Solution Brief

NetApp ONTAP SAN Solution for Hadoop

Speed time to insights, scale without sacrifice, and enable hybrid cloud deployments

Key Features

Lower Five-Year Cost of Operations by Up to 52%

- Independently scale storage and compute.
- Lower initial hardware and software investment by up to 61%.
- Dramatically shrink dataset space requirements with patented storage efficiency technologies.
- Reduce data copies from three to two.

Increase Business Agility with Seamless Hybrid Cloud Deployments

- Move data seamlessly anywhere in the Data Fabric.
- Lower expenses by automatically moving aging data to more cost-effective storage.
- Leverage cloud compute for analytics while keeping your data safely on the premises.

Maximize Availability for Greater Insight and ROI

- Leverage enterprise-grade backup, restore, and compliance capabilities.
- Maintain performance and data availability during controller failures.

Increase Efficiency of IT Infrastructure Staff by Up to 52%

- Unleash the power of your data with NetApp® OnCommand® software.
- Reduce IT support costs with a single interface management view of all global data repositories.
- Back up, restore, and replicate terabytes of data in seconds.

The Challenge

Harness the power of big data

Organizations collect and analyze increasingly large amounts of raw data, such as point-of-sale data, credit card transactions, log files, and machine and security data. IT leaders and analytics teams are under tremendous pressure to harness, analyze, and leverage this data to maximize the business value across their organizations.

Apache Hadoop and its growing ecosystem of products enable organizations to extract valuable insights from large volumes of diverse data that cannot be analyzed with relational databases. With these insights, people across the organization can ask the right questions and get better answers, supporting more informed decisions that help promote business transformation.

However, because initial Hadoop deployments often rely on commodity servers with internal drives, infrastructure resilience and agility issues prevent organizations from realizing the full benefits of their Hadoop deployment. For example, a single disk failure can degrade performance of the entire cluster. Managing disk replacements is continual and error-prone. In addition, triple file replication and failure redistribution models increase network costs and complexity.

The Solution

NetApp ONTAP SAN solution for Hadoop

The NetApp ONTAP® SAN solution for Hadoop features enterprise storage that is independent of the compute servers to offer an enterprise-class deployment with lower cluster downtime, higher data availability, and linear scalability. This solution leverages the robust and proven features of NetApp ONTAP 9 software, the enterprise data management software that powers the NetApp engineered systems of All Flash FAS (AFF) and FAS, as well as software-only ONTAP Cloud. ONTAP supports independent scaling of compute and storage resources, reducing costs and maximizing utilization of compute resources. In addition, ONTAP provides integrated data protection to safeguard operations with near-instant backup and recovery using the highly efficient NetApp Snapshot™, SnapRestore®, and SnapCenter® technologies.

The Data Fabric from NetApp weaves together data storage regardless of location—on the premises, in the near cloud, or in a private or public cloud—into a unified data system, enabling tremendous flexibility. Data moves seamlessly wherever needed within the hybrid cloud, supporting data security and governance, resiliency, and data management efficiency. A unified management console provides a single, global view of all data repositories. ONTAP supports multitenancy and adaptive quality of service (QoS), allowing Hadoop to run on the same enterprise-grade IT storage infrastructure already used for traditional core applications, including Oracle and SAP.

NetApp®

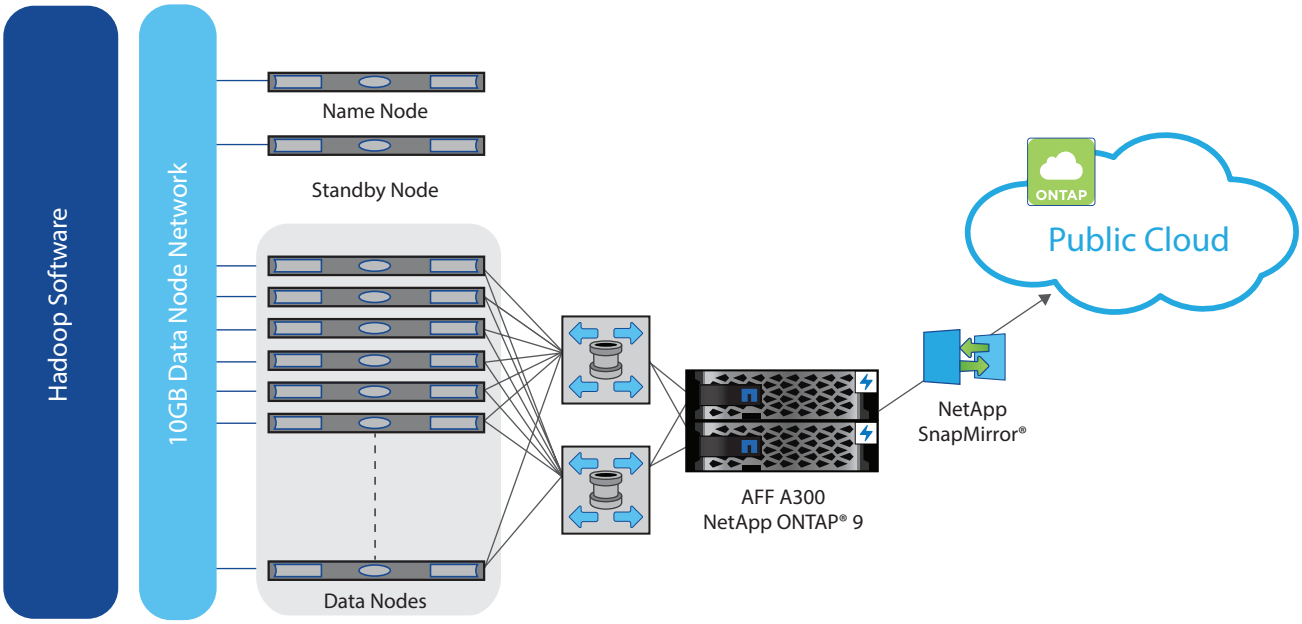

Figure 1) NetApp analytics in the hybrid cloud with Data Fabric.

Reduce operational costs by up to 52%

A big data platform based on ONTAP software enjoys significant cost advantages compared to the traditional server-based storage. ONTAP solutions require far fewer and less costly servers (with fewer storage bays) and consequently fewer OS and application licenses. These efficiencies can lower initial hardware and software investments by up to 61% and five-year cost of operations by up to 52%. In addition, data is protected by the ONTAP system, which allows the replication count for HDFS to be reduced from three to two. This reduction in the number of data copies plus the data reduction provided by deduplication and compression enable ONTAP to store the same data with a much smaller footprint.

Increase business agility with seamless hybrid cloud deployments

The big data platform can take on many forms, depending on the needs of each enterprise. With ONTAP, choosing between an on-premises installation, one in the near cloud, or one in a private or public cloud is not necessary. ONTAP supports all options, whether a single option is chosen or more typically any combination of options.

Place your analytics compute tier on cloud architectures such as Amazon EC2 while keeping your data safely on storage that you control on the premises. The decoupled design allows independent scaling of compute and storage layers. This capability provides the flexibility to add storage capacity without adding compute nodes. With this design, only two data copies are required, unlike HDFS deployed with server-based DAS, which requires three copies of data. In addition, the compute resources used by a Hadoop cluster in a public cloud can be shut down when not used, because the data is stored in NPS. This configuration can result in a huge cost savings on cloud resources.

Reduce storage requirements

Many enterprises choose to locate the core system on the premises, leveraging AFF for hot data, while using hybrid FAS (a combination of flash and HDD) for the warm and cold tiers. With most traditional enterprise applications also located in on-premises data centers, Hadoop can access data stored there without duplicating the data into a dedicated data lake. With the Data Fabric, the big data infrastructure can span data centers in multiple on-premises locations if required. Other choices include NetApp Private Storage (NPS), a near-cloud offering, or one or more public clouds.

Maximize availability for greater insight and ROI

Storage media is the most fault-prone component in the Hadoop architecture. It is simply a matter of time before a disk fails. The larger the cluster, the more likely it is that a disk failure will occur. In an HDFS implementation, a disk failure initiates a restart of a job task, which negatively affects the predictability of the job completion time.

With the NetApp ONTAP SAN solution for Hadoop, if a disk fails, the RAID DP® protection in AFF eliminates the need for job task restarts, thereby maintaining the predictability of job run times. ONTAP allows many critical operations to take place without disruption. These operations include assigning, promoting, and retiring storage resources with zero downtime. Unplanned failures of controllers and media can also be dealt with while all applications continue to operate normally.

Leverage multitenancy

ONTAP fully supports multitenancy, making it possible to share storage for the big data platform with traditional enterprise applications without compromise. By sharing a consolidated infrastructure with multiple tenants and workloads that have different performance, capacity, and security requirements, enterprises can realize substantial savings of time and money. ONTAP also creates secure partitions within the cluster, allowing each to be governed by specific rights and permissions.

Enable linear scalability

In Hadoop deployments that use commodity servers with internal storage, performance is negatively affected as more data nodes are added. Because the NetApp ONTAP enterprise shared storage design separates compute from storage, capacity and performance can scale independently. Servers are only added when more compute resources are required, and performance scaling is linear.

Increase efficiency of IT infrastructure staff by up to 52%

ONTAP provides a complete and powerful set of tools to manage the enterprise data infrastructure. The OnCommand management software family covers the entire range of data management activities. NetApp OnCommand Unified Manager provides a single interface view of the entire data infrastructure, no matter where the data is located. It allows you to easily monitor the health, availability, capacity, performance, and data protection status of your NetApp clustered storage and provides alerts and vital information for proactive management. The OnCommand Performance Manager is tightly integrated, delivers comprehensive storage performance monitoring, and assists in proactive management.

NetApp technologies such as NetApp Snapshot and SnapMirror® help create efficient backup and replication of Hadoop data. ONTAP also helps provide rapid NetApp FlexClone® thin-cloning technology to support DevOps.

About NetApp

NetApp is the data authority for hybrid cloud. We provide a full range of hybrid cloud data services that simplify management of applications and data across cloud and on-premises environments to accelerate digital transformation. Together with our partners, we empower global organizations to unleash the full potential of their data to expand customer touchpoints, foster greater innovation and optimize their operations. For more information, visit www.netapp.com. #DataDriven