Accelerate insights with SAS 9.4 deployed on IBM POWER9 and IBM FlashSystem 9150

IBM FlashSystem 9150, IBM Spectrum Scale, and IBM POWER9 deliver excellent performance for SAS 9.4

Highlights

- IBM Flash System 9150 with NVMe drives from the IBM FlashSystem 9100 product family delivers exceptional performance for SAS Mixed Analytics.
- IBM Power E980 server shows high performance analytics capability.
- Peak file access from storage was over 28 GBps for a workload with a read and write ratio of 65:35.

This technical brief demonstrates the test results from SAS 9.4 software deployed on IBM® POWER9™ processor-based servers and IBM FlashSystem® 9150 storage from the IBM FlashSystem 9100 product family using IBM's new Non-Volatile Memory Express (NVMe) attached disks based on IBM FlashCore® Modules technology.

The high bandwidth and low latency of the FlashCore Modules in the FlashSystem 9150 system clearly result in improved SAS workload real times. Teams from SAS and IBM applied a methodology to tune each component in the infrastructure which allowed to achieve optimal performance. The goals, results, and supporting information about the solution implemented are documented in this technical brief.

Goals

- Demonstrate the overall performance solution of POWER9 and FlashSystem 9150 running a demanding SAS Mixed Analytics (MA) workload.
- Show the high performance and low latency of IBM NVMe FlashCore Modules
- Use IBM Spectrum Scale, a shared file system running with FlashSystem 9150, to optimize storage demands in a multi-host environment when applications are demanding I/O storage requests needing low latency and high read/write bandwidth.
- Measure SAS Mixed Analytics workload throughput when running a 20-session and a 30-session workload concurrently on each of the four POWER9 logical partitions (LPARs) running the IBM AIX® operating system.
- Show concurrent execution of a 30-session SAS Mixed Analytics workload on POWER9 with four nodes. Each node has 18 cores, 128 GB memory, and four 16 Gbps Fibre Channel (FC) ports. The Fibre Channel is attached using two IBM SAN64B-6 switches.

Reference architecture

Software

- SAS V9.4 M5
- IBM AIX V7.2
- IBM PowerVM®
- IBM VIOS V2.2.6.21
- IBM Spectrum Scale V5.0.1.2

Hardware

- IBM Power E980 server (Model 9080-M9S)
- IBM FlashSystem 9150

SAN

- Dual IBM SAN64B-6 switch
- Eight 4-port 16 Gbps FC adapters (two dedicated per AIX node) on Power E980
- Six 4-port 16 Gbps FC adapters on FlashSystem 9150

Architecture

The infrastructure selected was the IBM FlashSystem 9150 storage system, IBM Power® System E980 server, IBM Spectrum Scale 5.0.1.2 file system, and two IBM SAN64B-6 Fibre Channel switches. The storage fabric was 16 Gbps Fiber Channel. Each node on the Power E980 server was connected with four FC ports (16 FC ports in total) to the switch. The FlashSystem 9150 system had 24 FC ports to the switch using six 4-port 16Gbps FC adapters. The FC fabric connectivity is shown in the architecture diagram.

The software building blocks are the IBM AIX operating system and the IBM Spectrum Scale shared file system. The test bed employed was the SAS Mixed Analytics workload based on the SAS V9.4 M5 platform. This combination creates a powerful system with enterprise capabilities allowing for an architecture providing high performance computing, storage, and storage fabric scalability.

Server

The **IBM Power System E980 server** is based on the IBM Power Architecture®. The architecture uses the concept of LPARs, which allow one or more cores in the system to be logically organized. These LPARs constitute the nodes used to run the workload. The diagram in Figure 1 has four LPARs, each with 18 cores and 128 GB memory. Each LPAR has four configured 16 Gbps FC ports (two each coming from two 16 Gbps 4-port FC adapters dedicated to each LPAR). The Power E980 server was pre-GA and running the AIX V7.2 operating system. The cores were in dedicated mode running SMT8.

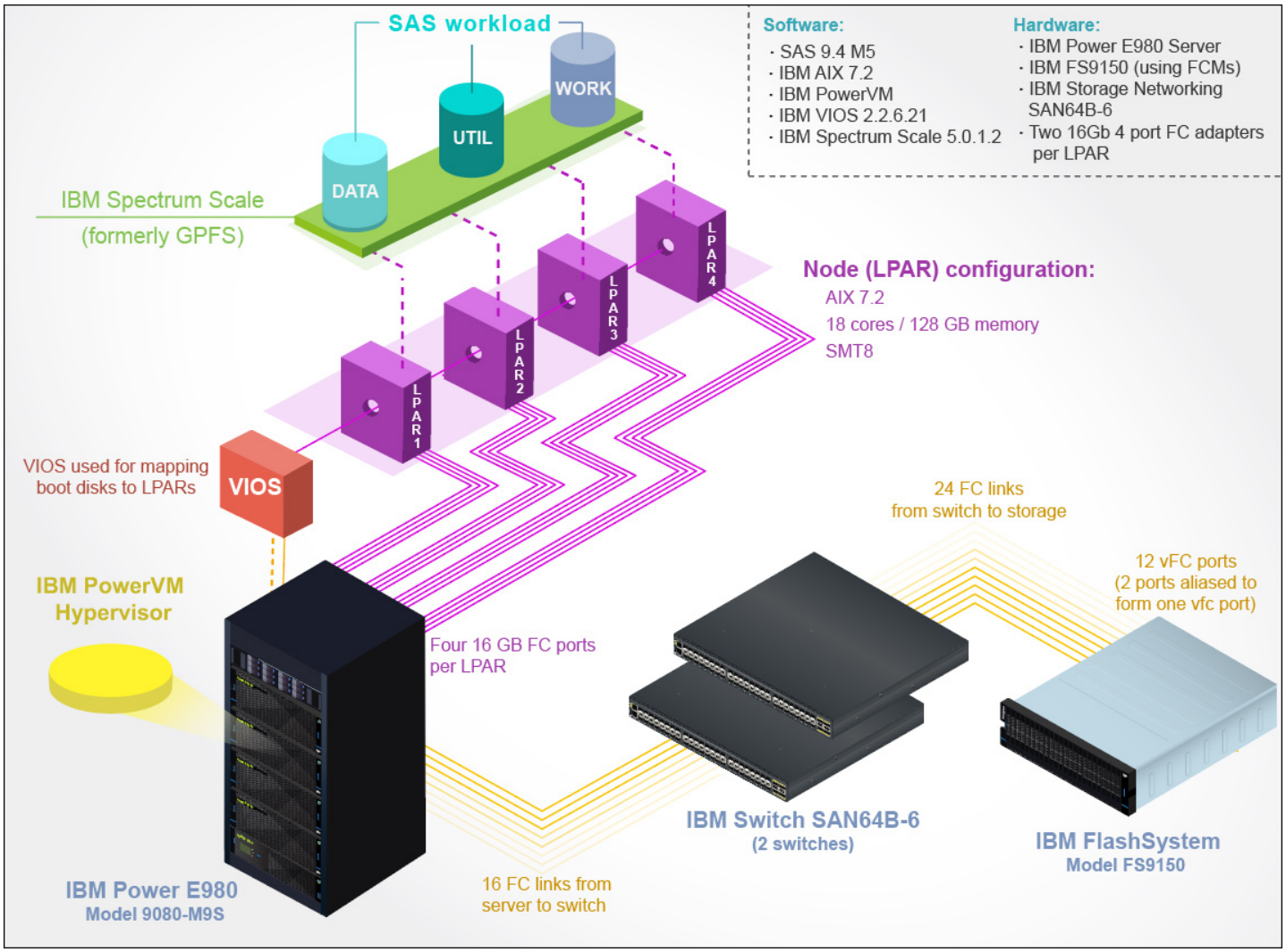

Figure 1. Architecture for SAS 9.4 with IBM FlashSystem 9150, IBM POWER9, and IBM SAN64B-6 switch

FlashSystem 9150 Storage

The IBM FlashSystem 9150 is a comprehensive enterprise-class storage system using NVMe drives and optional flash drives. It provides a rich set of software-defined storage (SDS) features, including data reduction, de-duplication and IBM Spectrum Virtualize — all in a powerful 2U configuration with up to 24 ports of FC (16 Gbps or 32 Gbps), or 12 ports iSCSI [**RDMA over Converged Ethernet (RoCE)]** / [Internet Wide Area RDMA Protocol) iWARP)].

As configured for this test, the FlashSystem 9150 system used the new IBM FlashCore Module NVMe disk drives. The IBM FlashCore Module NVMe dual ported disk drives have the unique features of built-in data reduction and encryption at the hardware level. It can provide up to a 2:1 compression factor with a 6 microsecond (or less) penalty per r/w cycle. Though the drives are available in 4.8 TB, 9.6 TB, and 19.2 TB the 19.2 TB density was used for the tests.

For testing, the FlashSystem 9150 system used 24 IBM FlashCore Module NVMe 19.2 TB drives. The array was configured with DRAID6, which tolerates up to two simultaneous drive failures and faster rebuild times. The base usable capacity was 384 TB. But with built-in 2:1 hardware compression it is possible to have as much as 772 TB of effective capacity without inhibiting performance.

IBM Spectrum Scale shared file system

IBM Spectrum Scale is a proven, scalable, high-performance data and file management solution based on IBM General Parallel File System (IBM GPFS™). IBM Spectrum Scale provides a world-class storage management with extreme scalability. IBM Spectrum Scale reduces storage costs while improving security and management efficiency in cloud, big data, and analytics environments. It is a powerful data management system that enables the unification of block, file, and object storage into a single comprehensive solution for a project or the entire data center.

A single Spectrum Scale file system was created with recommended parameters to create the DATA (permanent data), WORK (working data), and UTIL (utility data) subdirectories under it. The SAS BUFSIZE was set to 256 KB. Testing the SAS workload was performed with various file system block sizes for best performance. A 4 MB block size with a 8 KB sub-block size was chosen. This was also the recommendation for good sequential performance.

Note: The most important factor when configuring storage for SAS performance is throughput and not capacity. Large block sequential I/O is a design goal for tuning performance.

SAS software

The software tested is SAS V9.4 M5 (the latest version at the time). The test suite that drove the work in order to measure performance is the SAS Mixed Analytics workload. The SAS Mixed Analytics workload is used to provide a means to run many tests on a system. The 20- and 30-session Mixed Analytics workload scenarios were best suited for the test goals.

SAS Mixed Analytics test suite

The SAS Mixed Analytics workload consists of a mix of jobs that run in a concurrent and back-to-back fashion. These jobs stress the compute, memory, and I/O capabilities of the infrastructure. The SAS test team described the test bed they employed as a *good average SAS Shop* set of workload mix.

The SAS Mixed Analytics workload concurrent tests can be scaled up to test the system performance bandwidth. For the tests discussed in this technical brief, the team ran a 20-session and a 30-session SAS Mixed Analytics workloads. As an example, the SAS Mixed Analytics 20-session workload consists of 20 individual SAS jobs: 10 compute-intensive, 2 memory-intensive, and 8 I/O-intensive. Some of the test jobs rely on existing data stores and some test jobs rely on data generated during the test run. The tests are a mix of short-running (in minutes) and long-running (in hours) jobs. The tests are repeated to run both concurrently and in a serial fashion to achieve a 20-session workload. For the tests run in this proof of concept, when the single node 20-session workload was completed it had run a total of 71 jobs. There is a similar scaling of the 30-session workload where 101 jobs in total were run.

The 20-session and 30-session workloads were also run concurrently on each of the four AIX LPARs.

Data and I/O throughput

A single instance of the SAS Mixed Analytics 20 simultaneous test workloads on each node drives an aggregate of about 300 GB of data for the I/O tests and about 120 GB of data for the computation tests. Much more data is generated as a result of test-step activity and threaded kernel procedures.

It is important to note that SAS I/O pattern is predominately large-block, sequential I/O. There is some random access but sequential is the dominant access pattern. When configuring for SAS I/O, there are multiple distinct patterns such as large sequential workloads in the multi-gigabyte to terabyte size, small file sequential, random access, and random data step activity. However, it is the large sequential block I/O that dominates these patterns. Keeping that in mind helps to configure the file systems.

Testing the infrastructure

Many test cases were performed (by varying the LPAR resources, number of SAS sessions, SMT levels and so on) and the tests were focused on the scenarios shown in this technical brief. Tests on a single node as well as multiple nodes were performed.

All tests completed successfully. There was no limitation seen on the Power E980 (in terms of processor usage, memory, or network) but the storage fabric was saturated due to limited FC ports. The number of ports was doubled from 16 to 32. A corresponding increase in bandwidth was not observed but led to increased read/write service times on the FlashSystem 9150 system as the test team was nearing its maximum supported bandwidth for the configuration used.

Single / Four-node 20-session and 30-session Mixed Analytics workload

It was observed that the workload, during its peak, exceeded 21 GBps read and 6 GBps write throughput.
The test suite is highly active for about 90 minutes and then very gradually finishes with two or three low-impact, long-running *trail out* jobs. This is what the SAS team describes as a good average *SAS Shop* throughput characteristic for a single-node instance that simulates the load of an individual SAS compute node. The throughput is depicted from all three primary SAS file systems DATA, WORK, UTIL. The estimated I/O ratio for read and write operations was 65:35.

Test results

SAS Mixed Analytics 20-session and 30-session tests were run on both single and four AIX nodes.

Figure 2 shows the workload statistics: aggregate SAS FULLSTIMER real time (total elapsed run time, summed from each of the jobs in the workload) and the summed CPU time (User + System), both in minutes. Also shown is ratio of CPU to real time.

Figure 3 shows the maximum read/write bandwidth captured in the FlashSystem 9150 system.

The maximum read and write bandwidth recorded at the application level (nmon data from AIX nodes) is shown. The last column shows the ratio of the total CPU time to SAS real time. Because some SAS procedures are multi-threaded, you can use more CPU cycles than wall-clock, or real time leading to a ratio greater than 1 as seen in the results. This is ideal for an I/O intensive SAS application set.

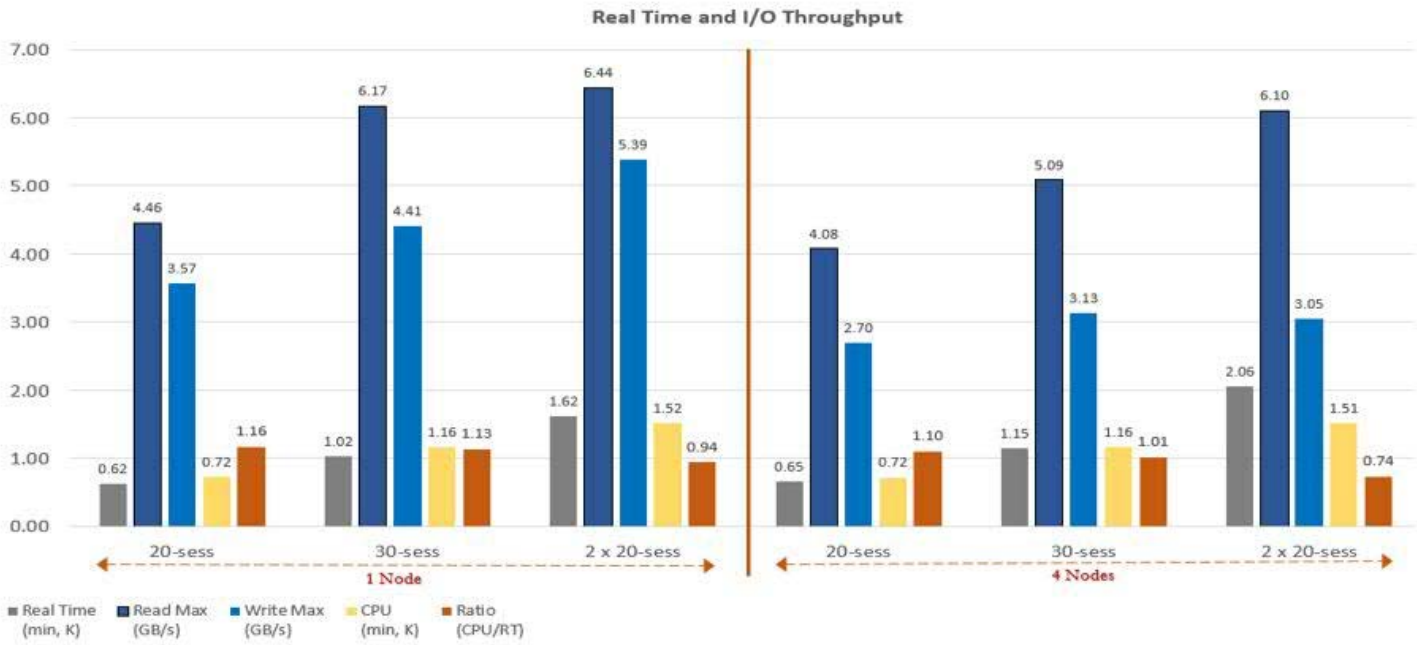

Figure 2: Graphical representation of workload performance on 1 and 4 nodes (averaged per node)

		Read	Write			Read	Write			Read	Write
Latency	11 ms	12 ms	8 ms	Bandwidth	20,483 MBps	15,011 MBps	5,470 MBps	IOPS	83,765	61,855	21,910
Latency	22 ms	22 ms	23 ms	Bandwidth	21,839 MBps	15,633 MBps	6,205 MBps	IOPS	88,320	63,484	24,836
Latency	26 ms	30 ms	8 ms	Bandwidth	27,279 MBps	22,172 MBps	5,105 MBps	IOPS	109,601	89,155	20,445
Latency	21 ms	25 ms	6 ms	Bandwidth	28,110 MBps	21,795 MBps	6,314 MBps	IOPS	115,064	89,791	25,271

Figure 3. FlashSystem 9150 bandwidth captured at various points during workload

Benefits of IBM FlashCore Modules

The chart in Figure 4 shows the results of the tests done using IBM FlashCore Modules versus generic NVMe modules:

- IBM FlashCore Modules with its always-on hardware built-in compression matches the performance of the generic modules where compression was **disabled.**
- Bandwidth drops significantly when compression is enabled on the generic modules due to higher latency as the compression is done by software.
  With FlashCore Modules, there is near-zero performance penalty as all the compression logic is built in to the hardware. This is valid up to a compression ratio of 2:1.

  Note: Two 20-session and two 30-session workloads used to generate high load on the FlashSystem 9150 system.

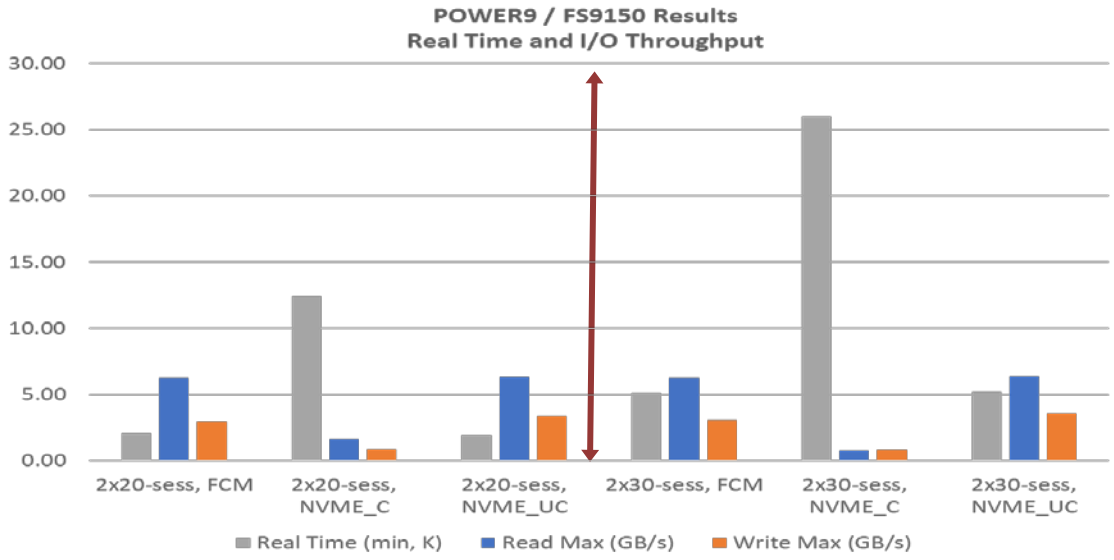

Figure 4. Comparison of IBM FlashCore NVMe modules over generic NVMe drives

Summary

The test results demonstrate that the high-performance combination of the IBM Power System E980 system and the IBM FlashSystem 9150 system proved to be a great choice for deploying SAS software which demands:

- Servers that can deliver high throughput per core
- Storage systems that can provide high bandwidth and low latency

The IBM Power E980 server proved to be a powerful work engine meeting the needs of the SAS Mixed Analytics workload. The processor, throughput, and memory limits of the system were not seen during the tests which shows the vitality of this system.

The IBM FlashSystem 9150 system with IBM Spectrum Scale proved to be a viable combination that provided a robust storage system with a shared file system capability.

The IBM FlashCore Module NVMe drives with their in-built always-on hardware compression had little to no impact on performance due to data reduction. The benefits of NVMe technology were demonstrated by delivering great performance.

Get more information

To learn more about the IBM and SAS products and capabilities, contact your IBM representative or IBM Business Partner, or visit the following websites:

- IBM FlashSystem 9100 Offering
- IBM FlashSystem 9100 Details
- IBM Power E980
- IBM Spectrum Scale
- IBM Storage Networking SAN64B-6
- SAS Institute

About the authors

Beth Hoffman is an IBM solution architect in IBM Cognitive Systems ISV Enablement organization. You can contact her at bethvh@us.ibm.com or www.linkedin.com/in/bethhoffmanibm

Abhijit Mane is a technical consultant in IBM Cognitive Systems ISV Enablement organization. You can contact him at abhijman@in.ibm.com or https://www.linkedin.com/in/abhijitmane

Brian Porter is an IBM storage solution architect in IBM Cognitive Systems Storage organization. You can contact him at bporter1@us.ibm.com or www.linkedin.com/in/brian-porter-a27a06b

Harry Seifert is an ISV technical sales support specialist in IBM Global Markets Sales organization supporting SAS deployments worldwide. You can contact him at seifert@us.ibm.com or www.linkedin.com/in/harry- seifert-329a336

Ben Smith is an IBM solutions architect in the IBM Systems Software Defined Infrastructure organization. You can contact him at smithbe1@us.ibm.com or www.linkedin.com/in/smithbe1

Acknowledgements

The authors would like to thank the following team members for their input and review of this technical brief:

Margaret Crevar
Manager, SAS Performance Research and Development Lab
You can contact her at Margaret.Crevar@sas.com or www.linkedin.com/in/margaret-crevar-78392112

Tony Brown
Distinguished Engineer, SAS Performance Research and Development Lab
You can contact him at Tony.Brown@sas.com or www.linkedin.com/in/tony-brown-1848753

IBM Systems
3039 Cornwallis Road
RTP, NC 27709

Produced in the United States of America

Please recycle