FUTURIOM
FUTURE OF CLOUD TECH

Cloud Market Trend Report

Data Management in the Age of AI

March 2026

Sponsored by:

Attimis

Highlights and Key Findings

- **Good data is essential to enterprise AI.** Reaping the rewards of AI requires data that is clean, clear, and ready to be transformed into effective machine-readable format. But many enterprises are struggling to meet these requirements.
- **HPC storage systems formed the basis for today's high-end AI data management systems.** DDN, VAST Data, and WEKA have extended their products to fit accelerated computing and AI workflows, realizing significant performance improvements.
- **Linkage to AI pipelines is key.** The capability to connect data stores directly to AI clusters of GPUs, XPUs, and CPUs is on the rise.
- **Data management platforms are rebranding as AI platforms.** Databricks and Snowflake have added vector search, integral models, and other features to support AI development.
- **Enterprise storage products are evolving into data management platforms.** Whether on-premises or in the cloud, or both, traditional storage vendors have adopted a range of technologies to prepare data for AI workflows.
- **No single solution is comprehensive.** While many products and services perform a range of data management functions, enterprises will need combinations plus their own efforts to address AI workflows.
- **Some companies highlighted in this report:** Attimis, Backblaze, Cloudian, Confluent, CTERA, DataCore Software, Databricks, Datadobi, dbt Labs, DDN, Dell, Everpure (formerly Pure Storage), F5, Fivetran, Hammerspace, Hitachi Vantara, HPE, IBM, Informatica, Komprise, Matia, Matillion, MinIO, Nasuni, NetApp, NVIDIA, Panzura, PEAK:AIO, Qumulo, Redpanda, Scality, Snowflake, UnifyApps, VAST Data, VDURA, Vultr, Wasabi, WEKA

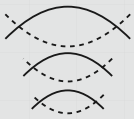
Don't Move Your Data. Use It.

Unify Silos and Defy Gravity with Attimis OneBucket™.

OneBucket™

AI success requires data mobility, but enterprises are stuck with data scattered across discrete, isolated silos, crippled by transfer and egress costs and long hardware backlogs.

Attimis OneBucket™ is a modern, real-time orchestration plane that unifies fragmented infrastructure into a global adaptive data fabric. It decouples compute from physical location, so the right data shows up in the right place at the right cost. Do not wait on logistics. Put your data in motion.



UNIFY DATA SILOS

OneBucket provides a single endpoint across on-prem, edge, and cloud. Applications get local access to data anywhere

BEAT CAPACITY CRUNCH

OneBucket instantly unlocks global available storage capacity with zero application or workflow changes.

ENABLE DATA MOBILITY

OneBucket eliminates risk, costs, and time of data movement, supporting hyperscalers, on-prem, and on-demand capacity providers.

OPTIMIZE COSTS

Move only the data needed. OneBucket optimizes data placement, reduces latency, and maximizes GPU utilization.

IMPACT SPOTLIGHT

Online Travel Company

AI Workload Performance	No Application Modifications Required	Storage & Data Transfer Costs
10× Faster	**0 App Changes**	**95% Reduction**

AI training and Inference, 100PB of data across AWS and on-premises Dell PowerScale — along with crippling egress fees — this global travel leader deployed OneBucket to eliminate "all-or-nothing" data transfers. OneBucket intelligently placed the 5% of data for each AI workload to the right place, at the right time and cost optimized.

Results: accelerated insights, 10× performance gains, and reduced storage and transfer costs by 95% — with no application or workflow changes.

Attimis

Request your OneBucket today

https://console.attimis.cloud/sign-up

attimis.co

TABLE OF CONTENTS

Introduction: AI Transforms Data Management

Enterprise AI feeds on corporate data—clean, clear, well-managed data. Whether the task is training or inference, the need is the same. Yet many enterprises face a challenge in supplying good data to AI pipelines.

Often, data is stored in silos throughout an enterprise, and much of it exists in a mix of formats, both structured and unstructured. Data is dispersed geographically, often around the globe. Some is restricted by sovereign AI or other regulations. Formats differ: Tabular data from accounting can't mix with documents, images, and social media messages from marketing. Even data within the same department might feature duplicated, outdated, or erroneous information that must be weeded out before any dataset is ready for processing.

The demands of AI have brought urgency to fixing these problems, particularly in the case of AI inference, which requires expanded data context to shape an effective AI application. Further, unless data can meet the requirements of AI processes within the new infrastructure of accelerated computing, costly delays can quickly erase any ROI from AI adoption.

The situation has been evolving as HPC storage, as well as enterprise storage systems, feature new data management tools and direct interaction with accelerated networks. Software-defined storage, though not new, has taken center stage. Data management software has been reworked to prepare data from data lakes, lakehouses, and lakebases for AI processing. Object management systems and global file systems have expanded to accommodate the speed and throughput needs of AI.

All this activity has resulted in more choices for enterprise customers, who are quickly learning that there is no single solution, but a series of interconnecting parts that require careful planning and expert engineering to realize ROI. In this report, we attempt to parse these issues, framing the market from a practical viewpoint.

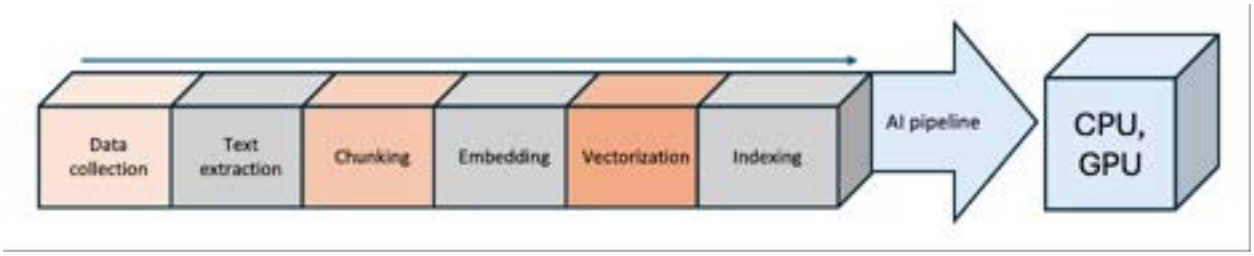
The Data Management Landscape for AI

The primary task of data management for AI is assembling data in a location and format that readies it for AI. This sounds simple, but it's a difficult challenge. AI workflows demand that data be clear, free of errors, easily convertible to vectors or labeling for model deployment, and capable of rapid throughput to accelerated computing clusters for processing. Unless these criteria are met, AI workflows can fail to process effectively, not only creating errors in AI applications but snarling expensive GPU-based systems, resulting in costly lags and restarts.

The good news is that there is a growing arsenal of products and services geared to data management for AI. But that's the bad news, too. Despite vendor claims for "platformization" and comprehensive solutions, most enterprise data management projects require a series of products, often from different vendors, along with substantial expertise (either within the organization or outside of it, or both) to get it all working together. There is no magic bullet.

Consider, for instance, the situation faced by industrial machinery market leader Caterpillar Inc., which leveraged in-house expertise and vendor technologies to develop Helios, a data management platform that is central to its AI strategy. Over a period of several years, the company refined Helios NVIDIA for AI, Microsoft Azure for cloud functionality, and Snowflake for data management. The system reportedly handles 16 petabytes of data from 1.5 million connected assets, ingesting, processing, and directing it through multiple data pipelines to support new AI applications, including a fleet of AI agents powering the Cat AI Assistant, which converses with operators to streamline efficiency and supports predictive maintenance.

Given that there is no single solution, it's helpful to review the data "chain of command" in AI workflows to get a feel for what so-called data transformation, or managing data for use in AI, is about. Theoretically, the process involves the following steps: data collection, text extraction, chunking, embedding, vectorization, and indexing. Note that these steps are generally applied to data for both training and inferencing workflows, though tasks differ within the actual AI pipeline (training creates models that recognize patterns and links within datasets, while inference generates information based on running specific data against a model).



Let's look at each of these steps in the AI process.

Data collection. This is the step that affects all the others after it. Data gathered from multiple locations and in multiple formats must be cleaned, deduplicated, and prepared for use in training and/or inferencing workflows. Metadata may also be gathered as part of the project.

Text extraction. This step may be part of data collection. In it, unstructured data is converted into machine-readable format. For instance, optical character recognition (OCR) may be applied to a PDF to glean text that is recognizable by a machine learning system.

Chunking. This is the process in which formatted text is cut up into chunks small enough to fit into a model's context space without sacrificing meaning. NOTE: Tokenization, or the breaking up of text into small units for model processing, may happen at this stage when the goal is model training.

Embedding. Vector embeddings transform chunks of data into numerical representations, ready for organization by vector databases. This is a step key to Retrieval Augmented Generation (RAG).

Vectorization. This is the process whereby embeddings are organized in vector databases for further processing. Within a vector database, vectors are grouped according to similarity, making data easier to retrieve for indexing.

Indexing. This step organizes vectorized data into a format that makes data items easy for models to retrieve, particularly for inference.

It's important to note that no single product supports all these steps. A storage system may gather and organize data for use in a data management system that may expose it to a third-party model for vectorization. There also may be third-party databases and foundation models evoked by the storage system itself for vectorization and indexing. Bottom line? There is overlap among storage systems and data management platforms, and there is no easy answer, as we will see in the following sections.

Data Management Taxonomy

Just as data management for AI can be broken down into a series of steps, so can the market for solutions. This isn't an easy task, since, as noted, there is a growing diversity—and overlap—among a variety of vendors at multiple levels. But it's helpful to view the data management for AI market according to basic functionalities. Following is our attempt at generalization:

HPC platforms. A handful of vendors who began life as HPC storage suppliers have evolved their products and services into high-end data management solutions through the addition of universal block, file, and object storage along with high-speed storage media, parallel file systems, data streaming, caching, and analytics. Vendors in this category include **DDN, VAST Data, VDURA,** and **WEKA.**

Enterprise storage systems. Companies in the large enterprise network-attached storage (NAS) market have eagerly embraced the need for data management features that favor AI workloads. These vendors have added data transformation functions such as ingestion, chunking, embedding, and vectorization to their on-premises storage systems, either as integral elements or through connectivity to a third-party vector database. Some also have expanded

their cloud support to become hybrid storage solutions. **Cloudian, Dell, Everpure** (formerly Pure Storage), **Hitachi Vantara, HPE, IBM, MinIO,** and **NetApp,** among others, fit into this category.

Software-defined storage. A growing group of software providers, such as **DataCore Software** and **PEAK:AIO,** to name just two, offer comprehensive storage management that works with third-party hardware, including storage arrays, servers, and NAS systems.

Object storage systems. Vendors specializing in object storage management have expanded their wares with a range of functions, including interaction with popular third-party AI models and direct linkage to NVIDIA GPUs. This category includes on-premises/hybrid vendors such as **Cloudian, MinIO, Qumulo,** and **VDURA,** along with object storage services from **AWS, Azure, Backblaze, Google Cloud, IBM Cloud, Vultr**, and **Wasabi.**

Data management platforms. This category is broad and incorporates systems from the likes of **Attimis, CTERA, DataCore Software, Datadobi, Fivetran, Hammerspace, Komprise, Matillion, Nasuni,** and **Panzura,** among others. Also included are data streaming platforms, such as **Confluent** (soon to be part of IBM), **LakeSail,** and **Redpanda. VAST Data** also streams data.

This is a field rife with innovation. **Attimis,** for example, provides a global namespace across heterogeneous storage platforms, unifying unstructured data across multiple silos for global AI and analytics workloads. Its product, OneBucket, doesn't deploy a database table to provide a single endpoint for data but uses a stateless approach that eliminates bottlenecks, improves performance, and enables dramatic scalability.

Overlapping the data management platforms category but with a different approach is **F5,** whose Application Delivery and Security Platform (ADSP) contains F5 BIG-IP software working on F5's rSeries appliances and F5 VELOS modular chassis systems (BIG-IP can also be deployed as managed software). These devices sit in front of any S3-compatible object storage system as the programmable control point for AI data ingestion. The F5 BIG-IP suite offers Local Traffic Manager (LTM), DNS, Advanced WAF, Application Firewall Manager (AFM), Access Policy Manager (APM), and Hybrid DDoS Defender to support the data traffic into and out of data and object storage, ensuring high-performance, resilient AI data delivery.

F5 is storage-agnostic but maintains strong alliances with storage providers such as **Dell, MinIO, NetApp,** and **Scality,** among others. F5 also integrates its F5 BIG-IP Next for Kubernetes software (which also is part of its F5 ADSP) with NVIDIA RTX PRO 6000 Blackwell Server Edition for enterprise and industrial AI. F5's software works with NVIDIA's BlueField-4 DPUs to allow F5 policies, security, and traffic management to run directly on the DPU at line rate. By delivering clean, high-throughput data into the storage systems that feed the GPU clusters, F5 ensures that GPUs stay optimally utilized.

No data management for AI summary is complete without reference to **Databricks, Informatica** (now part of Salesforce), and **Snowflake.** These platforms differ in functionality while contributing to AI workflows. **Databricks** offers data engineering, including interaction with third-party models for development of a variety of AI applications. **Informatica** offers data integration across multiple silos, along with master data management, including for data from Databricks and Snowflake. **Snowflake** also provides access to models for AI development and analytics within its secure, SQL-oriented environment.

As described above, these categories of data management tools often overlap and are not strictly circumscribed. In the next sections, we take a closer look at the challenges enterprises face and how these categories fit these needs.

Challenges of Data Management for AI

Enterprises report a range of issues when it comes to bringing data into AI workflows:

Data quantity. Business intelligence, analytics, and an array of enterprise applications generate enormous quantities and types of data, resulting in areas of heavy data gravity, such as customer relationship management (CRM) or sales data.

Siloed data. Most companies are geographically dispersed, which can lead to isolated application fiefdoms with corresponding isolated data stores. Even in one location, data is spread across cloud environments, on-premises systems, and a variety of applications.

Poor data quality. Data that is outdated, duplicated, inconsistently presented, or broken can't meet AI's stringent requirements. Categorization of data is also essential to ensuring its success in AI.

Lack of good metadata. Complete and accurate information about where data is stored, its format, appropriate dates, chain of custody, and other parameters is essential to good data management for AI.

Multiple data formats. Organizations face inconsistency in how data is formatted: Some is tabular or structured, but the majority is unstructured, including documents, videos, images, and other items that don't fit a standard database schema.

Data governance issues. Regulations abound about data, including privacy rules such as HIPAA in U.S. healthcare, the EU's General Data Protection Regulation (GDPR), and the California Consumer Privacy Act (CCPA), to name just a few.

Security problems. Data safety, including access control, ransomware protections, malware

attack strategies, and insider data breach avoidance must overlay any data management effort.

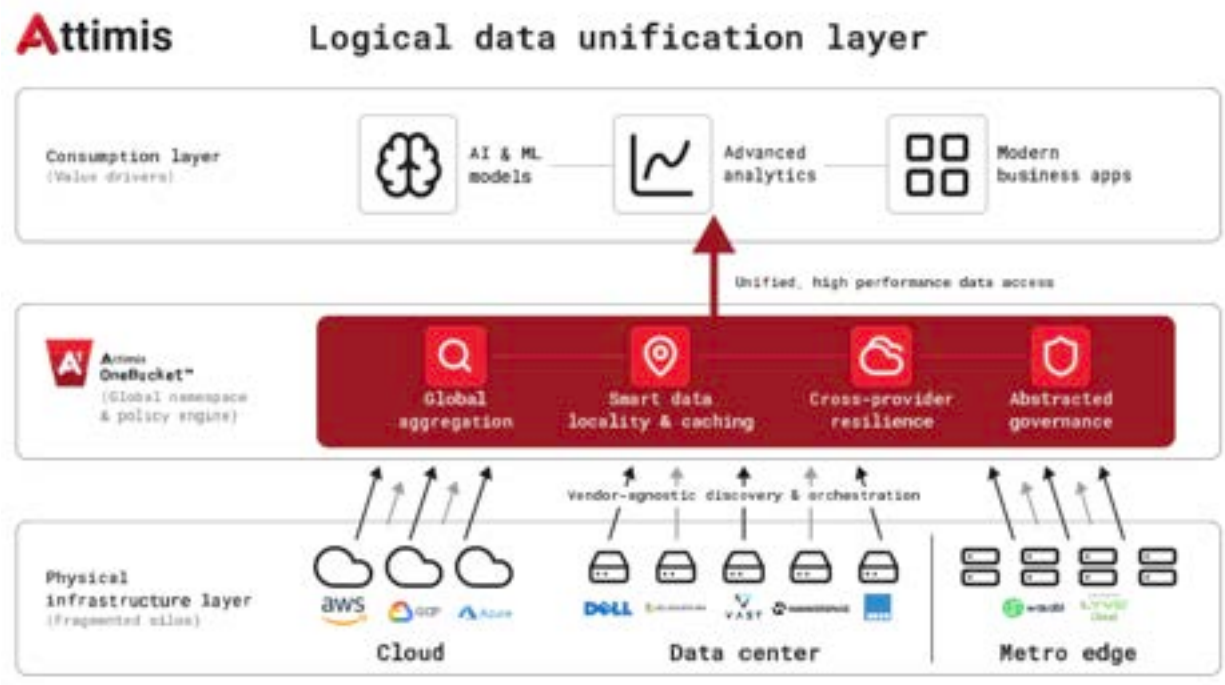
Linkage to AI pipelines. Getting data into the AI workflow can be daunting. Once data is clean, it needs to be delivered quickly to AI models.

Notably, these are generalizations and don't cover all the challenges facing enterprises looking to manage data for AI. For instance, there are problems of staffing, of communicating the need for better data and finding the talent to implement changes. Further, no single product can solve for all these issues, though some can tackle more than one. In the next section, we look at how enterprise storage solutions match up to these challenges.

Data Management Platforms for AI

This category of products includes a range of suppliers that offer different approaches to managing data in AI infrastructure, regardless of where the data is stored. Included are vendors such as the following:

Attimis provides OneBucket, a product that operates as a stateless S3 proxy and data orchestration layer. OneBucket does not own, convert, or relocate customer data into a proprietary format. Instead, it provides a unified, high-performance namespace across existing heterogeneous object and file storage systems, optimizing access, placement, and cost without requiring data migration or application rewrites. OneBucket optionally supports Dell ObjectScale, PowerScale, and OneFS platforms. Seagate is a global partner. TCS is a global integrator partner for AI and analytics.



Source: Attimis

CTERA's Data Intelligence platform continuously retrieves metadata from a range of object file stores, indexes it, and links it to third-party AI models from the likes of OpenAI, Anthropic, or Microsoft. As a bonus, CTERA has added specialized AI agents for information retrieval from those models.

Datadobi moves data between multiple storage systems and platforms (NAS, SAN, cloud, object storage), ensuring that data for AI workflows will be delivered safely and without corruption. The vendor provides StorageMAP for policy-driven lifecycle management, replication, data protection, data cleaning, "golden copies," risk protection and ransomware recovery, as well as a cost calculator for storage optimization.

F5's Application Delivery and Security Platform (ADSP) protects applications and APIs in a single platform that can function across a range of locations without requiring hardware adjustments, including on DPUs, in Kubernetes environments, as software on prem and as SaaS. It offers visibility and AI insights into all areas of apps and APIs to boost performance. Its load balancing, intelligent traffic routing, and autoscaling control based on GPU utilization help to keep AI workflows efficient and secure.

Hammerspace provides a global namespace for file and object data regardless of where it's stored—on premises or in any cloud. This streamlines access to data for use in AI workflows. The vendor's Hammerspace Data Platform leverages metadata and custom metadata to implement data protection and add policies to data, such as caching frequently accessed AI datasets.

Komprise offers a Global File Index to manage unstructured object and file data regardless of the type of underlying storage and its location. The vendor's Komprise Smart Data Workflows lets customers tag and enrich data for use in AI or for isolation in secure storage.

Nasuni prepares data for AI workflows by creating a single source of unlimited object-based file storage in a cloud environment. The vendor offers access controls, usage analytics, and intelligent edge caching for frequently accessed data. Nasuni also supports multi-site collaboration, global file synchronization, and global file locking to prevent multiple versions from confusing the data store.

Panzura's Symphony platform establishes AI pipelines for a range of third-party cloud and on-premises file and object storage systems, including Panzura's own CloudFS as well as ones from AWS, Microsoft Azure, Google Cloud Platform, NetApp, Backblaze, and Wasabi, among others.

In addition to the options above, data streaming platforms, such as **Confluent** (soon to be part of IBM) and **Redpanda** provide AI workflows with continuous, live data—a key element in applications such as financial services. Confluent is built on Apache Kafka, an open-source event-streaming platform, and Redpanda is compatible with it.

UnifyApps moves data to and from multiple SaaS and application sources, including SQL databases, cloud object storage services, and APIs, in real time. It cleans data to prepare it for use in AI pipelines. UnifyApps supports Change Data Capture (CDC), an approach that allows any changes to data to be streamed in an AI pipeline without requiring the entire dataset to be reconstructed.

As noted earlier, **Databricks** and **Snowflake** have adapted quickly to fulfill the demands of AI workflows. Below are some details:

Databricks supports structured and unstructured data in its lakehouse architecture, which is also equipped with Lakeflow Connect, a feature that provides access to AI models within Databricks. Since Databricks uses open-source Delta Lake technology with ACID guarantees, the AI models are served the freshest, cleanest data with low latency. Databricks also supports a Unity Catalog to govern all data, files, tables, models, and vectorized data in the Lakehouse, including embeddings and indexes, which are generated within the Lakehouse. Governing includes access control and auditing of all data. Tools to build AI agents are also governed by the Unity Catalog.

Snowflake supports structured, semi-structured, and preprocessed unstructured data in a table-oriented architecture. It provides a SQL analytics platform in its Cortex AI product, which runs LLMs, embedding models, and ML inference directly inside Snowflake, using SQL or Python, with Snowflake's governance and security applied automatically. Serverless Snowpipe Streaming ingests event-driven and row-oriented data at high throughput rates, providing a way to build real-time data applications.

Enterprise Storage Adapts to AI

One of the largest markets for data management for AI is emerging from the realm of traditional HPC and enterprise storage. These systems are quickly adapting to fit the challenges of AI. But getting to the heart of what each product or service can achieve is not easy because vendors have numerous claims about their wares. Only by digging deeper can we begin to see the finer points emerge. Let's take a closer look at the innovations storage systems are now featuring.

Universal storage support. Use of AI requires all types of data to be unified in storage, which has led to a widening of compatible data types in leading platforms. While there is widespread support for object storage, particularly related to AWS's S3 format, file and block storage aren't always a given in a single platform. **DDN's** Infinia and the **VAST Data** Platform support file, object, and block storage on their platforms. **Hitachi Vantara's** VSP One comes in three versions for object, file, and block.

Alternative cloud provider **Vultr** offers block, file, and object storage services. Block storage is available on HDD or NVMe SSD and can be expanded as needed. The service comes with AES-256 encryption and complies with global standards, including GDPR. The Vultr File System allows shared file storage volumes to be accessed via multiple virtual machines or by the Vultr Kubernetes Engine (VKE) CNCF-certified container orchestration service. Vultr's object storage service is S3 API-compatible.

It's notable that **DDN** supports block storage via NMVe over Fabrics (NVMeoF), as well as Cinder block storage for OpenStack. **VAST Data** has opted to support block storage via NVMe over TCP. Both NVMe-based protocols are faster and more flexible than older block storage protocols such as iSCSI and SATA. **VDURA** provides a unified namespace for file and object data via its parallel file system and uses NVMe flash to accelerate data-intensive workloads.

Data quality. This term can take on a range of meanings, but here we are focused on semantic data cleaning, which isolates data according to its meaning and context, not just its format. For instance, a system capable of semantic cleaning can eliminate duplications such as "NYC" for "New York City." It also can correct data based on business rules: e.g., each customer must have a single account, or an invoice must be categorized as late if it's dated after a specified date. Semantic data cleaning also eliminates white space in data entries and ensures that no entries are duplicated. Companies capable of semantic data cleaning include **Informatica** (now part of Salesforce), **IBM, Oracle,** and **SAP,** as well as specialty firms such as **DQLabs.**

Informatica establishes a "golden record" within a specific domain. Suppose one company has a database with customer information, another with supplier information, and yet another with information about the customer. Informatica supports a function called master data management (MDM) that puts it all into one schema that can be analyzed. Informatica also can integrate data from multiple silos (e.g., from ERP systems, ServiceNow, SAP, Oracle, etc.) and ensure that an entry that appears across all applications is handled as just one entity.

Despite all this capability, Informatica can't automatically generate business rules. That's the job of the data manager within the organization, who can use Informatica's CLAIRE AI engine to obtain suggested rules but must ultimately ensure they're accurate manually. Once that's done, Informatica allows rules to be applied across domains.

UnifyApps, whose software is aimed at developing AI agents and low-code applications, also performs MDM, which is key because it integrates data from over 700 enterprise applications and SaaS services and provides tools to perform semantic data cleaning and prepare data for AI.

Many other vendors in the data management space claim data quality capabilities that aren't classed as semantic cleaning. **Cloudian,** for example, supports similarity reach, defined as the ability to manage data according to type (e.g., keeping cats and dogs in different locations).

Data quantity/distributed data handling. Nearly all data management vendors we've surveyed focus on scale as a key differentiator. Whether in the cloud or on-premises, enterprise storage systems such as those from **Dell, Everpure, IBM, NetApp,** and **Qumulo** are easily extended across geographies by adding arrays and cloud software that ensure manageability via public or private clouds. Still, not all clouds are supported; compatibility depends on product.

Metadata management. The feature of discovering and managing data according to tags that identify its location, access controls, type, and other parameters is universally supported by most data management vendors. Enterprise storage firms and data orchestration suppliers typically support operational metadata such as date and location that appears in stored data infrastructure. Companies such as **Informatica** that support semantic data cleaning can discern metadata based on business rules, context, or wording. **CTERA** and **Hammerspace** also manage metadata based on content. And **IBM** understands data content in its Storage Scale software, which works with its storage appliances.

Data format normalization. This can mean different things, but for our purposes it refers to the ability to extract, transform, and load (ETL) data or to extract, load, and transform (ELT) it into a consistent format for analytics and AI. In ETL, data is transformed once it is extracted from a source and then it is loaded into a data store or lake. Reverse ETL describes the process of shifting data put into a warehouse or data lake via ETL back into applications. In ELT, the transformation happens within the data store.

Databricks provides both functions. **Matillion** does ELT by transforming data loaded into Snowflake, Redshift, Azure Synapse, or BigQuery tables using SQL. **Fivetran** does ELT by loading data into Snowflake, BigQuery, Redshift, Databricks, and other data stores, then allowing transformation using SQL or open-source Data Build Tool (dbt), which is also offered in a commercial version from **dbt Labs,** which recently merged with **Fivetran. UnifyApps supports** ETL by taking data from a data warehouse or data lake for loading into a variety of applications.

Startup **Matia** combines ETL and reverse ETL with observability in an approach called DataOps that eliminates the need for multiple data management tools. Matia's software connectors link over 100 data sources to enterprise data warehouses and data lakes, then push data back to applications. The platform catalogs data and discovers anomalies as well.

Some vendors claim to provide a form of ETL by organizing data according to file format. This isn't the same as data format normalization, but it allows for content to be classified according to metadata. In this camp, **Komprise**, **Nasuni,** and **Panzura** are examples. **Komprise** can also do minimal transformations by converting CSV files to Parquet columnar format. Both **Komprise** and **Panzura** can be used to load data into an ETL or ELT application. **Nasuni** isn't categorized as ETL/ELT but maintains a global file system with metadata and can group or filter data by file extension and other metadata characteristics for analytics.

Apache Iceberg integration. Apache Iceberg is an open-source table format for managing large-scale tabular data in data lakes and lakehouse architectures. It enables SQL-based access to data stored on object storage and supports ACID transactions and schema changes. Iceberg table data is stored as objects (e.g., Parquet files) and is distributed across the same object storage cluster as unstructured data. It allows multiple query engines—such as Spark, Dremio, and Trino, and increasingly platforms like Snowflake—to access the same data through shared table metadata, avoiding vendor lock-in and eliminating the need to migrate or rewrite data when switching platforms.

MinIO offers a native Iceberg implementation in its AIStor platform. Every AIStor node runs identical software, which distributes tables across a storage cluster alongside objects. The vendor also offers an AIStor Tables API, which allows for an embedded Iceberg catalog service within MinIO AIStor. This can help eliminate the need for proprietary catalogs from data management vendors, since storage owns the metadata.

CSI drivers. Most of the storage platforms we cover in this report support the Container Storage Interface (CSI), which provides a standardized "bridge" in the form of an API specification between storage data and container orchestrators such as Kubernetes. Notably, though, CSI is primarily used to enable software-based provisioning and management of block and file storage, not object storage.

Another specification is in the works, namely the Container Object Storage Interface (COSI), which is meant to integrate object storage into Kubernetes, streamlining application development. So far, COSI is still in its early innings, though **Cloudian** and **Scality** support it for object storage delivery to containerized apps.

Software and hardware disaggregation. There's a trend among enterprise storage players to separate storage software from underlying hardware, something that's been around for a while but is garnering more attention as storage increases its role in AI.

Software-defined storage systems from **DataCore Software, MinIO,** and **PEAK:AIO** provide comprehensive storage functionality on commodity hardware from the likes of Dell, HPE, Lenovo, and Supermicro.

IBM's Storage Scale software provides a range of content-aware data management functions in tandem with its own appliances, including the IBM Storage Scale System 6000, as well as for on-premises NAS filers and storage systems from the likes of Dell and NetApp.

VAST Data's Disaggregated, Shared Everything (DASE) architecture keeps stateless CPUs separate from storage, letting customers add CPUs as needed while storing data in a single pool of commodity flash SSD. According to VAST Data, this approach, along with comprehensive deduplication and compression of data, delivers consistent performance for any workload. It

also eliminates the need for tiering, or the practice of shifting data from one type of storage to another, typically SSD to HDD. Taken together, VAST claims its capabilities comprise an "AI operating system" that efficiently supports the data management "chain of command" from ingestion to AI workflows, including the development of agents.

VDURA, in contrast, supports storage tiering between SSD and HDD, but does so autonomously via AI, a feature called intelligent real-time tiering. The vendor, which started out as Panasas with a PanFS parallel file system featuring storage media optimization combined with data management and security was common in HPC environments. As AI gained traction, it shifted to a microservices architecture and completely reworked its platform to fit AI. VDURA implements its VDURA Data Platform by separating control and data planes. Director Nodes govern the control plane, orchestrating and managing metadata operations and coordinating Storage Nodes and a DirectFlow Client, a Linux-compatible driver for the vendor's parallel file system.

To help customers determine how much storage is required for a given AI project, VDURA has created a GPU Storage Calculator. Users enter their GPU footprint to see how many VDURA flash nodes are required to meet or exceed production targets. The calculator is based on a reference design that includes VDURA Storage Nodes and AMD Instinct GPUs.

Cloud integration. All the major storage platforms mentioned here offer integration with various public cloud hyperscalers. Typically, this means that data stored in the cloud can be accessed by the on-premises storage system, and vice versa. If storage tiering (the ability to shift data from one type or area of storage to another to save space and improve throughput), is supported, cloud and on-prem interaction can be significant. Also, cloud support usually means a storage system can be managed from the cloud.

The **Nasuni** File Data Platform is designed to leverage cloud infrastructure to optimize data management and object storage. It integrates with cloud hyperscalers (AWS, Azure, GCP, IBM Cloud, and others) at the storage layer and integrates with cloud platforms for analytics and deployment.

Related to cloud integration is the availability of object storage in cloud-based services, which provide alternatives to services from the public cloud hyperscalers due to reduced costs (typically, these services charge no fees for data egress or API transactions. Amond providers are **Backblaze**, which offers an object management service worldwide to over 500,000 customers. Among its products is B2 Overdrive, an S3-compatible, high-throughput service supporting rates to 1 Tbit/s. Backblaze has prebuilt integrations with dozens of partners, including Commvault, CoreWeave, Fastly, MinIO, Snowflake, Vultr, and many others. This allows customers to pair their B2 data with hardware, software, and services from a variety of providers.

Wasabi offers global services based on S3-compatible object storage infrastructure. It can serve

as an origin server for CDNs. Security features include encryption of data in transit and at rest. Wasabi offers industry-specific solutions, such as HIPAA-compliant "hot" storage for healthcare and life sciences. The company has over 200 technology alliance partners, including Akamai, Dell, Google Cloud, VAST Data, and VMware; and it boasts tens of thousands of managed service provider (MSP) partners.

KV caching. During the inferencing process, keys and values are generated for individual tokens in a model. If not cached, these elements must otherwise be recomputed repeatedly to provide context to each token, slowing down the system. If the GPU high bandwidth memory (HBM) cache is full, some storage systems can take the overflow. **DDN, VAST Data,** and **WEKA** all accommodate KV caching overflow, and **VDURA** plans to release the capability soon. It's notable, though, that this procedure requires integration with frameworks such as NVIDIA's Dynamo, which specifically offloads GPU cache to attached storage as needed.

At the recent CES conference in Las Vegas, NVIDIA revealed a new capability associated with its new BlueField-4 DPUs titled the Inference Context Memory Storage Platform. In this approach, KV caching is shared with the DPU, and only the most important tokens are processed by HBM in the GPU. Among the storage vendors developing products with BlueField-4 are **AIC, Cloudian, DDN, Dell Technologies, Everpure, HPE, Hitachi Vantara, IBM, Nutanix, Supermicro, VAST Data,** and **WEKA,** according to NVIDIA.

Support for parallel file systems. This type of system allows multiple clients or applications to address the same dataset spread across multiple storage nodes to improve performance. **VDURA,** which grew out of **Panasas,** was a pioneer in this technology. VDURA has rearchitected its platform to incorporate parallel filing in a distributed environment. VDURA's PanFS file system removes the single-node bottlenecks that plagued older NAS systems and allows client machines to interface directly with storage nodes.

Some vendors work with external parallel file systems: **Cloudian** partners with **Hammerspace,** which has its own parallel file system based on NFS that is integral to its orchestration platform, which works with any brand of storage. **IBM's** General Parallel File System (GPFS) has been morphed into today's IBM Storage Scale software. **VAST Data** claims to have eliminated the need for parallel file systems with its DASE architecture.

Vectorization capabilities. Data management vendors such as **Databricks** and **Snowflake** have integral vectorization capabilities. And several storage providers, including **Dell, DDN** (via Infinia), **NetApp, VAST Data,** and **VDURA,** have added vector embeddings and vector databases to their wares. **IBM** supports data extraction, chunking, and vectorization, enabling data to be then assigned to a third-party vector database such as Milvus. This is a common tack.

MCP support. A range of storage vendors have added integral Model Context Protocol (MCP) servers to their software to allow AI agents to view and access data within the storage system.

Dell, Everpure, HPE, Hitachi Vantara, IBM, MinIO, NetApp, and **VAST Data** have this capability.

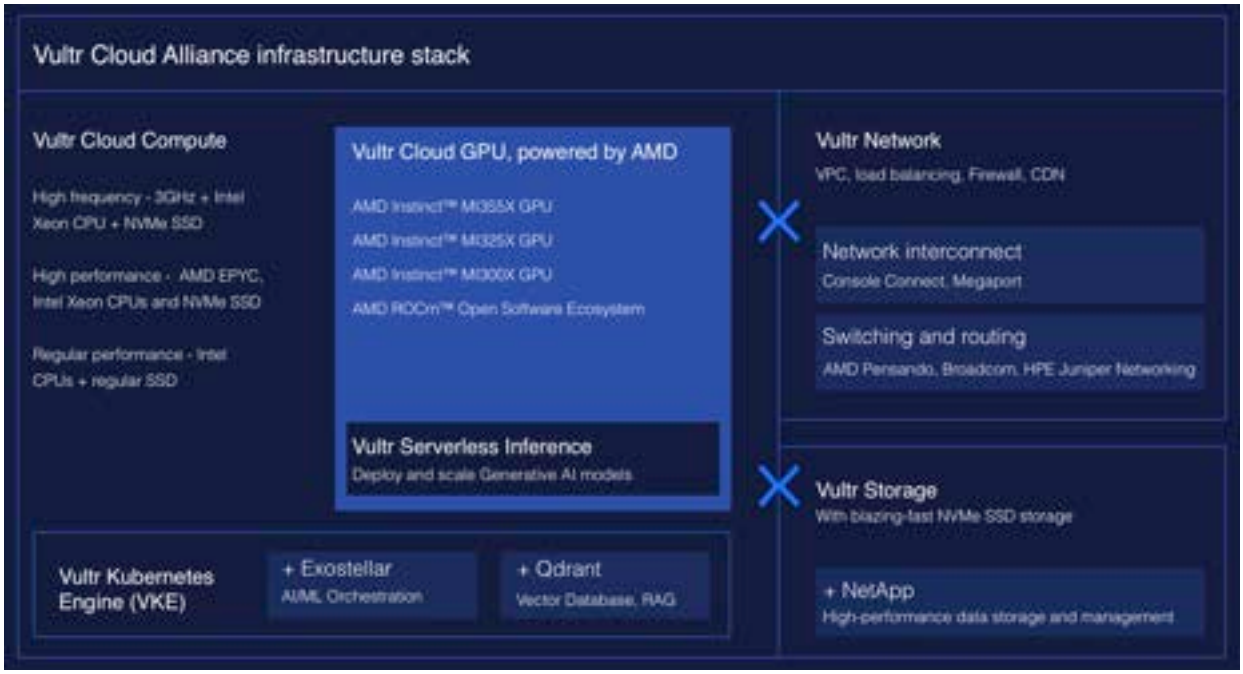
Security. Safety is table stakes for all data management vendors. Many offer end-to-end encryption, access controls, and data immutability (the capability to protect data from any changes or updates). Many, including the HPC players and most of the enterprise storage ones, offer secure multi-tenancy, which separates data for user groups within an organization. A few, including **Cloudian, DataCore,** and **Qumulo,** offer adaptive or multi-level erasure coding, which allows data to be split into fragments and stored across multiple sites to ensure protection and the ability to recover data in the event of a breach. **VAST Data** offers a feature called locally decodable erasure coding that the vendor says substantially lowers RAID overhead.

Linkage to AI pipelines. The capability to connect data stores directly to AI clusters of GPUs, XPUs, and CPUs is on the rise. **NVIDIA** has led the charge here with its GPUDirect technology based on RDMA. NVIDIA also has a reference architecture it called its AI Data Platform (AIDP) that combines NVIDIA Blackwell GPUs, NVIDIA BlueField DPUs and Spectrum-X networking, and NVIDIA AI Enterprise software with the products of storage and data management suppliers. **Cloudian, DDN, Dell, Everpure, H3C, HPE, Hitachi Vantara, IBM, NetApp, Nutanix, VAST Data,** and **WEKA** pledged support for the NVIDIA AIDP. **Cloudian** and **DDN** are among the vendors offering supported products and services for AIDP now.

Separately, NVIDIA is also working to standardize AWS's S3 object storage over RDMA, a move that substantially speeds up traffic to GPUs from AWS-compatible storage systems. So far, **Cloudian, Dell**, and **HPE** have built NVIDIA's server libraries supporting RDMA for S3 API-based object storage into their storage solutions.

NVIDIA isn't alone in offering reference architectures that incorporate storage into accelerated networks. **Vultr** has created a hybrid cloud/on premises reference architecture incorporating distributed storage from **NetApp** with **AMD** Instinct GPUs, which run with AMD's Enterprise AI Suite and its AMD ROCm software platform. ROCm competes directly with NVIDIA's CUDA framework. Vultr maintains that because the reference design can be deployment within Vultr's sovereign cloud regions, enterprises can maintain data compliance and residency controls.

Vultr also partners with **DDN and SQream** to accelerate data throughput to its GPU-driven cloud services via the Vultr Cloud Alliance, an infrastructure stack that partners Vultr services with storage and data management offerings, including with NVIDIA's Run:ai AI workload and GPU utilization orchestrator and the Qdrant vector database, as indicated in the chart below:



Source: Vultr

Selected Enterprise Storage Systems: An Overview

Vendor	Product	Data types supported	Cloud integration	Parallel file system	Integral vectorization	MCP support	KV cache offload	Secure multi-tenancy	NVIDIA GPUDirect support	Integration with NVIDIA AI Data Platform
Cloudian	Cloudian HyperStore AI Data Platform	Object, file	AWS, Azure, GCP	No	Via partnership with Milvus	No	Planned	Yes	Yes	Yes
DDN	EXAScaler	File	AWS, Azure, Google Cloud	Yes	No	No	Yes	Yes	Yes	Yes
	Infinia	Object, file, block	Google Cloud, Oracle OCI	No; works with EXAScaler's system	Yes	No	Yes	Yes	Yes	Yes
Dell	Dell AI Data Platform	Object, file, block	AWS, Azure	Pending	Yes	Yes	Yes	Yes	Yes	Yes
Everpure	Everpure Platform	Object, file, block	Azure	Yes	Via external systems	Yes	Pending	Yes	Yes	Yes
Hitachi Vantara	Virtual Storage Platform One (VSP One) and VSP 360 with Hitachi iQ Studio	Object, file, block	AWS, Azure, GCP	No but can partner with WEKA	Via iQ Studio and partnership with Milvus	Hitachi iQ offers MCP connectors	No	Yes, for VSP One Object	Yes	Yes
HPE	HPE Alletra Storage MP X10000	Object	AWS, Azure, GCP	No	Via partnership with Milvus	Yes	No	No	Yes	Yes
	HPE Alletra MP B10000	File, block	AWS, Azure	No	No	No	No	Yes	Yes	No
IBM	IBM Storage Scale System	Object, file	AWS, Azure, GCP, IBM Cloud	Yes	Yes	Yes, via IBM Storage Insights software	Yes	Yes	Yes	Yes

Vendor	Product	Data types supported	Cloud integration	Parallel file system	Integral vectorization	MCP support	KV cache offload	Secure multi-tenancy	NVIDIA GPUDirect support	Integration with NVIDIA AI Data Platform
MinIO	AIStor	Object, file	AWS, Azure, GCP	No	Via external systems	Yes	Planned	Yes	Yes	No
NetApp	NetApp AFX with NetApp AI Data Engine (AIDE)	Object, file	AWS, Azure, GCP	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Nutanix	Nutanix Unified Storage	Object, file, block	AWS, Azure	No	No	No	No	Yes	Yes	Yes
PEAK:AIO	AI Data Server and Open pNFS	Object, file, block	N/A	Yes	No	No	Yes	Yes	Yes	No
Qumulo	Qumulo Cloud Data Platform	Object, file	AWS, Azure, GCP, Oracle	Yes	No	Via external systems	No	Yes	Planned	No
Scality	Scality RING and RING XP	Object, file	AWS, Azure, GCP, Wasabi	Yes	No	No	No	Yes	No	No
VAST Data	VAST Data Platform	Object, file, block	AWS, Azure, GCP	No	Yes	Yes	Yes	Yes	Yes	Yes
VDURA	VDURA Data Platform	Object, file, block	AWS, Azure, GCP	Yes	Yes	No	Yes	Yes	Planned	No
WEKA	WEKA NeuralMesh	Object, file	AWS, Azure, GCP, Oracle	Yes	Via external systems	No	Yes	Yes	Yes	Yes

A Multifaceted Market

The above sections indicate that no data management project for AI can be supported by a single product. Instead, there is a large and complicated web of interconnected partnerships among vendors, often starring NVIDIA. And so far, no project can be completed automatically without human-in-the-loop and very human hands-on data engineering.

The complexities of data management for AI also have led to a dizzying array of point products from vendors claiming to have a single “platform.” Often, the reality entails a series of products layered onto one another as new capabilities are required.

Conclusion: As Data Management Progresses, Challenges Remain

AI's rapid advance has fundamentally redefined the role of data management. No longer a background IT function, data management now sits at the center of AI success or failure. As this report has shown, the accuracy, relevance, and trustworthiness of AI inference—and increasingly, the viability of agentic AI systems—depend less on model novelty than on the quality, accessibility, and orchestration of enterprise data.

The market response to AI's demands has been rapid but fragmented. HPC storage vendors have moved decisively into AI data platforms, leveraging their performance advantages and deep integration with GPU-centric architectures. Traditional enterprise storage providers have expanded their capabilities to incorporate faster throughput and cleaner, vectorized data. Data management platforms have rebranded themselves as AI platforms, embedding vector search, models, and orchestration directly into their environments.

Still, no single solution today addresses the full spectrum of AI data management requirements. Enterprises must instead assemble ecosystems of interoperable tools. Human expertise—data engineering, domain knowledge, and governance discipline—remains essential. AI has not eliminated the need for careful data management; it has amplified it.

Looking ahead, the competitive advantage will accrue to organizations that can adequately evaluate their data management needs and balance those against what vendors are really offering. This may mean enlisting external and internal expertise in integrating ETL/ELT products, RAG vectorization and MCP servers, enterprise storage systems, and third-party orchestrators, among other elements of a comprehensive data management pipeline.

Beyond these challenges, the advent of agentic tools that permit enterprise users to create applications that formerly required specialized software can't be ignored. It's likely that at least some functions of storage and data management will be taken over by AI models with agentic capabilities.

Still, successful use of these tools for a deep level of understanding on the part of enterprise staff. Indeed, investing in internal expertise may deliver better ROI than gearing up a single vendor's multiproduct architecture, only to find it falling short in a key area. When it comes to data management for AI, no single product is meeting all needs, and integration is essential to delivering results.

Appendix: Leading Companies to Watch

Attimis

Attimis provides a global namespace that spans heterogeneous object and file storage systems, offering stateless S3 proxy and data orchestration layer. OneBucket does not own, convert, or relocate customer data into a proprietary format. Instead, it optimizes access, placement, and cost without requiring data migration or application rewrites. OneBucket optionally supports Dell ObjectScale, PowerScale, and OneFS platforms. Seagate is a global partner. TCS is a global integrator partner for AI and analytics.
https://attimis.co

Backblaze (Nasdaq: BLZE)

Backblaze offers a range of S3-compatible object storage services designed to free businesses from barriers of lock-in, complexity, and cost. Its B2 Overdrive service accelerates AI workflows, powers data-heavy applications, streamlines media management, and protects critical data. Backblaze has dozens of integration partners, offers industry-specific solutions, and claims over 500,000 customers in 175 countries around the world. Backblaze went public in 2021.

CTERA

CTERA unifies unstructured data across edge, core, and cloud environments. The CTERA Intelligent Data Platform's global file system ensures that AI workloads can access current, semantically indexed data—regardless of where it lives—under a single namespace. Founded in 2008, CTERA has been adopted by over 50,000 businesses in 110 countries worldwide.

Databricks

Databricks offers a Data Intelligence Platform built on an open-source foundation built by the creators of Apache Spark, Delta Lake, MLflow, and Unity Catalog. Functionality is supported by a lakehouse architecture, which provides a unified approach to building, deploying, sharing, and maintaining enterprise-grade data, analytics, and AI solutions at scale. Databricks has over 20,000 customers, including AstraZeneca, Block, Comcast, Condé Nast, Rivian, Shell and over 60% of the Fortune 500.

Datadobi

Datadobi ensures that data for AI workflows will be delivered safely and without corruption. The vendor provides StorageMAP for policy-driven lifecycle management, replication, data protection, data cleaning, “golden copies,” risk protection and ransomware recovery, and a cost calculator for storage optimization.

DDN

DDN’s AI-native design unifies and accelerates AI pipelines, maximizing performance, cutting costs and accelerating ROI for leading organizations driving AI breakthroughs. A key partner to NVIDIA, DDN claims to achieve up to 99% utilization with its Data Intelligence Platform. DDN says 8 out of 10 of the world’s leading automotive companies trust DDN and that 7 out of 10 of the top banking and securities firms rely on DDN.

Everpure

Formerly Pure Storage, Everpure changed its name recently. The company offers a a single, intelligent storage and data management system virtualized across on-prem, cloud, and edge into a global fabric. Centralized intelligence automates provisioning, governance, and optimization. Integral cyber resilience protects data everywhere. These capabilities are delivered through storage-as-a-service with business guarantees.

F5 (Nasdaq: FFIV)

F5, Inc. is the global leader that delivers and secures every app. Backed by three decades of expertise, F5 has built the industry’s premier platform—F5 Application Delivery and Security Platform (ADSP)—to deliver and secure every app, every API, anywhere: on-premises, in the cloud, at the edge, and across hybrid, multicloud environments. F5 is committed to innovating and partnering with the world’s largest and most advanced organizations to deliver fast, available, and secure digital experiences. Together, we help each other thrive and bring a better digital world to life.
https://www.f5.com

Fivetran

Fivetran offers data integration that enables enterprises to accelerate cloud migration and power AI and ML workloads to drive innovation. The Fivetran platform reliably and securely centralizes data from hundreds of SaaS applications and databases to destinations on-premises, in the cloud, or in a hybrid environment. Over 10,000 global brands, including OpenAI, LVMH, Pfizer, Verizon, and Spotify trust Fivetran to move their most valuable data assets to fuel analytics, drive operational efficiencies and grow their businesses.

Komprise

Komprise delivers an analytics-driven SaaS platform to manage and mobilize all enterprise unstructured data. With Komprise Intelligent Data Management, enterprise IT teams can easily analyze, search, move and use unstructured data across silos and save 70% on storage, the vendor says. They can create policy-driven automated workflows to find, tag, and move data to the right storage at the right time to meet a variety of needs.

MinIO

MinIO provides high-performance, hyper-scale object storage. The company's commercial offering, AIStor, is built for the exascale data infrastructure challenges presented by modern AI workloads. AIStor features native Apache Iceberg support to unify structured and unstructured data. MinIO is deployed by AT&T, Broadcom, Comcast, Palantir, PNB Parabas, the U.S. Air Force, and Verizon, among many others.

Nasuni

Nasuni provides cloud-based object storage system for organizing unstructured data for use in AI and distributed enterprise applications. Its multisite synchronized system features an open-source MCP server and integration with AWS and Azure AI models. If needed, Nasuni supports physical and virtual edge appliances that interact with cloud storage continuously and with local caching of active files for optimal performance. Backup and disaster recovery are included, providing a single platform where other providers require two or more. Customers include Dow, Barnes and Noble, E-Trade, and Pernod Ricard.

Qumulo

Qumulo is a database as-a-service company targeting hybrid data management. Scale Anywhere includes Azure Native Qumulo (ANQ), a file-based cloud storage launched in April with Microsoft Azure. New additions to ANQ include Global Namespace (Q-GNS), a unified data plane for an organization’s entire unstructured data, from edge to core and cloud. Qumulo believes that Q-GNS can access remote data as if it were local for all workflows. It includes a set

of data services (multi-protocol access, enterprise security integrations, snapshots, quotas, replications, etc.) across geographically distributed systems—edge, core, or cloud.

Redpanda

Redpanda offers an Agentic Data Plane - a new category in AI infrastructure that makes it simple and secure to connect AI agents with enterprise data and systems. Built on a multi-modal data streaming engine, Redpanda empowers agentic applications that reason and act in real-time with speed, autonomy, and precision. Global leaders including Activision Blizzard, Cisco, Moody's, Texas Instruments, Vodafone and 2 of the top 5 banks in the U.S. rely on Redpanda to process hundreds of terabytes of data a day.

UnifyApps

UnifyApps is the Enterprise Operating System for AI, helping large global organizations turn fragmented data, knowledge, and processes into intelligent, production-ready automation. UnifyApps unifies enterprise records, activity, and knowledge into a single source of truth—powering autonomous agents, human-in-the-loop applications, and AI-native workflows at scale. Built for complex enterprises, UnifyApps supports all layers of the modern AI stack, from integrations and governance to orchestration and agentic AI. Organizations across financial services, retail, manufacturing, aviation, and the public sector use UnifyApps to compress processes from days to minutes.

VAST Data

VAST Data is an AI-focused software company that provides a unified, high-performance data platform, known as the VAST AI Operating System, which is designed to store, manage, and process massive datasets for deep learning and AI applications. It consolidates storage, database, and computation into a single, scalable engine to eliminate data silos and accelerate AI workflows.

VDURA

VDURA has created a data storage infrastructure that merges the velocity of flash-based parallel file systems with the durability of modern object storage. Born from two decades of PanFS heritage, it addresses the insatiable data demands of AI and high-performance computing environments. The platform's architecture eliminates traditional bottlenecks through intelligent separation of control and data planes, enabling thousands of clients to access petabytes simultaneously. Organizations from federal organizations and national laboratories to Fortune 500 manufacturers rely on VDURA.

Vultr

Vultr delivers enterprise-grade AI and GPU cloud infrastructure featuring globally decentralized compute clusters, robust networking, top-tier security, and regulatory compliance. Developers can instantly access dedicated or on-demand AMD and NVIDIA GPUs for training, deployment, and edge-serving AI applications. Vultr removes vendor lock-in, offers transparent pricing, and integrates easily with tools and APIs. Advanced orchestration (via Run:ai and Kubernetes) ensures optimal resource use, lifecycle management, and rapid AI project development across regions. This approach enables secure, compliant, cost-effective scaling and redefines AI architecture as the new public cloud standard.
https://www.vultr.com

Wasabi

Recognized as one of the technology industry's fastest growing companies, Wasabi is on a mission to store the world's data by making cloud storage affordable, predictable and secure. With Wasabi, visionary companies gain the freedom to use their data whenever they like without being hit with unpredictable fees or vendor lock-in. Instead, they're free to build best-of-breed solutions with the industry's fastest-growing ecosystem of independent cloud application partners. Customers and partners all over the world trust Wasabi to help them put their data to work so they can unlock their full potential.

WEKA

Founded in 2013, WEKA provides a high-performance, software-defined data platform optimized for AI, machine learning, and GPU-intensive workloads. It uses a parallel file system to accelerate data pipelines across on-premises, cloud, and edge environments, eliminating storage bottlenecks to improve efficiency, speed, and sustainability. WEKA says it is transforming from a data platform approach to a revolutionary containerized microservices architecture, pioneering a new category of adaptive AI infrastructure, designed to power the age of reasoning with its technology NeuralMesh.