Healthcare Association Chooses NetApp for Big Data Analytics

Customer Profile

The organization is a national federation of independent, community-based, locally operated healthcare insurance companies. With access to data from nearly 105 million people, the company is focused on transforming its robust data into rich insights and meaningful analytics that change how healthcare is delivered and paid for in the U.S.

Customer Situation

To accelerate its movement toward smart, affordable, data-driven healthcare, the organization embarked on a three-part initiative that included determining how to best use company and partner data, assessing new capabilities to analyze data, and using data insights to increase customer retention.

The company chose a NetApp Solution for Hadoop to help harness the power of its data. However, concerns about reliability and scalability prompted the company to request a proof of concept (POC) before fully committing to the solution.

Proof of Concept

The NetApp team quickly configured a 200TB POC. Performance and failover tests run during the POC gave the customer confidence in the NetApp Solution for Hadoop. The POC also helped the customer better prepare for larger amounts of data that would run in production.

POC Testing Results

- **Better performance.** The NetApp E-Series solution performed twice as fast on ingest and MapReduce jobs compared to the speed of internal drives.
- **Higher availability.** The E-Series solution had less downtime than whitebox servers with internal drives. Even with failed DataNodes, service remained uninterrupted during storage failover. NameNodes recovered faster. A key advantage of the E-Series is the ability to hot-swap disk drives while the cluster is running.

Why NetApp

- **Improved efficiency.** Two copies of data instead of three (Hadoop default) reduces the capacity needed and accelerates application performance—without compromising data availability.
- **Easier deployment**. Modular design of the reference architecture makes it simple to add nodes and expand, especially for storage-intensive workloads. SANtricity software enables management tasks to be performed while the storage remains online, so configuration changes, maintenance, or storage capacity expansion can be made without disruption.
- **Lower cost.** The customer is on track for potential savings of $2.1M in capex over five years, compared to the cost of whitebox servers with internal drives.

IN THE SPOTLIGHT

Customer
Healthcare Association, U.S.

Solution Goal
"We needed a reliable, fast, and cost-effective Hadoop solution for our new data initiatives. In the proof of concept, NetApp's Hadoop solution proved to be faster and more reliable than the commodity server approach. The 200TB proof of concept went smoothly and we have already started making progress toward the goals of our data initiatives."

NetApp Products and Services
- NetApp® E5560 storage array
- NetApp SANtricity®
- NetApp FAS2540
- NetApp FAS3250 (Disaster Recovery)
- Clustered Data ONTAP®

Environment
- File System: Hortonworks Data Platform (Hadoop)
- Cluster: 4 Hadoop DataNodes
- OS: RHEL 6.2
- Server Platform: HP
- Data Storage: NetApp E5560
- NameNode Storage: FAS2540 and FAS3250
- Network: Brocade 5100 switches
- Protocols: HDFS, SAS

NetApp®