Which Data Platform Is Right for Your Most Demanding Mission Workloads?

Introduction

As artificial intelligence, machine learning, and technical computing become mainstream workloads, federal agencies are demanding solutions to meet the high-performance requirements for these and more types of applications that can provide mission and program success.

Yesterday's supercomputer power is today's IT ecosystem powered by GPUs, and the GPUs, like supercomputers, are very hungry for data. The faster you can feed the GPU the faster you will get to results. That's why selecting the right data platform is critical. You don't want the GPUs running idle waiting for data, and you don't want your data platform to be so fragile that it requires a team of admins to manage and keep it running.

We will look at the most common solutions available to address these high performance, next generation workloads and compare the feature/functionality of each so you can make the most informed decision possible.

Stop Thinking Storage, Start Thinking Data Pipeline

Rear Admiral Grace Murray Hopper famously said, "The most damaging phrase in the English language is: we've always done it this way." She also said humans are allergic to change, which was why she had a clock on her wall that ran backwards. It was her way of fighting against complacency and addressing the problem with the same old tools. It wasn't that she thought everything had to be addressed with something new and fresh, it was the thought process she wanted to instill to consider the options around you rather than the routine you have become comfortable in, and the same applies when addressing which solutions best meet your needs for the next generation workloads.

When considering a data platform for your workloads, if you take an old school, "we've always done it that way" approach, it will invariably lead you back to a file-based storage conversation. But is storage really what should be at the heart of your decision criteria? It should be about how you will enable your high-performance ecosystem to efficiently, and concurrently serve data to the application workloads fueling the GPUs. This is what is referred to as a data pipeline. These GPU-accelerated, and data-intensive workloads consume data significantly faster than CPU-based, which is why a storage solution designed for legacy workloads will often create a data bottleneck. Studies have shown these bottlenecks may leave GPUs idle for up to 70% of the time. Not only that but from a sustainability perspective, an idle server can still draw as much as 50% of maximum power[1], which is wasted energy and wasted compute cycles. Again, the faster data can reach the cores of the GPU, the faster it can be processed and deliver the results to your business.

> "The most damaging phrase in the English language is: we've always done it this way."
>
> **REAR ADMIRAL GRACE MURRAY HOPPER**

Revolutionary vs. Evolutionary Change

Throughout computing's brief history it seems there has always been a clear delineation between scientific computing and enterprise IT computing needs. Going back to the early to mid 90s, while IT was learning all about the newly launched Microsoft Windows 95 and placing bets on either Microsoft NT Server or Novell 4 with Directory Services for its shared file services, IBM introduced the Vesta Parallel File System used exclusively for the IBM SP2 RISC based supercomputer. Ultimately, over time, enterprise IT filesystems naturally evolved as application and workloads shifted while parallel file systems experienced their own evolutionary change with the introduction of Lustre back in 1999 at the request of the Department of Energy (DoE). These two paths continued to evolve separately and at their own pace while maintaining the demarcation between scientific/research and enterprise IT. That is until the mid-2000s when GPUs were being used for applications requiring complex, simultaneous processing in what would be considered an enterprise IT environment. Two researchers at Stanford University wrote a paper[2] about the technological gains in machine learning applications. Over time, GPUs have been adopted for massively parallel processing of complex computations across demanding operations within the workloads of artificial intelligence, machine learning, and technical computation, just to name a few.

While the enterprise file system continued its evolutionary journey, IT professionals questioned whether this legacy scale-out/scale-up NAS architecture could actually meet the performance requirements of the GPUs. At the same time parallel file system providers saw a potential opportunity to reach down-market and support those GPU driven workloads.

While vendors on both sides recognized the opportunity that GPU fueled workloads would present, neither were necessarily ready for the requirements of IT to achieve their desired business outcomes. Revolutionary, not evolutionary change is what the market truly needed to keep pace with not only the technological advancements but customer expectations.

Good Enough, Is Not Enough

It is a fact that storage solutions have gotten faster over the years using a variety of methods such as accelerator caching cards, filesystem tweaks, CPU and RAM upgrades, and simply a tech refresh such as moving from HDD to SSD. Even though these traditional storage solutions have gained performance, many of these GPU fueled workloads still wait for data, as mentioned earlier. Even some of the recent entrants to the market fail because they are based on an old legacy architecture masked by the promises of flash. And while the parallel file systems may have the capability of delivering on performance, it is the complexities of management, reliability, usability, and overall customer experience that has left many enterprise IT organizations hesitant to deploy into their own data center or cloud. This dilemma has led enterprise IT to have serious trade-off discussions internally in order to meet company expectations.

On the one hand you have what we'll call the general purpose needs for Central IT, and on the other side the emerging next generation workloads for data analysis , simulations, AI/ML, deep learning, etc. with a completely different set of must-have needs and compromises.

The parallel file systems were created out of necessity and immediacy for many private large-scale operations and national labs. Not much concern was given to the user experience since it was also run by the engineers that coded the filesystem. It was simply accepted that the file system would be tuned as needed based on the data profile, and characteristics since these large organizations and labs had the staff available to do so. Where central IT has been faced with tighter budgets and seemingly fewer staff members to manage the environment, those solutions were developed with features and user experience in mind to deliver as much value with as little hands-on management as possible. This crossroad is where the industry must meet IT with an uncompromising solution, taking full advantage of the available GPU powerhouses in the IT ecosystem to deliver results, while at the same time offer a rich set of features that will support the more everyday enterprise IT workloads, thus eliminating multiple points of manage, support, and expense as it relates to maintenance and overhead.

In the next few pages are the products selected for this comparison and are considered by many to be representative of the leaders targeting the next generation workloads.

	VAST Data	IBM GPFS—Spectrum Scale	Lustre	WEKA
Capacities	500TB-1EB	48TB-8EB	48TB-8EB	8EB
All Flash/Object Hybrid	No	Yes	No	Yes
Specialized HW	Yes	Yes, IBM Appliances	No	No, Industry Standard
Read	40GB/sec[3]	40GB/sec	48GB/sec[4]	56GB/sec
Write	5GB/sec[5]	32GB/sec	34GB/sec[4]	20GB/sec
Read IOPs	<400,000	No data from vendor	3,000,000[4]	5,800,000
Write IOPs	<400,000	No data from vendor	No data from vendor[4]	1,600,000
Protocol Support				
NFS	Y	Y	Y, export via ZFS	Y
LDAP	N	Y	Y	Y
SMB	Y, SMB2	Y	Y, Samba	Y, SMB2/3
GPUDirect	Y	N	Y	y
POSIX	N	Y	Y	Y

	VAST Data	IBM GPFS—Spectrum Scale	Lustre	WEKA
Protocol Support				
S3	Y	Y, limited	Y	Y, as single namespace
Data Encryption	Y, with limitations. The entire cluster must be configured for encryption at installation.[6]	At-Rest & In-Flight	Linux kernel fscrypt	At-Rest & In-Flight
Data Protection	N+4	N+3, Reed Solomon only on ESS Appliance	EC	N+2, N+4[7]
Data Reduction	Y	Limited	N	Y
Snapshots	Y	Y, with perf impact	Y, on ZFS	Yes
Max Snapshots per file system	1,000	256	Native snapshots not supported	4,096
Snapshot to S3	Y	N	N	Y, instantaneous
SW Only	Yes, Requires Specialized HW	Yes	Yes, Open Source	Y
Which cloud platform(s)?	None	IBM Private Cloud, AWS	AWS	AWS, Google, Oracle, Azure
Tiering/Hybrid Disk Storage	N, Caches data to SCM. Requires multiple copies before writing to QLC. Inefficient.	No tiering, impacts performance	Yes, complex policy based HSM	Y
Ease of Use	Easy	Hard	Hard	Easy

Compare and See

In this document four solutions, including WEKA, are compared based on the most common selection criteria customers use to determine which solution best fits their particular needs. Some of the solutions selected for comparison have been in the market for several years, having been developed prior to new standards such as NVMe Flash, Cloud, and de facto standard cloud protocols like S3. We understand in some cases the older, legacy solutions may be a better fit based on unique and monolithic data characteristics, however research indicates the data profiles in most enterprise IT environments are rarely monolithic, and depending on the workload may require independent filesystems configured specifically to support those monolithic data sets. This is where enterprise IT has generally reached a breaking point between managing multiple silos containing individually configured file systems for optimal performance outcomes or simply averaging out the results across one or two file systems and settling for more of a general purpose solution that is good enough.

VAST Data

VAST DATA is a relatively new company in the storage solution space having been initially designed and positioned as a backup target, presumably competing in the backup appliance market, and then shifting its focus on more general IT primary storage. VAST's initial product to market consisted of non-standard storage class memory (SCM) by Intel called Optane (3D XPoint), that would cache all of the writes to a broad deployment of SCM SSDs in the VAST system and then evacuate the data to low-end laptop grade QLC SSDs on the back end when more front-end cache space is needed. This design requires a great deal of copying of data across and between SCM devices in order to be as optimally compacted as possible prior to writing to QLC, which has the lowest endurance of any of the SSDs on the market.

IBM Spectrum Scale (GPFS)

Spectrum Scale or GPFS is a parallel file system which began as the Tiger Shark file system, a research project at IBM's Almaden Research Center back in 1993. Tiger Shark was initially designed to support high throughput multimedia applications, such as streaming video from VHS tapeswhich was also well suited for scientific computing at the time. GPFS was and still is a complex, brittle solution requiring a high level of expertise to plan, install, configure, and operate given its many configuration options.

Lustre

Similar to GPFS, Lustre's origins can be traced back to a research project that began at Carnegie Mellon University in 1999. By 2001 the company, Cluster File Systems, Inc was formed and work on what would become the Lustre file system had begun under a program funded by the US Department of Energy (DoE). Lustre is a parallel file system, like GPFS, but has struggled to find mainstream adoption outside of national labs and impromptu builds for spot testing and evaluations. Lustre has changed hands several times over the 20+ years it has been in existence, from Sun, Oracle, Whamcloud, and Intel. Intel abandoned Lustre in pursuit of its own file system, selling it off to DDN who has now released it back to open source and the user community. Given its age, little development has been done with Lustre over the years which is reflected in its current version number, 2.15..

SUPPORT FOR MONOLITHIC DATA SETS

This is where enterprise IT has generally reached a breaking point between managing multiple silos containing individually configured file systems for optimal performance outcomes or simply averaging out the results across one or two file systems and settling for more of a general purpose solution that is good enough.

Why WEKA for Federal IT

WEKA was founded on the idea that current storage solutions have only provided incremental improvements to legacy designs, allowing for a widening gap between compute performance and data storage performance. Storage remains a bottleneck to application performance, and with the continued densification of compute in areas such as GPU-based applications, has become even more problematic. In today's hyper-competitive market, federal organizations need flexible infrastructure; application workloads are becoming increasingly complex and data sets are continuing to grow unchecked, forcing agencies to architect overly complicated and costly systems that reduce IT agility. As a result, important mission and program insights remain locked away, out of reach of decision makers.

Federal IT organizations are adopting cloud technology for its fluid, on-demand scalability that supports diverse workloads at scale. However, while network and compute can be virtualized to operate at scale very effectively, storage remains largely isolated in silos based on system performance profiles. Consequently, federal organizations are forced to architect a storage system that is highly customized for their environment and workloads from building blocks that do not scale. The result is a storage solution that is complex, temperamental, expensive, and slow. WEKA has built a software-only, high-performance file-based storage solution that is highly scalable and easy to deploy, configure, manage, and expand. The design philosophy behind the WEKA file system (WekaFS™) was to create a single storage architecture that runs on-premises or in the public cloud with the performance of all-flash arrays, the simplicity and feature set of network-attached storage (NAS), and the scalability and economics of the cloud.

OPTIMIZED FOR NVME	MULTI-PROTOCOL READY	BUILT-IN DURABILITY	ADVANCED SECURITY	BUILT FOR THE CLOUD
Achieve lowest possible latency and hightest performance	Supports Linux, Windows and Native POSIX access to data	Uses distributed data protection and instant backup to S3 cloud for rapid recovery	Keeps your data completely safe with integrated encryption, key management, and access control	Seamlessly run on-premises, in the cloud and burst between platforms

FIG. 1 WEKA Benefits Summary

1 Characteristics and Energy Use of Volume Servers in the United States. https://escholarship.org/content/qt8bb5j7ww/qt8bb5j7ww.pdf
2 "Large-scale deep unsupervised learning using Graphic Processors" http://www.machinelearning.org/archive/icml2009/papers/218.pdf
3 VAST Data Read/Write Performance https://docs-dev.nersc.gov/filesystems/vast/
4 Lustre.org does not publish performance numbers, any that are published are based on very specific file types, sizes, and specific hardware. Numbers in chart are based on PCIe Gen3, PCIe Gen4 numbers 90GB/sec Reads, 65GB/sec Writes on DDN hardware according to Blocks and Files: https://blocksandfiles.com/2021/11/10/ddn-doubles-performance-of-high-end-ai-array/
5 VAST Data Read/Write Performance https://docs-dev.nersc.gov/filesystems/vast/
6 VAST Data Encryption Limitations https://support.vastdata.com/hc/en-us/articles/4414803734812-Encryption-at-Data-at-Rest#
7 Industry's first server-level N+4 data protection https://www.weka.io/wp-content/uploads/files/2017/06/WEKA-DDP-WP-W01R1WP202009.pdf

WEKA FEDERAL
weka.io | 844.392.0665

WKA363-01 08/2023