Success Story

NetApp Accelerates AutoSupport Analytics with NetApp Open Solution for Hadoop

KEY HIGHLIGHTS

Industry
Data storage

The Challenge
Contend with growing volumes of customer storage data that created IT complexity and processing challenges.

The Solution
Deploy NetApp® Open Solution for Hadoop to drive superior AutoSupport™ operations.

Benefits
- Reduced database query on 24 billion records from 4 weeks to less than 10.5 hours, accelerating the team's ability to respond to customer needs
- Enabled team to conduct previously impossible database query on 240 billion records in less than 18 hours

Customer Profile

NetApp creates innovative storage and data management solutions that deliver outstanding cost efficiency and accelerate business breakthroughs. NetApp Global Support provides cost-effective support scaled and priced to meet the diverse needs of NetApp storage customers, whether large enterprises, classified government agencies, or midsize organizations.

A component of NetApp Global Support services, NetApp SupportEdge Services, help maintain system health and optimize performance for NetApp customers. The majority of NetApp customers leverage advanced AutoSupport monitoring and reporting functions.

The Challenge

Processing massive volumes of data

NetApp AutoSupport provides a vital service to NetApp end-user customers determined to optimize the performance and utilization of their storage solutions while maintaining the health of their storage environments. Customers leverage the My AutoSupport portal on the NetApp Support site for proactive systems management capabilities and insight into storage configuration, capacity, utilization, efficiency, and health-check information. Storing, managing, and analyzing customer storage system information are at the core of all of the services that AutoSupport provides.

Unstructured log and system diagnostic information is transmitted to the NetApp Technical Support Center from customer sites, with over 600,000 incoming data transactions weekly, resulting in rapidly expanding storage requirements for the AutoSupport team. Approximately 40% of the weekly incoming data is transmitted during an 18-hour period each weekend, creating the potential for I/O bottlenecks that could affect service-level agreement (SLA) windows.

As the NetApp end-user customer base continues to grow, AutoSupport data is growing at approximately 7TB per month, with related storage requirements on track to double every 16 months. The AutoSupport team proactively identified the need to upgrade its storage environment in order to tackle its big data challenges and accommodate continued growth.

"Running the NetApp Open Solution for Hadoop gives us the ability to turn an unwieldy data explosion into a highly manageable environment. It also will allow us to perform deeper analytics than before, which will provide better monitoring and troubleshooting of NetApp customer storage systems."

Marty Mayer
Senior Manager, AutoSupport, NetApp Global Support

"NetApp AutoSupport data is extremely valuable to us," says Marty Mayer, senior manager, NetApp AutoSupport. "Our customers depend on us to utilize the data to respond in a timely manner when potential problems and issues arise. Additionally, we actively analyze customer storage system data for fitness and health checks that help optimize the investment our customers have made in NetApp." To continue to deliver exceptional support, AutoSupport needed a big data storage solution that would allow it to store, manage, and analyze growing stores of unstructured data; enrich its understanding of these large, complex datasets; and scale for growth.

The team focused on storage solutions that support the Apache Hadoop open-source software designed for data-intensive distributed applications. Other storage criteria included scalability, high performance, and rich analytics capabilities. AutoSupport conducted a proof of concept on numerous technologies, evaluating them for key functions including parsing; extract, transform, and load (ETL); and data warehousing. Following in-depth due diligence, the team concluded that the NetApp Open Solution for Hadoop surpassed the other solutions in providing the ability to quickly process massive datasets for deeper insight into customer storage environments. The solution also provided an overall lower total cost of ownership.

The Solution

Delivering customer service applications 24x7

The NetApp AutoSupport team deployed a NetApp Open Solution for Hadoop, which includes a 20-node Hadoop cluster on 4 NetApp E2600 storage systems and a NetApp FAS2040 system. The team chose to leverage Hadoop for its high-performance parallel data processing and reliable data storage provided by the NetApp Open Solution for Hadoop.

"The NetApp Open Solution for Hadoop system offers us the scalability and flexibility we need to effectively support our growing client base and rapidly expanding data stores," says Mayer. "In addition, because the NetApp system addresses our parsing, ETL, and data warehousing needs in a single, comprehensive solution, it reduces our total cost of ownership, freeing up budget for other customer-focused projects."

The NetApp Open Solution for Hadoop offers high availability and high performance for even the most demanding AutoSupport workloads. The solution's balanced performance will sustain the high read-and-write throughput requirements of the system's data-intensive, high-bandwidth applications such as the weekend reporting that offers visibility into the health of hundreds of thousands of customer storage systems.

As a customer-facing organization, the NetApp AutoSupport team depends on the reliability of its storage environment 24x7x365. Data ONTAP® 8 on the FAS2040 storage system eliminates the single point of failure common in traditional Hadoop clustered deployments and instead offers full redundancy and automated path failover, along with online administration.

Business Benefits

High performance for big bandwidth applications

AutoSupport predicts significant performance gains with Hadoop running on its NetApp Open Solution for Hadoop. By supporting even the most bandwidth-intensive applications, the solution will enable the NetApp AutoSupport team to help meet stringent SLAs for parsing and loading data.

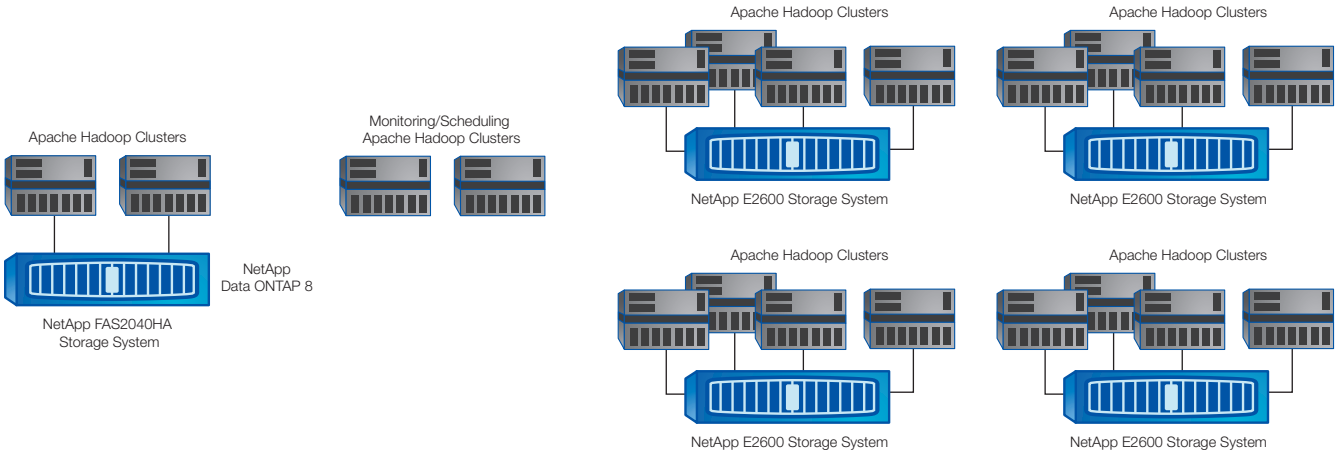

Figure 1) The NetApp AutoSupport team deployed a NetApp Open Solution for Hadoop, which includes Hadoop clusters on NetApp E2600 storage systems and a NetApp FAS2040 system running Data ONTAP 8.

The team has proven the performance of running queries on 24 billion records on the NetApp Open Solution for Hadoop. Loading the data and processing the queries through the data warehouse previously took four weeks to complete. The team can now load the expansive unstructured data into the Hadoop clusters on NetApp, producing query results in just 10.5 hours.

"The productivity and customer service benefits enabled by the NetApp solution are significant," says Mayer. "Running the NetApp Open Solution for Hadoop gives us the ability to turn an unwieldy data explosion into a highly manageable environment. It also will allow us to perform deeper analytics than before, which will provide better monitoring and troubleshooting of NetApp customer storage systems."

Another NetApp AutoSupport query met bandwidth hurdles, given the growing complexity and scale of customer storage data. The 240-billion-record query used for evaluating the performance of storage drives across the NetApp installed base exceeded the performance capabilities of the former storage infrastructure, making it impossible to process. Through performance testing on the NetApp Open System for Hadoop, the team was able to run the high-bandwidth query and achieve results in less than 18 hours.

Rich analytics for large datasets

AutoSupport provides skilled resources focused on supporting NetApp customers through analyzing log data for insight into system health. The NetApp Open Solution for Hadoop system that supports Hadoop workloads enables parallel processing for structured and unstructured data and allows AutoSupport applications to work with thousands of nodes and petabytes of data. "This highly efficient Hadoop processing will enable our AutoSupport team to quickly mine hundreds of terabytes of data, for real-time analysis of customer system event and performance data and rapid resolution of any issues," says Mayer.

Weekly AutoSupport logs composed of large, complex datasets offer the team insight into information such as storage capacity trending by customer, country, and other criteria. The team also runs ad-hoc queries such as investigating an error alert across an entire system set to identify the source, allowing the team to take proactive measures to prevent impact to customer systems. In addition, through automatic AutoSupport system analysis of deep pools of data, the NetApp Technical Support team is immediately notified of critical alerts. The NetApp Open Solution for Hadoop is designed to accelerate these and other AutoSupport queries and processes.

NetApp customers can also benefit directly from the high performance of the NetApp Open Solution for Hadoop system, which offers deeper analytics capabilities than before. With more in-depth visibility into configuration and system health data through the My AutoSupport portal, customers can be more proactive and achieve greater efficiencies in their NetApp storage environments.

Managing a holistic storage environment

The AutoSupport team will efficiently manage its big data environment centrally on the NetApp Open Solution for Hadoop system. The solution will provide high levels of throughput and scalability to meet the growing data demands and intensive performance requirements of AutoSupport applications. Storage managers can efficiently manage huge stores of data in the integrated, unified AutoSupport storage environment, which helps NetApp customers optimize the performance and utilization of their storage solutions.

“By design, the NetApp Open Solution for Hadoop offers tight integration between our Hadoop applications and NetApp storage, thereby allowing us to maximize storage utilization and reduce our capex and opex expenditure.”

Kumar Palaniappan
Enterprise Architect, NetApp

“The NetApp Open Solution for Hadoop is a holistic rather than point solution,” says Kumar Palaniappan, enterprise architect, NetApp. “By design, the NetApp solution offers tight integration between our Hadoop applications and NetApp storage, thereby allowing us to maximize storage utilization and reduce our overall capex and opex expenditure. Hadoop analytics will empower us to transform data specific to NetApp storage systems into business insight for our team and our customers.”

SOLUTION COMPONENTS

NetApp Products
NetApp Open Solution for Hadoop
E2600 storage systems
FAS2040 HA storage systems
Data ONTAP 8.0

Protocols
NFS
SAS

NetApp

www.netapp.com

NetApp creates innovative storage and data management solutions that deliver outstanding cost efficiency and accelerate business breakthroughs. Discover our passion for helping companies around the world go further, faster at *www.netapp.com*.

Go further, faster®