



NetApp®

# Alliance Solution Brief

## High-Performance Architecture for Large-Scale Data Analytics



**KEY BENEFITS**

**Improves on the Basic Apache Hadoop Architecture**

- Improves storage availability, scalability, performance, and manageability
- Improves fault tolerance

**Increases Network Speed and Efficiency**

- Supports 10Gb/s converged Ethernet and beyond
- Supports provisioning of the network to carry flows such as storage, data, or voice with unique prioritization and quality-of-service profiles
- Improves application performance across the network

**Flexible and Efficient Systems with Dense, Scalable Storage Modules**

- Scalable modules with 180TB per 4U allow precise scaling and growth

### The Data Analytics Challenge

We are in the midst of a data storm. Increased business process automation, larger scientific datasets, widespread adoption of social-networking applications, and the need to provide communications services anywhere, anytime on smart mobile devices are all contributing to skyrocketing volumes of unstructured data.

To address these challenges, many IT departments are using Apache™ Hadoop™, an open source software framework that processes large datasets in a parallel, distributed, environment over a server cluster. Hadoop is designed to scale from a single server to thousands of servers with a high degree of fault tolerance controlled by the software layer. However, the basic Hadoop architecture brings two limitations:

- Data protection is handled in software. Every time a dataset is written, the Hadoop File System makes two more copies in case a disk drive or data node fails. This reduces disk utilization capacity and network throughput performance.

- The Hadoop metadata repository is a single point of failure.

Alcatel-Lucent and NetApp developed a platform architecture using Cloudera Distribution of Hadoop, a leading Hadoop distributor, to address the two major limitations of the basic Hadoop architecture while improving manageability and efficiency.

### The Solution

#### Alcatel-Lucent and NetApp Open Solution for Hadoop architecture

The Alcatel-Lucent and NetApp Open Solution for Hadoop architecture combines the power of the Hadoop framework with NetApp storage, Alcatel-Lucent OmniSwitch™ LAN switches, and enterprise-class commodity compute resources to deliver a resilient, scalable, and efficient solution for high-performance data centers.

#### NetApp Open Solution for Hadoop Architecture Benefits

Improves fault tolerance

- NetApp FAS2040 with attached network file system (NFS) improves fault tolerance if the primary name node server fails. This is in contrast to the

basic Hadoop architecture, in which the primary name node server is a single point of failure.

Improves storage

- NetApp E2660 storage array with swappable disks and redundant array of independent disks (RAID) technology improves storage availability, scalability, and manageability. This is in contrast to the basic Hadoop architecture, in which internal hard drives impede scalability and availability.

Improves performance

- Using multiple disk drives over high-speed serial-attached SCSI (SAS) interfaces for each data node improves performance compared to that of traditional Hadoop implementations that use internal serial ATA (SATA) drives.

Increases network speed

- The OmniSwitch LAN switches provide 10Gb/s and beyond of converged Ethernet.

Increases network efficiency

- With OmniSwitch LAN switches, the network can be provisioned to carry flows such as storage, data, or voice with unique prioritization and quality-of-service profiles.

Testing confirms that the simplified NetApp Open Solution for Hadoop architecture scores high in scalability, performance, fault tolerance, and efficiency to create a solid foundation for analytics on extremely large datastores. All tests were carried out using open-source software, workload generators, and monitoring tools.

#### Supports large datasets

NetApp Open Solution for Hadoop architecture delivers the performance required for large analytic datasets. Testing was performed with cluster sizes of two, four, and eight data

## SOLUTION COMPONENTS

### Alcatel-Lucent Products and Technologies

Alcatel-Lucent OmniSwitch 10K modular LAN chassis

Alcatel-Lucent OmniSwitch 6900 stackable LAN switch

Alcatel-Lucent OmniSwitch 6850E stackable LAN switch

### NetApp Products and Technologies

NetApp FAS2040 with shared NFS

NetApp E2660 RAID storage array with hot swappable disks

### Third-Party Products and Technologies

Cloudera distribution for Hadoop (CDH3u3)

Emulex 10Gb/s server Ethernet network interface card (OCE11102-FX PCIe 10Gb/s)

Dell R710 compute data node (dual Intel® Xeon® E5645 six-cores @ 2.40Ghz with 64GB RAM)

LSI 9200-8e 6Gb/s SCSI (SAS) host bus adapter

nodes keeping the Hadoop replication count at two. Corresponding with the increasing data node cluster, the dataset size was increased to 1TB, 2TB, and 3TB.

- As the number of data nodes increased and the volume of the load increased linearly, the job completion time improved steadily, showcasing scalability in storage utilization rates.
- A job completion time of approximately 1 hour and 10 minutes for the creation of a 3TB dataset indicated that an eight-data-node cluster delivers an aggregate throughput of 5.85Gb/s.
- This modular building block approach can be used with increased dataset sizes to provide consistent performance.

#### Scales with performance

NetApp Open Solution for Hadoop architecture continues to deliver scalable performance even as data volumes increase. The tests were executed on the data nodes and datasets with a map count of 56 and a reducer count of 40. Testing was performed with cluster sizes of two, four, and eight data nodes and corresponding dataset sizes of 1TB, 2TB, and 3TB.

The test results demonstrate that as the number of data nodes increased and the volume of data generated increased linearly, the job completion time improved steadily, showcasing a scalable solution for compute performance.

#### About Alcatel-Lucent Enterprise

Alcatel-Lucent Enterprise provides communications solutions and services to businesses of all sizes. Our secure and open portfolio and worldwide industry expertise help you interconnect networks, people, processes and knowledge to transform your business into a dynamic enterprise. Discover more at: <http://enterprise.alcatel-lucent.com/>.

#### About NetApp

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](http://www.netapp.com).

Go further, faster®

