■ NetApp

NETAPP FOR
DATA LAKES

Harnessing the innovation in your data

The use of data lakes is motivated by the need to manage and analyze vast amounts of diverse data, the demand for advanced analytics and AI capabilities, cost efficiency, improved data integration, and accessibility, scalability, and regulatory compliance. By addressing these critical needs, data lakes have become an essential component of modern data management strategies. They enable organizations to unlock the full potential of their data and drive business success, illustrating that innovation is truly in your data.

Data lakes are anything but "niche" now. In fact, they are used by a wide range of organizations and industries to support their data-driven initiatives:

- **Enterprises.** Large corporations use data lakes to integrate and analyze data from various departments, advancing strategic decisions and innovation.
- **Healthcare.** Medical institutions use data lakes to store and analyze patient data, research results, and use data to improve patient care and operational efficiency.

- **Financial services.** Banks and financial institutions use data lakes for risk management, for fraud detection, and to gain insights from transaction data.
- **Retail.** Retailers use data lakes to analyze customer behavior, optimize supply chains, and personalize marketing efforts.
- **Technology companies.** Tech firms use data lakes to support product development, customer analytics, and operational efficiency.

Unmatched scalability

NetApp® StorageGRID® object-based storage is designed to handle the exponential growth of data, making it an ideal solution for data lakes. With the ability to scale smoothly from petabytes to exabytes, your storage infrastructure can grow alongside your data needs. This scalability is crucial for data lakes, which must accommodate vast amounts of structured and unstructured data from diverse sources. Whether you're dealing with IoT (Internet of Things) data, social media streams, or enterprise application logs, StorageGRID can efficiently store and manage it all.

Seamless hybrid cloud integration

In today's multicloud world, the ability to integrate seamlessly with hybrid cloud environments is a significant advantage. StorageGRID offers robust hybrid cloud integration, allowing organizations to manage and move data across on-premises and cloud environments effortlessly. This flexibility enables businesses to leverage the benefits of both private and public clouds, such as scalability, cost savings, and disaster recovery options. Additionally, StorageGRID integrates with leading analytics partners like Dremio and Snowflake and is compatible with open table formats, providing enhanced data processing and analysis features. With StorageGRID, you can avoid vendor lock-in and maintain control over your data, regardless of where it resides.

High performance

In the realm of data lakes, performance is paramount. Analytics and AI workloads require fast data retrieval and processing to deliver timely insights. StorageGRID excels in this area, offering high throughput and low latency for data access. This performance lets data scientists and analysts quickly access the data they need, run complex queries, and derive actionable insights without delays. The high performance of StorageGRID not only accelerates data processing but also enhances the overall efficiency of data lake operations.

Cost efficiency

Traditional storage solutions can become prohibitively expensive as data volumes increase. StorageGRID addresses this challenge through automated data tiering, which moves infrequently accessed data to more cost-effective storage tiers. This optimization reduces the need for expensive high-performance storage and significantly lowers overall storage costs. With StorageGRID, organizations can achieve a balance between performance and cost, ensuring that they get the most value out of their storage investments.

Robust data durability and protection

Data durability and protection are critical considerations for any storage solution, especially for data lakes that store valuable, sensitive information. StorageGRID employs advanced data protection features such as erasure coding and georedundancy. Erasure coding provides strong protection against data loss by distributing data across multiple nodes, and georedundancy replicates data across different geographic locations. These features enable high data availability and integrity, safeguarding your data against hardware failures, disasters, and other potential threats.

Regulatory compliance and data governance

Meeting regulatory requirements and enforcing data governance policies are essential for many organizations. StorageGRID provides secure, long-term data retention capabilities that help businesses comply with industry regulations and standards. Features such as object lock, encryption, and access controls mean that data is stored securely and can be retrieved as needed for audits and

KEY BENEFITS

- **Scalable.** Easy scaling from petabytes to exabytes accommodates growing data volumes.

- **Integrated.** Seamless integration with hybrid cloud environments and analytics tools provide flexible data management.

- **High-performing.** Fast data retrieval and processing capabilities enhance analytics and AI workloads.

- **Efficient.** Tiered storage—moving infrequently accessed data to cost-effective tiers—optimizes storage costs.

- **Durable.** Advanced data protection features like erasure coding and geo-redundancy maintain data integrity.

- **Compliant.** Secure, long-term data retention helps meet regulatory requirements and enforce data governance policies.

compliance checks. By using StorageGRID, organizations can confidently manage their data in accordance with regulatory mandates, reducing the risk of noncompliance and associated penalties.

Simplified management and automation

Managing a data lake can be complex, but StorageGRID simplifies this task through its intuitive management interface and automation. The solution offers policy-driven data management, which automates tasks such as data tiering, replication, and lifecycle management. This automation reduces the administrative burden on IT teams, allowing them to focus on more strategic initiatives. With the simplified management that StorageGRID offers, your data lake operates smoothly and efficiently, with minimal manual intervention.

Future-proof technology

As data storage technologies continue to evolve, it's crucial to invest in a solution that's future-proof. StorageGRID is built on a modern, software-defined architecture that can adapt to changing technological landscapes. This adaptability means that your investment in StorageGRID will continue to deliver value as new storage innovations emerge. By choosing StorageGRID, you're not only meeting your current data lake needs—you're preparing your organization for future data challenges.
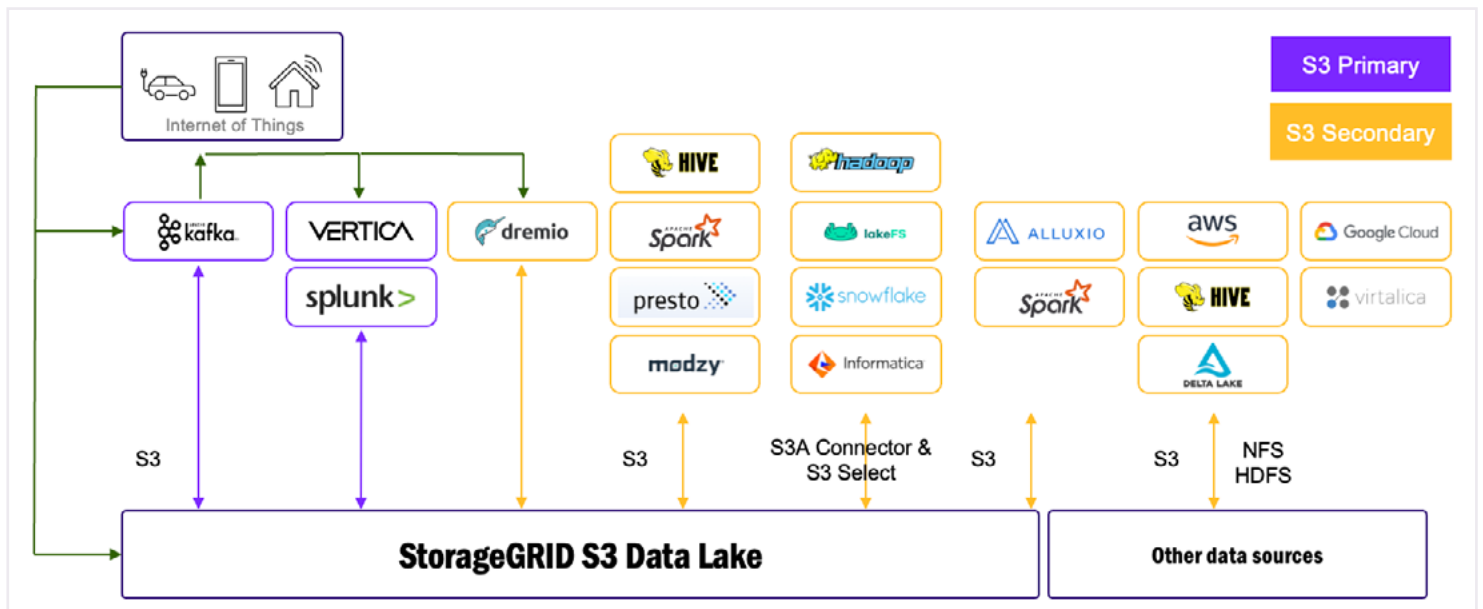
Figure 1: StorageGRID S3 data lake illustrated with partner solutions.

A comprehensive solution

NetApp StorageGRID stands out as the industry-leading object storage solution for data lakes due to its unmatched scalability, cost efficiency, high performance, robust data durability, seamless hybrid cloud integration, regulatory compliance, simplified management, and future-proof technology. By leveraging StorageGRID, organizations can harness the full potential of their data, drive innovation, and achieve business success.

Whether you're looking to enhance your analytics capabilities, support AI initiatives, or enable long-term data retention, StorageGRID provides the comprehensive solution you need to manage your data lake effectively and extract the innovation from your data.

▶ To learn more, visit the NetApp StorageGRID page.

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**

Contact Us

■ NetApp