White Paper

Enterprises Looking to Optimize Storage Environments for Artificial Intelligence Workloads Should Consider a Data Platform Approach

Sponsored by: WEKA

Eric Burgener
May 2022

IDC OPINION

Artificial intelligence (AI)–driven workloads are growing at an 18.0% compound annual growth rate (CAGR) through 2025, and enterprises will be deploying them at a rapid rate to support the big data analytics that are increasingly driving business decisions in digitally transformed organizations. AI-driven workloads generally have multiple stages in their data pipeline that can require very different I/O profiles; as a result, enterprises often find themselves using different storage platforms and moving data around between them to meet these varying requirements. This siloed approach to storage for AI environments is time consuming, inefficient, and very costly (due to the need to purchase and maintain separate storage systems).

Fragmented storage infrastructure is not the only issue enterprises will encounter when deploying AI-driven workloads. IDC has noted that while AI pilot projects can be successfully run with a variety of different file system platforms, problems often arise when the workloads are deployed in production and start to scale. Getting the right storage infrastructure in place up front is important to avoid disruptive changes later. This means that the storage for AI must be not only flexible and able to cost-effectively meet the variety of different I/O profiles associated with the different data pipeline stages but also able to perform at scale.

Several AI data pipeline stages require the performance of parallel file systems (in particular, the use of a highly parallel client whose capabilities scale far beyond that of traditional NFS access methods). It makes compelling economic sense to deploy a single data platform for AI, provided that the solution can meet all data pipeline stage requirements and perform at the scale required by successful AI projects, where data tends to grow by more than the 30–40% each year (which enterprises generally experience during digital transformation). This suggests that enterprises should be looking at scale-out, file-based platforms that support a variety of different access methods (NFS, SMB, S3, and an intelligent [parallel] client) and offer the multitenant management capabilities enterprises need to meet performance, availability, recovery, and security requirements for mission-critical workloads.

SITUATION OVERVIEW

With more than 90% of global enterprises now actively pursuing digital transformation (the evolution to much more data-centric business models), large-scale data analytics workloads requiring high levels of performance to support are being widely deployed. Data is growing at 30-40% per year for most organizations, and more and more enterprises are working with artificial intelligence, machine learning (ML), and/or deep learning (DL) technologies to drive better business insights.

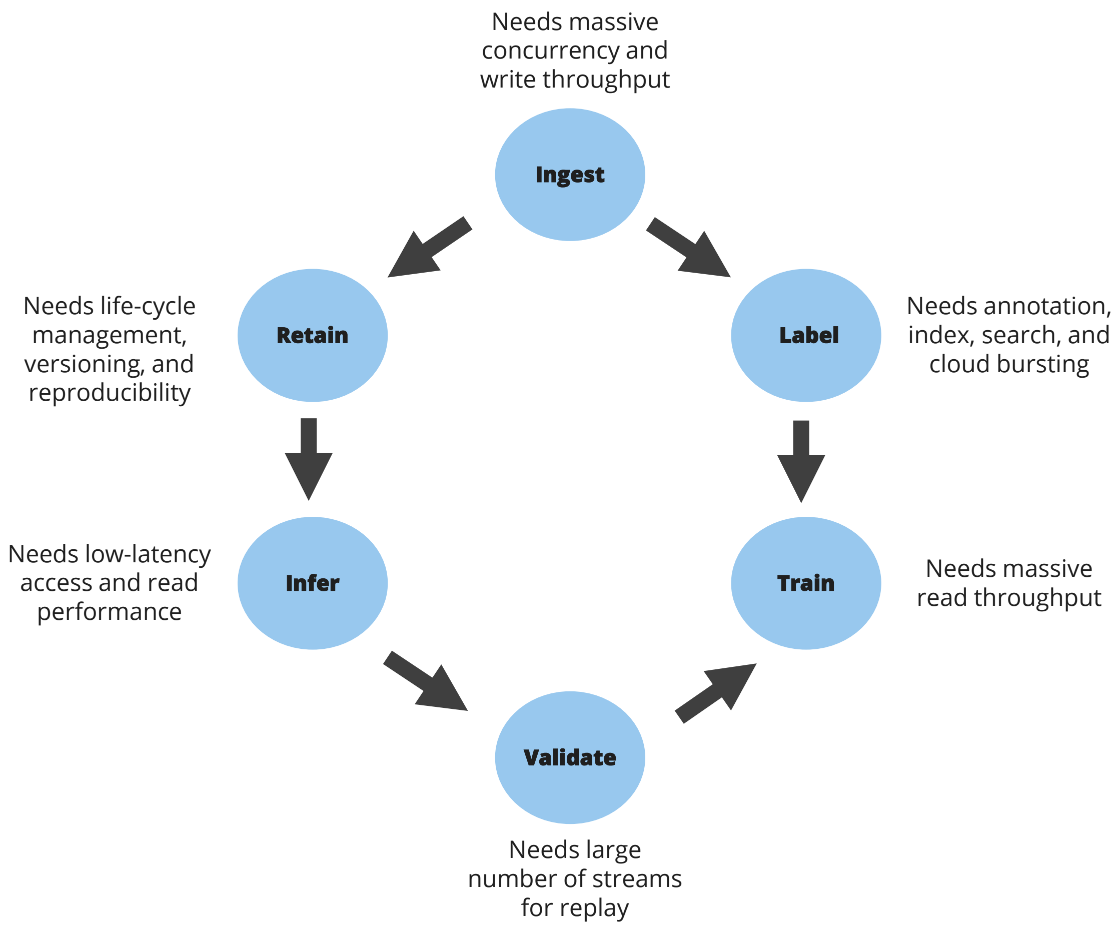
AI life-cycle workloads are now growing faster than any other storage workload except for industry-specific business applications, driving an 18.0% CAGR through 2025. The business intelligence/data analytics and AI life-cycle workloads markets combined are expected to reach $15.0 billion by 2025, and the storage capacity required to support them will be 1.795EB. Although many enterprises are managing multiple petabytes of data used for analytics today, by 2025, nearly all Global 2000 companies will face this same challenge.

A growing need to operate data analytics workloads at exabyte scale has exposed issues with more traditional approaches to enterprise storage. The largely hardware-defined history of enterprise storage has necessitated a siloed approach: most IT organizations today employ multiple storage systems, and 87% of IT organizations deal with three or more different enterprise storage vendors (source: IDC's *Portfolio Survey,* September 2020). A primary reason for having multiple storage silos was the requirements for different workloads in terms of performance, availability, scalability, storage management functionality, and cost could not be easily met by a single system. As a result, IT managers have been forced to buy and configure a variety of disparate systems — an expensive proposition that has also complicated administration.

The deployment of AI-driven workloads in the enterprise exacerbates this problem. Figure 1 shows a typical multistage data pipeline for enterprise AI workloads.

FIGURE 1

Typical AI Workload Data Pipeline



Source: IDC, 2022

For any given AI workload, the data sets across these stages will likely share a lot of commonalities; ingested data will be placed into data stores that are used in each of the subsequent stages (label, train, validate, etc.). Each of these stages have different I/O profiles, and often organizations must move the data sets to different storage systems to get the price/performance they need, even though the data set remains largely the same throughout. In addition to having to purchase and maintain separate systems, moving large data sets between systems takes time and can require manual intervention on the part of administrators. The use of multiple storage silos for AI workloads causes delays to data processing, which impacts time to value and adds to both cost and complexity for enterprises deploying these multistage applications.

Many enterprises already owned some type of enterprise file storage platforms even before they first began working with AI-driven workloads. For these organizations, it is quite common that they will first run pilots on those systems. When operating AI workloads at small (i.e., pilot) scale, existing scale-up and scale-out NAS using NFS/SMB access methods can perform quite well; once the workloads deploy in production, however, the relevant data set sizes start to grow, and these more traditional storage architectures can run into scalability limitations that can be exceptionally difficult to manage. At scale, the varying I/O profiles and requirements of the different AI pipeline stages are often what prompts an organization to deploy additional storage silos to support them — this is where costs may start to grow and spiral out of control.

Newer storage architectures being used for AI workloads are invariably more software defined than legacy architectures have been, giving them a flexibility of configuration not possible with older, more hardware-defined approaches. The advent of newer technologies like NVMe and faster and denser storage media types works synergistically with the agility of software defined to build single storage systems that can simultaneously accommodate a much wider variety of I/O profiles than in the past. When combined with the right enterprise-class storage management features, different workloads can be much more densely consolidated. This results in much higher infrastructure efficiency and can lower costs significantly because fewer storage platforms are needed. Given that many AI-driven workloads work on the same data set across a variety of different pipeline stages, there are compelling ease-of-use and economic benefits that accrue if a single storage system can be used to service all pipeline stages.

The Varying Storage Requirements for AI-Driven Workloads

There are five areas that must be considered by technical staff to ensure that storage selected for use with AI-driven workloads meets defined business requirements: performance, availability, flexibility, ease of use, and cost.

Performance

Before starting to look at different storage options, it is necessary to understand the performance requirements of the relevant set of AI workloads, including those of all the separate data stages in each workload pipeline. Review latency, read, write (ingest), and concurrency requirements as well as the random versus sequential nature of each workload, investigating where there may be any potential conflicts between workloads that may be running at the same time (i.e., the "noisy neighbor" problem) rather than serially. Take note of the different access methods required by the applications that will be accessing the AI data sets. Meeting capacity requirements is part of the performance consideration since systems that may perform well with 300TB of data may not perform well with 10PB of data — ensure that relevant "performance at scale" requirements are also well understood.

Broad access method support can be important. The most popular access methods in enterprises for big data environments have been NFS and SMB, along with HDFS (for older analytics workloads built around hard disk drive [HDD]-based architectures). Many enterprises are migrating legacy Hadoop environments away from HDFS to S3-based access methods. Workloads in technical computing that require very high serial throughput to large files commonly use intelligent (parallel) clients specifically designed to overcome the performance limitations of NFS. It is interesting that stages of the AI data pipeline also require very high throughput to large files, requiring more enterprises deploying these types of workloads to consider what have traditionally been referred to as "parallel file systems." There are, however, scale-out file system platforms that support NFS, SMB, S3, and intelligent clients at the same time, allowing workloads that use different access methods to access the same underlying big data store residing on the same storage system.

For AI workloads, which are more demanding in terms of performance, support for NVIDIA's GPUDirect Storage can be helpful. The NVIDIA API enables highly efficient RDMA-based communication directly between external storage and NVIDIA GPUs, offloading general-purpose CPUs from managing I/O and driving much higher efficiencies and lower latencies on big data analytics workloads. As GPUDirect Storage uses RDMA, it will be easier to leverage in environments that have at least some degree of POSIX compliance (in particular, where memory mapping occurs).

Storage systems that can deliver across a variety of performance metrics allow more stages in the AI data pipeline to be consolidated onto a single storage system. When data can be shared in this manner without limiting performance, availability, or storage management capabilities, it presents a compelling economic argument. Fewer systems need to be purchased and less administrative time and effort is spent migrating data between different storage systems.

Availability and Resiliency

Enterprises that have more traditionally deployed parallel file systems tend to have very different requirements for failure impacts: fast recovery, nondisruptive upgrades, data resiliency, and overall availability than the technical computing environments. In enterprises that are digitally transforming, AI-driven analytics can regularly inform real-time operational decisions, requiring that the systems performing the analyses cannot be down. Vendors taking their parallel scale-out file system designs into the enterprise need to enhance their platforms to meet these requirements. Customers should look for configurable on-disk data protection options (e.g., RAID and/or erasure coding) that let administrators set protection levels as needed, scalable snapshots (both immutable and read/write), replication to support disaster recovery (DR), designs that support nondisruptive expansion and software upgrades, and methods to minimize failure impacts (with host multipathing, redundant power and cooling, and rack awareness within and across datacenters).

Machine learning workloads tend to deliver much better results when working from larger data sets. Many enterprises will soon be growing to a petabyte and beyond and, therefore, should consider a recovery strategy that can meet both recovery point objective (RPO) *and* recovery time objective (RTO) requirements. In the event of a site disaster, restoring a petabyte-scale system from backups is not likely to meet RTO requirements for critical workloads. Erasure coding that can be spread across sites as well as replication features can help enterprises design the right recovery strategies for large-scale data environments.

When purchasing storage infrastructure for critical AI-driven workloads, technology decision makers should ensure they understand the impacts of various failure scenarios on performance and availability to avoid surprises. It is important to evaluate for these scenarios not only in the context of small-scale pilot projects but also when the environment is operating at scale. This is a common mistake many enterprises have made as they deploy their first AI workload — just because a storage system performs well in a pilot does not guarantee its continued ability to meet storage requirements at scale.

Flexibility

For customers looking to consolidate multiple AI applications and data pipeline stage processing onto the same system, the flexibility to simultaneously meet the needs of various I/O profiles and workload requirements is critical. Support for different access methods is important since applications will be using a variety of interfaces on a common underlying data set. NFS, SMB, S3, and an intelligent parallel client tend to be the access methods of most interest in enterprise AI workloads. For file access, NFS can be easier to deploy and use if it can meet requirements, but there are stages of the AI data pipeline that require the parallelism of intelligent clients such as ingest, training, or real-time inferencing as data sets get larger. In environments where Windows-based devices are used for data collection, SMB is generally required.

Mixed node and media support are also important. Within a single namespace, enterprises often want to tier data between a higher-performance tier used for time-sensitive analytics and an archive tier used for cost-effective, long-term data retention. Vendors that offer all-NVMe, hybrid, and HDD-only nodes give customers options to configure a system to best meet their needs, and the ability to tier to external platforms (such as an object storage platform or cloud storage target) can also drive costs out of required big data analytics support infrastructure. When tiering is involved, check whether tiers are included as part of a unified namespace or must be searched and managed separately. Within a single node, support for multiple media types (such as Intel Optane or NAND flash–based SSDs) as well as devices from different suppliers at different capacities can offer additional flexibility and makes it easier to accommodate newer device technologies as they become available.

The ability to granularly apply data services (erasure coding, replicas, compression, deduplication, snapshots, encryption, quality of service, replication, etc.) to different data sets on the same storage platform is important for efficiency reasons. When data services are applied on an all-or-nothing basis at the system level, more CPU resources may be consumed than are necessary, a factor that can impact application performance and infrastructure efficiency. While this may be less of a concern for smaller systems, enterprises dealing with multi-petabyte data sets need this kind of granularity.

Ease of Purchase, Deployment, Management, and Scaling

Many enterprises will expect ease-of-use features that were always a requirement for technical computing customers, which were among the earliest adopters of parallel file systems. Vendors targeting enterprise customers are responding and now support many of these capabilities in their parallel file systems. For ease of purchasing, look for "converged infrastructure stacks" that include a set of pre-validated (to work together) compute (or accelerated compute), storage, and networking technologies. Converged infrastructure stacks go beyond just reference architectures (which are also handy) in that they tend to offer single SKU purchasing and a single point of support contact. And because these systems are preconfigured in the factory, they typically install much faster and more reliably than systems where the enterprise buyer must qualify, purchase, assemble, and deploy components separately.

When evaluating a storage system, buyers should consider associated workflow requirements and investigate how to perform those operations. How easy is it to deploy and maintain intelligent clients? Does it offer the multitenant management capabilities needed to meet performance, availability, security, and cost objectives? Will technology refreshes be easy, flexible, and nondisruptive? What is the process to create a comprehensive snapshot of the entire namespace and replicate it to a remote site for disaster recovery purposes?

For ease of management, look for systems that have simplified common workflows (e.g., storage provisioning, copy creation, versioning, backup and restore, cluster expansion) and use defaults during these workflows that do not negatively impact performance, availability, security, or other aspects of system operation.

For high-growth environments, scalability considerations are important. When AI workloads are expected to grow rapidly (as most successful ones do), it is critical to ensure that the storage system will not be outgrown or present issues when operating at scale that limit its capabilities or increase costs. Ask vendors about their average and maximum cluster sizes and talk to customer references to better understand whether the system can scale to support workloads that are similar to your own.

Infrastructure Efficiency and Cost

Different approaches to AI storage can result in very different costs. For example, a system using HDFS may require 20 times the number of application CPUs and significantly more disk devices (Hadoop was originally developed for use with HDDs, not SSDs) to meet throughput and capacity requirements than one of the newer AI storage platforms that is more "POSIX compliant" (and therefore more compatible with the modern AI frameworks that can benefit from POSIX-compliant memory mapping capabilities). Systems that require more hardware to meet performance and/or capacity requirements not only cost more but also draw more power and have a larger datacenter footprint. With environmental, social, and governance (ESG) initiatives on the rise, many companies are concerned about energy and floorspace consumption. On this point, IDC's 2021 *Datacenter Operational Survey* indicated that one in four U.S. organizations are reporting delays in IT deployments due to power or space constraints.

Certain features impact infrastructure efficiency. AI workloads, which require high-performance density, can benefit from NVMe and NVMe over Fabrics technologies that provide more compact, efficient systems to meet any given performance requirements. Data reduction features like compression and deduplication, when they are applicable, can lower the amount of raw capacity required to store a given data set. Erasure coding approaches that enable broader distribution of data across more devices can require less redundant capacity to meet any given data durability requirement. Space-efficient snapshot technologies can enable broader use of snapshots for data protection, test and development, and other tasks with less raw capacity utilization.

While HDDs provide the lowest-cost raw capacity, they consume more wattage per terabyte than solid state media options (which tend to be more expensive on a raw price-per-gigabyte basis to buy). For workloads with any performance sensitivity, solid state devices can meet performance requirements with generally far fewer devices, while archival workloads whose data is not often accessed after it is moved to an archive may be better served by HDDs (due to their lower cost of raw capacity).

Systems that are more performance and/or capacity dense may result in lower purchase and operating costs initially, but they may also cause a bigger blast radius (scope of impact) in the event of device or node failures. Most enterprises will want to weigh cost considerations against failure impacts and recovery times — rebuild times are a classic example of this. Use of larger devices may mean fewer devices need to be purchased but can also result in longer rebuild times in the event of a disk failure. What steps (if any) has a vendor taken to minimize device rebuild times, and how does a very performance- and/or capacity-dense system behave while operating in a failure mode?

THE WEKA DATA PLATFORM FOR AI WORKLOADS

WEKA is a data platform software vendor that has been shipping parallel file system technology since 2017. The company's data platform is built on a high-performance, scale-out file system optimized for NVMe-based nodes and supports a variety of different access methods including a POSIX-parallel client, NFS, SMB, S3, and NVIDIA's GPUDirect Storage all to the same data. After turning heads with high water mark performances on the io500 over several years, WEKA noted the applicability of its technology to AI-driven workloads that were being deployed by enterprises. As more enterprises have embraced digital transformation, they began to deploy more large-scale data analytics workloads, many of which are using AI technology. Serving the needs of enterprises with its data platform for AI, WEKA has entered a period of very rapid growth.

In 2021, the company tripled revenue and won more deals in its FY 4Q21 alone than it had in all of 2020 combined. It grew customer acquisitions by more than 125% and surpassed an exabyte of data under management in its customer installed base. The vendor is experiencing significant channel traction with a major new OEM deal with Hitachi Vantara, new joint solution bundles with HPE and Supermicro, and reference architectures with object storage partners such as Amazon Web Services, Hitachi, IBM, Dell, NetApp, Quantum, Cloudian, Seagate, and Scality. WEKA also now counts 8 of the Fortune 50 among its customers, along with many others. For a more detailed look at WEKA's recent performance and evolution, see *Storage Software Vendor WEKA Announces Oversubscribed Series C2 Round and Positions Itself for Rapid Growth* (IDC #US48944322, March 2022).

While WEKA continues to sell WekaFS in technical computing markets, the company is now focused on AI-driven workloads in enterprises. WekaFS forms the core of the WEKA Data Platform for AI, which includes several capabilities designed to meet enterprise requirements for ease of deployment and management, availability, scalability, flexibility, security, and infrastructure efficiency for high-performance AI-driven workloads:

- **Deployment and configuration flexibility.** As a software-defined data platform, WEKA supports deployment across a variety of different x86-based storage servers with Intel and AMD CPUs, including Cisco, Dell, Hitachi Vantara, HPE, Lenovo, and Supermicro as well as bare metal, containerized, virtual, and cloud-based options. It can be deployed as a hyperconverged appliance (with compute and storage services running on the same server-based storage node), a dedicated storage appliance, or natively in the cloud. With hybrid cloud deployments, WEKA instances support the same exact capabilities regardless of deployment location and are managed through a single pane of glass. Available access methods include WEKA's own POSIX-intelligent client, NFS, SMB, S3, and NVIDIA's GPUDirect Storage. A CSI plug-in for Kubernetes is also available to provide persistent storage to container-based high-performance workloads. WEKA has published reference architectures with a number of storage OEM and public cloud providers (as mentioned previously) and has NVIDIA POD certification as well as installations with multiple SuperPOD-sized deployments.

  WEKA supports multigenerational technology refresh at both the storage device and the node levels. Matched capacity or generation SSDs and nodes are not required, allowing different generations and capacity points to coexist in a single cluster.
- **Ease of management.** WEKA has its own integrated volume manager that deploys as part of the file system and supports thin provisioning, handles free space management, supports easy capacity expansion, and enables seamless integration with EC2 auto-scaling groups in AWS. Integrated tiered data management places data across storage types based on policies, which use a "heat index" that is tracked and maintained for all data. Metadata is always stored on the flash tier storage nodes, and select file systems can be locked exclusively onto NVMe SSDs when desired.

  WEKA proactively monitors and manages the performance, resiliency, and capacity health status. As necessary, data can be automatically rebalanced in real time to avoid hot spots. Rebalancing can also occur as new storage devices or nodes are added to the cluster, all of which requires no downtime for data migration. Quality-of-service controls allow administrators to establish both a preferred and a maximum throughput for per-application performance management.

A special "snap to object" feature allows an entire namespace to be committed to an object store with a single click or API call. The snap to object includes both data and metadata, which allows it to be recovered to any WEKA cluster (not just the one from which the snapshot originally came) and also to clusters that are not a mirror copy (in terms of hardware type and node count) of the source cluster. The WEKA cluster can send multiple copies of the data to remote clouds and/or datacenters at exabyte scale. These capabilities make this feature ideal for cloud bursting use cases as well as backup, archive, async mirroring, DR, and ransomware protection.

Security features include 256-bit encryption for data both at rest and in flight, authentication and access control, and support for LDAP and Active Directory (AD). Role-based access control establishes five privilege levels. WEKA also supports capacity quotas, which can be set as either advisory, soft, or hard quotas, and can allow up to 64 different "organizations" to manage their own file systems and capacity in this way within the WEKA Data Platform's single namespace.

The WEKA Data Platform provides three different ways to manage a file system instance: a graphical user interface (GUI), a command line interface (CLI), or a Representational State Transfer (REST) API. Complete reporting, visualization, and overall systems management functions are accessible using any of the approaches.

- **Availability and resiliency.** For on-disk data protection, the WEKA Data Platform allows administrators to set configurable data protection levels from 4 + 2 to 16 + 4. This erasure coding is supplemented by configurable failure domains, inline end-to-end checksums for data integrity, metadata journaling, and space-efficient local snapshots (for fast file-level recoveries) and space-efficient clones (read/write snapshots). Data is protected at the file level, so WekaFS only needs to rebuild active files and data on a failed component, speeding recovery times. Optional virtual hot spares, which are spread evenly across all failure domains, provide reserved capacity so a system can undergo a complete data rebuild and still maintain the same net capacity. Using snapshot to object capabilities allows the creation of asynchronous-like replication, enabling data to be efficiently mirrored to remote locations for DR purposes, as mentioned previously.
- **High performance and scalability.** WEKA's data platform for AI uses a two-tier design with a high-performance, NVMe-based, front-end scale-out file system and a back-end data lake that supports HDDs, as well as cloud-based storage targets — all under a single namespace. NVMe technology in the front end gives the file system an ability to handle high degrees of data concurrency with low latency and high IOPS, while the back-end data lake provides massive low-cost capacity scalability. Intelligent data placement algorithms ensure data is where it needs to be to meet established objectives for low latency, high throughput and bandwidth, or low cost. Data and metadata are widely distributed throughout the cluster storage nodes to avoid hot spots.

  WEKA provides massive scalability with support for up to 4PB in a single file, 6.4 trillion files or directories, 1,024 file systems, and up to 14EB of managed capacity in a single global namespace that can span both on- and off-premises (i.e., public cloud) deployment locations. Given its scalability and ability to provide multiprotocol access to the same data sets, WEKA can be used to consolidate multiple data lakes into a single massively scalable data warehousing instance under a single namespace for easy search and simplified management. Migrating from other platforms can be challenging and put performance at risk, but WEKA provides the option to mount directly to the back-end infrastructure that passes ingested data directly to the object store without consuming file system (i.e., flash) tier capacity. This provides a "lazy migration" option, which maintains normal performance and operations for workloads that need the full performance of the file system–based front end.

With its flexible NVMe-based front-end and back-end data lake architecture, the WEKA Data Platform can cost-effectively meet the I/O profile requirements of every stage in the AI data pipeline, while massive scalability supports the very large data sets and rapid data growth that characterize so many successful enterprise AI deployments. Its rich multitenant management capabilities give administrators the tools they need to safely and securely host workloads used by different constituencies within an organization while meeting a variety of different performance, availability, security, and cost objectives. What this effectively means for enterprises is that they can consolidate most of or all their AI-driven workloads onto a single, highly scalable data platform, dispensing with the time and effort involved with moving large data sets around as part of AI-driven big data analytics workflows. This not only simplifies operations but also offers compelling economic benefits due to high infrastructure efficiencies.

CHALLENGES/OPPORTUNITIES

The fact that small pilot AI workloads can run effectively on many different types of file storage infrastructure presents a challenge for vendors that solve "performance at scale" problems that prospective customers have not yet experienced. Although IDC has seen many AI projects run into problems that are directly related to storage infrastructure limitations as they begin to scale, the awareness of this problem is low among enterprises that are new to working with AI-driven big data analytics workloads. Vendors such as WEKA need to raise awareness of this issue, showing how making the right data platform purchase up front can help avoid potentially disruptive issues down the road that can significantly delay time to value and drive up associated costs. Vendors with dense workload consolidation capabilities will need to address potential concerns with IT managers who are only familiar with the more traditional limitations of legacy storage architectures in this area.

CONCLUSION

With the transition to data-centric business models prompted by digital transformation, enterprises are deploying big data analytics workloads as never before. As part of this evolution, enterprises are rapidly deploying AI-based applications, making AI workloads one of the fastest-growing segments of the storage workload market tracked by IDC over the next five years. This rapid buildout of infrastructure to support AI-driven workloads is changing storage requirements and prompting significant interest in data platforms with proven multi-petabyte scalability.

To meet the needs of the next-generation workloads being deployed as part of digital transformation, enterprises are refreshing their server, storage, and/or data protection infrastructure at a rapid rate. Moving to more efficient infrastructure is part of this evolution, and enterprises are actively looking for solutions that can enable denser workload consolidation without putting performance, availability, and other business requirements at risk. Traditionally, big data analytics workloads have generally been spread across multiple storage systems because of an inability to meet a wide range of I/O profiles with a single platform, but new scale-out file system architectures and storage technologies are changing that. WEKA offers a massively scalable data platform with the performance capabilities that AI workloads demand while also meeting enterprise requirements for high availability, fast data recovery, ease of management, and security — at a cost-competitive price point. WEKA customers can consolidate their AI-driven workloads onto a single storage cluster and dispense with data movement between storage platforms. As a result, they are driving faster time to value with these applications, even as they enjoy compelling economic benefits from much more efficient storage infrastructures.

About IDC

International Data Corporation (IDC) is the premier global provider of market intelligence, advisory services, and events for the information technology, telecommunications and consumer technology markets. IDC helps IT professionals, business executives, and the investment community make fact-based decisions on technology purchases and business strategy. More than 1,100 IDC analysts provide global, regional, and local expertise on technology and industry opportunities and trends in over 110 countries worldwide. For 50 years, IDC has provided strategic insights to help our clients achieve their key business objectives. IDC is a subsidiary of IDG, the world's leading technology media, research, and events company.

Global Headquarters

140 Kendrick Street
Building B
Needham, MA 02494
USA
508.872.8200
Twitter: @IDC
blogs.idc.com
www.idc.com