

USING NETAPP STORAGEGRID FOR AI



Unlocking the power of AI with NetApp StorageGRID

Artificial Intelligence is transforming industries by enabling advanced data analysis, predictive modeling, and automation. However, the success of AI initiatives depends heavily on the quality and volume of data used for training models. To truly harness the innovation in your data, you need a storage solution that can handle large volumes and varieties of data, give you high data quality, and deliver with speed, scale, and efficiency.

NetApp® StorageGRID® stands out as a premier object storage solution that's designed to meet the unique needs of AI data preparation and model training. With its advanced capabilities, StorageGRID helps AI practitioners to efficiently manage, store, and retrieve vast amounts of data, thereby accelerating the AI development lifecycle.

Integration and compatibility

Seamless integration with AI ecosystems. StorageGRID is designed to integrate seamlessly at different phases of AI/ML workflows as well as with popular AI and ML frameworks, such as TensorFlow, PyTorch, Apache® Spark™, and NVIDIA NeMo. This compatibility means that data scientists and engineers can use their preferred tools without facing integration challenges. The support for standard protocols like S3 and RESTful APIs further simplifies the integration process, enabling smooth data flow between storage and compute environments.

Hybrid and multicloud capabilities. In this era of hybrid and multicloud deployments, StorageGRID offers the flexibility to operate across on-premises and cloud environments. This capability is especially beneficial for AI workloads that may require data to be stored and processed in different locations. The ability of StorageGRID to tier data to public clouds such as Amazon Web Services, Azure, and Google Cloud means that organizations can take advantage of cloud scalability and cost efficiency while maintaining control over their data.

Optimized appliances or software defined. StorageGRID offers flexible deployment options, supporting both appliance-based and software-only delivery models. This flexibility allows organizations to choose between preconfigured, high-performance appliances for ease of deployment and management, or software-only solutions that can be installed on the existing infrastructure, providing maximum flexibility and cost efficiency. This versatility enables StorageGRID to meet diverse operational and budgetary requirements.

Scalability and performance

Unmatched scalability. AI data preparation and model training require handling large datasets that can span petabytes. StorageGRID offers exabyte scalability, allowing organizations to start relatively small and grow seamlessly as their data requirements increase. This scalability means that AI projects are not constrained by storage limitations, enabling continuous data ingestion and processing without interruptions.

High performance. The performance of data storage solutions is critical in AI workflows, where data needs to be accessed and processed rapidly. StorageGRID leverages a distributed architecture that optimizes data placement and access patterns. This optimization results in high throughput and low latency, which means that AI models can be trained efficiently. The system's ability to handle concurrent read and write operations without degradation in performance is crucial for the iterative nature of model training.

Cost efficiency

Optimized storage costs. Managing storage costs is a significant concern for organizations that are dealing with large AI datasets. StorageGRID offers cost-efficient storage solutions through its intelligent data placement and tiering capabilities. By automatically moving infrequently accessed data to lower-cost tiers, StorageGRID helps organizations optimize their storage expenses. This cost efficiency allows organizations to allocate more resources to AI development and innovation.

KEY BENEFITS

- **Integrated.** Easily integrates with artificial intelligence and machine learning (AI/ML) tools and frameworks, facilitating efficient data workflows.
- **Performant.** Delivers high throughput and low latency for rapid data access.
- **Scalable.** Seamlessly scales to accommodate large volumes of AI data.
- **Efficient.** Optimizes storage costs with tiered storage options and lifecycle policies.
- **Simplified.** Enables data organization and retrieval with metadata tagging and search.
- **Secure.** Provides security with encryption, access control, object lock, authentication, audit logging, and data protection.
- **Durable.** Provides data integrity with robust replication and erasure coding.

Reduced operational overhead. StorageGRID automated management features reduce the operational overhead associated with managing large-scale storage environments. It's easy to operate from setup to expansion and simple to upgrade. The system's self-managing and self-healing capabilities minimize the need for manual intervention, allowing IT teams to focus on higher-value tasks. This operational efficiency translates to reduced total cost of ownership and improved return on investment for AI projects.

Security and compliance

Comprehensive security features. AI data often includes sensitive and proprietary information, making security a top priority. StorageGRID offers comprehensive security features, including encryption at rest and in transit, access controls, and audit logging. These features protect data against unauthorized access and breaches. Additionally, StorageGRID compliance with industry standards and regulations helps organizations meet their legal and regulatory obligations.

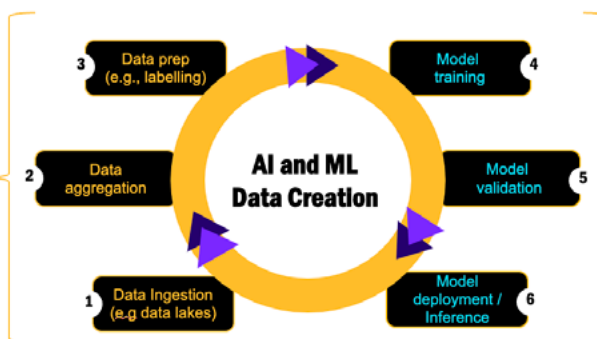
Data sovereignty and governance. For organizations operating in regulated industries, data sovereignty and governance are crucial considerations. StorageGRID enables organizations to define and enforce data residency policies, so that that data remains within specified geographic boundaries. This capability is essential for complying with data protection regulations such as GDPR and CCPA, giving organizations confidence that their data governance requirements are met.

StorageGRID in AI

Object usage in AI/ML

StorageGRID & Data Prep

- Scalable storage for massive datasets
- Low \$/GB for data lake & repositories
- Cost-effective storage scalability
- Metadata tagging for easy data organization



StorageGRID & Training

- High-performance data access and retrieval
- Low-latency access to critical data
- Seamless integration with AI workflows
- Enhanced data security and compliance

Best AI Content Repositories

Some needs will converge

Extreme Performance for AI/ML

Figure 1: StorageGRID and AI workflow.

Data management and durability

Intelligent data management. AI data preparation involves several stages, including data collection, cleaning, transformation, and annotation. StorageGRID provides advanced data management capabilities that simplify these processes. Features such as metadata tagging, policy-driven data lifecycle management, and automated data tiering help organizations to organize and manage their data effectively. These capabilities mean that relevant data is readily accessible, while optimizing storage costs.

Robust data durability. Data durability is a critical factor in AI projects, where data loss can have significant repercussions. StorageGRID maintains data durability through its erasure coding and replication mechanisms. These techniques protect data against hardware failures and maintain data integrity. Additionally, StorageGRID self-healing capabilities automatically detect and repair data inconsistencies, giving you peace of mind that your AI datasets are secure and reliable.

Innovation is in your data

NetApp StorageGRID stands out as an industry-leading object storage solution for AI data preparation and model training due to its scalability, performance, data management capabilities, integration with AI ecosystems, security features, and cost efficiency.

With StorageGRID, organizations can accelerate their AI initiatives, giving them the robust and reliable storage foundation they need to drive innovation and achieve competitive advantage in the AI-driven world.

► To learn more, visit the [StorageGRID web page](#).



Contact Us

About NetApp

NetApp is the intelligent data infrastructure company, combining unified data storage, integrated data services, and CloudOps solutions to turn a world of disruption into opportunity for every customer. NetApp creates silo-free infrastructure, harnessing observability and AI to enable the industry's best data management. As the only enterprise-grade storage service natively embedded in the world's biggest clouds, our data storage delivers seamless flexibility. In addition, our data services create a data advantage through superior cyber resilience, governance, and application agility. Our CloudOps solutions provide continuous optimization of performance and efficiency through observability and AI. No matter the data type, workload, or environment, with NetApp you can transform your data infrastructure to realize your business possibilities. www.netapp.com



© 2025 NetApp, Inc. All Rights Reserved. NETAPP, the NETAPP logo, and the marks listed at <http://www.netapp.com/TM> are trademarks of NetApp, Inc. Other company and product names may be trademarks of their respective owners. SB-4339-0525