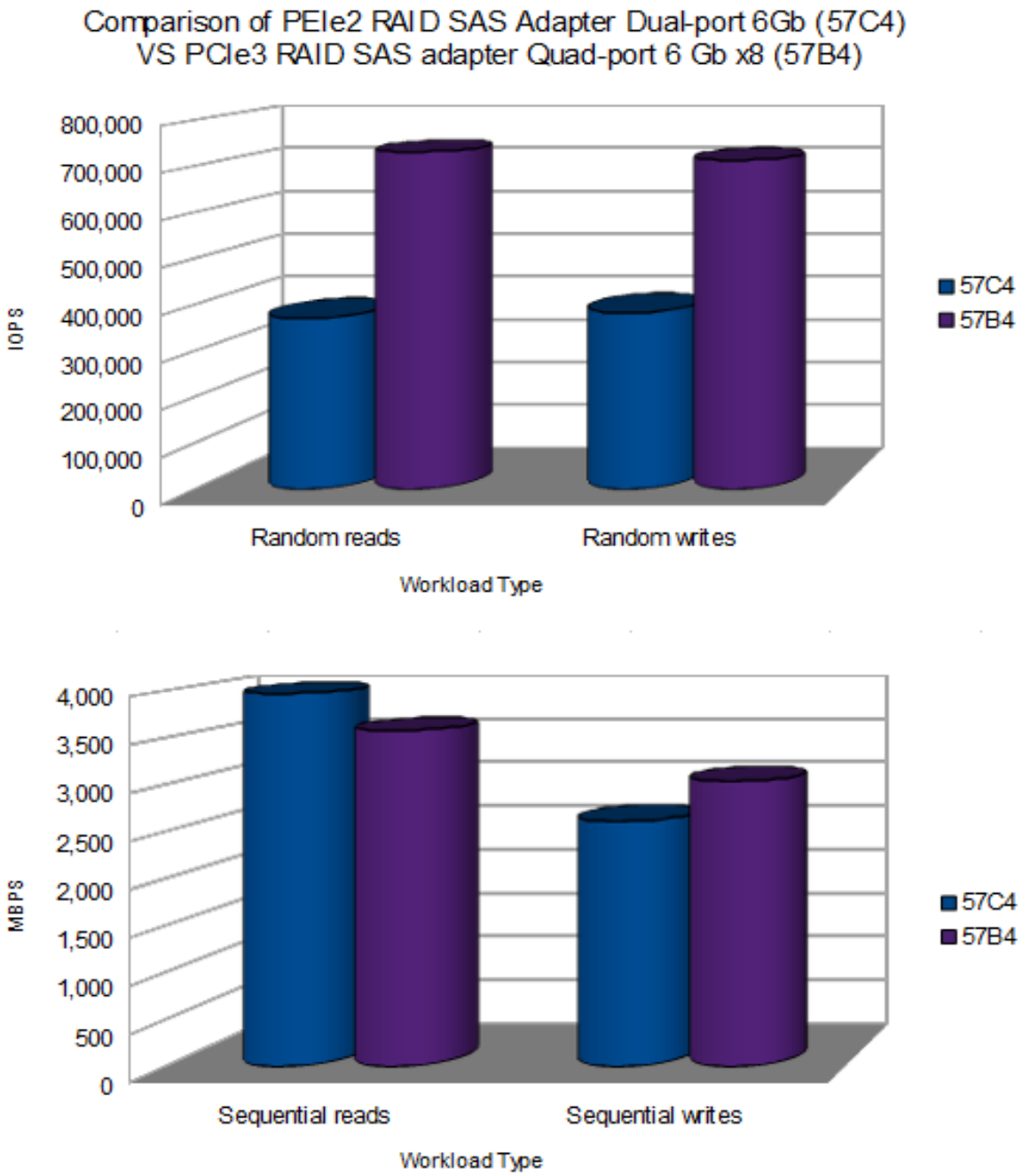
Four Corners results:

The following graphs show performance gain over the older generation of SAS adapters for the four corner cases of the workload in RAID 0 protection configurations.

Drawing 3: Cacheless Raid-0 4-Corners Results compares older generation cache-less 57C4 with the newer generation cache-less 57B4 adapter.

The top graph compares random read/write workloads for cache-less adapters. The newer 57B4 has almost 2X the throughput when compared to the older generation adapter.

The bottom graph compares sequential workloads. On the POWER7 systems, the newer generation adapters were initially supported only in Gen1 slots whereas the older generation are supported in Gen1 and Gen2 slots. For this study, to get best performance from the older generation, the cards were placed in Gen2 slots. Hence, sequential reads for the older generation appear faster than the newer generation, as the 57B4 hit the Gen1 slot limit and not an internal adapter limit.

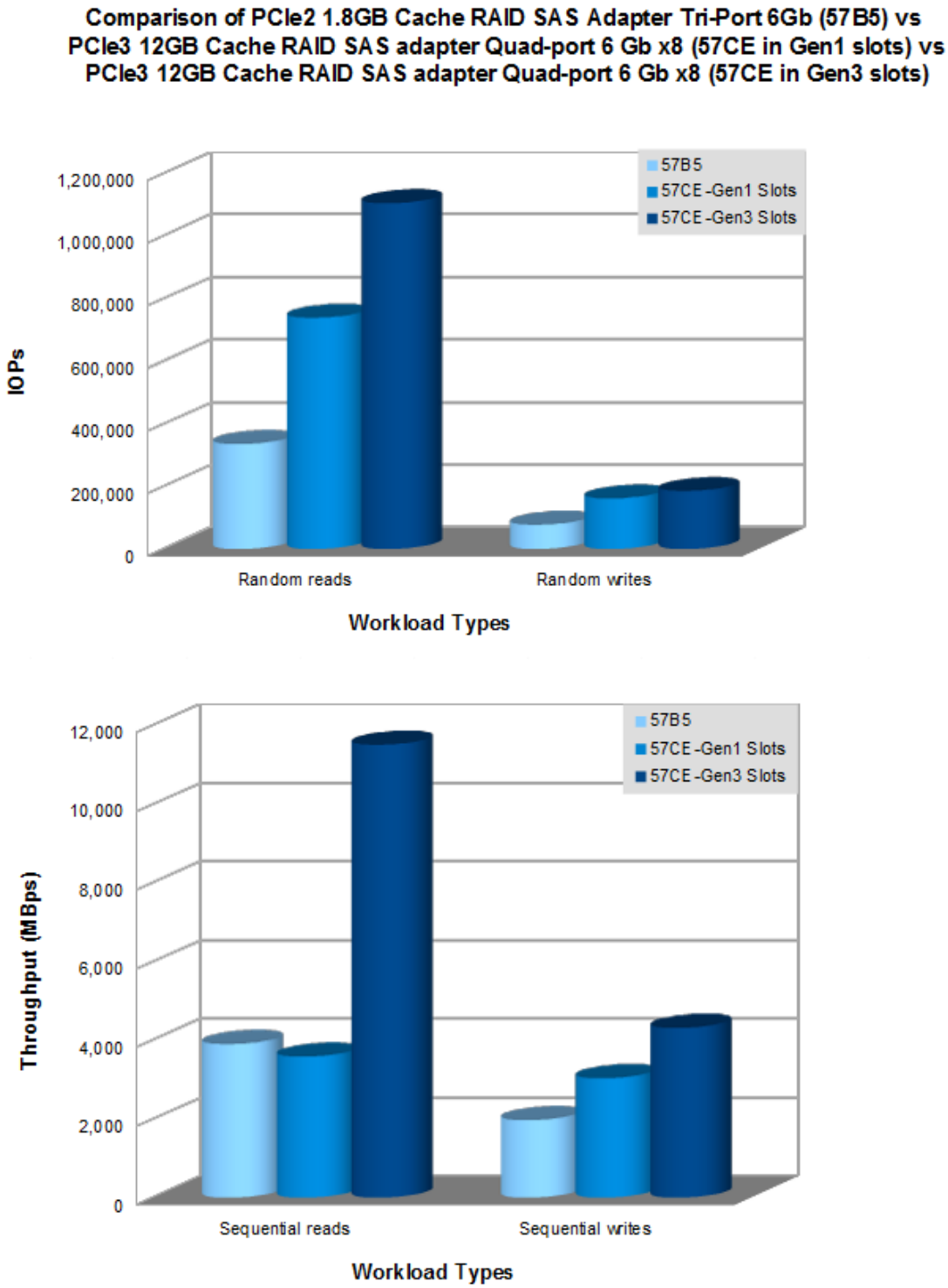


Drawing 3: Cacheless Raid-0 4-Corners Results

Drawing 4: Cached Raid 0 4-Corner Results top graph gives a comparison for the random read/write workloads for the caching adapters. We see almost 2X throughput with the newer generation adapter. This performance improvement can be attributed to several factors like the use of POWER8 systems higher speed PCIe Gen3 slots, that are able to take advantage of the increased SAS bandwidth available from the use of three IO drawers.

Drawing 4: Cached Raid 0 4-Corner Results bottom graph compares sequential workload performance, when the 57CE cards were placed in Gen1 slots in POWER7 systems and the same cards placed in Gen3 slots in POWER8 systems. Since the older generation cards (57B5) are supported in Gen1 and Gen2 slots, to get best performance from the older generation, the cards were placed in Gen2 slots for this study. Hence, sequential reads for the older generation appear faster than the newer generation cards in Gen1

slots, as the 57CE hit the Gen1 slot limit and not an internal adapter limit. However, there was a performance improvement when the 57CE cards were placed in Gen3 compared to placing them in Gen1. Sequential write performance increased too when 57CE cards were placed in Gen3 slots.



Drawing 4: Cached Raid 0 4-Corner Results

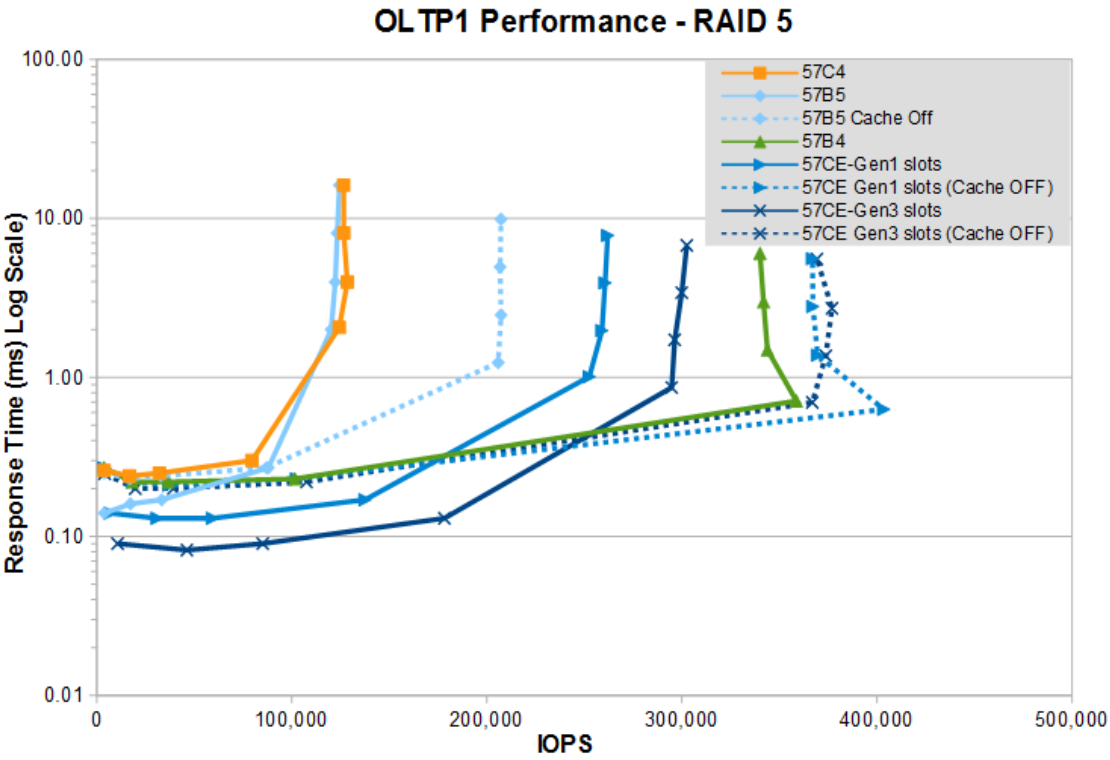
The following table provides a 4-corners throughput comparison of all the RAID protection types. The 57CE adapters in Gen3 slots provided the best performance with 1.1 million IOPs for Random Reads, ~12GBps on Sequential Reads and 4.4GBps on Sequential Writes for Raid0 with 48 SSDs. 48 SSDs were used to support the maximum SSD capacity configuration (Refer to table on page 18), while the adapters throughput could have been saturated with ~26 SSDs .

Work load Type	57CE on Gen1 Slots(CacheOn)				57CE on Gen3 Slots(CacheOn)				57B4 on Gen1 Slots (Cacheless)			
	RAID 0	RAID 5	RAID 10	RAID 6	RAID 0	RAID 5	RAID 10	RAID 6	RAID 0	RAID 5	RAID10	RAID 6
Random Read (IOps)	762K	824K	710K	571K	1.10M	1.07M	1.09M	893K	764K	734K	728K	398K
Random Write (IOps)	161K	141K	171K	138K	184K	158K	192K	150K	713K	747K	230K	221K
Sequential Read (GBps)	3.75	3.75	3.75	3.75	11.7	9.4	9.5	9	3.75	3.75	3.75	3.75
Sequential Write (GBps)	3.2	3	3	1.36	4.4	3.2	3.9	2.03	3.2	0.85	2.2	1

The reason for the random write IOPS with 57CE being lower than the 57B4 is due to the cost of managing the cache not being offset by cache benefit. The cache benefit to throughput due to cache hits is low with a highly random workload that exceeds the back end storage's ability to destage the cache, like the purely random write workload seen in the above table. Thus, the advantage of 57CE cache can be seen for most practical workloads which have cache hits,Example OLTP3 kind of workload which have some cache hits. This advantage is not seen for 100% Random Write workload since there are no cache hits.

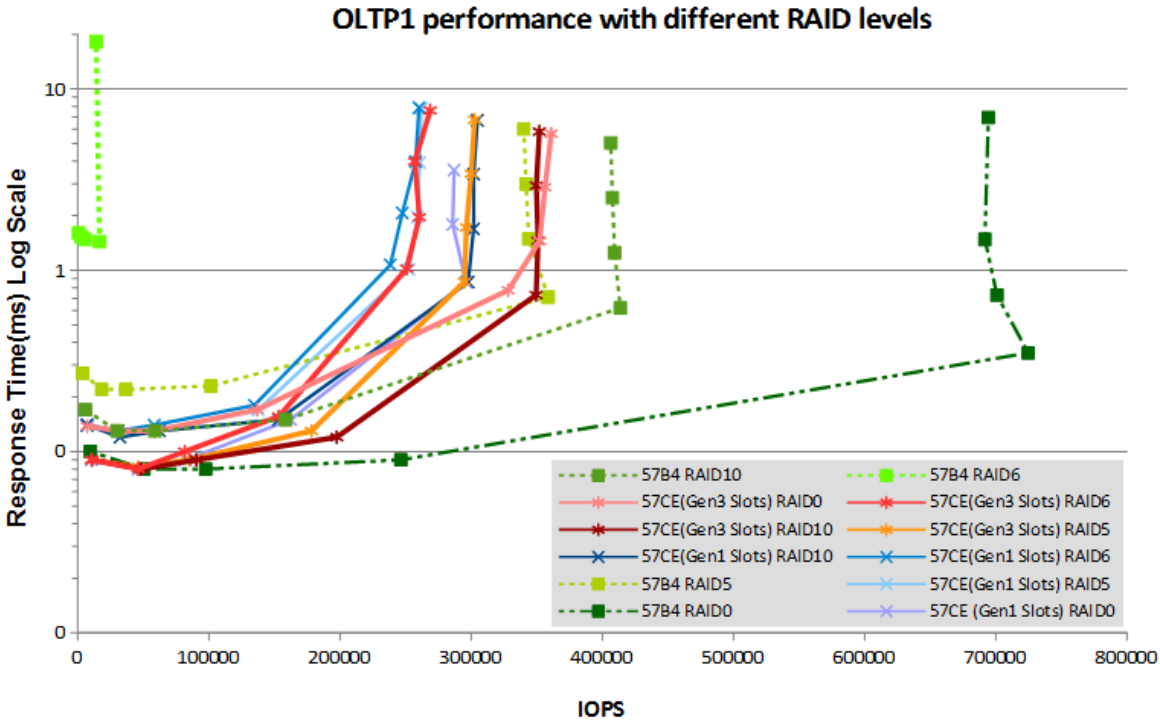
OLTP1,2,3 results

Drawing 5: OLTP1 Generation Comparison shows performance of the ASIC generation SAS adapters against the older FPGA based adapters. The maximum IOPS that we can get with the newer generation adapters is almost twice that of the older generation adapter cards. The IOPS with the new cached adapters 57CE is also twice as that of the older generation cached adapters. Notice that the 57CE's in Gen1 and Gen3 slots response time stays much lower than the 57B5's through much higher throughputs. This is due to the larger effective cache size. Since the OLTP1 workload has some writes and writes to cache occur much faster than writes that must make their way to the back-end storage devices, at low throughputs, where caches are able to be kept below their full threshold, both caching adapters have lower response times than both non-caching adapters. Also, the performance was even better when the 57CEs were placed in Gen3 slots compared to Gen1 slots.

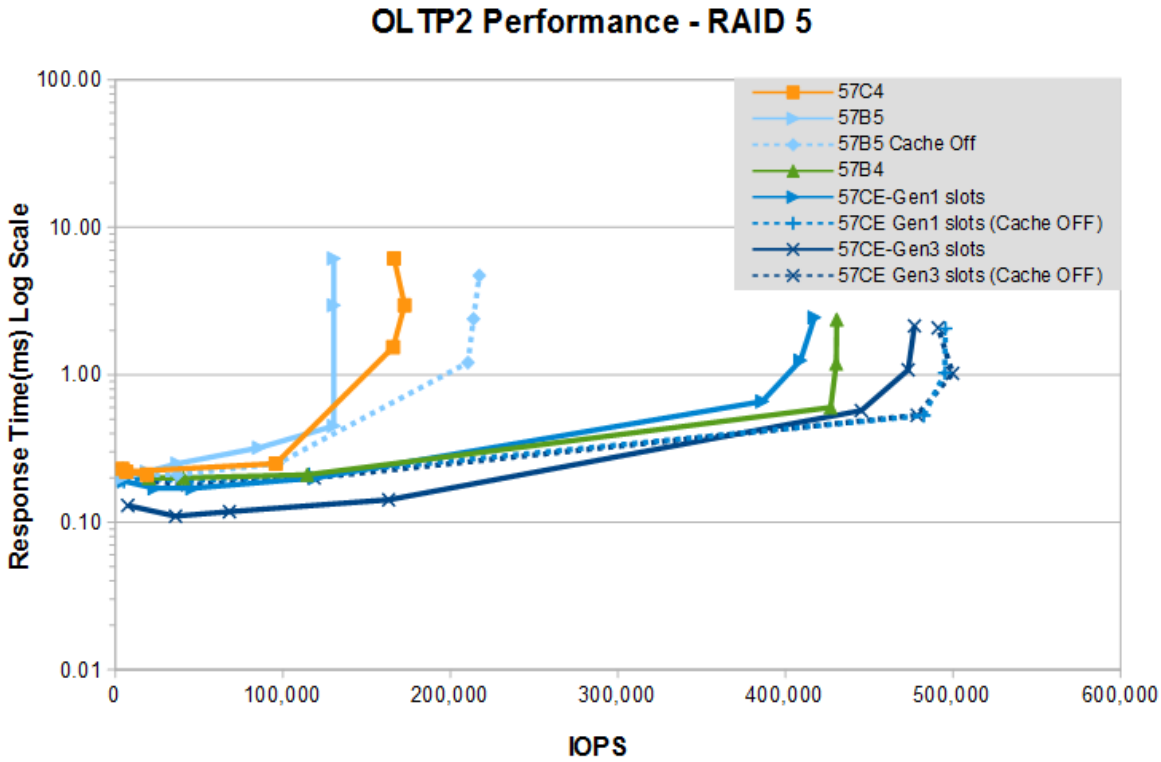


Drawing 5: OLTP1 Generation Comparison

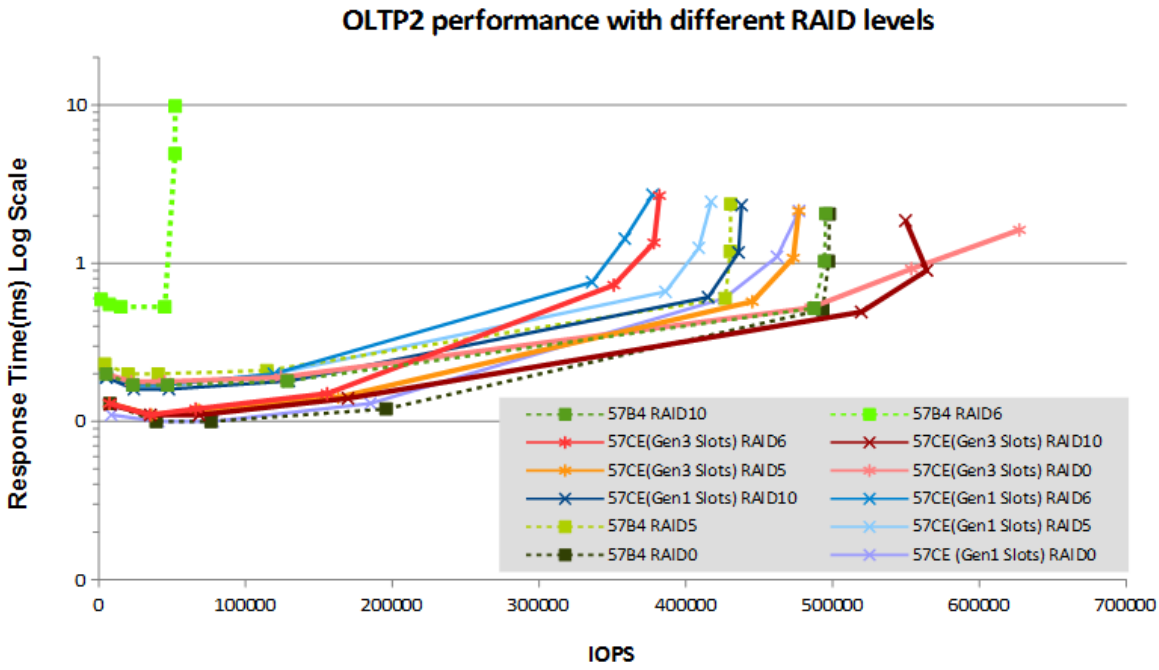
The next figure shows the relative differences of all of the supported RAID types of the new generation adapters to help you make protection/speed tradeoffs with workloads that may be similar to the OLTP1 workload.



Drawing 6: OLTP1 RAID Comparison

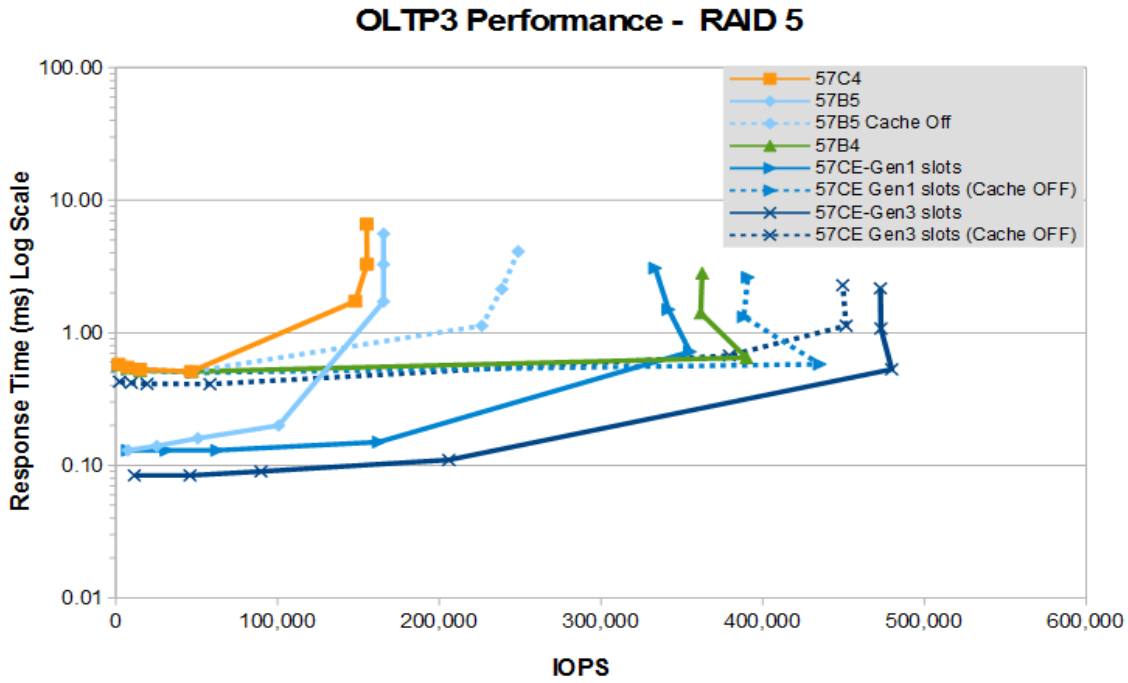


Drawing 7: OLTP2 Generation Comparison

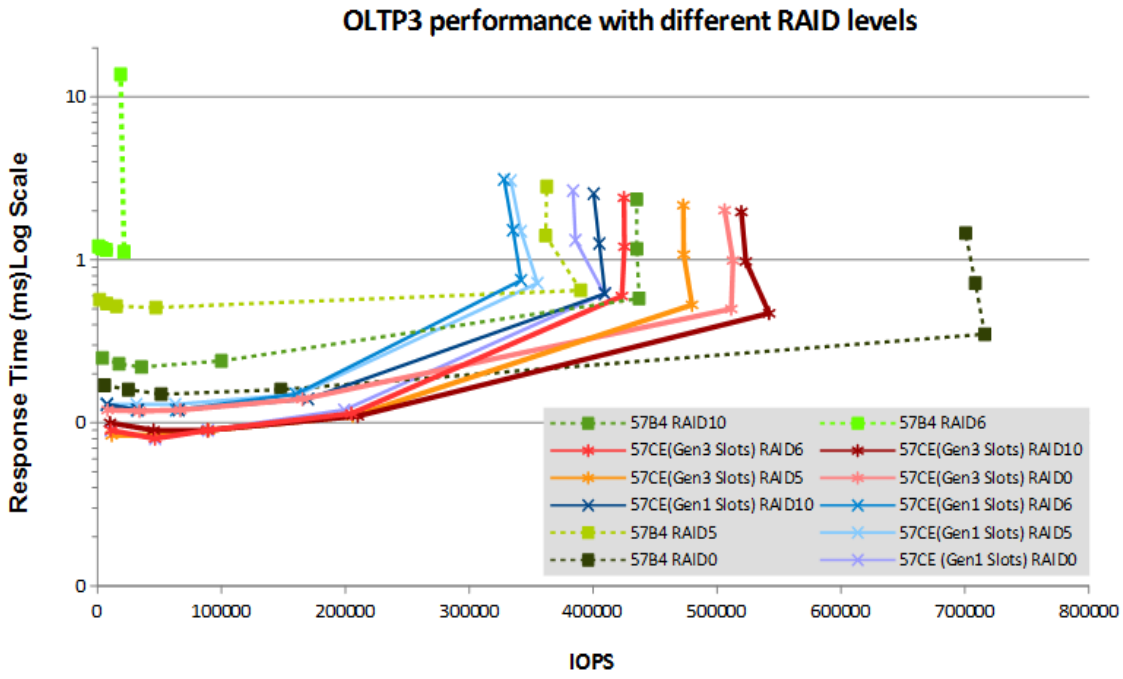


Drawing 8: OLTP2 RAID Comparison

We do not recommend using RAID 6 without cache. The performance is drastically impacted without a cache. Performance is not impacted on adapters with cache.

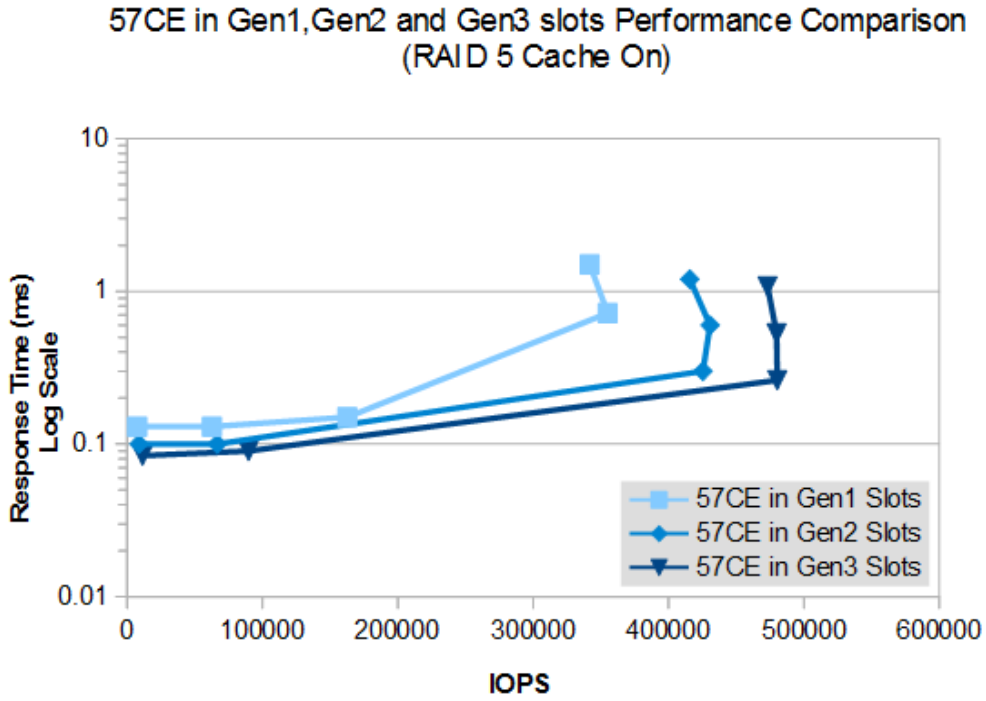


Drawing 9: OLTP3 Generation Comparison



Drawing 10: OLTP3 RAID Comparison

The below graph shows the OLTP3 performance comparison of 57CE in Gen1,Gen2 and Gen3 slots. The Gen3 hardware performed the best.



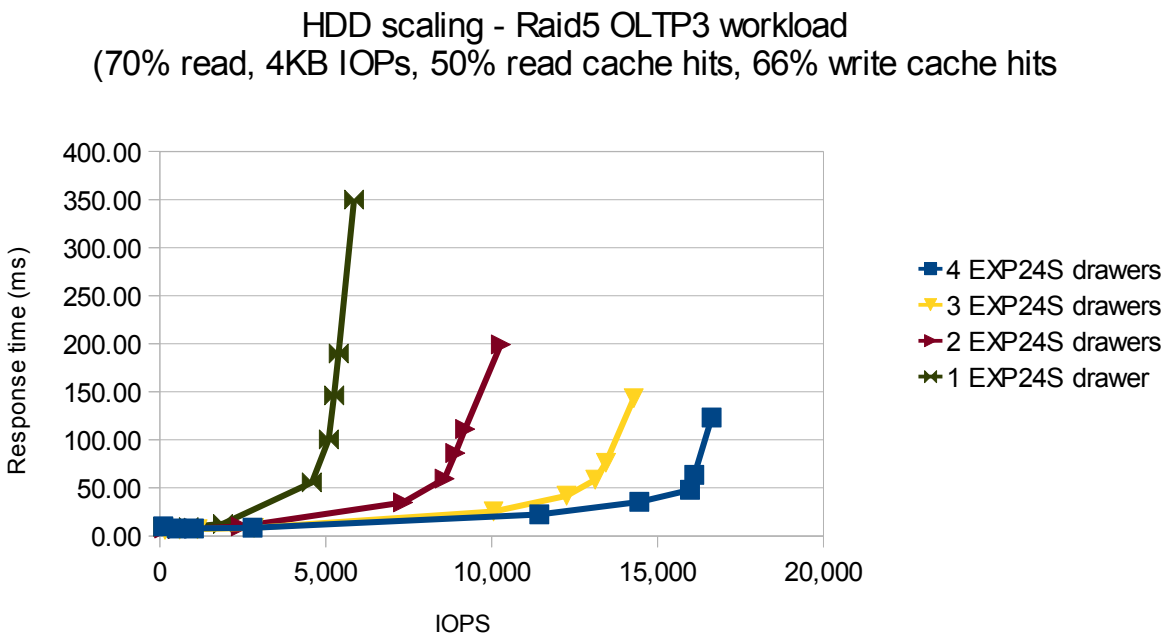
Drawing 11: 57CE Slots Generation Comparison

Advantage of 57CE Cache:

On average, maximum IOPS with 57CE are lower than what the 57B4 can support, due to the cost of managing the cache not being offset by cache benefit. The cache benefit to throughput due to cache hits is low with a highly random workload that exceeds the back end storage's ability to destage the cache, like OLTP1 and OLTP2. However, the lower response time advantage of cache can be seen at lower queue depths for those workloads due to the write-back nature of the caches. The throughput of 57CE for an OLTP3 kind of workload, which has a lot of cache hits that eliminate device ops, is almost twice that of the older generation SAS adapters. For many workloads, the queue depth is generally closer to 24. Thus, the advantage of 57CE can be seen for most practical workloads which have cache hits. This advantage is not seen for OLTP1 kind of workloads since there are no cache hits.

HDD scaling with 57B4

The next graph shows the HDD scaling with 57B4 adapter cards. We recommend using a caching SAS adapter with HDDs to avoid large response times. However, the advantage with 57B4 is that you can connect up to 4 EXP24S drawers to a pair of 57B4s, thus up to 96 HDDs. Although we did not test this, one other low cost possible solution would be running the 57B4s in single controller mode, at the cost of no redundant connection resulting in possible lower device availability.



Drawing 12: HDD Scaling with 57B4s

The graph above shows close to linear scaling of performance with increase in number of HDDs.

Conclusion

The following rules of thumb should guide you, along with the throughput and response time numbers shown in this paper and detailed knowledge of application subsystem requirements, to size the impacts of the new ASIC-based PCIe Gen3 SAS RAID adapters in numerous server applications.

- 57CE I/O adapter transaction workload Tp is approximately 2 times that of previous-generation Power

Systems DAS I/O adapters. The same can be said for the 57B4 adapters. When installed in PCIe Gen2 and Gen3 slots the 57CE adapter sees even more gains.

- There are too many combinations of possible supported SSD-only, HDD-only and SSD/HDD hybrid configurations to mention here. So the following table shows the extreme configurations (SSD only) that can drive the adapters to their limits. The table shows an approximation of how many devices can be fully utilized with a single pair of adapters, where the blue columns describe the latest generation ASIC-based family, yellow columns show the previous FPGA-based family optimized for SSDs and red columns show the adapter family available at the time the Gen1 SSDs were first made available.

	Workload	57B5 & 5887 [3]	57C4 & 5887 [3]	57B5 & 5887 [4]	57C4 & 5887 [4]	57CE & 5887 [3]	57B4 & 5887 [3]	57CE & 5887 [4]	57B4 & 5887 [4]	57CE & 5887 [5]	57B4 & 5887 [5]
Max SSD attach	**N/A**	**24**	**24**	**24**	**24**	**48**	**48**	**48**	**48**	**48**	**48**
with Gen1 eMLC SSD	**W1**	**24**	**24**	**24**	**24**	**48**	**48**	**48**	**48**	**48**	**48**
	W2	**24**	**24**	**24**	**24**	**~36**	**~36**	**48**	**48**	**48**	**48**
with Gen2 eMLC SSD	**W1**	**~19**	**~17**	**~22**	**~20**	**~38**	**~35**	**~45**	**~41**	**48**	**~41**
	W2	**~14**	**~14**	**~15**	**~15**	**~14**	**~14**	**~29**	**~21**	**~47**	**~21**
with Gen3 2.5" eMLC SSD	**W1**	**~10**	**~9**	**~12**	**~10**	**~20**	**~18**	**~24**	**~21**	**~26**	**~21**
	W2	**~8**	**~8**	**~9**	**~9**	~8	**~8**	**~17**	**~13**	**~28**	**~13**
with Gen4 2.5" eMLC SSD	**W1**	**~8**	**~7**	**~10**	**~9**	**~17**	**~15**	**~20**	**~18**	**~23**	**~18**
	W2	**~7**	**~7**	**~7**	**~7**	**~7**	**~7**	**~14**	**~10**	**~22**	**~10**

W1 = transaction/command/DB2 type workload, smaller block, IOPS sensitive, RAID5

W2 = save/restore/large-file type workload, throughput sensitive, RAID5.

Also assumes #5887 is in Mode2 using two SAS ports for higher bandwidth.

[1] This is a simple rule of thumb. Actual reasonable maximum depends on many factors.

[2] It is possible to use #5802 or 5803 I/O drawer instead of #5887 EXP24S Drawer.

Max SSD attach may reduce for W2-like workloads.

[3] When adapters are installed in PCIe gen1 slots.

[4] When adapters are installed in PCIe gen2 slots.

[5] When adapters are installed in PCIe gen3 slots.

- Besides showing the relative differences in speed between the generations of adapters in various PCIe slot types, the above table also shows that the third-generation eMLC SSDs are roughly 2 times the speed of second-generation eMLC SSDs. While fourth-generation eMLC SSDs can be as much as another 30% faster than third-generation models.
- If desired, remaining device slots could be populated with slower drives for storing less frequently accessed data or hot spares.
- It is recommended for the I/O adapter write cache to be kept enabled for response time sensitive applications and/or if there is a high likelihood to get fast write cache hits or IO coalescence. This recommendation is especially important with RAID 5 and RAID 6 protection schemes. Otherwise, it may be worth trying an experiment with the cache disabled running the actual I/O workload the storage subsystem will be supporting.

To fully utilize the rapidly increasing growth rate in SSD speed and capacity, in the most cost efficient manner, by keeping purchase price, real estate, power consumption, cooling and maintanence costs as low as possible, the controllers used to attach mass storage to application servers must improve speed at an equal or greater rate. The tests described here suggests the IBM newer generation 57CE and 57B4 RAID SAS adapters do exactly that.

Acknowledgments

The authors would like to give thanks to Mark Olson, who provided content for this paper, also Mark Schreiter, Sue Baker, Jeff Meute, David Navarro, Robert Galbraith and others who contributed during the testing and review processes.

Legal Information

IBM Corporation
Systems and Technology Group
Route 100 Somers, NY 10589

Produced in the United States
September 2015

This publication could include technical inaccuracies or photographic or typographical errors. This publication was produced in the United States. IBM may not offer the products, services or features discussed in this document in other countries, and the information may be subject to change without notice. References herein to IBM products and services do not imply that IBM intends to make them available in other countries. Consult your local IBM business contact for information on the products or services available in your area.

Information about non-IBM products is obtained from the manufacturers of those products or their published announcements. IBM has not tested those products and can not confirm the performance, compatibility or any other claims

related to non-IBM products. Questions on the capabilities of non-IBM products should be addressed to the suppliers of those products.

Some information in this document addresses anticipated future capabilities. Such information is not intended as a definitive statement of a commitment to specific levels of performance, function or delivery schedules with respect to any future products. Such commitments are only made in IBM product announcements. The information is presented here to communicate IBM's current investment and development activities as a good faith effort to help with our customers' future planning.

IBM, the IBM logo, **ibm.com**, AIX, Power Systems, Storwize and System Storage are trademarks or registered trademarks of IBM Corporation in the United States, other countries or both. If these and other IBM trademarked terms are marked on their first occurrence in this information with a trademark symbol (® or ™), these symbols indicate U.S. registered or common law trademarks owned by IBM at the time this information was published. Such trademarks may also be registered or common law trademarks in other countries. A current list of IBM trademarks is available on the Web at "Copyright and trademark information" at http://www.ibm.com/legal/us/en/copytrade.shtml.

Linux is a trademark of Linus Torvalds in the United States, other countries, or both. InfiniBand, InfiniBand Trade Association and the INFINIBAND design marks are trademarks and/or service marks of the InfiniBand Trade Association. Other company, product and service names may be trademarks or service marks of others.

When referring to storage capacity, 1 TB equals total GB divided by 1000; accessible capacity may be less. 1GB equals 10*9 Bytes. 1Gb = 10*9 bits.

MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE. Some jurisdictions do not allow disclaimer of express or implied warranties in certain transactions; therefore, this statement may not apply to you.

This publication may contain links to third party sites that are not under the control of or maintained by IBM. Access to any such third party site is at the user's own risk and IBM is not responsible for the accuracy or reliability of any information, data, opinions, advice or statements made on these sites. IBM provides these links merely as a convenience and the inclusion of such links does not imply an endorsement.

Information in this presentation concerning non-IBM products was obtained from the suppliers of these products, published announcement material or other publicly available sources. IBM has not tested these products and cannot confirm the accuracy of performance, compatibility or any other claims related to non-IBM products. Questions on the capabilities of non-IBM products should be addressed to the suppliers of those products.

Performance results set forth in this document are based on measurements and projections using standard IBM benchmarks in a controlled environment. The actual performance that any user will experience will depend on considerations such as the amount of multiprogramming in the user's job stream, the I/O configuration, the storage configuration and the workload processed. Therefore, no assurance can be given that an individual user will achieve performance improvements equivalent to the performance ratios stated here.