September 2017

Transforming real-time insight into reality

IBM FlashSystem 900 provides the all-flash foundation for today's "big memory" solutions

Highlights

- Build high-performance, big memory solutions using IBM® FlashSystem® 900
- Accelerate big data analytics, cognitive applications, and mobile and social systems of engagement
- Reduce cost and complexity compared to traditional clustered server architectures

Technologies as diverse as Apache Spark, SAP HANA and NoSQL databases all share a single architectural feature—the use of in-memory data stores. Moving data sets from traditional storage media into memory can dramatically reduce system latency and increase application performance. But building effective in-memory data stores brings plenty of challenges. Chief among these are cost and scalability.

IBM has been a leader in researching, engineering and bringing to market ultra-low latency data storage solutions. IBM FlashSystem 900 provides the all-flash storage system foundation for many application workloads and business use cases that demand extremely fast response times, mission-critical reliability, and simple, essentially unlimited scalability.

Big memory

Data management tools have evolved to address the increased performance and scalability requirements of cognitive applications, mobile and social systems of engagement, real-time analytics, and the Internet of Things (IoT), among many others. NoSQL databases, for example, employ data structures such as key-value stores that differ from those used in relational databases and help make big data analytics and real-time web applications faster. Cognitive use cases such as the algorithms for machine learning systems were among the original drivers behind the

development of Apache Spark.[1] SAP HANA provides business intelligence to enterprises worldwide by enabling analysts to query large volumes of data virtually in real time. Like NoSQL databases and Apache Spark implementations, SAP HANA achieves high performance by storing data sets in server memory. But these data sets are growing ever larger, leading to requirements for what might be described as "big memory."

Constructing big memory systems is not easy. The most common solution, until recently, has been to aggregate memory by replicating multiple commodity servers filled with costly DRAM. This approach has a number of limitations, including cost, resource utilization and energy consumption. Typical DRAM costs can be literally hundreds of times more than hard disk drive storage and up to 10 times the cost of flash,[2] making the use of DRAM for very large data sets prohibitively expensive. Also, DRAM-based solutions still require disk- or flash-based persistent storage, and when a cache miss fails over to disk, the performance of these next-generation systems plummets.

To address these challenges, IT architects are turning to flash storage. Flash serves as a lower cost, persistent extension of the existing DRAM while providing exceptional performance compared to spinning disk storage. For these reasons, among others, big memory solutions built with flash storage are quickly gaining popularity.

In the past, flash-based big memory solutions have been built using large clusters of commodity servers filled with solid-state drives (SSDs). But these scale-out models typically lead to more hardware and software, greater cost and lower reliability, among other liabilities. New technologies such as the Non-Volatile Memory express (NVMe) protocol and the Coherent Accelerator Processor Interface (CAPI) from IBM are enabling the construction of effective big memory solutions using all-flash, array-based architectures. These solutions are much simpler to build, can significantly reduce system costs, can offer higher performance and much greater reliability, and can scale without adding redundant server hardware and software.

The heart of big memory

IBM FlashSystem 900 offers an excellent storage platform on which to construct cost-efficient, highly reliable, easily scalable and extremely low-latency big memory solutions. For more than three decades, the IBM FlashSystem 900 lineage has been purpose-engineered to address ultra-low latency requirements. Long before the advent of Apache Spark or NoSQL, low-latency use cases existed in the scientific and intelligence communities, among others. And there to serve them well were the ancestors of IBM FlashSystem 900. Until the summer of 2007, the precursors to IBM FlashSystem 900 used RAM instead of NAND flash. Those storage arrays were engineered to

accommodate 10 to 20 microsecond latencies, with internal buses and interfaces operating at essentially bare-metal speeds. That heritage and engineering focus has carried over to this day in IBM FlashCore® technology.

Figure 1. IBM FlashSystem 900

IBM FlashSystem 900 delivers the extreme performance, enterprise reliability and operational efficiencies required by big memory, real-time analytics workloads and use cases. For storage environments already equipped with robust storage management capabilities such as those provided by IBM Db2®, IBM Spectrum Virtualize™ or IBM Spectrum Scale™ software, the ultra-low-latency, market-leading performance and ease of implementation offered by IBM FlashSystem 900 make it an ideal choice for transforming business-critical applications into engines of business growth and competitive advantage.

IBM FlashSystem 900 is composed of up to 12 massively parallel IBM MicroLatency® modules that provide very high storage density with latency in the 100 microsecond range. The systems can scale usable capacity from as low as two terabytes (TB) to as much as 57 TB in a single array. The MicroLatency modules also support an AES-256 encryption engine, high-speed internal interfaces, and full hot-swap and storage capacity scale-out capabilities, enabling organizations to achieve lower cost per capacity without compromising enterprise reliability.

IBM FlashSystem 900 offers some powerful advantages over other ultra-low latency solutions in the marketplace today:

- First, it's a system already thoroughly proven for years in the most demanding environments. The technology doesn't come from a recent startup with little experience or track record in supporting critical applications.
- Second, it doesn't need special application programming interface protocols or complementary systems in order to be effectively deployed. The arrays offer industry-standard Fibre Channel, InfiniBand and Ethernet connectivity.
- Finally, it's not designed with only a very limited set of use cases in mind. Instead, IBM FlashSystem 900 arrays have already been deployed in almost every conceivable IT environment, from standard databases, through high-performance computing, to leading-edge big memory proofs-of-concept.

Purpose-engineered matters

Originally, most all-flash storage systems began essentially as software projects. Ingenious engineers wrote code with storage-tiering capabilities, then loaded this into a box filled with SSDs, and thus an all-flash array was born. But not all have chosen this path to market, and some, such as IBM FlashSystem and its precursors, have used purpose-engineered flash storage instead of SSDs from the beginning, while more and more vendors are now pivoting toward this orientation for very good reasons. All-flash arrays employing optimized rather than commodity flash resources offer a number of advantages over both the clustered commodity server architectures with direct-attached SSDs often employed for big memory solutions and the all-flash arrays that are essentially "boxes of SSDs." These advantages include:

- *Lower latency.* Because the flash resources and capacity of purpose-engineered all-flash arrays such as IBM FlashSystem 900 are consolidated in a single array rather than distributed between many individual commodity server nodes, performance, latency and system resiliency can all be maximized far beyond the capabilities of these distributed systems. Also, the ratio of controllers to flash capacity can be optimized within purpose-engineered all-flash arrays, eliminating many throughput and performance bottlenecks often encountered by commodity server/SSD architectures.
- *Higher density.* Removing packaging and other redundancies from individual SSDs can sharply increase storage density.
- *Greater reliability.* Most enterprise storage systems employ some kind of system-wide data protection regime, such as various RAID configurations uniting individual SSDs in an array or in distributed servers for that matter. But failures within each SSD are hidden from these systems and their data protection monitoring software. By contrast, IBM FlashSystem 900 has built-in flash management monitoring and mitigation down to the individual flash chip level and below, making data protection much more robust, granular and efficient.
- *Lower cost.* Deploying flash storage can lower costs in big memory environments because it can significantly reduce the amount of DRAM required to support high-performance in-memory implementations. In fact, when much less complex all-flash array architectures are used instead of distributed systems, the number of server nodes required to provide enough memory and storage capacity for big memory implementations is reduced because sharing an all-flash array reduces stranded resources on servers, thus significantly cutting overall implementation costs.[2] Finally, hardware accelerated all-flash storage can actually be less costly per gigabyte (GB) than SSDs because there are fewer individual packaging and controller redundancies, among the many other advantages noted above.

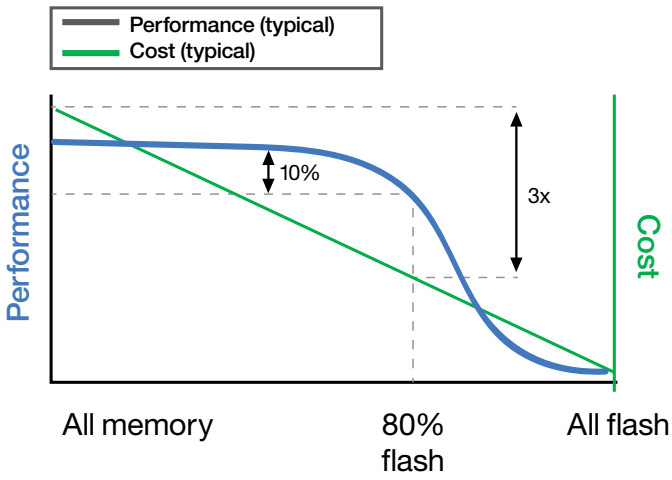

Figure 2. Relative performance and cost as a function of memory:flash ratio

Building efficient big memory solutions

The question becomes how to build ultra-low latency big memory solutions that take advantage of the reliability and performance offered by IBM FlashSystem 900 arrays. As it happens, the key is in the interface. Data traffic to and from the CPU and storage traverses a number of components, all of which add some latency, including the processor to backplane latency, the operating system, the file system, the interface protocol, the network and the storage system latencies. Various technologies are in play across each of these components, with different options affecting the total latency, cost and efficiency of the overall solution. To address this challenge of aggregate system latencies, one technology receiving plenty of attention these days is NVMe.

By design, NVMe allows host hardware and software to more fully exploit the microsecond latency provided by all-flash arrays. It reduces input/output (I/O) overhead between CPUs and storage, resulting in performance improvements compared to previous interfaces that were originally developed for use with far slower hard disk drives. NVMe is already being utilized for internal PCI express (PCIe)-based storage. Driven by the need for much larger memory capacities and better reliability and data protection in order to perform even bigger analytics jobs using Spark, SAP HANA and other related technologies, the IT industry has developed NVMe over fabrics (NVMe-F) to map the NVMe protocol onto storage network fabrics—either IP or Fibre Channel—providing faster storage access for applications. NVMe-F, when fully realized, can enable IT architects to connect external storage resources such as all-flash arrays with hundreds of TBs and even petabytes of capacity to analytics engines over an ultra-low latency storage network fabric. Essentially, the size of in-memory applications can be unlimited, without compromising any of the reliability and data services expected from enterprise-class storage systems.

Beyond NVMe

Complementing NVMe, IBM has developed CAPI, a protocol offering even lower overhead and thus lower latency than NVMe technologies. CAPI allows storage devices to interact with processors and system memory like another processor would. To make CAPI technology available to developers worldwide, IBM helped found the OpenCAPI Consortium in October 2016. CAPI provides a way to attach storage devices with essentially no software overhead or device drivers.

It demonstrates that there has been significant, ongoing IBM innovation in big memory for ultra-low latency workloads. CAPI offers significant advantages over NVMe technologies:

- CAPI allows IBM FlashSystem 900 to interact with processors and system memory like another processor would, enabling applications to access flash directly, bypassing elements such as the device driver, kernel, pinned pages or memory copies.
- This eliminates the majority of CPU cycles associated with moving data to and from flash, freeing up CPUs to do useful work (thus avoiding one of the pitfalls associated with other flash-as-memory solutions)
- The removal of code path length from flash access significantly reduces application-visible latency compared to accessing flash through standard I/O subsystem pathways.

Big memory use cases

These days, big memory is big business. For example, market research suggests that by 2020, NoSQL sales will reach USD3.4 billion that year alone, driven by a 21 percent annual growth rate.[3] Scaling up web infrastructure on NoSQL databases has proven successful for Facebook, Digg and Twitter, among many others. Clearly, the market for large, high-performance data stores and big data analytics is enormous.

The cybersecurity arena provides an excellent real-world example. A number of enterprises are currently deploying the cybersecurity product SparkSecure from IBM Business Partner SparkCognition.[4] SparkSecure helps find cyber threats faster and optimize cybersecurity workforces. The solution combines an on-premises SparkSecure client, IBM FlashSystem 900 storage to help ensure maximum performance, and IBM Watson™ in the cloud. SparkSecure analyzes structured and unstructured data in the form of server logs and natural language sources such as social media in order to identify possible attacks or vulnerabilities in a cloud environment. It can automatically detect patterns present in new threats and provide recommendations to help protect against them. All of this knowledge is delivered in context, at the right time, to the right person.

Tests conducted by IBM and SparkCognition confirm that a simple, cost-effective, high-performance IBM POWER8® and IBM FlashSystem-based infrastructure provides multiple benefits compared to the typical clustered x86-based server configuration currently used for SparkSecure deployments. When deployed in large enterprises, SparkSecure can generate more than 300 GB of data per day from millions of events that are parsed, analyzed and reported. Such implementations previously required multiple servers and a massive number of hard disk drives to achieve acceptable performance and throughput. Conversely, the solution architecture tested by IBM and SparkCognition consisted of a single POWER8 server and IBM FlashSystem 900 array. The test results confirm that not only is this solution more efficient and much less complex, but it also provides significantly better performance.[4]

Additionally, the performance test results don't directly speak to the reduced complexity and increased efficiency provided by the IBM FlashSystem 900-based architecture. The x86-based test bed utilized a cluster of seven servers, a typical infrastructure solution for SparkSecure. Even with SSDs added, this complex configuration couldn't muster the performance easily provided by one POWER8 server and one IBM FlashSystem 900 array. For IT administrators, the reduced complexity offered by the IBM FlashSystem 900 solution architecture translates into increased system resiliency and availability, with less time spent managing infrastructure.

Making real-time possible

Real-time analytics power 21st-century business. Enterprises as diverse as large research institutions, energy utilities and online dating sites use in-memory solutions based on technologies such as NoSQL data stores, Apache Spark, and SAP HANA to handle the rapid-fire, high-volume, variable application I/O from their systems of engagement. Originally, the "big memory" systems required to support these cognitive, cloud, web-based and big data analytics applications were constructed using clustered commodity server architectures. But new technologies such as NVMe and IBM CAPI are enabling IT architects to design big memory solutions that are much less complex and lower in cost, but that also offer even higher performance. IBM FlashSystem 900 offers an excellent all-flash array foundation for these leading-edge big memory solutions. It provides the ultra-low latency, proven reliability, simple scalability and high-density flash cost efficiency needed to transform real-time insight into reality.

For more information

To learn more about IBM FlashSystem 900, please contact your IBM representative or IBM Business Partner, or visit: **ibm.com**/storage/media

Additionally, IBM Global Financing provides numerous payment options to help you acquire the technology you need to grow your business. We provide full lifecycle management of IT products and services, from acquisition to disposition. For more information, visit: **ibm.com**/financing

© Copyright IBM Corporation 2017

IBM Systems
New Orchard Road
Armonk, NY 10504

Produced in the United States of America
September 2017

IBM, the IBM logo, ibm.com, IBM FlashSystem, IBM FlashCore, Db2, IBM Spectrum Virtualize, IBM Spectrum Scale, IBM Watson, MicroLatency, POWER, and POWER8 are trademarks of International Business Machines Corp., registered in many jurisdictions worldwide. Other product and service names might be trademarks of IBM or other companies. A current list of IBM trademarks is available on the web at "Copyright and trademark information" at **ibm.com**/legal/copytrade.shtml

This document is current as of the initial date of publication and may be changed by IBM at any time. Not all offerings are available in every country in which IBM operates.

The performance data discussed herein is presented as derived under specific operating conditions. Actual results may vary. It is the user's responsibility to evaluate and verify the operation of any other products or programs with IBM products and programs.

THE INFORMATION IN THIS DOCUMENT IS PROVIDED "AS IS" WITHOUT ANY WARRANTY, EXPRESS OR IMPLIED, INCLUDING WITHOUT ANY WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE AND ANY WARRANTY OR CONDITION OF NON-INFRINGEMENT. IBM products are warranted according to the terms and conditions of the agreements under which they are provided.

Actual available storage capacity may be reported for both uncompressed and compressed data and will vary and may be less than stated.

[1] Derrick Harris, "4 reasons why Spark could jolt Hadoop into hyperdrive," *Gigaom*, June 28, 2014. https://gigaom.com/2014/06/28/4-reasons-why-spark-could-jolt-hadoop-into-hyperdrive/

[2] "IBM Data Engine for NoSQL—Power Systems Edition" (white paper), *IBM Corp.*, October 2014. https://www-304.ibm.com/webapp/set2/sas/f/capi/CAPI_FlashWhitePaper.pdf

[3] "NoSQL Market Forecast 2015-2020," *Market Research Media*, January 2016. https://www.marketresearchmedia.com/?p=568

[4] "Accelerating a more cognitive enterprise" (technical white paper), *IBM Corp.*, June 2016. https://www-01.ibm.com/common/ssi/cgi-bin/ssialias?htmlfid=TSW03445USEN

Please Recycle

TSW03555-USEN-00