

Eight Storage Requirements for AI and Machine Learning

Data is the life-blood of artificial intelligence and machine learning (AI and ML). Vast quantities of training data enhance accuracy in the search for potentially predictive relationships.

Here are eight specific storage requirements of AI and ML applications and why they demand the data management capabilities supplied by enterprise object storage solutions.

1. SCALABILITY

Artificial intelligence systems can process vast amounts of data in a short timeframe—an essential attribute since large data sets are required to deliver accurate algorithms. This data volume drives significant storage demands. Microsoft, for example, required five years of continuous speech data to teach computers to talk. Tesla is teaching cars to drive with 1.3 billion miles of driving data. Managing these data sets requires a storage system that can scale without limits.

HOW CLOUDIAN HELPS

Object storage is the only storage type that scales limitlessly within a single namespace. Plus, the modular design allows capacity to be added at any time. You can scale with demand, rather than ahead of demand.

2. COST EFFICIENCY

A useful storage system must be both scalable and affordable, two attributes that don't always co-exist in enterprise storage. Historically, highly-scalable systems have been more expensive on a cost/capacity basis. Large AI data sets are not feasible if they break the storage budget.

HOW CLOUDIAN HELPS

Object storage is built on the industry's lowest cost hardware platform. Combine that with low management overhead and space-saving data compression features, and the result is 70% less cost than traditional enterprise disk storage.

3. SOFTWARE-DEFINED STORAGE OPTIONS

Vast data sets will sometimes require hyperscale data centers with purpose-built server architectures already in place. Other deployments may benefit from the simplicity of pre-configured appliances.

HOW CLOUDIAN HELPS

Object storage keeps your deployment options open, with your choice of storage appliances or software-defined storage.

4. HYBRID ARCHITECTURE

Different data types have varying performance requirements, and the hardware must reflect that. Systems must include the right mix of storage technologies to meet the simultaneous needs for scale and performance, rather than a homogeneous approach that will fall short.

HOW CLOUDIAN HELPS

Object storage employs a hybrid architecture, with spinning disk for user data and SSDs for performance-sensitive metadata, thus optimizing cost and performance.

5. PARALLEL ARCHITECTURE

For data sets that grow without limits, a parallel-access architecture is essential. Otherwise, the system will develop choke points that limit growth.

HOW CLOUDIAN HELPS

Object storage employs a shared-nothing cluster architecture, which means that all parts of the system work in parallel. Data throughput grows continuously as the system expands.

6. DATA DURABILITY

Backing up a multi-petabyte training data set is not always feasible; it would often be cost and time prohibitive. But you can't leave it unprotected either. Instead, the storage system needs to be self-protecting.

HOW CLOUDIAN HELPS

Object storage is designed with redundancy built-in, so data is protected without requiring a separate backup process. Furthermore, you can select the level of data protection needed for each data type to optimize efficiency. Systems can be configured to tolerate multiple node failures, or even the loss of an entire data center.

7. DATA LOCALITY

While some AI/ML data will reside in the cloud, much of it will remain in the data center for a variety of reasons: performance, cost, and regulatory compliance are three of them. To be competitive, on-prem storage must offer the same cost and scalability benefits as its cloud-based counterpart.

HOW CLOUDIAN HELPS

Object storage is the storage of the cloud. In fact, Cloudian supplies object storage solutions to many cloud providers for use as public cloud infrastructure. The scalability and economics of cloud storage are now available to you on-prem.

8. CLOUD INTEGRATION

Regardless of where data resides, integration with the public cloud will still be an important requirement for two reasons. First, much of the AI/ML innovation is occurring in the cloud. On-prem systems that are cloud-integrated will provide the greatest flexibility to leverage cloud-native tools. Second, we are likely to see a fluid flow of data to/from the cloud as information is generated and analyzed. An on-prem solution should simplify that flow, not limit it.

HOW CLOUDIAN HELPS

Cloudian is cloud-integrated in three ways. First, it employs the S3 API, the de-facto standard language of cloud storage. Second, it facilitates tiering to Amazon, Google, and Microsoft public clouds, and lets you view local and cloud-based data within a single namespace. Third, data stored to the cloud from Cloudian is directly accessible by cloud-based applications. This bi-modal access lets you employ both cloud and on-prem resources interchangeably.



Realizing the full potential of AI/ML requires an infrastructure that supports innovation. Cloudian enterprise object storage delivers the scalability, cost efficiency, and interoperability that enhances the capabilities of these emerging technologies.



Cloudian, Inc.
177 Bovey Road, Suite 450
San Mateo, CA 94402
Tel: 1.650.227.2380
Email: info@cloudian.com
Web: cloudian.com