Redbooks
ibm.com/redbooks

SAS Business Analytics Deployment on IBM POWER8 Processor-Based Systems

With IBM XIV Storage System and IBM FlashSystem

Beth L. Hoffman

Narayana Pattipati

Storage

IBM

Redpaper

Abstract

This IBM® Redpaper™ publication describes the performance tuning guidelines that can help you successfully deploy SAS business analytics on IBM POWER8® processor-based servers. The paper describes the test environment, the testing that was performed, and SAS analytics performance results. The test environment includes the SAS business analytics solution running on an IBM Power System S822 (8284-22A) server with IBM AIX® 7.1 operating system and IBM XIV® Storage System Gen3 with IBM Spectrum™ Scale (formerly IBM GPFS™).

This paper also describes the deployment of SAS analytics on IBM POWER8 processor-based servers with IBM XIV and IBM FlashSystem™ in a hybrid-storage model, along with configuration and tuning guidelines.

Introduction

SAS analytics provide an integrated environment for predictive and descriptive modeling, data mining, text analytics, forecasting, optimization, simulation, experimental design, and more.

The IBM Power Systems™ family of servers includes proven workload consolidation platforms that can help clients to control costs while improving overall performance, availability, and energy efficiency. With these servers and IBM PowerVM® virtualization solutions, an organization can consolidate large numbers of applications and servers, fully virtualize its system resources, and provide a more flexible and dynamic IT infrastructure.

In 2014, IBM introduced the next generation of Power Systems with IBM POWER8 technology. Some models are purpose-built for big data and analytics, scale-out, and cloud-based solutions. The new POWER8 processor-based servers are built with the IBM POWER8 processor technology and offer enhancements such as on-chip transactional memory, increased threading, Coherent Accelerator Processor Interface (CAPI), PCIe Gen3 I/O slots, and enhanced reliability, availability, and serviceability (RAS) features.

POWER8 supports up to 12 cores per socket compared to 8 cores per socket on the IBM POWER7® processor. The POWER8 processor-based servers have increased core density and compute power within the same form factor. POWER8 also supports SMT8 mode, which increases the maximum amount of execution work per cycle. SMT8 mode with eight hardware threads can improve the performance of workloads that use parallelism by allowing additional instructions to run at the same time.

POWER8 has improved cache and memory bandwidth compared to IBM POWER7+™ or POWER7. Compared to POWER7, the POWER8 L2 cache has doubled from 256 KB per

core to 512 KB per core and L3 cache has doubled from 4 MB per core to 8 MB per core (total L3 cache 96 MB per chip). In addition, POWER8 also has 128 MB L4 cache in memory controller to reduce memory access latency, which benefits applications that use large caches.

POWER8 uses an industry-standard PCIe Gen3 I/O bus, which has much better internal I/O bandwidth compared to GX++ buses used in POWER7. Also, PCIe Gen3 doubles the I/O from Gen2 attached devices.

SAS analytics solutions are compute-, I/O-, and memory-intensive and they stress the system to a large extent by running concurrent analytical jobs from multiple business users. The enhancements in the POWER8 processor with respect to compute, I/O bus, memory, and caches can positively impact performance of SAS analytics workloads.

IBM POWER8 processor-based servers with IBM Spectrum Scale™ (formerly IBM GPFS) and IBM XIV Storage Systems provide an integrated analytical environment that business analytics solutions such as SAS require.

IBM FlashSystem storage systems store the data in flash memory; they are designed for dramatically faster access times and support very large amounts of input/output (I/O) operations per second (IOPS) and throughput with significantly lower latency than hard disk drive (HDD) based solutions. Due to their macro-efficiency design, FlashSystem storage systems also consume lower energy and have significantly lower cooling and space requirements, all while allowing server processors to run SAS analytics more efficiently. Deploying SAS analytics on a hybrid storage model with XIV (disk-based storage) and FlashSystem (all-flash storage) helps customers to get best of both. The persistent data of SAS workloads (SASDATA), which has larger physical space requirements can be deployed on XIV storage system and temporary SAS work area (SASWORK and SASUTIL) can be deployed on FlashSystem.

This paper describes the suggested SAS business analytics solution deployment on an IBM Power S822 server, which is based on POWER8 processor technology. The paper also describes the test environment, configuration and tuning, and performance test results of SAS workloads. Some of the new POWER8 features mentioned earlier were tested with SAS analytics and their results are mentioned throughout the paper.

The paper also describes the deployment of SAS analytics on IBM POWER8 processor-based servers on hybrid-storage environment, also with configuration and tuning guidelines.

SAS analytics deployment on IBM Power servers with XIV storage

This section describes SAS analytics deployment on the IBM Power S822 (8284-22A) server with IBM XIV Storage System and IBM Spectrum Scale (formerly IBM GPFS). The logical deployment architecture is described in Figure 1 on page 3.

IBM Power S822 (8284-22A) is a scale-out server with two sockets, 20 cores, and 256 GB memory. The server is configured with a Virtual I/O Server (VIOS) and five client logical partitions (LPARs). The VIOS helps in sharing the Fibre Channel (FC) adapters among the client LPARs by virtualizing the physical FC adapters using N_Port ID Virtualization (NPIV). The test configuration uses a single VIOS, however, the preferences is to use dual-VIOS in production deployments for high availability.

One of the LPARs is used for testing SAS analytics. Volumes are created on the XIV storage system and are mapped to the LPARs. The SAS workload Spectrum Scale file systems are created on the mapped volumes for running SAS analytics jobs.

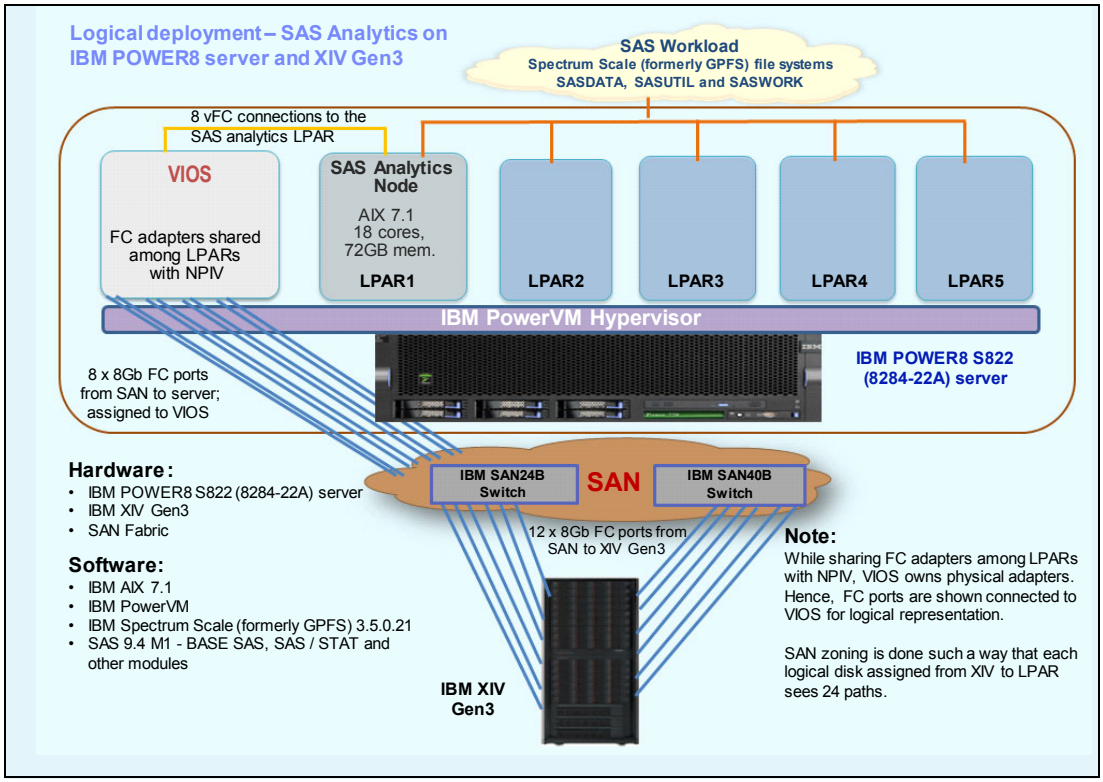

Figure 1 SAS analytics deployment architecture on IBM Power S822 server with XIV Storage System and Spectrum Scale

Test environment

The hardware, software, and Spectrum Scale file systems are listed in this section.

Hardware

The hardware configuration details described here are for Power S822, XIV, and SAN.

Power S822 configuration details

Configuration details for Power S822 are as follows:

Model	8284-22A
Firmware version	FW810.02 (061)
Processor architecture	POWER8
Clock speed	4116 MHz
SMT	OFF, 2, 4, 8 (SMT4 is default)
Cores	20 (18 cores for the SAS analytics LPAR and 2 cores for VIOS)
Memory used	256 GB (72 GB for the SAS analytics LPAR and 8 GB for VIOS)
Internal drives	Four 600 GB drives (used for booting VIOS and LPARs)
FC connectivity	Four quad-port 8Gb FC ports (16 ports) attached to the server; used 8 ports during the testing

XIV configuration details

Configuration details for XIV are as follows:

XIV machine type	2810
Machine model	114
System version	11.5.0.x
Drives	180 SAS drives each with 2 TB capacity and 7200 rpm speed
Usable capacity	161 TB
Modules	15
Cache	DDR3 360 GB
SSD cache	6 TB
Connectivity	Six 8 Gb dual-port Fibre Channel (FC) adapters (12 ports) connected to storage area network (SAN)
Stripe size	1 MB (default)
SSD cache	Enabled (by default) for all volumes used in workload

SAN configuration details

Configuration details for SAN are as follows:

- Two FC switches: IBM System Storage® SAN24B-4 Express (24 ports) and IBM System Storage SAN40B-4 (40 ports). Both support NPIV.
- Sixteen 8Gb dual-port FC ports connected from Power S222 server to the SAN fabric. Eight ports are connected to the first switch and eight more ports connected to the second switch.
- Twelve 8 Gb FC ports are connected from SAN Switches to XIV Gen3.
- Switch zoning is performed on the SAN switches such that each logical disk assigned from XIV to the LPAR has 12 or 24 paths.

Software

The following software was used:

- SAS 9.4 M1 64-bit software
- IBM AIX 7100-03-03-1415
- VIOS 2.2.3.3
- IBM PowerVM Enterprise Edition
- IBM Spectrum Scale (formerly IBM GPFS) 3.5.0.21

Spectrum Scale file systems

SAS workload contains the following three file systems. The block size used in the testing was 1 MB.

SASWORK	4 TB with 24 volumes of 172 GB each
SASDATA	4 TB with 16 volumes of 256 GB each
SASUTIL	1.6 TB with 16 volumes of 100 GB each

Suggested performance settings

The following list describes the tuning guidelines for SAS workloads to perform optimally on POWER8 processor-based servers with the AIX 7.1 operating system. For more details, read the SAS on IBM AIX 5L™, AIX 6, and AIX 7 Tuning Guides:

http://ibm.com/support/techdocs/atsmastr.nsf/WebIndex/WP101529

- AIX tuning:
  - Virtual Memory Manager (VMM); use the **vmo** command to tune:

    ```
    maxfree = 73728  (default 1088)
    minfree = 8192 (default 960)
    nokilluid = 10 (default 0)
    ```

    Default values were used for other virtual memory tunable.
  - I/O; use the **ioo** command to tune:

    ```
    j2_dynamicBufferPreallocation = 256 (default 16)
    j2_maxPageReadAhead = 1024 (default 128)
    j2_minPageReadAhead = 16 (default 2)
    j2_nPagesPerWriteBehindCluster = 64 (default 32)
    ```

    Default values were used for other I/O tunable.
  - Network; use the **no** command to tune:

    ```
    tcp_nodelayack = 1
    ```

    Default values were used for other network tunable.
  - Changed `maxuproc` to `2000` from default of `512`:

    ```
    #chdev –l sys0 maxuproc=2000
    ```
  - Changed user limits (ulimit) to unlimited to run the SAS workloads.

> **Note:** The remaining AIX performance settings use the default values, that is, no change was made to them during performance optimization. See “Complete list of performance settings” on page 18.

- XIV Storage System:
  - No specific tuning was required on XIV Storage Systems. It works using the default environment.
  - XIV, by default, uses wide stripe and stripes data across all the available disks. The XIV Gen3 system used in testing has 180 disks.
  - XIV uses 1 MB as the stripe size, by default.
- IBM Spectrum Scale:
  - Version 3.5.0.21 (file system version 13.23).
  - File system block size was 1 MB.
  - File system block allocation type was Cluster (default is Scatter if the number of disks in a file system is greater than 8).
  - Spectrum Scale tunable:

    ```
    pagepool 8G (default 1G)
    seqDiscardThreshhold 1G (default 1MB)
    maxMBpS 8000 (default 2048)
    prefetchPct 40 (default 20)
    ```

```
maxFilesToCache 20000 (default 4000)
scatterBuffers no (default yes)
stealFromDoneList no (default yes)
```

- FC host bus adapter (HBA):
  - Virtual adapters at the client LPAR:

    ```
    lg_term_dma: 0x800000 (default)
    max_xfer_size: 0x200000 (default 0x100000)
    num_cmd_elems: 256 (default 200)
    ```

  - Physical adapters at the VIOS:

    ```
    lg_term_dma: 0x800000 (default)
    max_xfer_size: 0x200000 (default 0x100000)
    num_cmd_elems: 500 (default)
    ```

- Logical disks on the client LPAR:

  ```
  max_transfer: 0x100000 (default 0x80000)
  queue_depth: 64 (default 40)
  ```

- Root disk for the client LPAR (set same tunable at client LPAR and VIOS):

  ```
  max_transfer: 0x100000 (default 0x40000)
  queue_depth: 128 (default 3)
  ```

- SAS software configuration:

  ```
  memsize: 2048 MB
  bufsize: 256 k
  sortsize: 256 MB or 1024 MB
  fullstimer
  ```

Workload and performance metrics details

The test workload used during the performance testing was SAS Mixed Analytics workload. The workload consisted of a mix of compute and I/O tests to stress compute, memory, and I/O subsystems and measure concurrent, mixed job performance on a given IT infrastructure.

The workload consisted of 20 individual SAS tests:

- 10 compute-intensive
- 2 memory-intensive
- 8 I/O-intensive

Some of the tests relied on existing data stores and some tests relied on generated data during the test run. The tests were a mix of short-running (minutes) and long-running (hours) jobs. The tests were repeated to run concurrently and in a serial fashion to achieve a 20-session or 30-session workload. The 20- and 30-session workloads consisted of 71 and 101 jobs respectively. During the peak load, the 20-session and 30-session workloads concurrently ran about 35 and 55 processor and I/O-intensive jobs respectively.

The performance measurement for these workloads was workload response time (in minutes), which is the cumulative real time of all the jobs in the workload. A lower response time is better. However, other performance metrics are also provided such as CPU time (user + system time), processor usage from the server, I/O throughput, and I/O latency.

The workload response time (real time) and CPU time (user + system time) are captured from the log files of SAS jobs. These statistics are logged with SAS fullstimer option. IBM Power systems, starting with the POWER7 processor architecture, use Processor Utilization Resource Register (PURR) accounting for accurate reporting of system usage. The PURR factor for POWER8 processor must be applied to the CPU time metrics, as described in “PURR accounting and interpreting performance results” on page 20.

SAS workload performance results

To understand SAS business analytics performance on IBM Power S822 server and XIV Gen3 storage, a SAS mixed analytics 20-session workload was run on a single LPAR. The 20-session workload was the appropriate size workload given the compute and I/O demands of the workload on the Power S822 server with two sockets and 20 cores. The Power S822 server with 20 cores and 256 GB memory can also support a more intensive 30-session workload; however, lower response times are expected compared to a 20-session workload.

20-session workload performance

Here is the configuration that was used for the workload:

- LPAR is configured with 16 cores in dedicated mode, 64 GB memory, and SMT4.
- VIOS is configured with 2 cores in dedicated mode and 8 GB memory; SMT4.
- Used eight 8 Gb FC full-duplex ports from server to SAN for I/O.
- SAN zoning is configured with 24 paths for each volume mapped from XIV.
- Performance tuning is described in “Suggested performance settings” on page 5.
- The workload was run with no other competing activity on the server or storage.

The performance summary of the workload is as follows:

- Workload response time is 1175 minutes, user time is 784 minutes, and system time is 37 minutes.
- Peak I/O throughput is 4.5 GBps and sustained I/O throughput is 2.7 GBps, which translates to 170 MBps per core sustained I/O throughput.
- Peak latency of the XIV system is 8 ms and the average latency is 5 ms.
- Average processor usage (user + sys) is 56% and wait time is 5%.
- At host, average disk service time is 7.5 ms and the peak disk service time is 15 ms.
- Total data transferred during the workload is 7.5 TB (5.5 TB read and 2.0 TB write).

Note: The PURR factor for POWER8 processor was applied on the CPU (user and system) times mentioned earlier. For more details, see “PURR accounting and interpreting performance results” on page 20.

The graphs in Figure 2 show details of the I/O throughput and latency measured at XIV Storage System. Figure 3 on page 8 shows processor usage measured on the POWER8 processor-based server.

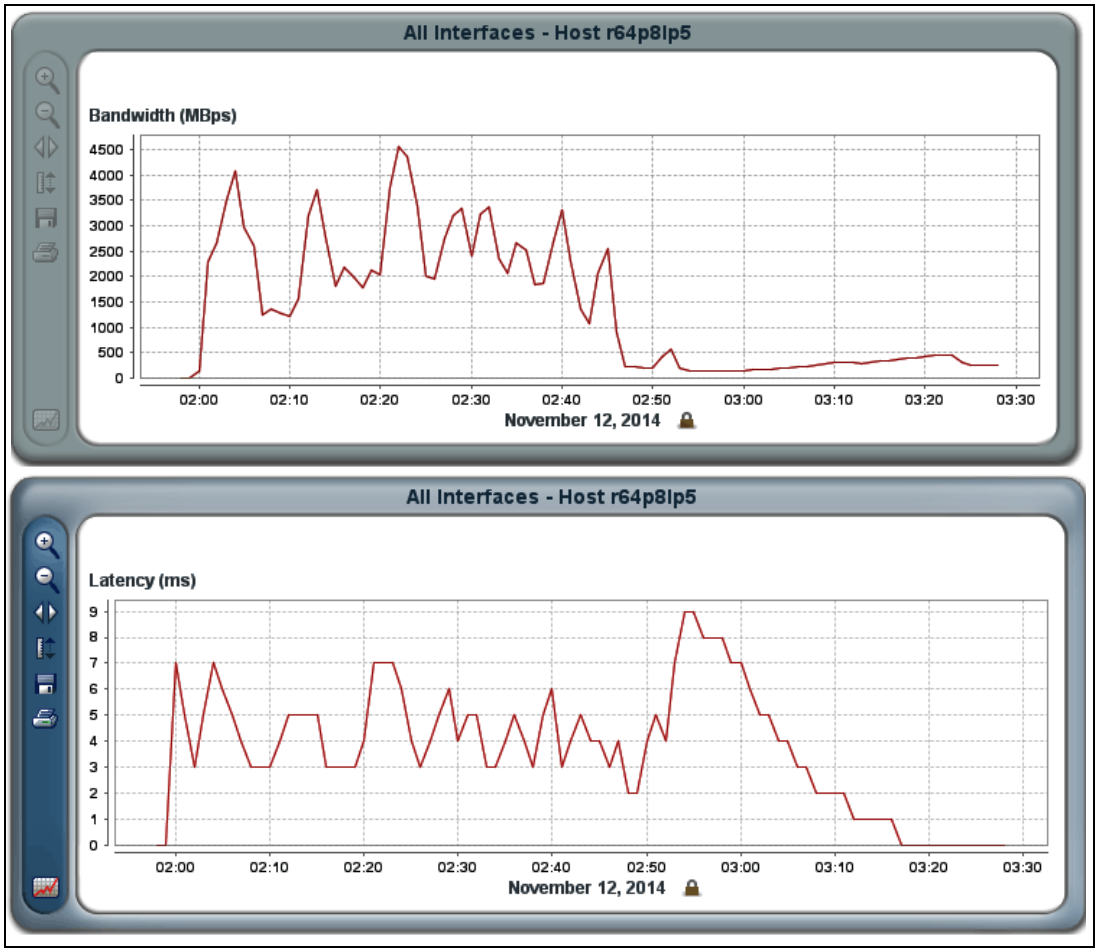

Figure 2 I/O throughput and latency for mixed analytics 20-session workload measured at XIV

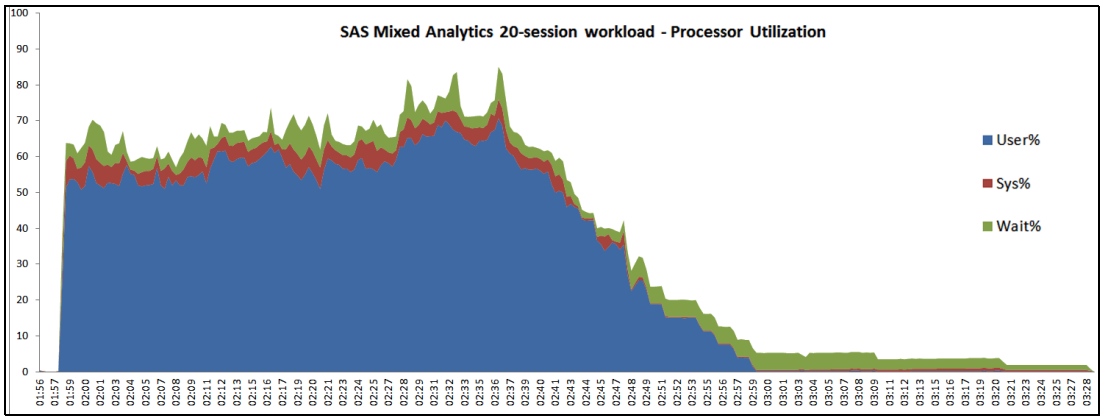

Figure 3 Processor usage for mixed analytics 20-session workload, measured at POWER8 server

The test was run with default configuration settings first (see “Complete list of performance settings” on page 18) and then the environment was tuned for optimal performance and the same tests were run again. The optimal performance was achieved after multiple iterations of

tuning and re-running the tests. for final tuning and configuration used for the test, see “Suggested performance settings” on page 5.

As shown in the earlier graphs, the workload quickly jumped to 60% system utilization and generated an I/O throughput of 4 GBps. As the jobs completed their run and new jobs were launched, the I/O throughput varied from 2 GBps to a peak of 4.5 GBps. The workload had about 35 I/O- and compute- and memory-intensive jobs running concurrently at peak. Overall, for the 20-session workload, the system utilization sustained at 60% and I/O throughput sustained at 2.7 GBps.

I/O throughput from POWER8 scale-out server and XIV Gen3

Although the 20-session was the appropriate size workload given, the compute and I/O demands of the workload on the Power S822 server with just 2 sockets and 20 cores, a more-intensive, 30-session workload was also run to understand the performance and the I/O throughput. As expected, with the same set of resources, the more-intensive, 30-session workload showed slower response time compared to 20-session workload. However, during the 30-session workload run, the Power S822 server (with 18 cores and 72 GB memory) with XIV Gen3 produced a good I/O throughput. As the graph in Figure 4 on page 9 shows, the 30-session workload produced peak I/O throughput of 5.4 GBps and sustained I/O throughput of 4.5 GBps. This translates to 250 MBps per core I/O throughput, considering the LPAR had just 18 cores.

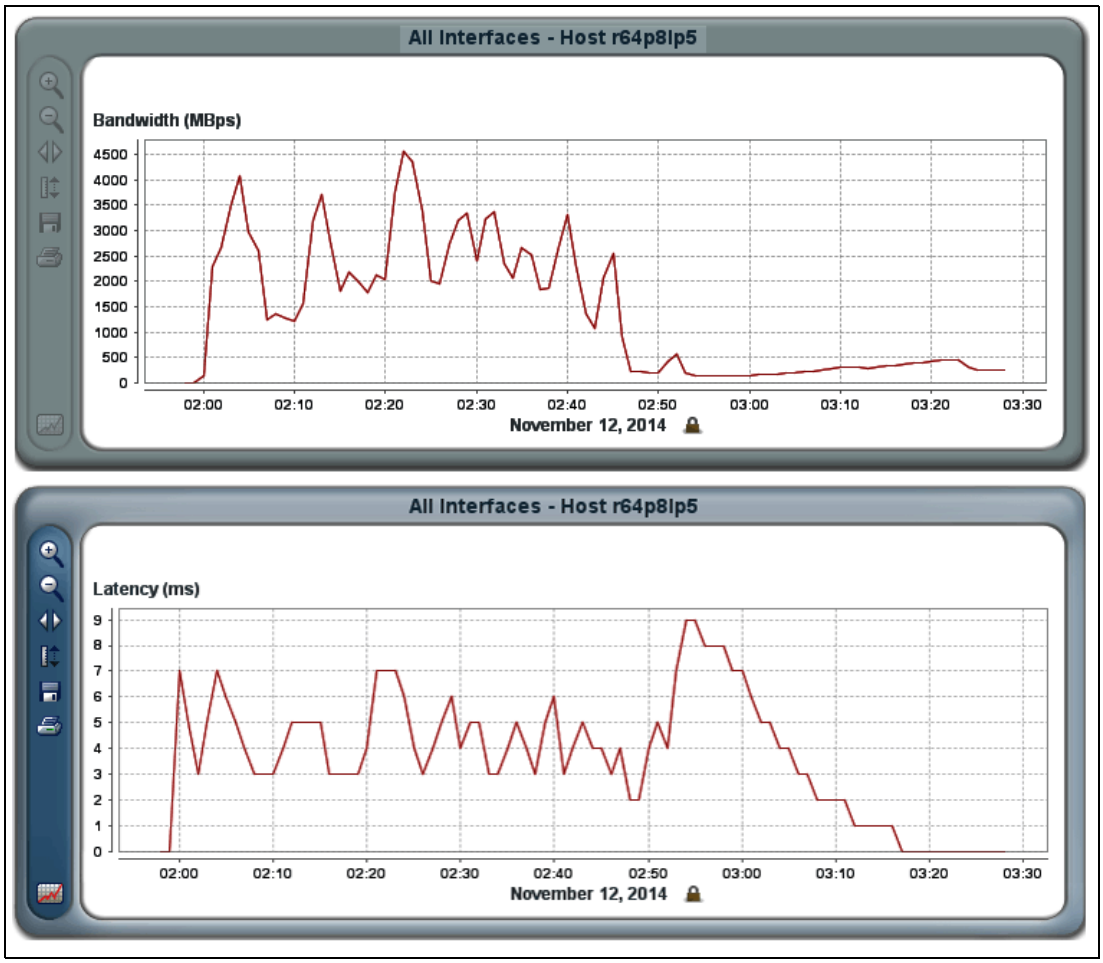

Figure 4 I/O throughput and latency for mixed analytics 30-session workload measured at XIV

SMT8 considerations

The POWER8 processor-based servers run with SMT4, by default, and support up to eight threads per core (SMT8). During the performance testing activity, the 20-session workload was also run with SMT8 to understand how the workload performs. During the test, the SAS 20-session workload did not have enough threads to take advantage of eight parallel threads with SMT8. Hence, the workload showed similar performance with SMT4 and SMT8.

Testing SMT8 while running with fewer cores showed that the workload used all eight threads of a physical core assigned to the LPAR where the test was run. It proves that the SAS application has sufficient parallelism to make use of all eight hardware threads. If you have enough parallel threads in your SAS workload and system utilization is high (>70%), consider using SMT8 to improve the performance of the workload and overall system utilization.

The suggestions and performance tunings described in this paper were based on the testing on a scale-out POWER8 processor-based Power S822 server and XIV Gen3 storage in a lab environment. The suggested approach is to start with these best practices and explore further optimization based on the actual workload and configuration of the server and storage being used.

Also note that the XIV Gen3 storage system used during the testing was a 2012 model with 11.5.0.x version firmware. The newer model of XIV storage system has many more features and enhancements, which have the potential to provide much better I/O throughput for SAS analytics workloads.

To summarize, consider the following key points to use from the performance testing on POWER8 processor-based Power S822 scale-out server (with 20-cores and 256 GB memory) and XIV Gen3:

- 20-session SAS mixed analytics workload, with a mix of 71 compute- and I/O-intensive jobs ran extremely well. The Power S822 server with 16 cores supported 35 concurrent jobs at peak during the workload.
- Extremely good I/O throughput of 250 MBps per core (with 30-session workload) was achieved during the testing with only 18 cores.
- If SAS workloads have enough parallel threads, the new SMT8 feature can be used to improve the performance of the workload and overall system utilization.

SAS analytics deployment on IBM Power servers in a hybrid-storage environment

This section describes the deployment of SAS analytics on IBM POWER8 processor-based servers and hybrid-storage environment involving IBM XIV Gen3 and IBM FlashSystem 840.

As described in the previous sections, SAS workloads typically have two sets of file systems. While SASDATA stores persistent data that includes input and output files, the SAS work area (SASWORK and SASUTIL file systems) stores temporary data that is deleted at the end of each SAS job. In typical customer deployments, SASDATA requires much more space for storing input files and output files. Hence, it is beneficial to deploy SASDATA file system on an XIV Storage System (which is disk-based storage) and the SAS work area on a FlashSystem. Figure 5 describes the deployment in detail.

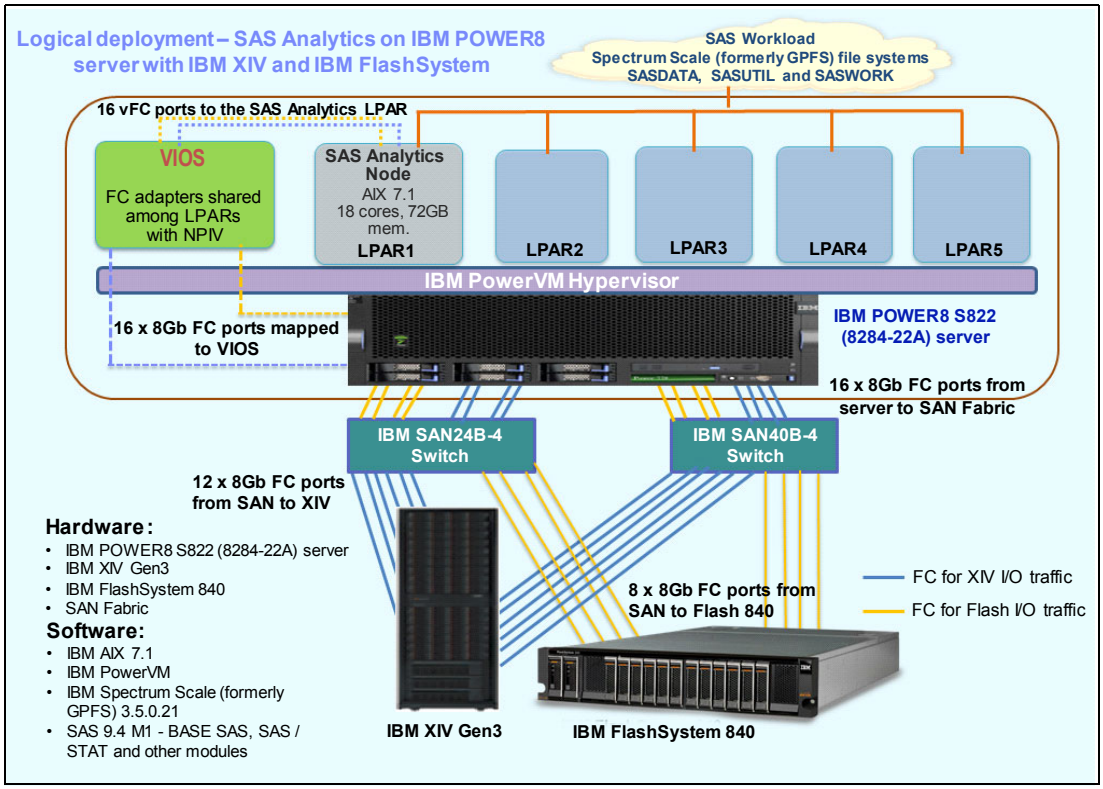

Figure 5 SAS analytics deployment on IBM Power S822 server in hybrid-storage environment

IBM Power S822 (8284-22A) is a scale-out server with two sockets, 20 cores, and 256 GB memory. The server is configured with a VIOS and five client logical partitions (LPARs). The VIOS helps in sharing the FC adapters among the client LPARs by virtualizing the physical FC adapters using NPIV. The test configuration uses a single VIOS, however, the suggestion is to use two VIOS in production deployments for high availability. I/O traffic from XIV and FlashSystem is segregated using separate FC ports (physical and also virtual) as shown in Figure 5.

One of the LPARs is used for testing SAS analytics. Volumes are created on XIV and FlashSystem and are mapped to the LPARs. The SAS workload Spectrum Scale file systems are created on the mapped volumes for running SAS analytics jobs. The file system layout on a hybrid-storage environment is shown in Figure 6. A SASDATA file system is created with 16 volumes mapped from XIV Gen3 storage. The 16 volumes proved to be optimal for the workload used during the testing. The 32 volumes were mapped for SASWORK and SASUTIL, each from FlashSystem 840. For AIX clients, 32 volumes are recommended in *Implementing IBM FlashSystem 840*, SG24-8189:

http://ibm.com/redbooks/abstracts/sg248189.html?Open

The file system layout in the hybrid-storage environment is shown in Figure 6.

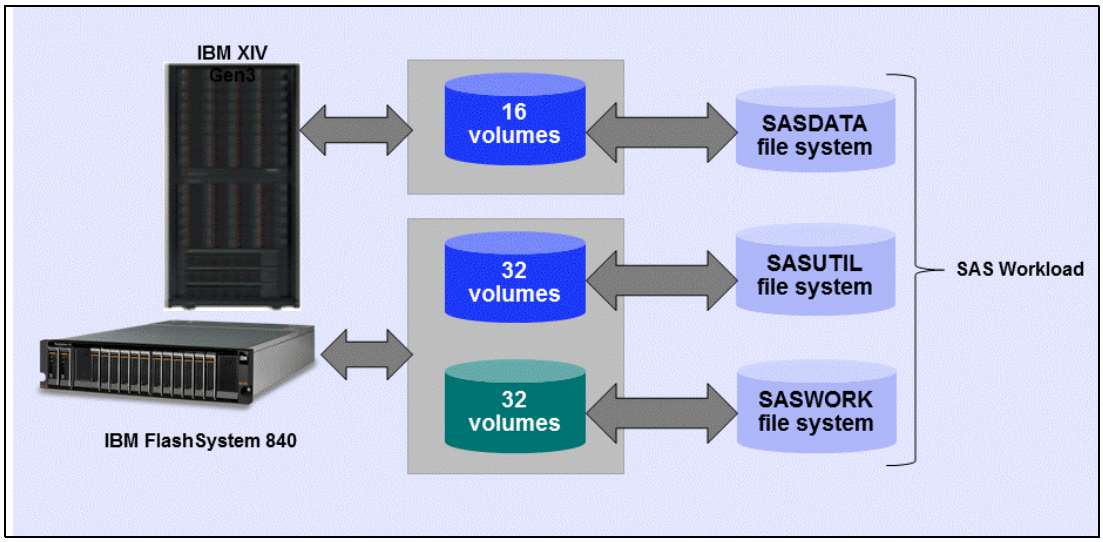

Figure 6 File system layout for deploying SAS analytics on hybrid-storage environment

Test environment for hybrid-storage deployment

This section describes the hardware, zoning information, and Spectrum Scale file systems for this test environment.

Hardware

The hardware configuration details are as follows:

- Power S822 server configuration details are listed in “Test environment” on page 3.
- XIV Gen3 storage configuration details are listed in “Test environment” on page 3.
- FlashSystem 840 configuration details:
  - Flash modules: Twelve 4 TB
  - Total capacity: 41.2 TB (after system-level RAID5 with 11 member drives and 1 hot spare and 4 KB stripe size)
  - Code Level (FW): 1.1.3.2
  - FC ports connected: Eight 8 Gb FC ports (four ports attached to each canister)

Zoning at switches to segregate XIV and FlashSystem I/O traffic

The LPAR used for testing has 16 virtual client FC ports that are mapped to 16 virtual server FC ports at VIOS. The virtual server FC ports in turn are mapped to 16 physical FC ports that are assigned to the VIOS from HMC. I/O traffic is segregated at switches by optimal zoning.

The zoning details are depicted in Figure 7 and in Figure 8 on page 13.

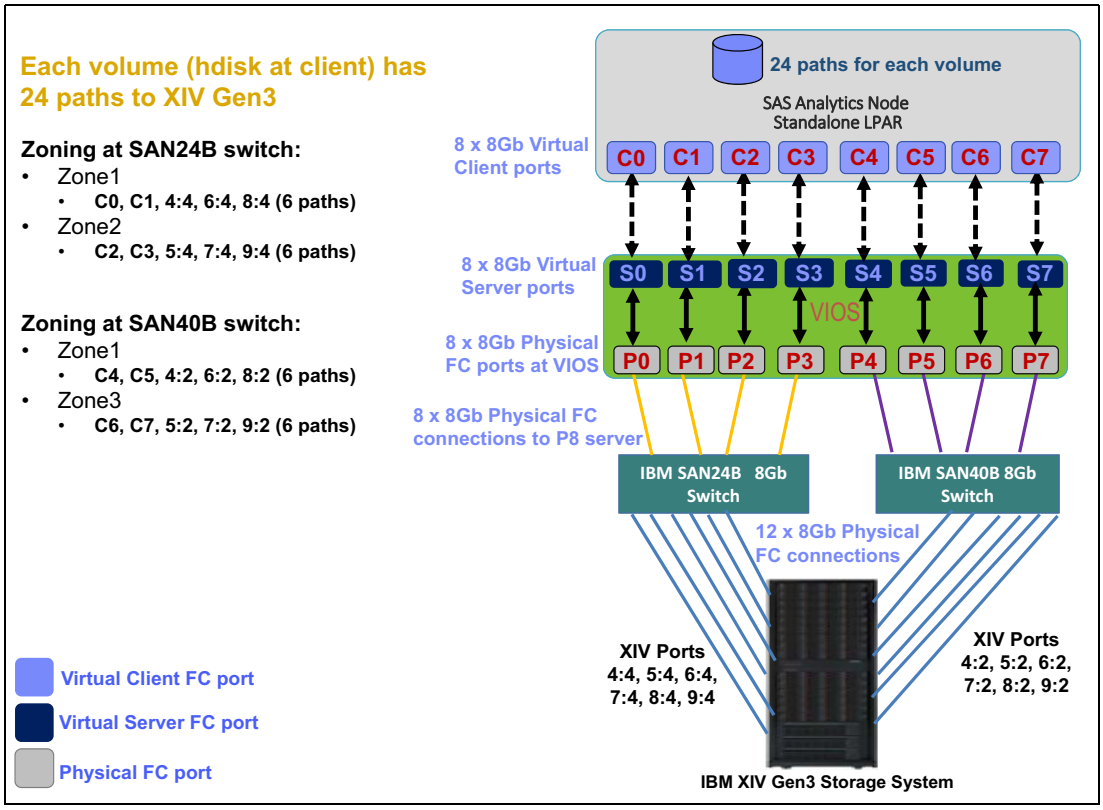

Figure 7 Zoning details for XIV I/O traffic in a hybrid-storage environment

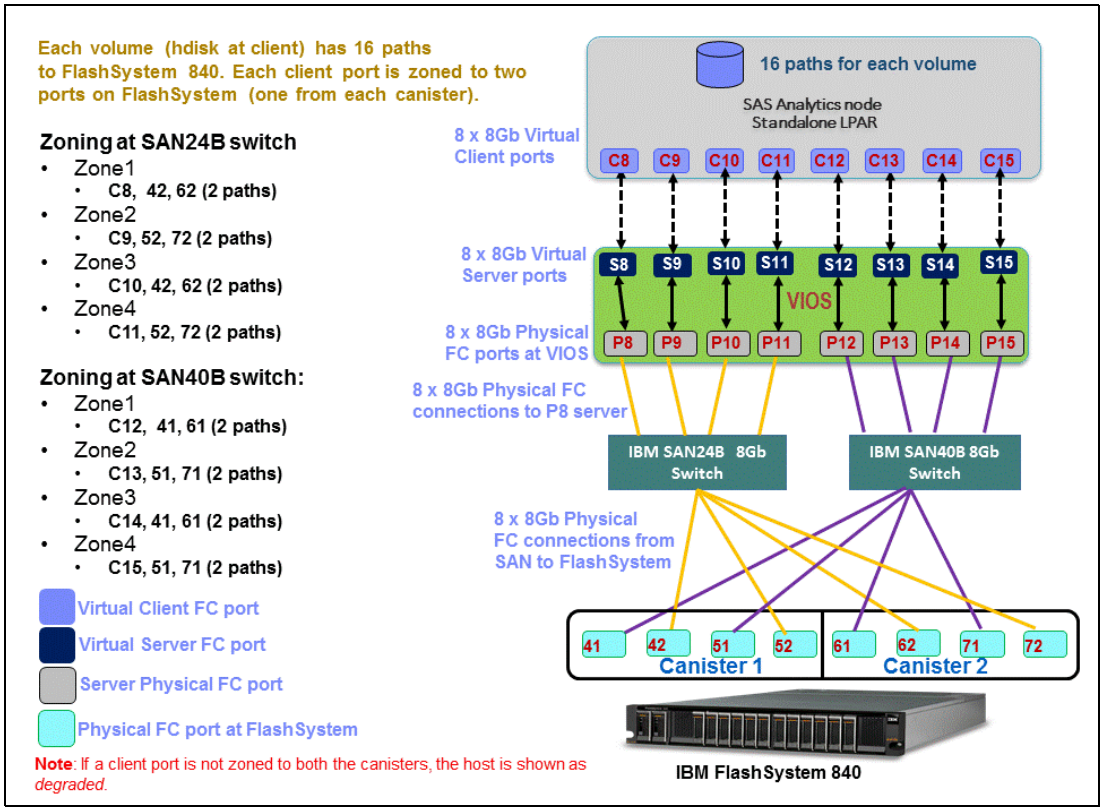

Figure 8 Zoning details for FlashSystem 840 I/O traffic in a hybrid-storage environment

Spectrum Scale file systems

SASDATA file system is created out of the volumes mapped from XIV storage, and SASWORK and SASUTIL file systems are created out of the volumes mapped from FlashSystem 840.

- SASDATA (on XIV Gen3):
  - 4 TB total size with 16 volumes of 256 GB each
  - File system block size: 1 MB
  - File system block allocation type: Cluster (default is scatter)
- SASWORK (on FlashSystem 840):
  - 4 TB total size with 32 volumes of 172 GB each
  - File system block size: 256 KB or 512 KB
  - File system block allocation type: Scatter (default)
- SASUTIL (on FlashSystem 840):
  - 1.6 TB total size with 24 volumes of 50 GB each
  - File system block size: 256 KB / 512 KB
  - File system block allocation type: Scatter (default)

During the testing in the lab environment, the scatter file allocation type proved to be optimal when using FlashSystem; the cluster block allocation type proved to be optimal while using XIV Gen3 storage. The allocation type must be chosen based on the number of volumes and cluster size in production environments.

Suggested performance settings for hybrid-storage environment

Because separate I/O paths are used for segregating I/O traffic from XIV and FlashSystem in a hybrid-storage environment, tuning the I/O paths differently is important for optimal performance.

The following list describes the tuning guidelines for SAS workloads to perform optimally on POWER8 processor-based servers with the AIX 7.1 operating system on hybrid-storage environment with XIV Gen3 and FlashSystem 840. For more details, read the SAS on IBM AIX 5L, AIX 6, and AIX 7 Tuning Guides:

http://ibm.com/support/techdocs/atsmastr.nsf/WebIndex/WP101529

- AIX tuning:
  - Same as the tuning described in “Suggested performance settings” on page 5.
- XIV Storage System:
  - No specific tuning required on XIV Storage Systems. It works using the default environment.
  - XIV, by default, uses wide stripes, and stripes data across all the available disks. The XIV Gen3 system used in testing has 180 disks.
  - XIV uses 1 MB as the stripe size, by default.
- Spectrum Scale (formerly GPFS) tunable:

```
pagepool 8G (default 1G)
seqDiscardThreshhold 1G (default 1MB)
maxMBpS 12000 (default 2048)
prefetchPct 40 (default 20)
maxFilesToCache 20000 (default 4000)
scatterBuffers no (default yes)
stealFromDoneList no (default yes)
```

- FC host bus adapter (HBA):
  - Virtual adapters at the client LPAR for XIV

    ```
    lg_term_dma: 0x800000 (default)
    max_xfer_size: 0x200000 (default 0x100000)
    num_cmd_elems: 256 (default 200)
    ```

  - Virtual adapters at the client LPAR for FlashSystem

    ```
    lg_term_dma: 0x800000 (default)
    max_xfer_size: 0x100000 (default)
    num_cmd_elems: 256 (default 200)
    ```

  - Physical adapters at the VIOS for XIV

    ```
    lg_term_dma: 0x800000 (default)
    max_xfer_size: 0x200000 (default 0x100000)
    num_cmd_elems: 500 (default)
    ```

  - Physical adapters at the VIOS for FlashSystem

    ```
    lg_term_dma: 0x800000 (default)
    max_xfer_size: 0x100000 (default)
    num_cmd_elems: 500 (default)
    ```

  - Also change the following fscsi parameters for FlashSystem FC adapters at VIOS and client LPAR

    ```
    fc_err_recov = fast_fail (default is delayed_fail)
    dyntrk = yes (default is no)
    # chdev -l fscsi0 –a fc_err_recov=fast_fail –a dyntrk=yes -perm
    ```

- FlashSystem specific tuning for volumes

  ```
  32 volumes per file system
  Sector or block size while creating volumes:  512 (default)
  Auto contingent allegiance support:  yes (default)
  ```

- Logical disk (volume / hdisk) on client LPAR from FlashSystem 840

  ```
  Algorithm: round_robin  (default is shortest_queue)
  max_transfer: 0x80000  (default)
  queue_depth: 64 (default)
  ```

- Logical disk (volume / hdisk) on client LPAR from XIV

  ```
  Algorithm: round_robin  (default)
  max_transfer: 0x100000 (default 0x80000)
  queue_depth: 64 (default 40)
  ```

- Root disk for the client LPAR (set same tunable at client LPAR and VIOS)

  ```
  max_transfer: 0x100000 (default 0x40000)
  queue_depth: 128 (default 3)
  ```

- SAS software configuration

  ```
  memsize: 2048 MB
  bufsize: 256 k
  sortsize: 256 MB / 1024 MB
  fullstimer
  ```

SAS workload performance in a hybrid-storage environment

To understand SAS business analytics performance on an IBM Power S822 server and a hybrid-storage environment with XIV Gen3 and FlashSystem 840, SAS mixed analytics 20-session workload was run on a single LPAR. The 20-session workload was the appropriate size workload given the compute and I/O demands of the workload on the Power S822 server with two sockets and 20 cores.

20-session workload performance

Here is the configuration used for the workload:

- LPAR is configured with 16 cores in dedicated mode, 64 GB memory, and SMT4
- VIOS is configured with two cores in dedicated mode, 8 GB memory, and SMT4
- Performance tuning is described in “Suggested performance settings for hybrid-storage environment” on page 14.
- Used 512 KB block size for SASWORK and SASUTIL Spectrum Scale file systems that are deployed on FlashSystem (512 KB block size proved optimal for Spectrum Scale file systems on FlashSystem 840).
- The workload was run with no other competing activity on the server or storage systems.

The performance summary of the workload is as follows:

- Workload response time is 1134 minutes, user time is 781 minutes, and system time is 41 minutes.
- Combined (FlashSystem and XIV) peak I/O throughput is 5.0 GBps and sustained I/O throughput is 3.25 GBps, which translates to 200 MBps per core sustained I/O throughput.
- At XIV, peak latency is 3 ms and average latency is 1 ms (for SASDATA file system). At FlashSystem 840, peak latency is 4 ms and average latency is 2 ms (for SASWORK and SASUTIL file systems). The latency at FlashSystem 840 is on the expected lines because of the large-block sequential I/O nature of the workload.
- Average processor usage (user + sys) is 56% and wait time is 6%.
- At host, average disk service time is 3.5 ms and the peak disk service time is 7 ms.
- Total data transferred during the workload is 11 TB (9 TB read and 2 TB write).

Note: The PURR factor for POWER8 processor was applied on the CPU (user and system) times mentioned earlier. For more details, see “PURR accounting and interpreting performance results” on page 20.

The graphs in Figure 9 and Figure 10 depict the I/O performance at XIV and FlashSystem.

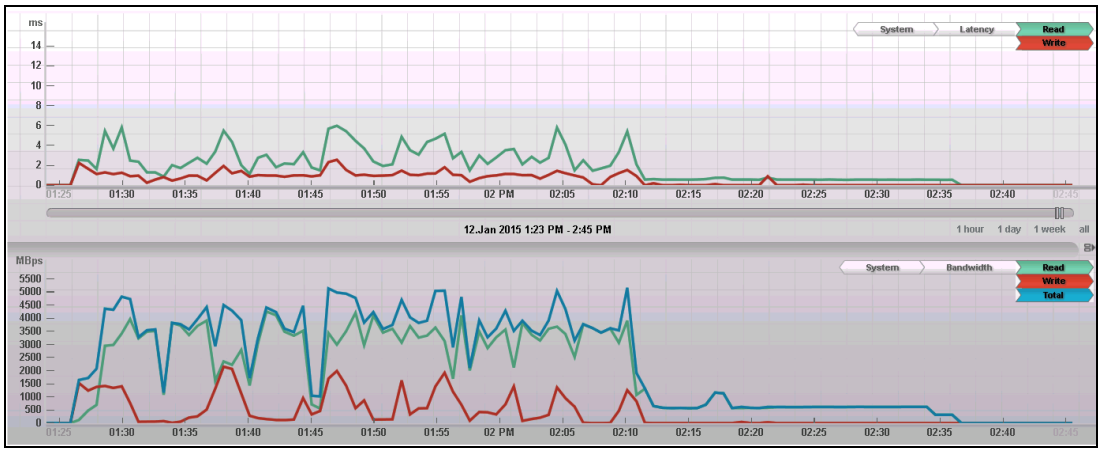

Figure 9 Mixed Analytics 20-session I/O performance at FlashSystem for SASWORK and SASUTIL

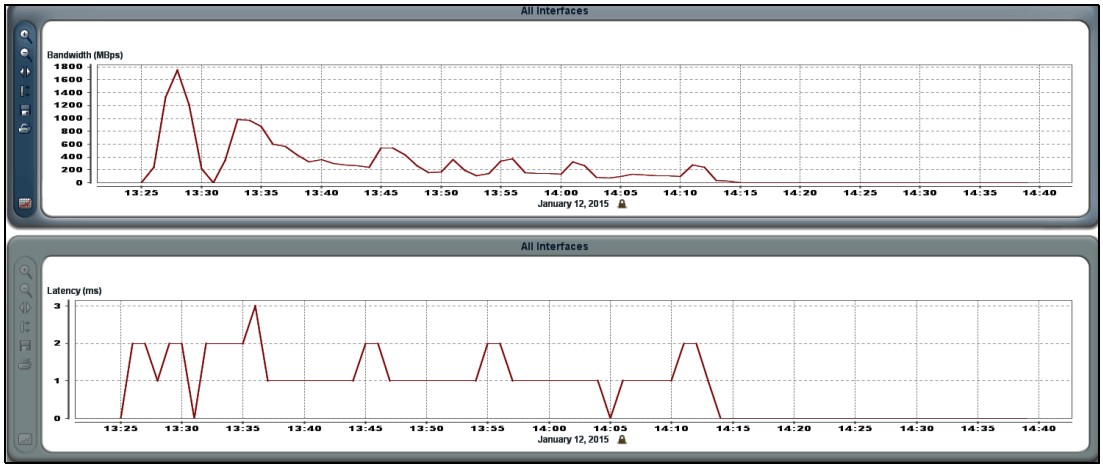

Figure 10 Mixed Analytics 20-session I/O performance at XIV for SASDATA

The workload used during the testing is large-block sequential in nature with 80:20 read-write ratios. This is true for many SAS analytics workloads. If you observe Figure 9 on page 17 and Figure 10 on page 17, the SASWORK and SASUTIL file systems, which are deployed on FlashSystem in hybrid-storage environment, contribute 85 - 95% of the I/O generated by the workload. During the testing, SAS analytics workloads proved to effectively use hybrid-storage architecture involving XIV and IBM FlashSystem 840. In the hybrid-storage environment, the XIV Storage System is used for deploying persistent data for optimal storage space utilization, and FlashSystem is used for deploying the SAS work area for optimal I/O throughput performance.

While XIV is a proven disk-based storage system for SAS workloads, customers can consider hybrid storage architecture to accelerate analytics jobs, keeping the overall total cost of ownership (TCO) low. In some cases, SAS analytics customers might have space and cooling constraints in their existing data center. To augment storage capacity, buying a full-rack-sized disk-based storage subsystem might not be an option for such customers. IBM FlashSystem 840 is a perfect choice for such environments; it provides superior performance in a 2U form factor.

Overall conclusions

IBM POWER8 processor-based Power S222 server, IBM XIV Gen3, and IBM Spectrum Scale (formerly IBM GPFS) together form a platform that is well-suited for running even the most demanding SAS business analytics workloads. The IBM Power S822 server performed extremely well for SAS workloads that support the compute and I/O requirements.

Also, IBM POWER8 processor-based servers performed extremely well for SAS workloads in hybrid-storage environments involving IBM FlashSystem. SAS analytics customers can explore the possibility of using hybrid-storage architecture involving XIV or other disk-based storage systems and IBM FlashSystem to accelerate SAS analytics while keeping the overall TCO low.

Complete list of performance settings

The following list of AIX performance settings that were *not* changed during the performance optimization exercise; that is, these settings use the default values.

- Virtual memory

```
ame_cpus_per_pool = n/a
ame_maxfree_mem = n/a
ame_min_ucpool_size = n/a
ame_minfree_mem = n/a
ams_loan_policy = n/a
enhanced_affinity_affin_time = 1
enhanced_affinity_vmpool_limit = 10
esid_allocator = 1
force_relalias_lite = 0
kernel_heap_psize = 65536
lgpg_regions = 0
lgpg_size = 0
low_ps_handling = 1
maxperm = 16378813
maxpin = 17052800
maxpin% = 90
memory_frames = 18874368
memplace_data = 0
memplace_mapped_file = 0
memplace_shm_anonymous = 0
memplace_shm_named = 0
memplace_stack = 0
memplace_text = 0
memplace_unmapped_file = 0
minperm = 545953
minperm% = 3
msem_nlocks = 0
npskill = 2048
npswarn = 8192
num_locks_per_semid = 1
numpsblks = 262144
pinnable_frames = 16407621
relalias_percentage = 0
scrub = 0
thrpgio_inval = 1024
thrpgio_npages = 1024
```

```
v_pinshm = 0
vm_mmap_bmap = 1
vmm_default_pspa = 0
vmm_klock_mode = 2
wlm_memlimit_nonpg = 1
```

- I/O

```
aio_active = 0
aio_maxreqs = 131072
aio_maxservers = 30
aio_minservers = 3
aio_server_inactivity = 300
j2_atimeUpdateSymlink = 0
j2_inodeCacheSize = 200
j2_maxRandomWrite = 0
j2_metadataCacheSize = 200
j2_nRandomCluster = 0
j2_recoveryMode = 1
j2_syncPageCount = 0
j2_syncPageLimit = 16
lvm_bufcnt = 9
maxpgahead = 8
maxrandwrt = 0
numclust = 1
numfsbufs = 196
pd_npages = 4096
posix_aio_active = 0
posix_aio_maxreqs = 131072
posix_aio_maxservers = 30
posix_aio_minservers = 3
posix_aio_server_inactivity = 300
spec_accessupdate = 0
```

- Scheduler

```
affinity_lim = 7
big_tick_size = 1
ded_cpu_donate_thresh = 80
fixed_pri_global = 0
force_grq = 0
maxspin = 16384
pacefork = 10
proc_disk_stats = 1
sched_D = 16
sched_R = 16
tb_balance_S0 = 2
tb_balance_S1 = 2
tb_threshold = 100
timeslice = 1
vpm_fold_policy = 1
vpm_throughput_core_threshold = 1
vpm_throughput_mode = 0
vpm_xvcpus = 0
```

The performance settings that were changed during the optimization exercise are listed in “Suggested performance settings” on page 5.

PURR accounting and interpreting performance results

IBM Power Systems, starting with the POWER7 processor architecture, use Processor Utilization Resource Register (PURR) accounting for accurate reporting of system usage.

On POWER8 processor-based servers, the PURR factor for SMT4 is 0.59 - 0.63, which means a single thread in SMT4 mode consumes a maximum of 59 - 63% of the physical core. In SMT8 mode, the PURR factor is 0.56, which means a single thread in SMT8 mode consumes a maximum of 56% of the physical core. In SMT2 mode, the PURR factor is 0.76 - 0.78. SMT2 mode is optimized for single thread performance on POWER8 processor-based servers.

SAS jobs report processor metrics (such as system time and user time) with the `fullstimer` SAS option. The numbers reported by the `fullstimer` option do not account for PURR. The PURR factor must be applied based on the SMT mode to get correct measurement for processor usage. For example, in the SMT8 mode, if a job reports 180 seconds as CPU time (user + system), the actual time spent by the processor in the user + system mode for the job is 321 seconds (180/0.56) as per PURR accounting.

Resources

The following resources are useful references to supplement the information in this paper:

- IBM Power Systems Information Center:

  http://publib16.boulder.ibm.com/pseries/index.htm
- IBM Power Systems at IBM PartnerWorld®:

  http://ibm.com/partnerworld/systems/p
- IBM AIX:

  http://www.ibm.com/systems/power/software/aix/
- SAS on IBM AIX 5L, AIX 6, and AIX 7 Tuning Guides:

  http://ibm.com/support/techdocs/atsmastr.nsf/WebIndex/WP101529
- IBM Power Systems scale-out servers:

  http://ibm.com/systems/power/hardware/scale-out.html
- *Implementing IBM FlashSystem 840*, SG24-8189:

  http://ibm.com/redbooks/abstracts/sg248189.html?Open
- *Accelerate insights with SAS Business Analytics and IBM FlashSystem*:

  http://public.dhe.ibm.com/common/ssi/ecm/ts/en/tsw03263usen/TSW03263USEN.PDF
- SAS 9.3 grid deployment on IBM Power servers with IBM XIV Storage System and IBM GPFS:

  http://ibm.com/support/techdocs/atsmastr.nsf/WebIndex/WP102192
- Understanding Processor Utilization on Power Systems – AIX:

  http://ibm.com/developerworks/community/wikis/home?lang=en#!/wiki/Not+AIX/page/Understanding+Processor+Utilization+on+Power+Systems+-+AIX

Authors

This paper was produced by a team of specialists from around the world working at the International Technical Support Organization, Poughkeepsie Center.

Beth L. Hoffman is an IBM Executive IT Specialist and is Thought Leader Certified in Business Partner Enablement. She has more than 25 years of experience at IBM. As a Solution Architect, Beth leads the ISV Technical Enablement for Big Data, Analytics, and NoSQL solutions for IBM Systems Unit. Her areas of expertise include solution architecture, reference architectures, application development, emerging technologies, Power Systems, and Linux. Beth holds a Computer Science degree and an MBA degree from Minnesota State University.

Narayana Pattipati is a Senior Technical Consultant at IBM India. He has 15 years of IT experience in systems software development, open source software, ISV technical enablement, and Big Data and Analytics. Narayana holds a Bachelor of Engineering degree in Mechanical Engineering from Birla Institute of Technology and Science (BITS) Pilani, India.

Thanks to the following contributor to this project:

Harry Seifert
IBM

Now you can become a published author, too!

Here's an opportunity to spotlight your skills, grow your career, and become a published author—all at the same time! Join an ITSO residency project and help write a book in your area of expertise, while honing your experience using leading-edge technologies. Your efforts will help to increase product acceptance and customer satisfaction, as you expand your network of technical contacts and relationships. Residencies run from two to six weeks in length, and you can participate either in person or as a remote resident working from your home base.

Find out more about the residency program, browse the residency index, and apply online at:

ibm.com/redbooks/residencies.html

Stay connected to IBM Redbooks

- Find us on Facebook:

  http://www.facebook.com/IBMRedbooks

- Follow us on Twitter:

  http://twitter.com/ibmredbooks

- Look for us on LinkedIn:

  http://www.linkedin.com/groups?home=&gid=2130806

- Explore new Redbooks publications, residencies, and workshops with the IBM Redbooks weekly newsletter:

  https://www.redbooks.ibm.com/Redbooks.nsf/subscribe?OpenForm

- Stay current on recent Redbooks publications with RSS Feeds:

  http://www.redbooks.ibm.com/rss.html

Notices

This information was developed for products and services offered in the U.S.A.

IBM may not offer the products, services, or features discussed in this document in other countries. Consult your local IBM representative for information on the products and services currently available in your area. Any reference to an IBM product, program, or service is not intended to state or imply that only that IBM product, program, or service may be used. Any functionally equivalent product, program, or service that does not infringe any IBM intellectual property right may be used instead. However, it is the user's responsibility to evaluate and verify the operation of any non-IBM product, program, or service.

IBM may have patents or pending patent applications covering subject matter described in this document. The furnishing of this document does not grant you any license to these patents. You can send license inquiries, in writing, to:
IBM Director of Licensing, IBM Corporation, North Castle Drive, Armonk, NY 10504-1785 U.S.A.

The following paragraph does not apply to the United Kingdom or any other country where such provisions are inconsistent with local law: INTERNATIONAL BUSINESS MACHINES CORPORATION PROVIDES THIS PUBLICATION "AS IS" WITHOUT WARRANTY OF ANY KIND, EITHER EXPRESS OR IMPLIED, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF NON-INFRINGEMENT, MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE. Some states do not allow disclaimer of express or implied warranties in certain transactions, therefore, this statement may not apply to you.

This information could include technical inaccuracies or typographical errors. Changes are periodically made to the information herein; these changes will be incorporated in new editions of the publication. IBM may make improvements and/or changes in the product(s) and/or the program(s) described in this publication at any time without notice.

Any references in this information to non-IBM websites are provided for convenience only and do not in any manner serve as an endorsement of those websites. The materials at those websites are not part of the materials for this IBM product and use of those websites is at your own risk.

IBM may use or distribute any of the information you supply in any way it believes appropriate without incurring any obligation to you.

Any performance data contained herein was determined in a controlled environment. Therefore, the results obtained in other operating environments may vary significantly. Some measurements may have been made on development-level systems and there is no guarantee that these measurements will be the same on generally available systems. Furthermore, some measurements may have been estimated through extrapolation. Actual results may vary. Users of this document should verify the applicable data for their specific environment.

Information concerning non-IBM products was obtained from the suppliers of those products, their published announcements or other publicly available sources. IBM has not tested those 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.

This information contains examples of data and reports used in daily business operations. To illustrate them as completely as possible, the examples include the names of individuals, companies, brands, and products. All of these names are fictitious and any similarity to the names and addresses used by an actual business enterprise is entirely coincidental.

COPYRIGHT LICENSE:

This information contains sample application programs in source language, which illustrate programming techniques on various operating platforms. You may copy, modify, and distribute these sample programs in any form without payment to IBM, for the purposes of developing, using, marketing or distributing application programs conforming to the application programming interface for the operating platform for which the sample programs are written. These examples have not been thoroughly tested under all conditions. IBM, therefore, cannot guarantee or imply reliability, serviceability, or function of these programs.

This document REDP-5288-00 was created or updated on October 23, 2015.

Send us your comments in one of the following ways:

- Use the online **Contact us** review Redbooks form found at:
  ibm.com/redbooks
- Send your comments in an email to:
  redbooks@us.ibm.com
- Mail your comments to:
  IBM Corporation, International Technical Support Organization
  Dept. HYTD Mail Station P099
  2455 South Road
  Poughkeepsie, NY 12601-5400 U.S.A.

Trademarks

REDP-5288-00

ISBN 0738454532

Printed in U.S.A.