

# FlexPod®

## FlexPod Select with Hadoop: Enterprise-Class Infrastructure to Accelerate Time to Value from Your Data

### IT Seeks Control and Business Seeks Insights and Value from Big Data

*Big data* is now a well-known term, with large amounts of data coming from newer sources such as sensors, videos, call logs, social media, and mobile devices. Big data is typically defined as data that is beyond an organization's ability to store, process, and analyze by using its current tools or infrastructure. Although there are challenges, big data also presents opportunities to get new and different insights. Many organizations are turning to Apache Hadoop, an open-source data management platform, to help manage their big data and start gaining some of these insights.

Although Hadoop is a powerful and essential tool to handle this data, as with new technologies, it introduces new challenges. For example, enterprises might have difficulty in deploying the solution by using existing staff and skillsets, in keeping up with open-source patches and updates, and in knowing how to optimize and fine-tune the Hadoop cluster. Typically, Hadoop also requires three copies of data, which lowers storage efficiency and strains existing resources. Therefore, a ready-to-deploy Hadoop solution that comes optimized and tuned can enable businesses to quickly obtain value and derive new, previously unknown insights.

### FlexPod Select with Hadoop Accelerates Time to Value

Prevalidated reference designs reduce risk and speed deployment of Hadoop

The FlexPod® solution, jointly developed by NetApp and Cisco, is a flexible, converged infrastructure for enterprise applications. FlexPod Select with Hadoop is targeted for big data workloads that require dedicated infrastructure based on Cisco UCS C-Series Servers and NetApp E-Series storage.

### Business Benefits

#### Faster Time to Value

- Get validated, preconfigured components to enable rapid deployment of your Hadoop cluster.
- Store, process, and analyze all data for new opportunities and business impact.
- Have more time to focus on data analysis rather than dealing with cluster downtime.

#### Easier Hadoop Experience

- Leading Hadoop distributions are tuned and tested on enterprise-class infrastructure.
- Monitoring and management tools employ NetApp® SANtricity® software, Cisco UCS® Manager, and Hadoop management software.
- High density and high capacity reduce your data center footprint.

#### Best-in-Class Open Platform

- Future-proof your data center against lock-in and benefit from an evolving ecosystem.
- Choose best-of-breed components from more than one vendor to build your analytic solution stack.
- Increase compatibility with existing applications and infrastructure.



FlexPod Select is a unique solution that provides:

- Simplified management with the unified Cisco UCS Manager
- Higher availability, with less downtime, faster recovery from failures, and robust protection of NameNodes
- A flexible design for workloads while integrating with existing enterprise applications
- Resiliency, with hot spares and RAID protection for faster recovery from disk failures
- Improved storage efficiency, with a Hadoop replication count of two for less network congestion, fewer system resources and less storage needed, and less strain on system resources
- Easy scaling as data needs grow, with the ability to scale up to 10 racks in a single Cisco UCS domain

### Availability and Scalability for Enterprise-Level SLAs

FlexPod Select provides higher resiliency by recovering more quickly from failed jobs. This allows tasks on the DataNodes to run during drive failures and offers robust NameNode protection (recovery in minutes

versus hours) compared with traditional DAS-based Hadoop nodes.

In addition, serviceability is higher because disk drives and nodes can be repaired while a cluster is running, which minimizes data loss. And with faster job completion times and data load performance, your organization can meet tight SLAs on your data applications and workloads.

### FlexPod Select with Hadoop Reference Architectures

FlexPod Select has been pretested and jointly validated with leading Hadoop vendors, including Cloudera and Hortonworks.

### FlexPod Select with Cloudera

Cloudera Enterprise is a revolutionary data management platform designed specifically to address the opportunities and challenges of big data. It combines Cloudera's Distribution Including Apache Hadoop (CDH), Cloudera's 100% open-source Hadoop distribution, with powerful management software to create a single, massively scalable system in which you can unite storage with an array of powerful processing and analytic frameworks. By uniting flexible storage and processing

under a single management framework and set of system resources, Cloudera delivers the versatility and agility required for modern data management. You can ingest, store, process, explore, and analyze data of any type or quantity without migrating it between multiple specialized systems.

### FlexPod Select with Hortonworks Data Platform

Hortonworks Data Platform (HDP) is a 100% pure open-source distribution of Apache Hadoop. As a data management platform, it helps you combine the power and cost-effectiveness of Hadoop with the advanced services and reliability that your enterprise needs. HDP operates alongside existing data center infrastructure, complementing existing ETL and data warehousing, and providing processing at a larger scale and more cost-effectively than ever before.

### Getting Started

To learn how FlexPod Select can help you set up and run a Hadoop cluster, contact your local NetApp or Cisco representative or partner. You can also get more detailed deployment information about the CDH reference design and the HDP reference design.

**FlexPod Select support:** NetApp works with Cisco to resolve infrastructure issues and works with Cloudera or Hortonworks to resolve Hadoop-related issues.

Customers get access to expert technical support for the full range of interoperable technologies involved.

