Accelerate insights with SAS 9.4 deployed on IBM Power10 and IBM FlashSystem 9200

IBM FlashSystem 9200, IBM Spectrum Scale, and IBM Power10 deliver excellent performance for SAS 9.4

Highlights

- IBM FlashSystem 9200 with NVMe FlashCore Modules from the IBM FlashSystem 9200 product family delivers exceptional performance for SAS Mixed Analytics workload.
- IBM Power E1080 server shows high performance analytics capability.
- Peak file access from storage was over 54 GBps for a workload with a read and write ratio of 65:35.

This technical brief summarizes the test results from SAS 9.4 M7 software deployed on IBM® Power10 processor-based servers and IBM FlashSystem® 9200 storage using IBM's Non-Volatile Memory Express (NVMe) attached disks embedded with IBM FlashCore® Modules with GEN2 technology.

The high bandwidth and low latency of the FlashCore Modules in the FlashSystem 9200 system result in vastly improved real times for this SAS Mixed Analytics workload. Teams from SAS and IBM applied methodologies that fine tune each component of the infrastructure, including the SAS software, Power servers, 32 Gb SAN fabric and all-flash array storage components. The tuning enabled optimal performance to be achieved. The solution goals, results, and conclusions are documented in this technical brief. The solution and tuning can easily be applied to almost any large scale SAS Grid deployment framework.

Goals

- Demonstrate the overall performance of Power10 and FlashSystem 9200 running a demanding SAS Mixed Analytics workload.
- Show the high performance and low latency of IBM NVMe FlashCore Modules
- Use IBM Spectrum® Scale 5.1.3, a high performance shared file system running with FlashSystem 9200 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 on a single node as well as concurrently on each of the four Power10 logical partitions (LPARs) running the IBM AIX® operating system.
- Measure the performance of running SAS 9.4 M7 with AIX 7.3 and Spectrum Scale to demonstrate the full performance potential the Power10 and AIX 7.3 enhancements can achieve.
- Show concurrent execution of 20-session and 30-session SAS Mixed Analytics workloads on Power10 with four nodes. Each node has 18 cores, 128 GB memory, and four 32 Gbps Fibre Channel (FC) ports. The Fibre Channel is attached using an IBM SAN64B-6 switch.

Reference architecture

Software

- SAS 9.4 M7
- IBM AIX V7.3
- IBM PowerVM®
- IBM VIOS V3.1.3.10
- IBM Spectrum Scale V5.1.3

Hardware

- IBM Power E1080 server (Model 9080-HEX)
- IBM FlashSystem 9200

SAN

- Dual IBM SAN64B-6 switch
- Eight 2-port 32 Gbps FC adapters (two dedicated per AIX node) on Power E1080
- Six 4-port 32 Gbps FC adapters on FlashSystem 9200

Architecture

The infrastructure selected was the IBM FlashSystem 9200 storage system, IBM Power® System E1080 server, IBM Spectrum Scale 5.1.3 file system, and an IBM SAN64B-6 Fibre Channel switch. The storage fabric was 32 Gbps Fiber Channel. Each node on the Power E1080 server was connected with four FC ports (16 FC ports in total) to the switch. The FlashSystem 9200 system had 24 FC ports to the switch using six 4-port 32 Gbps FC adapters. The FC fabric connectivity is shown in the architecture diagram in Figure 1.

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 9.4 M7 platform. This combination creates a powerful system with enterprise capabilities allowing for an architecture providing high performance computing, storage, and storage fabric scalability and resiliency.

Power10 processor-based server

The **IBM Power System E1080 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 32 Gbps FC ports (two each coming from two 32 Gbps 2-port FC adapters dedicated to each LPAR). The Power E1080 server was running the AIX Version 7.3 operating system. The cores were in dedicated mode running SMT8.

FlashSystem 9200 storage

IBM FlashSystem 9200 is a comprehensive enterprise-class storage system using NVMe drives and optional flash core modules. 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)].

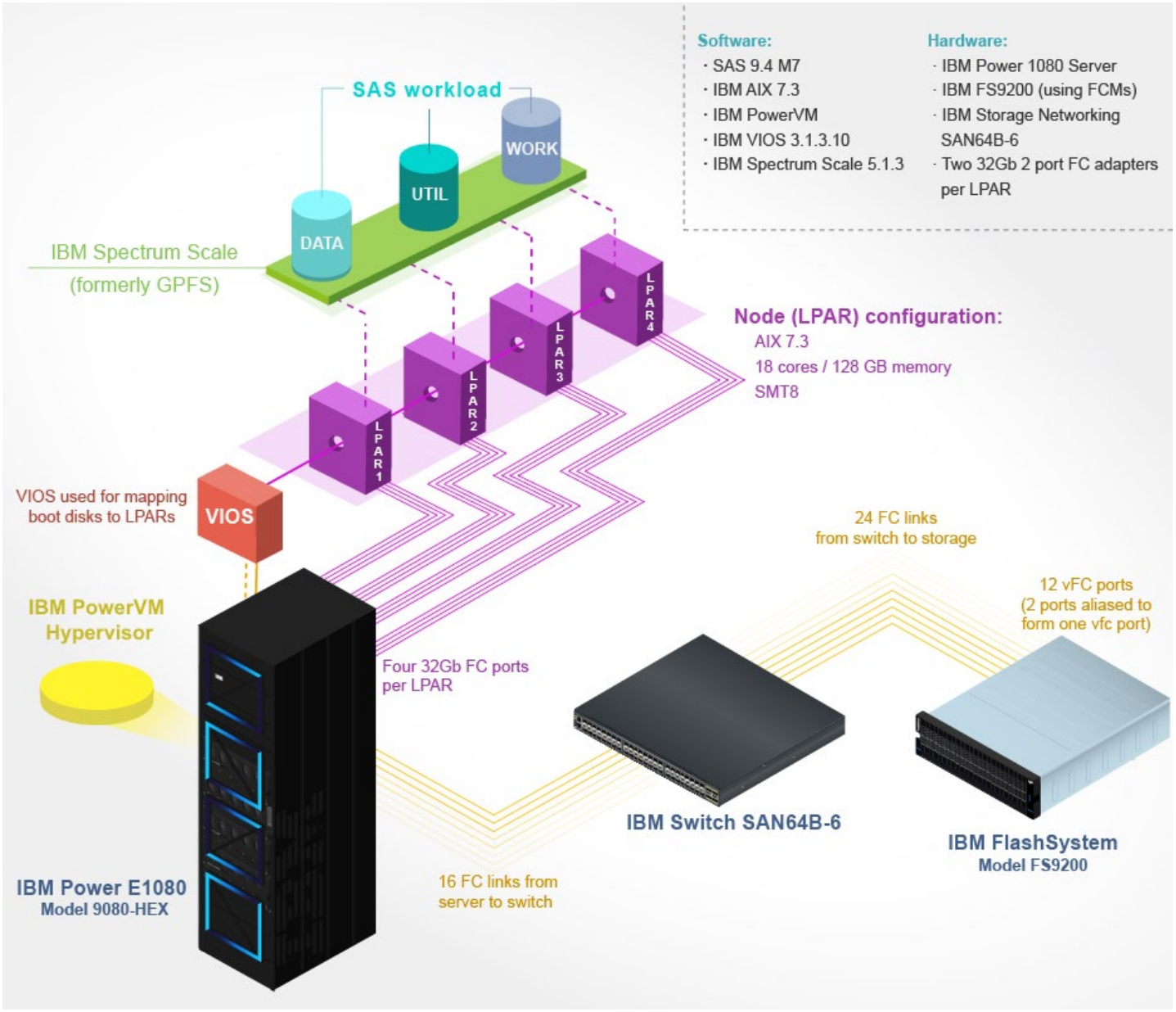

Figure 1. Architecture for SAS 9.4 with IBM POWER10, FlashSystem 9200, and IBM SAN64B-6 switch

As configured for this test, the FlashSystem 9200 storage system used 2nd generation 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 read/write cycle. The drives are available in 4.8 TB, 9.6 TB, 19.2 TB, and 38.4 TB density sizes.

For testing the SAS Mixed Analytics workload, the FlashSystem 9200 system used 24 IBM FlashCore Module NVMe 9.6 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 230 TB. However, with built-in 2:1 hardware compression, it is possible to have nearly 480 TB of effective capacity without inhibiting performance. With IBM FlashCore Modules, there is near-zero performance penalty as all the compression logic is built into the hardware.

IBM Spectrum Scale shared file system

IBM Spectrum Scale is a proven, scalable, high-performance data and file management solution based on the IBM General Parallel File System (IBM GPFS™). IBM Spectrum Scale provides a world-class storage management framework with extreme scalability. IBM Spectrum Scale reduces storage costs while improving performance, 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/temporary data), and UTIL (utility data) file sets under it. The SAS BUFSIZE was set to 256 KB. Testing the SAS Mixed Analytics workload was performed with various file system block sizes for best performance. A 4 MB block size with an 8 KB sub-block size was chosen. This was also the recommendation for good sequential performance. The file system and the metadata were deployed using dedicated logical unit numbers (LUNs) for each to minimize contention for the system pool resources.

Note: The most important factor when configuring storage for SAS performance is I/O throughput and not input/output operations per second (IOPS). Large block sequential I/O is a design goal for tuning performance.

SAS software

The software tested is SAS 9.4 M7 (the latest version at the time). The workload 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 SAS jobs on a system. The 20- and 30-session Mixed Analytics workload scenarios were best suited for the test goals.

SAS Mixed Analytics workload

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 workload they employed as a *good average SAS 9.4 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. When the single node 20-session workload test 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 jobs 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, 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 E1080 server (in terms of processor usage, memory, or network), however, the storage fabric was saturated at 55+ GBps due to limited FC ports.

Single node 30-session Mixed Analytics workload

It was observed that the single node SAS Mixed Analytics workload running on a Power10 processor-based server with 18 cores achieved 14 GBps of I/O throughput (approximately 11 GBps read throughput and 3 GBps write throughput). Figure 2 shows nmon data for a single AIX node. The green line illustrates the actual peak throughput of 14 GBps (or 1400 KBps) from a single LPAR. The WAvg in orange includes the trailing idle time. Along the bottom of the figure, you can see the throughput for each of four I/O read adapters and four write adapters with the average shown in blue. Figure 2 distinctly shows perfectly distributed I/O throughput across the available adapters for both read and write operations.

After the tests approached the CPU utilization of 75%, additional nodes were added to validate linear scalability.

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

It was observed that the four-node SAS Mixed Analytics workload, during its peak, exceeded 42 GBps read and 12 GBps write throughput. The workload is highly active for about 70 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, and UTIL. The estimated I/O ratio for read and write operations was 65:35.

Comparing test results

Figure 3 shows the SAS Mixed Analytics workload statistics for the Power10 four-node test results compared to the identical testing done previously on IBM Power9 with FlashSystem 9150. The left half of the chart shows the comparison results for two 20-session test and the right half shows the comparison results for two 30-session tests. There are three colored bars and each represent a key performance measurement for SAS workloads. The two time-related measurements are shown in minutes and the smaller the number the better. The ratio is shown as a percentage and the smaller the number the better.

SAS real-time results

The orange bar in Figure 3 measures aggregate SAS FULLSTIMER real time (total elapsed run time, summed from each of the jobs in the workload). The real time for two 20-session SAS workload was reduced by 300 minutes (which calculates to 5 hours) when running on Power10 with FS9200 versus the same tests on Power9 with FS9150. The real time for two 30-session SAS workload was reduced by 724 minutes (which calculates to 12 hours). For both scenarios this is a 29% real time reduction.

CPU time results

The gray bar in Figure 3 measures the summed CPU time (user + system). The CPU time for two 20-session was reduced by 576 minutes (or more than 9 hours and 30 minutes) when running on Power10 cores with FlashSystem 9200 versus the same tests on Power9 cores with FlashSystem 9150. The CPU time for two 30-session was reduced by 1,021 minutes (or almost 17 hours). The CPU time was reduced by 76% and 78% respectively!

Ratio results

The blue bar in Figure 3 shows the ratio of CPU time to SAS real time which reflects the efficiencies in processor core utilization. The two 20-session ratio and the two 30-session ratio both saw a 65% reduction with Power10 and FlashSystem 9200 compared to Power9 with FlashSystem 9150. 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.

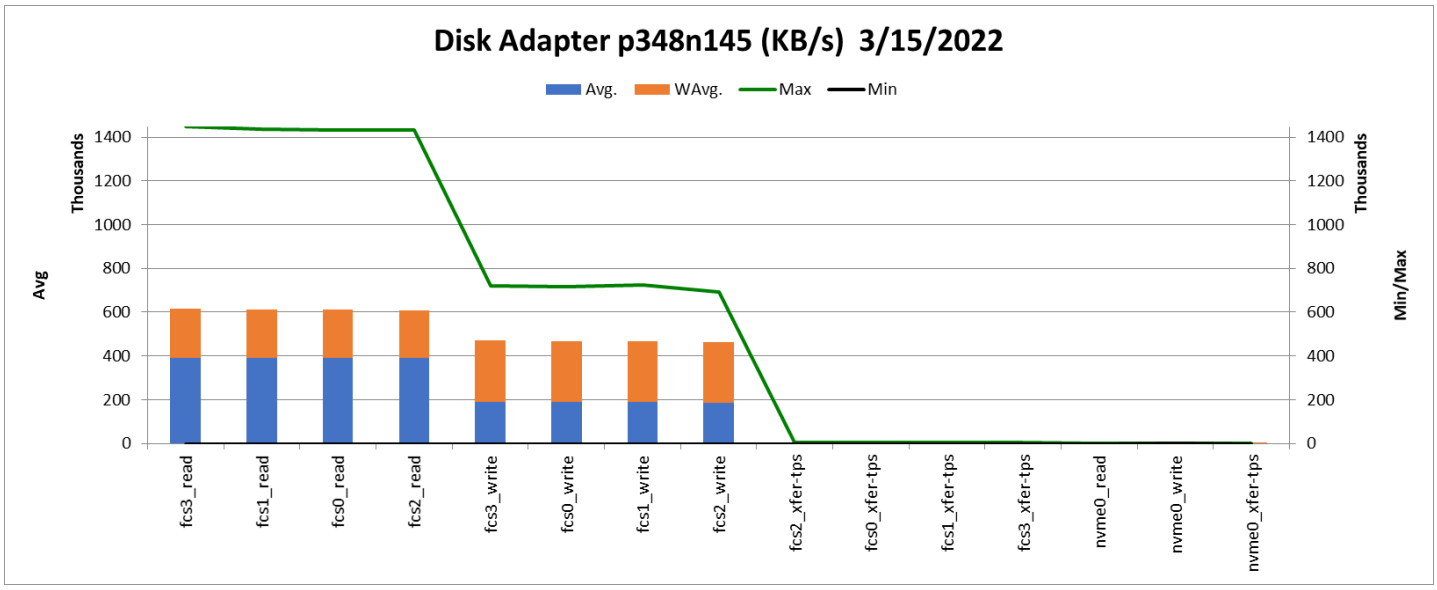

Figure 2. SAS Mixed Analytics workload single-node Power10-FlashSystem 9200 disk adapter I/O throughput rate and peak

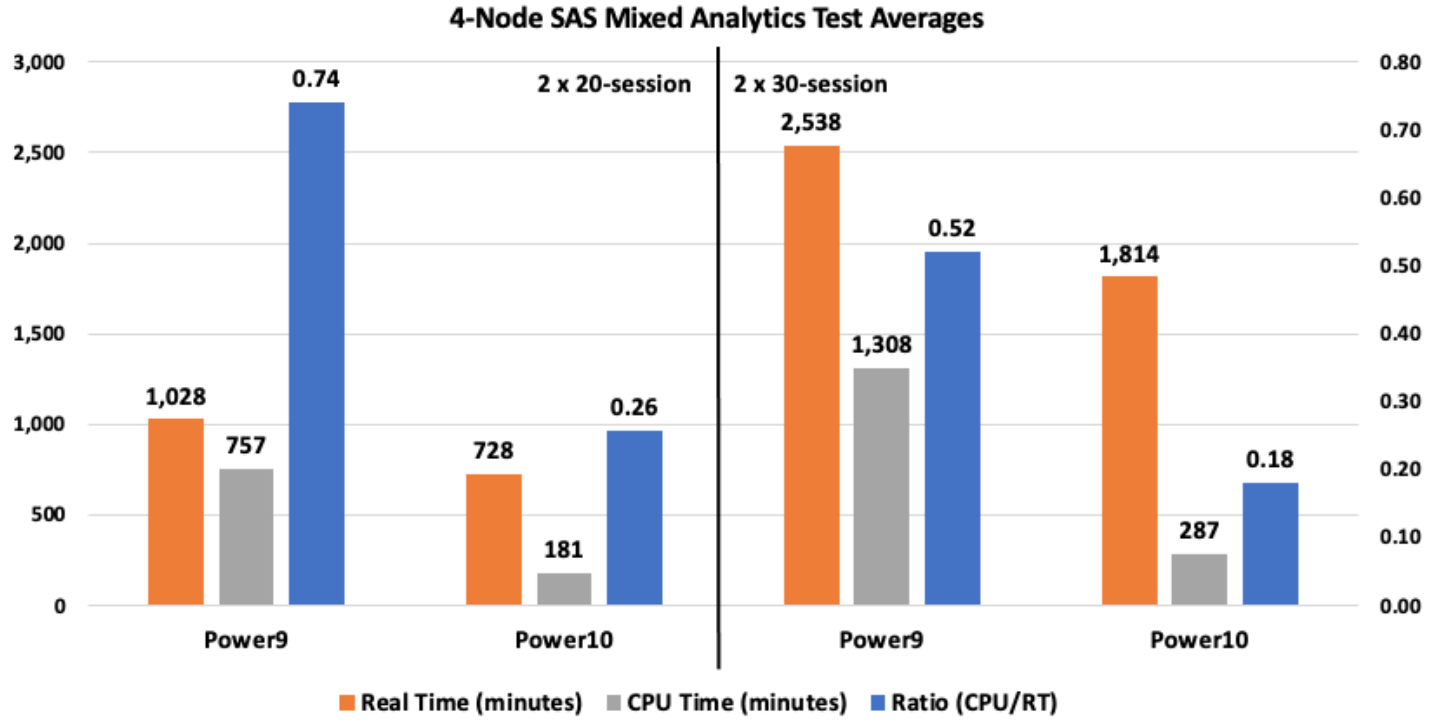

Figure 3: Graphical representation of SAS Mixed Analytics workload performance on four nodes (averaged per node)

Summary

The test results demonstrate that the high-performance combination of the IBM Power E1080 server and the IBM FlashSystem 9200 system proved to be an excellent 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
- Servers with utmost resiliency

The IBM Power E1080 server proves to be a powerful work engine meeting the needs of the SAS Mixed Analytics workload. No processor, throughput, or memory limits of this system were observed during the extensive testing, which clearly demonstrates the vitality of this end-to-end reference architecture design. The IBM FlashSystem 9200 system with IBM Spectrum Scale proves 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 with high resiliency while placed under significant stress.

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 9200 Offering
- IBM FlashSystem 9200 Details
- IBM Power E1080
- 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

Harry Seifert is a Principal Power Systems Technical Sales Specialist and ISV Solution Architect in IBM Advanced Technology Group, Americas. You can contact him at seifert@us.ibm.com or www.linkedin.com/in/harry-s-329a336

Glen Corneau is a Senior Power Partner Technical Specialist and IBM Spectrum Scale on AIX SME in IBM Advanced Technology Group, Americas. You can contact him at gcorneau@us.ibm.com or www.linkedin.com/in/corneau

Byron Grossnickle is a Senior Brand Technical Specialist and IBM Flashcore NVMe Storage SME in IBM Advanced Technology Group, Americas. You can contact him at byrongro@us.ibm.com or www.linkedin.com/in/byron-grossnickle

Vlad Sorokin is a Certified Technical Specialist and Cloud Infrastructure Engineer in IBM Client Engineering for Systems in Global Sales. You can contact him at vvsorokin@us.ibm.com

Acknowledgments

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

Jim Kuell
Software Performance Engineer, SAS Compute Services Research and Development
You can contact him at Jim.Kuell@sas.com or https://www.linkedin.com/in/jim-kuell-43a42354

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