NetApp Solution for In-Place Analytics

Run analytics using existing data stored on NFS-based systems and data in the hybrid cloud

Key Benefits

Reduce Operational Costs
- Run analytics natively on NFSv3 data without adding infrastructure
- Use a single storage back end to manage data for both traditional and analytics workloads

Enable Hybrid Cloud Deployments
- Place your analytics compute tier in the cloud while keeping your data safely on storage that you control on the premises
- Use independent scaling of compute and storage

Manage Enterprise Data
- Back up and replication of Hadoop data efficiently
- Clone data rapidly and space efficiently

Maintain Flexibility and Choice
- Benefit from Hortonworks certification and compatibility with Hortonworks Data Cloud
- Support Apache Ranger™, YARN, MapReduce, Spark, HBase, Pig, and Hive

The Challenge
Harness the power of big data

Apache Hadoop and its growing ecosystem of products enable organizations to extract valuable insights from large volumes of diverse data. 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.

Historically, Hadoop has been used primarily on incoming, external data; however, there's been a need to use Hadoop on existing data, typically stored in network-attached storage (NAS). Until now, this required setting up another Hadoop Distributed File System (HDFS) storage silo to host the data in order to run the Hadoop analytics. This resulted in additional data management, more inefficiencies, and additional costs of moving the data between NAS and HDFS.

The Solution
In-place analytics with NetApp

You can use NetApp® in-place analytics to run analytics on NFSv3 data without moving the data, creating a separate analytics silo, or setting up a separate HDFS cluster. You can switch from HDFS to NFS or run NFS alongside HDFS. Different Hadoop services are supported, including Apache Ranger, YARN, MapReduce, Spark, HBase, Pig, and Hive, as well as being certified by Hortonworks for Hortonworks Data Platform.

These capabilities mean that you can support many types of workloads: batch, in-memory, streaming, and more. An in-place analytics configured cluster can support S3 targets such as NetApp StorageGRID® or as Amazon Simple Storage Service (Amazon S3) using the Hadoop S3A connector.

Hybrid Cloud Deployments

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 or in NetApp Cloud Volumes Service. 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, just one copy of data is required, unlike HDFS, which requires three copies of data. In addition, the compute resources in a Hadoop cluster in a public cloud can be shut down when not used, because the data is stored in NetApp Cloud Volumes Service.
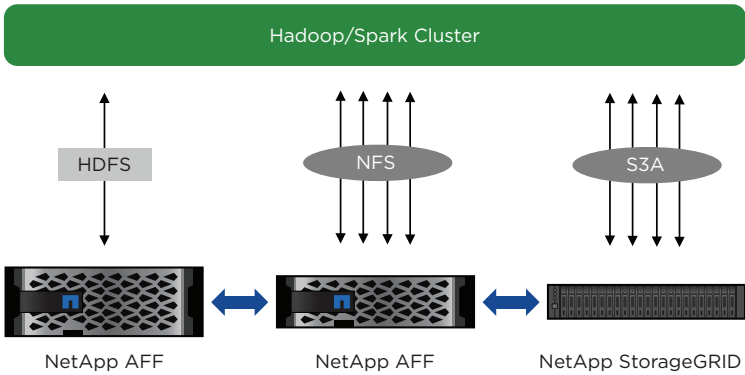
NetApp®

Figure 1) NetApp in-place analytics.

Enterprise Data Management

NetApp ONTAP® data management capabilities from technologies such as NetApp Snapshot™ and SnapMirror® help create efficient backup and replication of Hadoop data. ONTAP also provides rapid NetApp FlexClone® thin-cloning technology.

Succeed with Big Data Analytics

From business owners and consumers of big data insights to data professionals, developers, and administrators, proven NetApp solutions for Hadoop can help everyone in your organization succeed with big data analytics.

If you need help in designing or deploying your NetApp solutions for Hadoop, NetApp Services experts and our certified partners can help.

Learn more about NetApp solutions for Hadoop at http://www.netapp.com/us/solutions/big-data/hadoop.aspx.

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
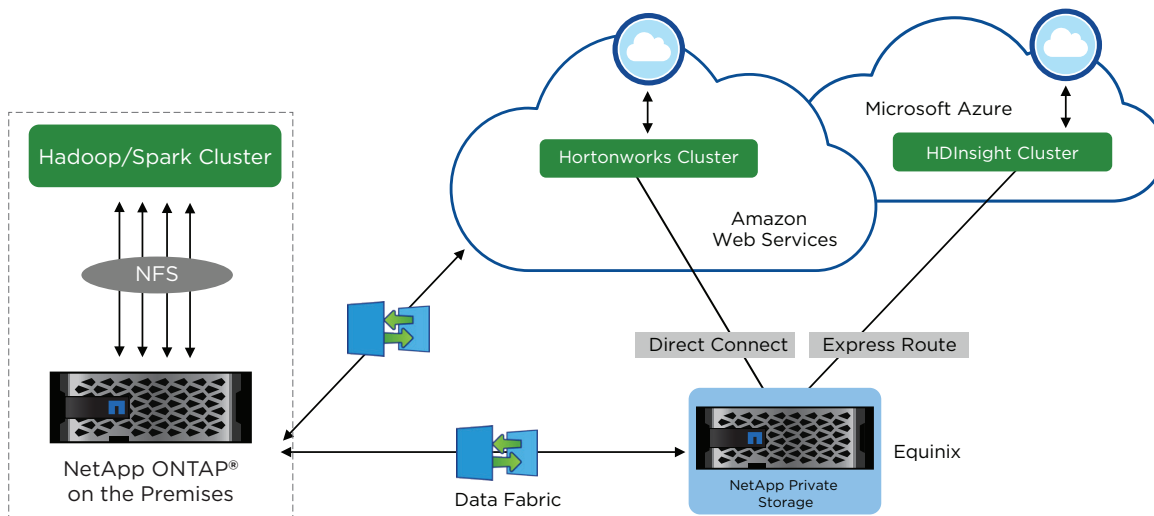

Figure 2) NetApp analytics in the hybrid cloud with Data Fabric.