HP StorageWorks P2000 G3 FC MSA Dual Controller Virtualization SAN Starter Kit Protecting Critical Applications with Server Application Optimization (SAO)

Technical white paper

Table of contents

The HP P2000 G3 FC MSA Virtualization SAN Starter Kit addresses the growing need for server and storage administrators to deliver end-to-end storage, virtualization, and management capabilities across their network infrastructure. The hardware bundle consists of six Host Bus Adapters (HBAs), two FC switches, an MSA storage device as well as two licenses per switch which enable Server Application Optimization (SAO) and Adaptive Networking (AN).

This white paper explains how SAO, an adapter technology that offers Quality of Service (QoS) capabilities, can be used in conjunction with fabric wide QoS (a part of AN) to enable a high performing, virtualized infrastructure. With SAO deployed, admins can protect a virtual machine (VM) and its application by assigning its IO flow to a specific priority level. The QoS priorities are enforced at the hardware level, thereby providing adequate bandwidth during periods of high congestion. This paper also describes how common applications such as email and databases achieve better performance when deployed in specific SAO and MSA configurations.

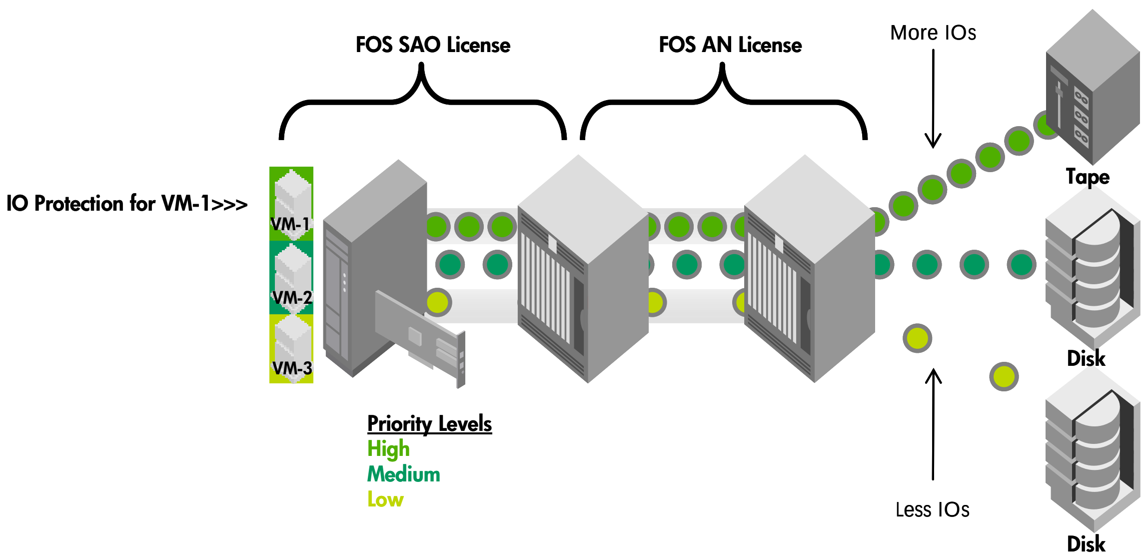
Introduction

As storage and network administrators continue to deploy IO intensive tier 1 applications on both physical and virtual machines, it becomes more and more important to PROTECT the IO flows for applications, thereby enabling the applications have sufficient bandwidth to access data. HP SAO adapter technology deployed in an HP P2000 G3 FC configuration provides hardware enforced priority levels across the entire adapter and switch infrastructure to allow storage and server administrators to meet guaranteed service level agreements (SLAs). The SAO technology extends B-series fabric based QoS algorithms to the adapter and VM level by allocating a percentage of the available bandwidth for given IO flows.

QoS: Protecting an IO flow

QoS can reduce the impact of network congestion by isolating different IO flows from one another, essentially creating isolated IO data paths for VMs and their related applications. QoS support in the server, switch, and storage hardware enables QoS solution for various types of IO workloads (reads/writes). Figure 1—QoS depicts how SAO can protect a VM and its IO flow by assigning a high priority across the adapter and switch infrastructure. In an IO flow configured for high priority more bandwidth is available during peak times, which allows more IOs to be transferred. Without QoS, network congestion can impact the performance of all IO flows within the infrastructure.

Figure 1: QoS. Protect a VM and its IO flow by configuring a high priority level.



SAO implements Virtual Channels for priority levels

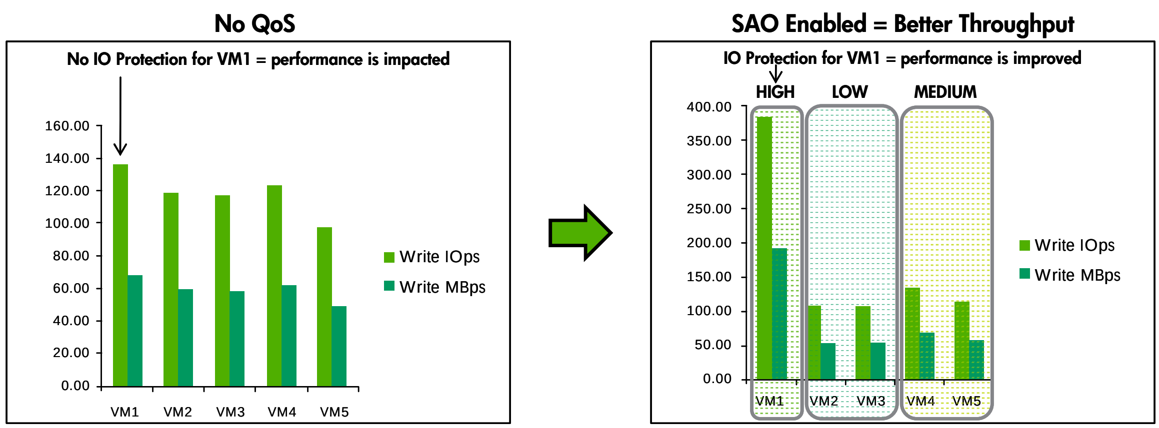
Virtual Channels (VCs) is the ASIC technology behind HP B-series SAO and AN FOS licenses. Together, the VCs support priority levels within the adapter and switch infrastructure to enable QoS solution. Without VCs and SAO deployed, VMs and related applications arbitrate for available bandwidth. In this instance, performance across all VMs and applications will degrade during high sustained throughput levels. Figure 2—Performance impact of SAO and QoS shows how SAO separates the 3 IO flows into distinct and isolated priority levels (high, medium, and low).

Figure 2: Performance impact of SAO and QoS. With SAO enabled, IOs are categorized into priority levels.

SAO improves throughput performance

The main advantage of having hardware enforced priority levels is that VMs and related applications can achieve better performance in terms of throughput and response times. Figure 3—Increased throughput performance with SAO depicts graphically how protecting an IO flow (that is, VM1) improves the throughput for that VM.

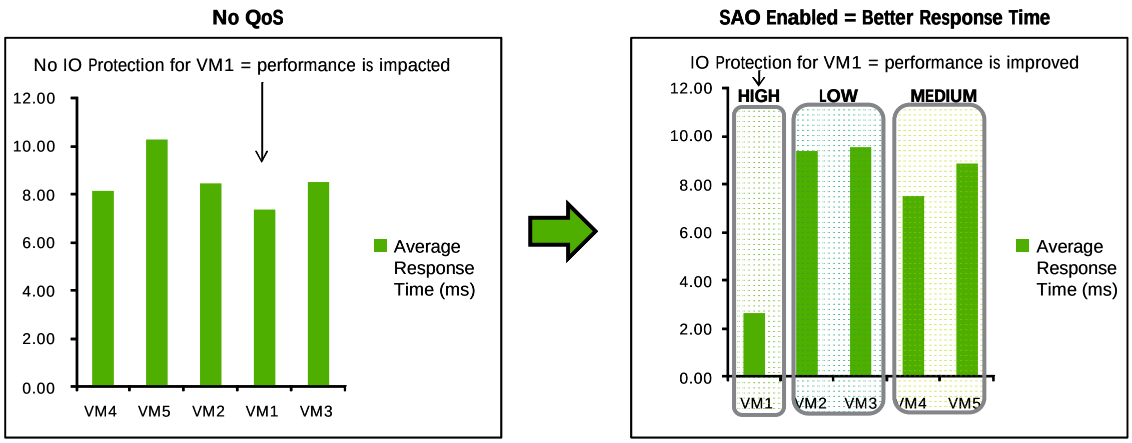
Figure 3: Increased throughput performance with SAO. High throughput for VM1 with SAO enabled.



SAO improves response time performance

Similarly, Figure 4—Increased application response time with SAO depicts graphically how protecting an IO flow (that is, VM1) improves the response time for that VM.

Figure 4: Increased response time performance with SAO. Low response time for VM1 with SAO enabled.



Optimizing applications with SAO: IO characteristics

Users can now deploy enterprise applications within a VM environment that historically required dedicated standalone servers. B-series Host Bus Adapters coupled with SAO and AN enable the user to dedicate connectivity resources down to the physical and virtual port level enabling application SLAs are met while providing industry leading performance capabilities.

Understanding the IO requirements for your applications is critical to enable you to take full advantage of your server's capabilities and properly load your VM servers. Outlined below are three key applications found in nearly all data centers, each having unique IO characteristics. Through SAO you will see that you can now deploy enterprise applications within a VM, prioritize traffic, and leverage unused system resources for lower tier applications.

Email Servers: Microsoft® Exchange

Exchange is a very IO-intensive application. The need for an infrastructure capable of producing high IO is essential—not only the ability to produce high IO but the ability to produce high IO at the block sizes that pertain to Microsoft Exchange. Most block sizes used by Exchange are 4 or 8 kilobytes (KB). Exchange writes data to disk in 8 KB blocks, but the server transfers data on the network at 4 KB random IO. Exchange is very transaction oriented, so it relies heavily on the ability to process random input and output and has constant streams of data going in and out.

Microsoft Exchange Server writes data in block sizes varying from 4, 8, and 32 KB to 64 KB. The block size is dependent on which database store is being written (EDB, STM, LOG). IO operations to a database can range from 8 KB to 1 megabyte (MB). HBAs servicing the disk IO need to be able to sustain the high throughput required by Exchange.

Due to stringent FC standards, no other technology can offer the stability or reliability of Fibre Channel. Only B-series HBAs can isolate individual database and transaction log IO streams for multiple instances of Exchange, maintaining QoS priorities from server to storage with B-series SAO and Adaptive Networking.

Databases: Oracle RAC

Data block size recommendations always vary depending on the application with which the database is being used, as shown in Table 1—Application IO profiles. In general, Oracle recommends that the database block size match or be multiples of the operating system block size. For an online transaction processing (OLTP) environment that drives many small transactions, a smaller block size is appropriate. For larger transactions seen in an online analytical processing (OLAP) environment, a smaller block size would impact the overall performance of the database.

Table 1: Application IO Profiles. Different block sizes for OLTP applications.

Application	R/W	Access Type	Block Size
OLTP Redo	Write	Sequential	2—64 KB
OLTP Database	Write	Random	8—256 KB
	Read	Random	8 KB
OLTP Archive	Write	Sequential	8—256 KB
OLTP Control	Read / Write	Random	2—256 KB
Data Warehouse	Read	Random	8—1024 KB

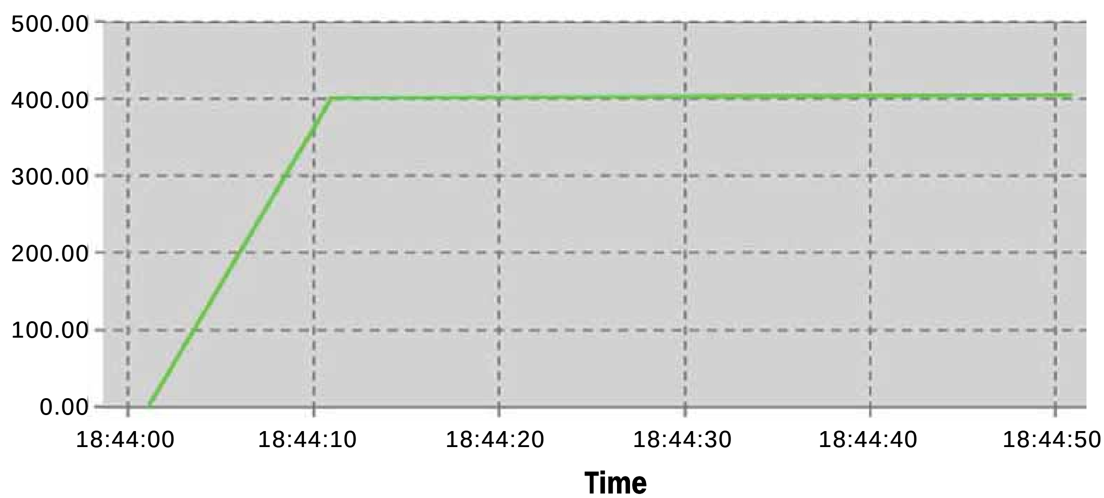
Typically, customers configure multiple smaller OLTP databases that support different business units, such as order entry, human resources, and accounting. These different databases have vastly different IO patterns and also SLAs, though all may share the same storage, storage area network (SAN), and possibly server resources via use of server virtualization technology. Only B-series HBAs can isolate individual database and transaction log IO streams, maintaining QoS priorities from server to storage by using B-series SAO and AN.

IO application comparison with and without SAO

A quantifiable and realized benefit can be seen by running application IO in real world block sizes, leveraging the hardware and software within the bundle. In the example shown we depict an environment that has QoS disabled as well as enabled on our server ports making use of application traffic to represent a real world scenario. This allows visibility into how the application traffic can be prioritized and protected on a given link. Multiple VMs have also been created and added to a running configuration to showcase how protection and prioritization of IO can occur on a link. The performance graphs below leverage the exact same hardware and software configuration. The only difference is enabling the QoS capability for the final two performance graphs shown in Figure 7.

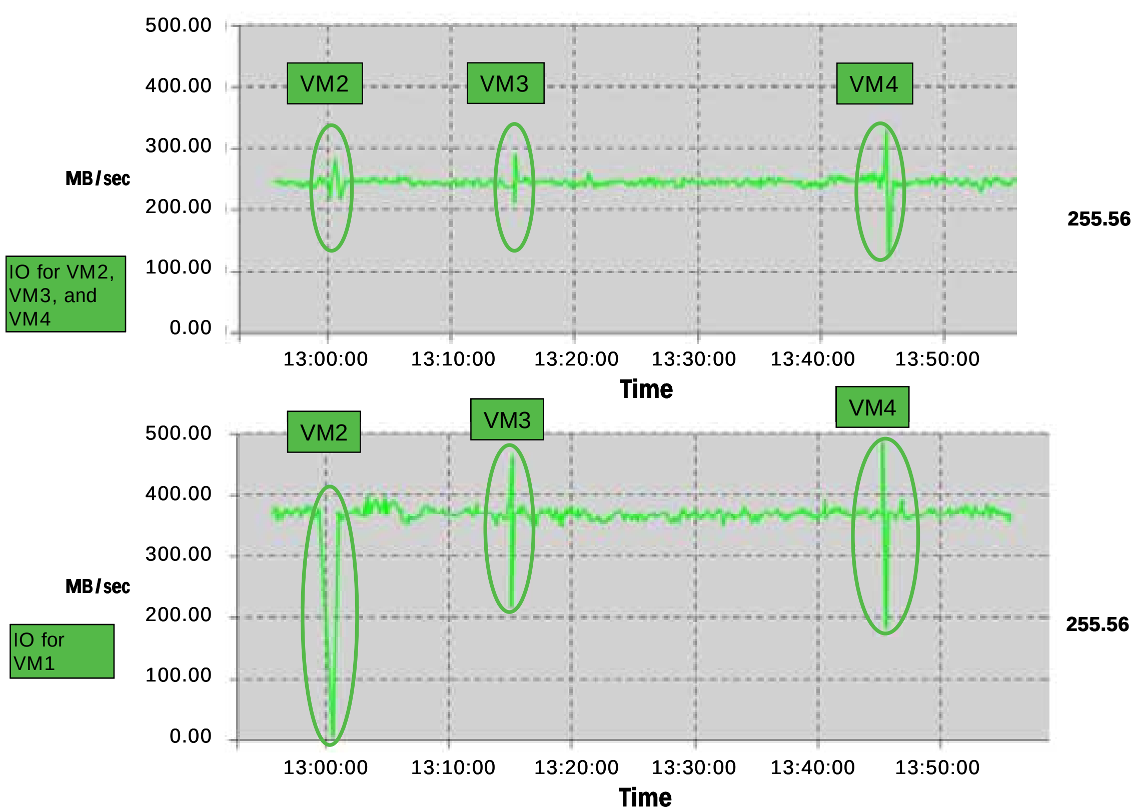
The first performance graph (Figure 5) shows a single virtual machine running an application driving 400.61 MB/s of IO at peak demand. The physical port on this HBA is capable of running at 800 MB/s, the server requires between an average of 360 MB/s to a peak of roughly 400 MB/s to deliver on the demands of the application and its users. SAO and QoS are not enabled within this configuration. The user in this configuration will realize no prioritization on the link nor will application or VM's IO be protected from any other traffic on the same link. You will notice a very flat line due to the short time period in this graph. This is also what a legacy HBA would offer to the user as far as an IO experience.

Figure 5: Single VM IO Performance. No QoS enabled at the server port.



The second and third performance graphs (Figure 6) show the same application running in VM1 within the bottom graph. The top graph shows the IO associated with 3 incremental VM's (VM2, VM3, and VM4) added to the server over a period of 45 minutes. As you can see in the graphs when VM2, VM3, and then VM4 are added to the server our application running in VM1 is impacted and now running at 369.88 MB/s. It is no longer realizing the full 400.61 MB/s it was driving when it had dedicated IO resources. We also can see that when VM2, VM3, and VM4 are brought on line there is a direct impact to the IO of our application server in VM1 as represented by the ovals around the IO glitches. No longer having dedicated resources, the IO in VM1 is now impacted from a performance perspective and subject to disruption of traffic when additional VM's are brought on-line within the server.

Figure 6: IO Performance for VM1. VM2, VM3, and VM4 added to the configuration.

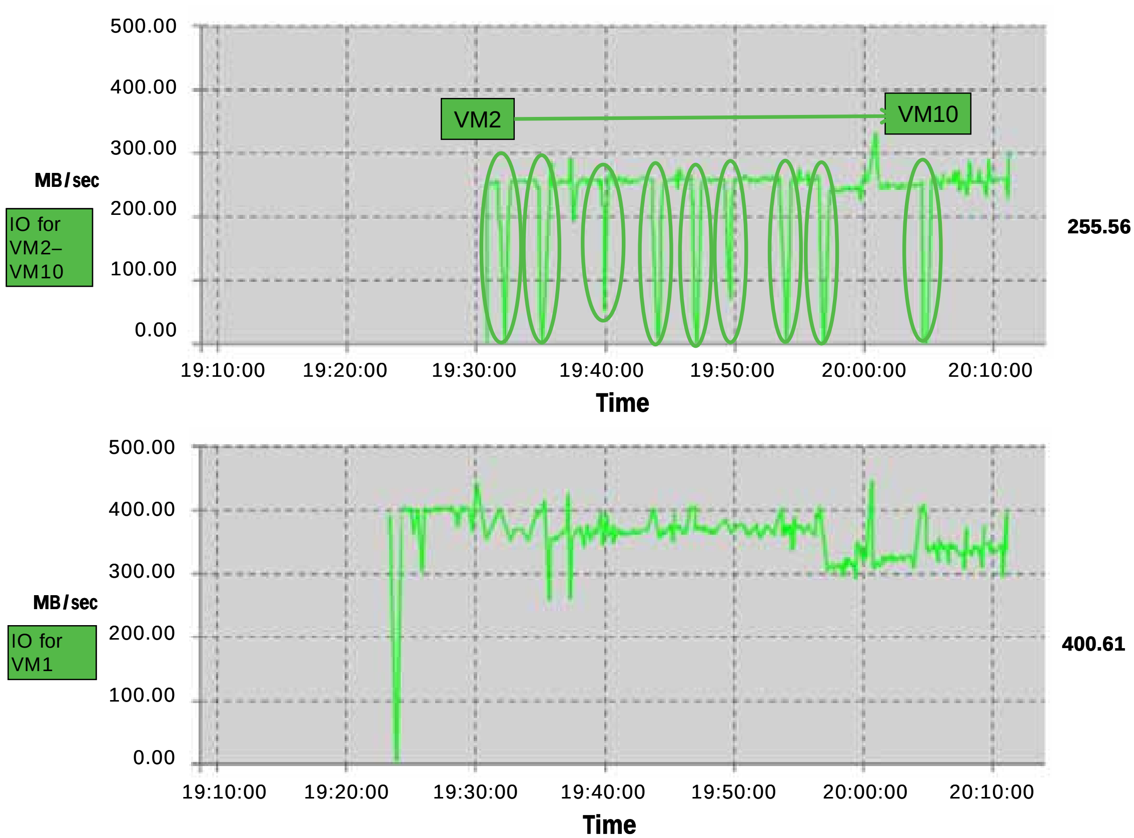


The fourth and fifth performance graphs (Figure 7) show the same application running in VM1 within the bottom graph. The top graph shows the IO associated with same 3 incremental VM's (VM2, VM3, and VM4) added to the server in Figure 6 as well as an additional 5 VM's added over a period of 45 minutes.

In this configuration we have now enabled SAO/QoS for VM1; this took place at 19:24 on the graphs timeline. VM1 is now running with a high QoS priority and will be protected from other IO on the wire as well as enabled to take full advantage of the bandwidth it requires. As VM2 through VM10 are added to the server, no QoS has been configured for them.

As you can see in the graph, VM1 is no longer impacted by the additional VM's and is running at the full Potential hitting peak IO in the 400.61 MB/s range like it was driving when it had dedicated IO resources as shown in Figure 5. We also can see that when VM2 through VM10 are brought on line there is no impact to the IO of our application server in VM1 (QoS High) and we continue to realize the full performance capability of the application server. We do though see the continued impact to VM2 though 10 as additional VM's are added within the same zone set, but given they are in a QoS medium profile they are able to continue to realize the resources allocated to them in QoS medium.

Figure 7: IO Performance for VM1 with SAO enabled. IOs for other VMs do not impact VM1.



Conclusion

The HP MSA virtualization bundle, which includes SAO, offers a unique solution that enables QoS capabilities and addresses the negative impact of network congestion for both non-virtualized and virtualized environments. By extending the fabric based AN capabilities from the fabric to the server interconnect, VMs and their related critical applications can be protected from other IO intensive applications. As long as there is an IO flow to be protected (either in a virtualized or non-virtualized environment), SAO can deliver the hardware enforced priorities across the infrastructure to optimize overall performance for that VM and IO flow.

To know how you can deliver end-to-end storage, virtualization, and management capabilities across your network infrastructure, please visit:

Administrator Guide: Implementation and Best Practices—
http://h20195.www2.hp.com/v2/GetPDF.aspx/4AA1-5600ENW.pdf

Optimizing the Data Center with SAO—
http://h20195.www2.hp.com/v2/GetPDF.aspx/4AA1-6589ENW.pdf

Best Practices for HP StorageWorks MSA2000 G1 or G2 and P2000 G3 FC—
http://h20195.www2.hp.com/v2/GetPDF.aspx/4AA0-8279ENW.pdf

4AA1-6640ENW, Created June 2010