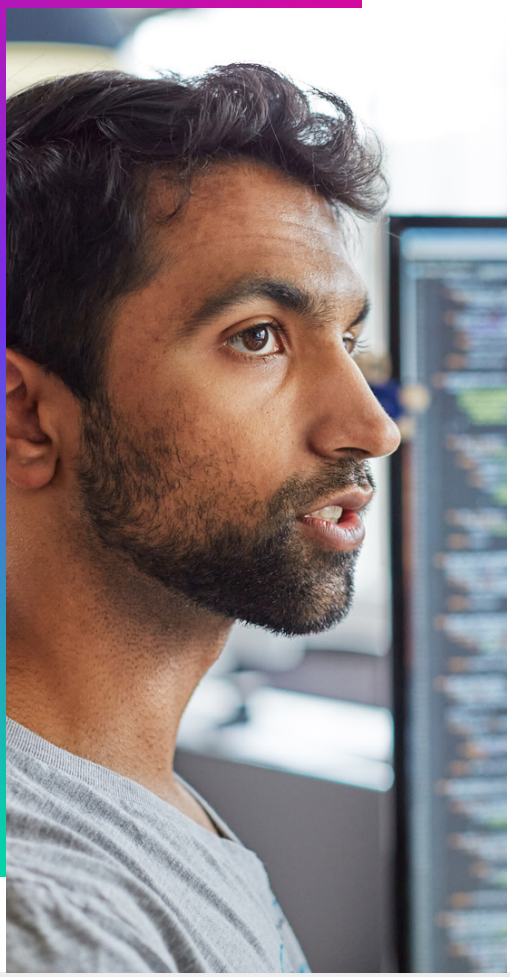
HPE Storage

Solution overview

Realize the full potential of AI with HPE Data Fabric Software

HPE GreenLake



80% of AI projects fail due to poor data management.[1] Fragmented, siloed, and inconsistent data leads to inefficient AI models, delayed insights, and high operational costs. But the efficacy of AI is interwoven with the quality of data that fuels its development.

Inaccurate, incomplete, or biased data can lead to flawed AI models, compromising their reliability and hindering their ability to support informed decision-making.

In today's data-driven world, ensuring data quality is a strategic imperative, not merely a technical requirement. The quality of data significantly affects the performance of AI models. When these models are trained on biased or incomplete data, they tend to make inaccurate predictions. Anomalies and inconsistencies in data can further mislead AI models, leading to operational inefficiencies and compromised decision-making. Moreover, data scientists often devote a disproportionate amount of their time, approximately 67%, to data preparation tasks.[2]

Data fragmentation, resulting from disparate data silos, hinders efficient data access, impedes data-driven decision-making, and can ultimately limit an organization's ability to discover valuable insights and achieve its full potential.

Seamless data landscape

HPE Data Fabric Software has empowered hundreds of organizations by providing a unified platform that seamlessly bridges the gap between diverse data silos. The software helps users access, manage, organize, and govern enterprise data, across a variety of formats, in a single, consistent, easy-to-use, edge-to-cloud data plane optimized for AI and analytics workloads. This centralized approach enables AI practitioners to effortlessly access, manage, and secure a diverse range of data assets—files, objects, tables, and streaming data—all in their original formats. No more grappling with scattered data locations or compatibility issues.

By federating data into a unified platform, HPE Data Fabric Software offers a compelling value proposition by streamlining data management processes, optimizing resource utilization, and improving data governance. By federating data across diverse environments and automating management, security, and governance, customers can leverage the following:

- **Unified data management**—HPE Data Fabric Software supports structured (Apache Iceberg, Parquet), semistructured (JSON, Avro), and unstructured data (object storage, file systems, streaming data)
- **Global namespace (GNS)**—HPE Data Fabric Software enables data federation, allowing users to query and access data across different storage locations **without data movement**
- **Hybrid cloud and multicloud deployment**—HPE Data Fabric Software can be deployed across on-premises, cloud, and edge environments, offering vendor-agnostic flexibility
- **Optimized for AI and analytics workloads**—HPE Data Fabric Software is designed to accelerate machine learning workflows with **Apache Iceberg support**, improving query performance and scalability
- **Advanced security and compliance**—Integrated governance, **policy-driven encryption, role-based access control (RBAC), and regulatory compliance** enables to meet stringent industry standards (SOC 2, HIPAA, GDPR)

[1] "The Root Causes of Failure for Artificial Intelligence Projects and How They Can Succeed," RAND, August 13, 2024.

[2] "New AI survey: Poor data quality leads to $406 million in losses," Fivetran, March 20, 2024.

Accelerating data-driven insights

Imagine a world where data science teams have unfettered access to their entire universe of data. No more wrestling with access rights to scattered locations or battling incompatible formats. Seamless access to any organization's data universe is possible from a single, consistent naming scheme known as a Global Namespace (GNS).

A GNS provides a unified, location-transparent view of your data, simplifying management and enabling seamless access. Unlike traditional methods that require physical data movement, HPE Data Fabric Software can ingest and manage data directly from diverse sources, including edge devices, multiple clouds, colocation facilities, and on-premises environments. By eliminating the need for unnecessary data movement, Data Fabric reduces costs and improves data accessibility. Users can easily access the datasets and automated security and governance policies help ensure data is secure, private, and in compliance with the GNS.

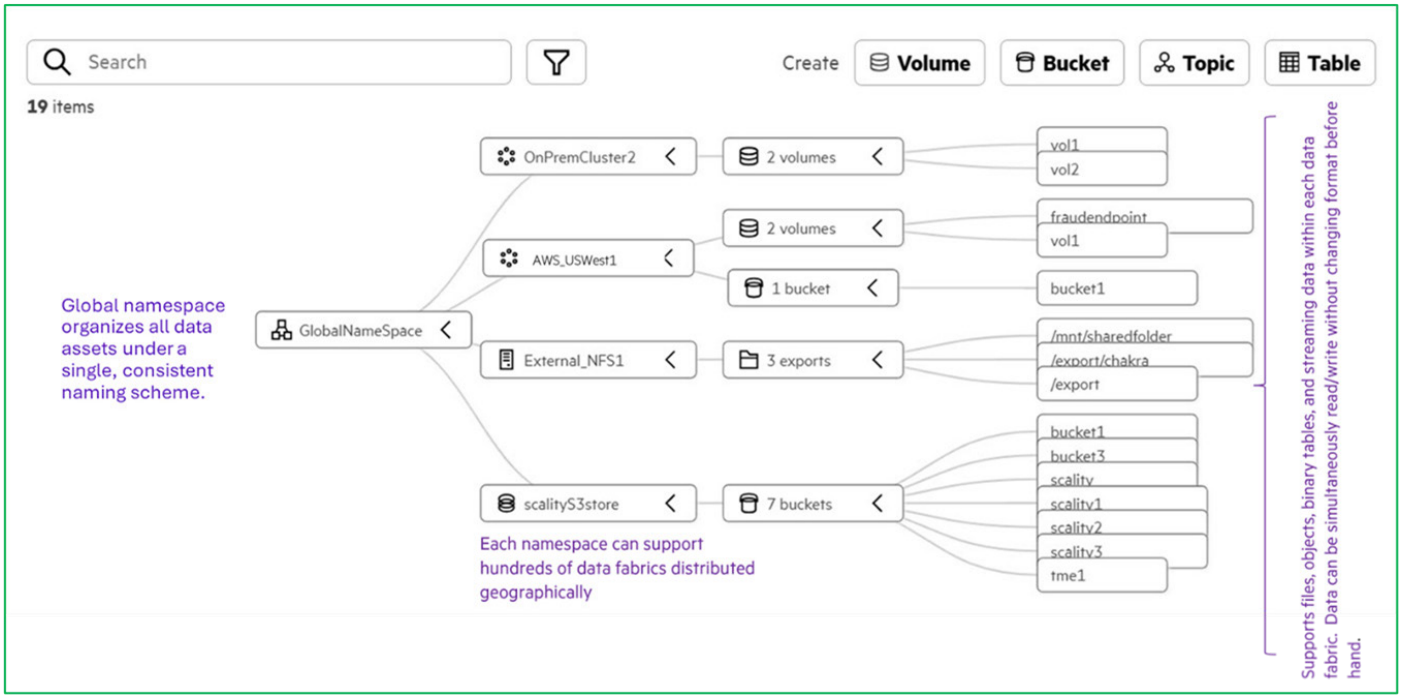
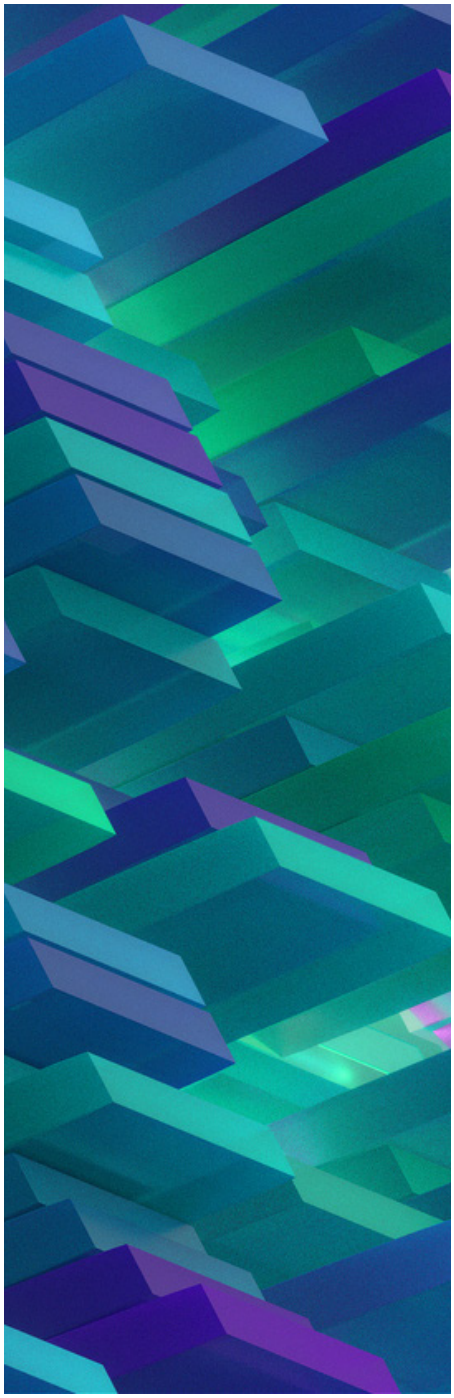


Figure 1. An illustration of GNS that ingests and manages data directly from any source to help eliminate data movement and associated costs

The result is a centralized control plane that simplifies data quality, access, and management, helping ensure that AI initiatives are fueled with high-quality data. It streamlines data operations and enables data science teams to focus on valuable insights instead of managing infrastructure.

Have existing AI and analytics architectures? No problem. HPE Data Fabric Software can ingest external data sources, allowing organizations to leverage existing data lakes or warehouse architectures, whether cloud based or on-premises, all under a unified and secure umbrella. As your data continues to grow, the GNS automatically scales to deliver smooth performance without downtime or manual intervention.

What makes HPE Data Fabric Software unique



1. **Transparent and secure access for AI, analytics and data lakes**
    a. **Comprehensive solution:** HPE Data Fabric Software provides a complete solution for data management and analytics needs, without relying on third-party integrations
    b. **Unified data view:** Data Fabric provides a unified view of data across multiple storage systems, simplifying data management and reducing complexity.
    c. **Bundled Analytics Tooling:** Analytics tooling is included with the platform, providing customers with real-time insights and decision-making capabilities
2. **Consistent data access and management across all data sets**
    a. **Unified data management:** Data Fabric's unified data management capabilities simplify the creation and execution of complex data pipelines, reducing latency and costs.
    b. **Simplified Integration:** Integrates with a wide range of tools and systems using NFS, POSIX, HDFS, and S3 support, making it easier to access and share data across the organization.
    c. **Scalable Infrastructure:** Grow the data lake house as needed, without worrying about performance or scalability issues
    d. **Software-Defined:** Data Fabric provides a software-defined platform that integrates with existing storage investments.
3. **Integrated governance, compliance, and data plane optimization**
    a. **Consolidated data and security**—HPE Data Fabric Software unifies data view and security across data landscape in a single platform.
    b. **Robust compliance**—Data Fabric helps organizations meet strict security standards and compliance requirements with robust access controls and encryption features.

While alternate solutions offer data management capabilities, HPE Data Fabric Software provides

Feature	HPE Data Fabric Software	Other solutions
Data federation	Unified **GNS** across storage platforms	Requires data ingestion into native platform formation
Hybrid cloud and multicloud	Fully deployable **on-premises, in the cloud, or at the edge**	Primarily cloud dependent
Multiformat support for modern AI and analytics workloads	**Apache Iceberg support**, real-time **query acceleration**	Limited support for multiformat AI workloads
Security and compliance	**End-to-end encryption, RBAC, and audit trails**	Requires third-party integrations for compliance
Storage optimization	**Tiered storage and intelligent data placement** for cost efficiency	Standard data storage models

Accelerate AI, enhance security, and streamline operations

HPE Data Fabric Software continues to lead the way in AI and analytics data management, addressing the growing challenges of data volume and complexity. The latest release introduces enhanced security and compliance features, streamlines operations for both on-premises and cloud environments, and accelerates AI and analytics initiatives. These enhancements include:

1. **Support PC-AI initiative through adoption of Apache Iceberg**—The new release is the data backbone of the HPE Private Cloud AI data lakehouse and provides an Iceberg interface for PC-AI users to data housed throughout their enterprise. This unified data layer allows data scientists to connect to external stores and query that data as Iceberg-compliant data without moving the data. Apache Iceberg is the emerging format for AI and analytical workloads. With this new release, HPE Data Fabric Software becomes an Iceberg end point for AI engineering. This makes it simple for AI engineering data scientists to easily point to the data lakehouse data source and run a query directly against it. The software can access the parquet files (open-source file format) in any file or object store, join them, and make them appear as Iceberg compliant stores. It takes care of metadata management, secure access, and joining files or objects across any source on-premises or in the cloud in the GNS.

2. **Integrating with HPE Alletra Storage MP X10000**—In this latest update, the HPE Data Fabric's Global Namespace has been expanded to support X10000. This enhancement provides organizations with even greater flexibility and efficiency in managing their data across diverse storage environments. Customers can now take advantage of heterogeneous read and write access to object systems as well as the option to integrate the X10000 as a high performance tier for advanced analytics projects.

3. **Expanding the hybrid data lakehouse vision** to deploy HPE Data Fabric Software easily in the cloud or at the edge in hardware-limited environments—It is a 100% software-defined platform and is highly configurable. Users can deploy the software easily through the UI or installer or even have the flexibility to build custom scripts like Chef, Bash, and such. Businesses can quickly deploy AI-ready data lakehouses anywhere without complexity.

4. **Integration with open-source tools**—HPE Data Fabric Software integrates with a collection of open-source tools. This enables users to perform analytics on data from any location. It reduces configuration time required to integrate and configure open-source tools along with improving productivity by preserving existing access mechanisms. We have expanded the set of tools including:

    a. **JWT authentication support** for secure access.

    b. **Apache Flink for real-time streaming analytics.**

    c. **Pre-integrated data connectors** to accelerate analytics

    This reduces time spent configuring and integrating third-party tools, improving developer productivity.

5. **New dashboards to expose metadata for governance and compliance**—HPE Data Fabric Software collects audit data from the GNS and makes it accessible as Iceberg-compliant parquet data. This allows any Iceberg-compliant query, dashboarding, or AI algorithm access to compliance data from the software. In this release, we include a dashboard providing out-of-the-box insights to compliance data through the software's user interface. This provides organizations with **end-to-end visibility, security, and compliance monitoring**—without additional tooling.

Summary

HPE Data Fabric Software offers a comprehensive and unified data platform that empowers organizations to overcome siloed data and accelerate AI initiatives. By addressing the challenges of data fragmentation, helping ensure data quality, and providing scalable and secure infrastructure, the software enables businesses to derive higher value from their data. With its ability to streamline data management, optimize resource utilization and enhance security, HPE Data Fabric Software is the ideal solution for organizations seeking to realize the full potential of AI.

Learn more at

HPE.com/datafabric

Visit **HPE.com**

Chat now (sales)

Hewlett Packard Enterprise

a00130639ENW, Rev. 7