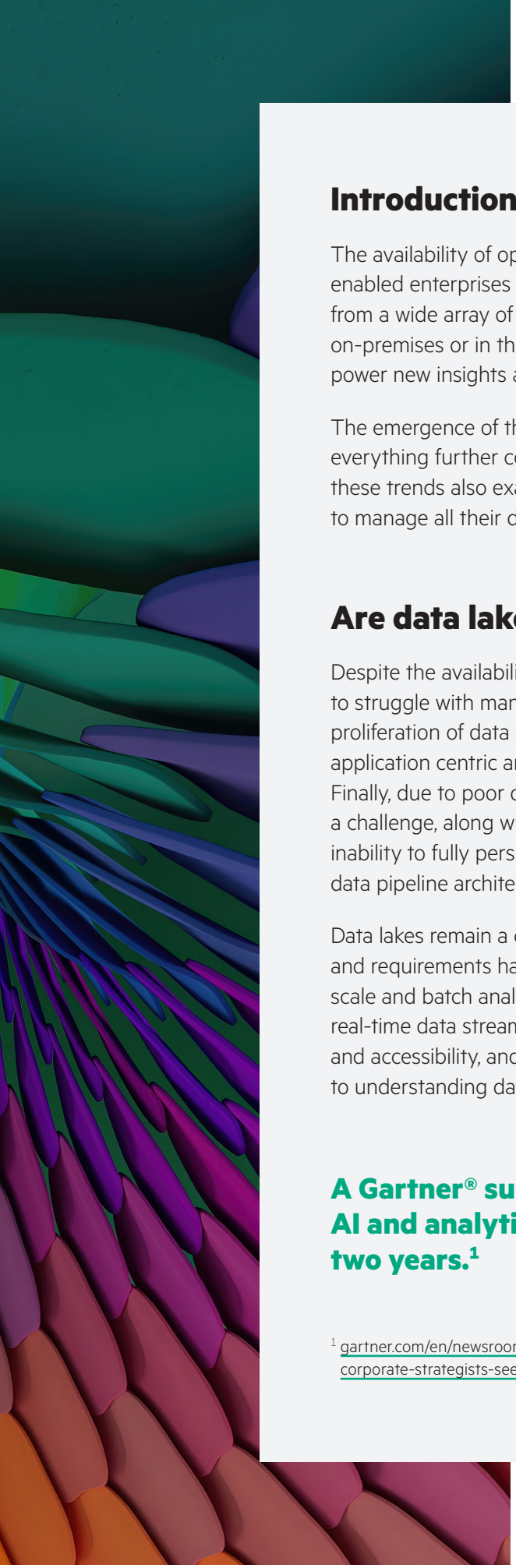
Solution brief

HPE Alletra Storage

The optimal foundation for a modern data lake

HPE Alletra Storage Servers

HPE GreenLake



Introduction

The availability of open-source frameworks and distributed file systems has enabled enterprises to integrate and utilize structured and unstructured data from a wide array of sources. Better equipped to create and manage data lakes on-premises or in the cloud, enterprises can now enable analytics on their data to power new insights and accelerate business decisions.

The emergence of the Internet of Things (IoT) and the instrumentation of everything further contributes to the exponential growth of data lakes. However, these trends also exacerbate enterprises' challenges as they search for better ways to manage all their data and use it efficiently to achieve critical business objectives.

Are data lakes still relevant?

Despite the availability of data lakes, data challenges persist. Enterprises continue to struggle with managing exponential data growth, which often results in the proliferation of data silos with little or no governance. Meanwhile, the legacy of application centric architectures inhibits the move to data-centric architectures. Finally, due to poor or missing metadata management, untrusted data is also a challenge, along with a lack of adequate governance and oversight, and the inability to fully persist and use data coming in off event streams due to inadequate data pipeline architectures.

Data lakes remain a crucial tool to answer these challenges, but the objectives and requirements have expanded. What was originally designed only for extreme scale and batch analytics must now support varying performance levels. It includes real-time data streaming and analytics, machine learning, horizontal data movement and accessibility, and the ability to store and catalog data securely, which is critical to understanding data lineage. The answer is a modern data lake.

A Gartner® survey finds 79% of corporate strategists see AI and analytics as critical to their success over the next two years.[1]

[1] gartner.com/en/newsroom/press-releases/2023-07-05-gartner-survey-finds-79-percent-of-corporate-strategists-see-ai-and-analytics-as-critical-to-their-success-over-the-next-two-years

What is a modern data lake?

Early data lakes for Hadoop were constructed using standardized models and symmetric node profiles, with the primary objective being quick and easy storage of structured and unstructured data at any scale while supporting batch analytics on top of this data. Initially, that worked, but with the introduction of new workloads and the need for real-time streaming and analytics, there was a demand for more flexibility and elasticity in the building blocks used to create data lake architectures.

A modern data lake must be built on infrastructure that can optimize performance, capacity, and cost while being extremely scalable, modular, reliable, and highly available. The infrastructure should support a variety of analytic engines, workloads, software defined storage (SDS), and interfaces such as S3 API, Hadoop Distributed File System (HDFS), NFS, and SMB. With these diverse requirements, many enterprises have begun adopting object storage, either cloud based or on-premises, as an alternative or a supplement to HDFS.

A new data management paradigm

AI and analytics are at the core of data-driven organizations and digital transformation. Thus, the modern data lake and the data management paradigm must focus on optimizing data accessibility for analytics and artificial intelligence (AI) applications while considering:

- Data lake architectures must flexibly support heterogeneous hardware environments, including hardware accelerators such as GPU and persistent memory (PMEM), as well as fast storage.
- Data mobility, no longer defined as simply from primary to secondary storage, must include the ability to persist data across the data pipeline, from the data lake to the optimal datastore (for example, when you want to save the results from model testing).

HPE Alletra Storage Servers are density and performance-optimized building blocks for the modern data lake

Constructing a modern data lake demands flexible infrastructure building blocks to support a variety of workloads, a centralized data repository, larger archives, and backup repositories, ideally using cost-effective storage capacity and leveraging dense architectures for smaller form factors. HPE Alletra Storage Servers meet all these requirements. They're architected to accommodate both ends of the data-centric workload spectrum, from deeper data lakes and archives to performance sensitive machine learning (ML), data analytics, hyperconverged infrastructure (HCI), and cache-intensive workloads.

HPE Alletra Storage Servers 4110 and 4120 support the latest technology and lead in performance with all-NVMe flash or hybrid models. They are designed for demanding, throughput-intensive workloads that can cache data and support flexible storage tiers while delivering a superior, balanced system architecture with fast I/O in a compact and dense chassis.

HPE Alletra Storage Server 4140 is a density optimized, high-throughput, bulk capacity platform in a 4U chassis ideal for deep data lakes. Unlike competitive products, HPE Alletra Storage Server 4140 supports NVMe drives, including advanced EDSFF flash media, easy drive access via front drawers, and low cost/GB density.

What's new with HPE Alletra Storage Server

HPE Alletra Storage Servers offer advanced capabilities for the most demanding data lakes and analytics workloads. They are available in an incredibly dense and all-NVMe flash model in 1U (HPE Alletra Storage Server 4110) or an NVMe hybrid model in a compact 2U that delivers affordable performance (HPE Alletra Storage Server 4120).

Figure 1. HPE Alletra Storage Server 4110

HPE Alletra Storage Server 4110 is all-NVMe and supports up to 20 new E3.s format EDSFF for performance and advanced thermals or SFF NVMe SSDs for outstanding performance and density.

Figure 2. HPE Alletra Storage Server 4120

HPE Alletra Storage Server 4120 is NVMe hybrid and is available as a 24 LFF disk with 12 EDSFF, or 6 SFF or 4 LFF in the rear, or in a larger chassis with 48 SFF disks with 12 EDSFF or 6 SFF in the rear, delivering greater capacity for deep data lakes.

Figure 3. HPE Alletra Storage Server 4140

HPE Alletra Storage Server 4140 is high-throughput bulk storage available in a base 68 LFF or expanded 92 LFF chassis with optional rear chassis with up to 8 EDSFF or 4 NVMe SFF for caching performance.

Conclusion

HPE Alletra Storage Servers provide storage-optimized, purpose-built solutions for modern data lakes and data pipelines, offering flexibility, elasticity, and performance tiers to accommodate a wide array of storage intensive workloads.

Learn more at

HPE.com/us/en/storage/Alletra-4000.html

Visit **HPE GreenLake**

Chat now (sales)

a50004344ENW, Rev. 1

Hewlett Packard Enterprise