Data lake to ONTAP NFS

NetApp Solutions

NetApp
September 23, 2021

Table of Contents

Data lake to ONTAP NFS 1
Customer challenges and requirements 1
Data mover solution 1

Data lake to ONTAP NFS

Previous: Customer scenarios.

This use case is based on the largest financial customer proof of concept (CPOC) that we have done. Historically, we used the NetApp In-Place Analytics Module (NIPAM) to move analytics data to NetApp ONTAP AI. However, because of recent enhancements and the improved performance of NetApp XCP as well as the unique NetApp data mover solution approach, we reran the data migration using NetApp XCP.

Customer challenges and requirements

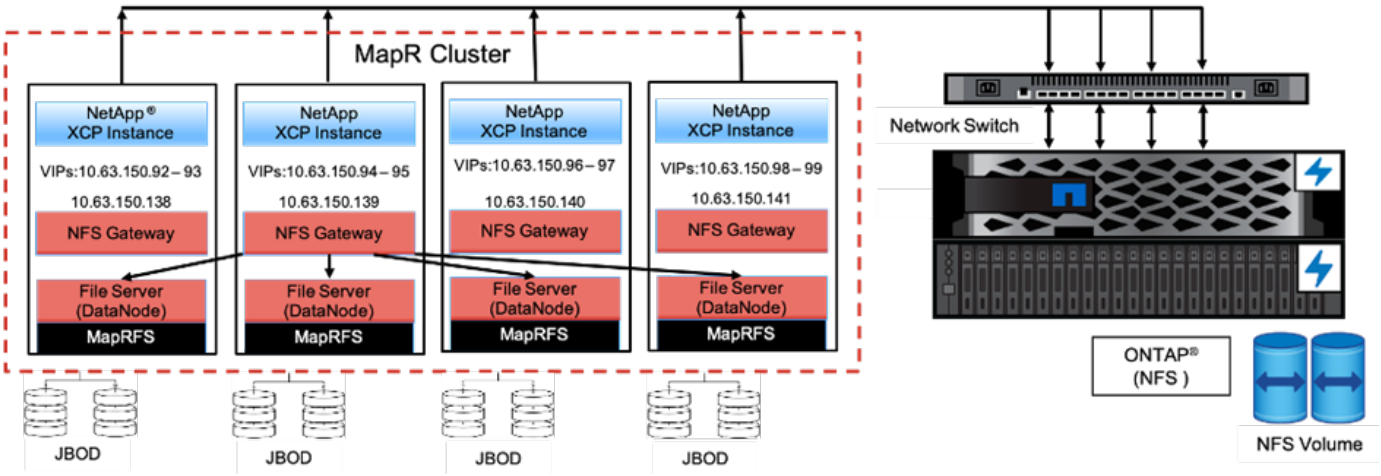
Customer challenges and requirements that are worth noting include the following:

- Customers have different types of data, including structured, unstructured, and semistructured data, logs, and machine-to-machine data in data lakes. AI systems require all these types of data to process for prediction operations. When data is in a data lake-native file system, it is difficult to process.
- The customer's AI architecture is not able to access data from Hadoop Distributed File System (HDFS) and Hadoop Compatible File System (HCFS), so the data is not available to AI operations. AI requires data in an understandable file system format such as NFS.
- Some special processes are required to move data from the data lake because of the large amount of data and high-throughput, and a cost-effective method is required to move the data to the AI system.

Data mover solution

In this solution, the MapR File System (MapR-FS) is created from local disks in the MapR cluster. The MapR NFS Gateway is configured on each data node with virtual IPs. The file server service stores and manages the MapR-FS data. NFS Gateway makes Map-FS data accessible from the NFS client through the virtual IP. An XCP instance is running on each MapR data node to transfer the data from the Map NFS Gateway to NetApp ONTAP NFS. Each XCP instance transfers a specific set of source folders to the destination location.

The following figure illustrates the NetApp data mover solution for MapR cluster using XCP.



For detailed customer use cases, recorded demos, and test results, see the Using XCP to Move Data from a Data Lake and High-Performance Computing to ONTAP NFS blog.

For detailed steps on moving MapR-FS data into ONTAP NFS by using NetApp XCP, see Appendix B in TR-4732: Big Data Analytics Data to Artificial Intelligence.

Next: High-performance computing to ONTAP NFS.