



Solution Brief

NetApp and IBM Partner on Comprehensive Analytics Platform

Reference architecture integrates IBM InfoSphere BigInsights with NetApp storage

KEY BENEFITS

Extract Business Value from Big Data

- Balanced and integrated IBM® compute and networking, NetApp storage and IBM InfoSphere BigInsights, an easy-to-use Hadoop application suite

IBM Big Insights, a full-featured, easy to use Hadoop application suite

- Sophisticated text analytics for big data and IBM BigSheets for discovery by data scientists, analysts, and business users
- Simplifies Hadoop application implementation, management, and development through a rich set of tools to accelerate return on investment (ROI)

Open Partner Ecosystem for Analytics Stack

- Open Partner Ecosystem for best of breed analytics
- Enterprise-grade, reliable Hadoop with compute, network, and storage
- Works well with other products in your existing infrastructure

The Enterprise Big-Data Challenge

Every day, sources like machines, sensors, videos, blogs, social media, and mobile devices are generating more data than ever, that businesses need to collect. Big data is about capturing value from this opportunity, and it's vital for businesses to have a strategy around analytic tools so they can find value in structured and unstructured data, to create differentiated benefits and gain competitive advantage. Hadoop is a powerful and essential tool that has extended business analytics beyond BI tools. However, there are challenges with Hadoop:

- It can take significant time to become productive
- Not easy to implement with existing staff and skillsets
- Can be difficult to make Hadoop data useful for analysis
- Reliability of NameNode as single point of failure
- Standard Hadoop makes three copies of data

There is also risk in not taking advantage of the business opportunities that big data presents. To address the opportunities and manage the challenges, companies need analytic and infrastructure solutions that:

- Provide a reliable infrastructure for Hadoop
- Work with other analytics tools used by businesses
- Scale compute and storage independently for flexibility when expanding clusters
- Combine high-performance Hadoop clusters with intelligent analytics applications
- Can gain insight from structured, unstructured data, as well as from data "in flight" and "at rest"

To address these challenges, IT departments are using Apache Hadoop, an open-source software framework for distributed processing of large datasets across clusters of servers. IBM and NetApp have partnered to offer NetApp storage with IBM BigInsights, a balanced and integrated reference design that delivers a full-featured, easy-to-deploy Hadoop solution.

The Solution

NetApp reference architecture for IBM InfoSphere BigInsights

The NetApp reference architecture for IBM BigInsights combines leading-edge technologies to deliver a complete analytics solution. This reference architecture helps enterprises gain control of

their data, so they can manage, process, and unlock the value of these new varieties of unstructured and semi-structured data. The NetApp reference architecture for BigInsights is based on the NetApp Open Solution for Hadoop, adding IBM x3630 M4 servers and networking resources to that solution.

The NetApp reference architecture for BigInsights answers the needs of big-data analytics by enabling you to spend your time extracting value from your data, rather than fixing or maintaining cluster operation. It provides enterprise-grade server and storage systems, offering data protection and addressing the single points of failure that plague current Hadoop deployments. Enterprise-grade firmware RAID provides near-nonstop transparent operations, reducing the need for server-based replication to two copies of data, rather than the three copies typically used in Hadoop.

Big-data analytics with Enterprise TCO

The NetApp reference architecture for IBM BigInsights is optimized for node-storage balance, reliability, performance, and storage capacity and density. Organizations that use Hadoop often use traditional server-based storage with inefficient, hard-to-scale, internal direct-access storage (DAS). The NetApp storage for IBM BigInsights employs the managed DAS model, with higher scalability and lower TCO. The lower cost comes from choosing the number of clusters and storage you want and not spending money for features you don't need. It also comes from non-proprietary lower-cost options and by decoupling compute nodes from storage, preventing unnecessary purchases of compute nodes for storage-intensive workloads.

Superior flexibility provides rapid ROI

NetApp and IBM focus on offering customers flexible choices that are driven by important factors such as number of clusters, number of servers, and support options. With NetApp and IBM you have a range of analytics applications to accelerate the processing of big data and for working with standard Hadoop.

SERVICE COMPONENTS

IBM Products

- IBM InfoSphere BigInsights
- IBM System x3630 M4 servers
- IBM BNT RackSwitch G8264 10Gb/ Ethernet switch
- IBM BNT RackSwitch G8052 1Gb/ Ethernet switch

NetApp Products

- NetApp FAS2240 with shared NFS running Data ONTAP®
- NetApp E2660 RAID storage arrays with SANtricity® storage management software

NetApp reference architecture for IBM BigInsights fosters an open ecosystem

The reference architecture includes testing on the open-source Apache Hadoop framework, not a proprietary Hadoop distribution. This allows greater interoperability with other Hadoop tools and enables businesses to use analytical tools they already have or tools from any number of advanced analytical vendors, whether they're established or on the leading edge of Hadoop development.

IBM BigInsights helps you start taking advantage of big data right away

IBM BigInsights offers an open and flexible Hadoop foundation that provides enterprise-class infrastructure for advanced analytics deployments. You can deploy the solution rapidly and benefit from the reliability of IBM servers and networks and NetApp storage. Because of the optimal price/performance metrics to deploy and administer the solution, you can save operational costs, deploy the solution cost effectively, and attain performance that is equal to or better than that of a distributed server with direct attached storage.

Why Choose IBM and NetApp for Big-Data Analytics?

- IBM and NetApp have extensive experience in deployments of over 50PB. IBM and NetApp have worked on joint big data customer deployments.

• Service and support

IBM and NetApp offer extensive global support with on-site production installation assistance and 24/7 support. IBM Global Technology Services (GTS) and NetApp Professional Services teams also coordinate project management at the outset of each project.

• Ongoing technological cooperation

IBM and NetApp technical teams provide certification of our joint solution and collaborate for solution enhancements.

About IBM

IBM InfoSphere BigInsights brings the power of Hadoop to the enterprise. With built-in analytics, extensive integration capabilities and the reliability, security and support that you require, IBM can help put your big data to work for you. <http://www.ibm.com/ibm/us/en/>

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.

Go further, faster®

