

AI IN THE CLOUD ERA



82% OF ENTERPRISES WANT TO LEVERAGE THEIR DATA WITH GENAI.¹

THE PROBLEM

Just about every enterprise is looking for ways to extract value from their data by using artificial intelligence and the cloud. However, as organizations move from GenAI pilots to retrieval-augmented generation (RAG) and agentic AI workloads, they put increasing demands on their storage infrastructures. Moving large AI datasets across services or clouds can be slow, complex, and expensive, often creating bottlenecks and exposing security vulnerabilities.

Training AI models at scale requires high bandwidth to rapidly ingest vast, diverse datasets. Preparing data for training, whether you're cleaning, labeling, or transforming the data, demands extremely high I/O performance and orchestration across CPUs, memory, and GPUs. At the inference stage, where AI is expected to deliver real-time answers and actions, you need ultra-low latency and fast memory access to GPUs to ensure that responses are instant and reliable.

Whether powering chatbots or agentic AI systems that adapt and act independently, delays in data access can have ripple effects that impact the user experience.

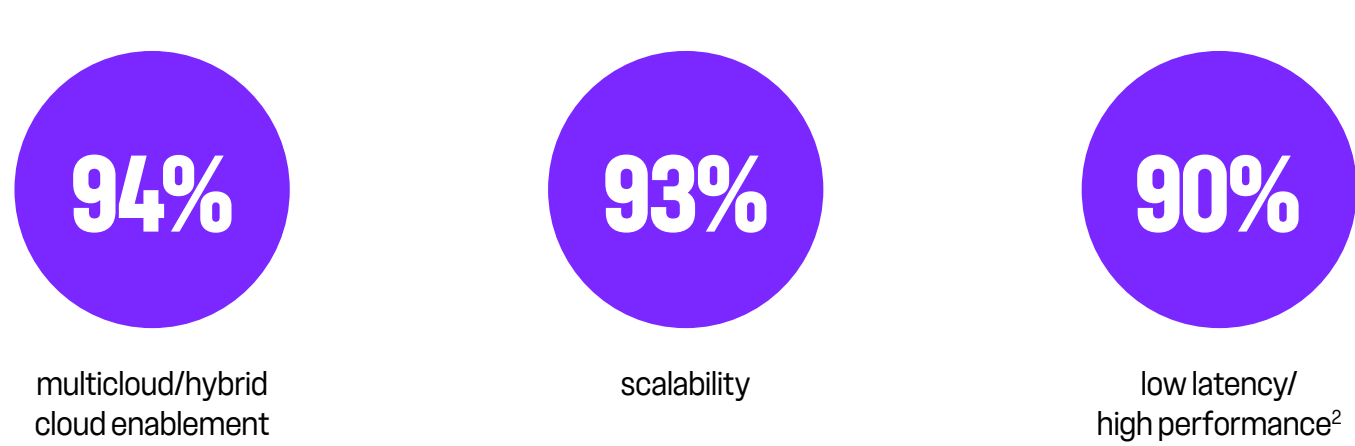
THE SOLUTION: BRING AI TO YOUR DATA WITH AZURE NETAPP FILES

Azure NetApp Files is a fully managed file storage service from Microsoft Azure that integrates seamlessly with Microsoft Azure AI and data services to bring AI to your data. By keeping enterprise data in place and making it directly accessible to Azure AI services, Azure NetApp Files accelerates AI adoption and eliminates the need for costly, complex data movement or replication.

Azure NetApp Files delivers the submillisecond latency, high throughput, and bandwidth required for every stage of the AI data pipeline, from ingest to preparation to inference. With built-in security, governance, and compliance controls, Azure NetApp Files makes AI in Azure faster, safer, and more cost effective for your most sensitive AI workloads.

Seamless integration with Microsoft AI tools, such as CoPilot and Azure AI Video Indexer, enables you to bring AI to your data to boost productivity without increasing management overhead. Azure AI Search can index and retrieve data from Azure NetApp Files via OneLake, supporting semantic search and vector-based retrieval for intelligent applications.

Important features of AI storage infrastructure



Benefits

- High performance for AI pipelines.** Accelerate ingest, training, and inference with submillisecond latency and massive throughput.
- Seamless Azure integration.** Keep your data in one place and bring AI to your data with direct connections to Databricks, Fabric, OneLake, Copilot, and AI Search.
- Real-time AI enablement.** Empower data scientists and developers with instant access to enterprise data for analytics, model training, and real-time inference.
- Deliver better user experiences.** Support chatbots, search, and interactive AI applications that feel immediate and responsive.
- Strengthen trust and compliance.** Protect sensitive data with built-in security, access controls, and governance for AI workloads.
- Control costs.** Eliminate unnecessary infrastructure, data movement, and storage duplication costs, while getting more value from your existing Azure investments.

Azure NetApp Files empowers critical AI workloads

Large Language Models

Enhance your LLMs with RAG at enterprise scale by using Microsoft Azure, NVIDIA's AI Blueprint for RAG, and Azure NetApp Files. Azure NetApp Files provides a robust foundation for enterprise RAG workloads through a comprehensive feature set that includes flexible service levels, enterprise-grade reliability, seamless integration with Azure services, and built-in security with encryption at rest and in transit—all managed through the Azure portal and APIs.

Copilot

Copilot responses are only as good as the data you provide to it. With the NetApp® Neo Copilot Connector, you can enhance Microsoft 365 Copilot by enabling the integration of your enterprise file data without complex migration. Securely unlock the value of your data to enable faster insights, richer context, and more accurate responses while respecting existing permissions.

Microsoft Discovery

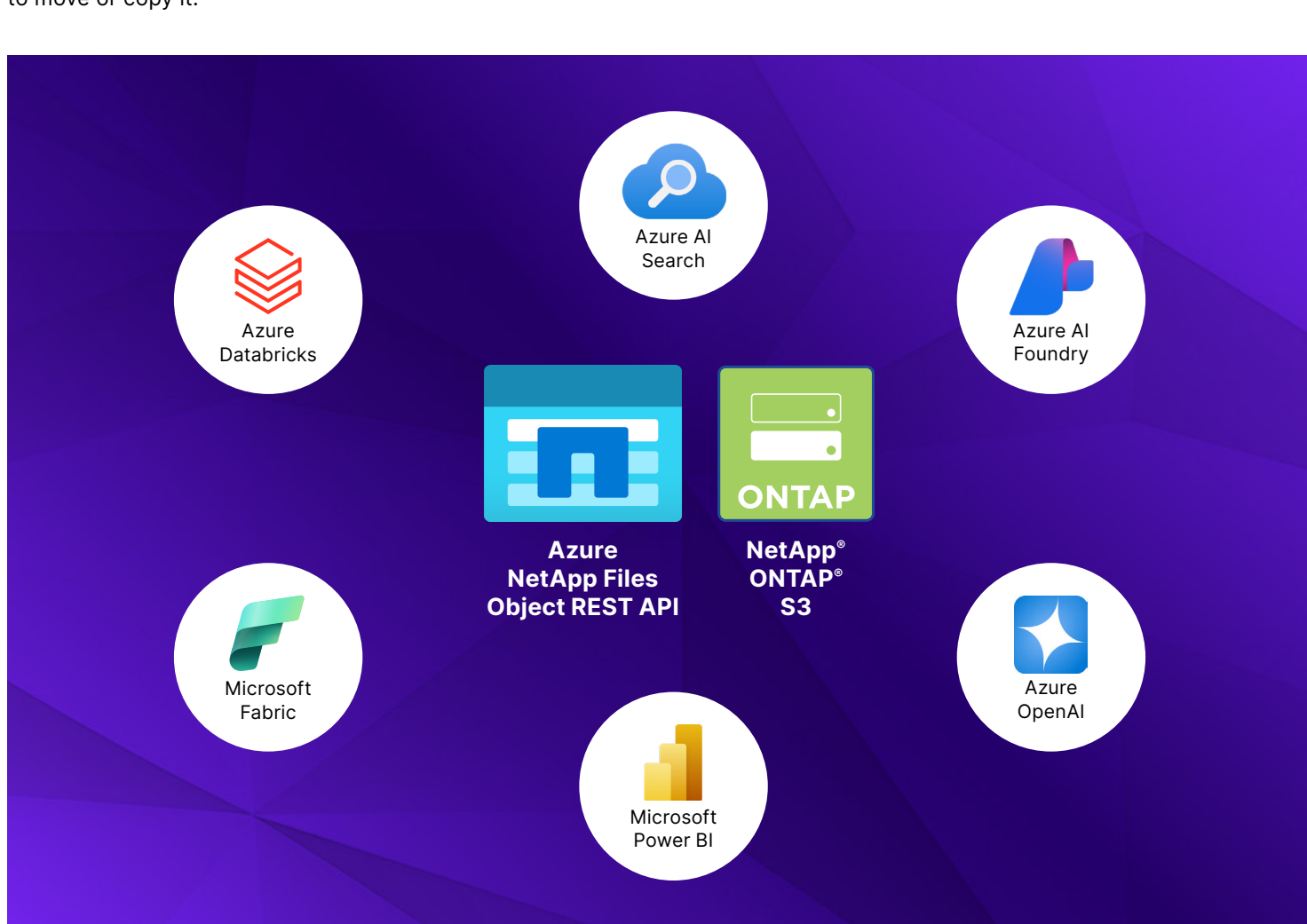
For exploratory data analysis (EDA) workloads, Microsoft Discovery provides the platform and capabilities you need to implement a complete agentic AI environment without management overhead. High-performance file storage from Azure NetApp Files delivers unmatched performance for metadata-heavy workloads, while the AI models, the data they use, and the results they produce all remain under your control within your environment.

Azure Databricks

Enable real-time analytics for AI/ML workloads by using Azure Databricks. Object REST API enables Azure Databricks to connect directly to Azure NetApp Files volumes so users can browse and query data in place, unlocking advanced analytics workflows, reporting, and visualizations instantly.

Integrations

Azure NetApp Files integrates with Azure AI services so you can bring AI to your data without having to move or copy it.



Coca-Cola Japan

Discover how Coca-Cola Bottlers Japan Inc. migrated their record-breaking SAP workload to the cloud, positioning them to explore the art of the possible for AI in the cloud. Coca-Cola Bottlers Japan can now leverage their data to make better decisions.

[Read the customer story >](#)

Additional resources

- [Azure NetApp Files >](#)

[Enhancing Microsoft 365 Copilot with External Data: NetApp Neo Copilot Connector | Microsoft Community Hub >](#)
- [Microsoft Discovery: The path to an agentic EDA environment | Microsoft Community Hub >](#)

[Building an Enterprise RAG Pipeline in Azure with NVIDIA AI Blueprint for RAG and Azure NetApp Files | Microsoft Community Hub >](#)

1 Source: IDC, August 2023
 2 Source: Forrester