FlexPod Select for Hadoop: Enterprise Class Infrastructure to Accelerate Time to Value from your Data.

IT Seeking Control, Business seeking 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 using their current tools or infrastructure. While there are challenges, big data also presents opportunities to get new and different insights, and many organizations are turning to Hadoop, an open source analytic platform to help manage their big data and start gaining some of these insights.

While Hadoop is a powerful and an essential tool to handle this data, as with new technologies, it introduces new challenges. It will be difficult for enterprises to deploy with existing staff and skillsets, they need to keep up with open source patches and updates, and know how to optimize and fine-tune the Hadoop cluster. Typically, Hadoop requires three copies of data that lowers storage efficiency and strains existing resources. Therefore, a ready to deploy Hadoop solution that comes optimized and tuned can enable businesses to get value and new, previously unknown insights.

FlexPod Hadoop Accelerates Time to Value

Prevalidated Reference Designs reduce risk and speed deployment of Hadoop

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

FlexPod Select is a unique solution that provides:

- Simplified management with unified UCS Management
- Higher availability with less downtime, faster recovery from failures, robust protection of NameNode
- Flexibility with A flexible design for workloads while integrating with existing enterprise applications
- Resiliency with Hot spares and RAID protection means, faster recovery from disk failures and less system resources
- Improved storage efficiency with replication count of two means less network congestion, system resources, less storage needed and less strain on system resources
- Easy scaling as data needs grow – can be scaled up to ten racks in a single UCS domain

BUSINESS BENEFITS

Faster time to value

Validated, preconfigured components enable rapid deployment of your Hadoop cluster

Store, process, and analyze all data for new opportunities and business impact

More time to focus on data analysis rather than deal with cluster downtime

Easier Hadoop Experience

Leading Hadoop distributions tuned and tested on enterprise-class infrastructure

Monitoring and management tools with SANtricity and Cisco UCS Manager

High density and capacity reduces data center footprint

Best of Breed, Open Platform

Future proof against lock-in and benefit from evolving ecosystem

Not dependent or locked in to entire stack from one vendor

Increased compatibility with existing applications and infrastructure.

FlexPod™

Availability and Scalability for Enterprise-level SLAs

FlexPod Select provides higher resiliency by recovering faster from failed jobs, allowing tasks on the Datanodes to run during drive failures, and robust NameNode protection (minutes vs. hours) compared to traditional DAS based Hadoop nodes.

In addition, serviceability is enabled as disk drives and nodes can be repaired while cluster is running, minimizing less data loss. In addition, with faster job completion times and data load performance, businesses can meet tight SLAs around their data applications and workloads.

FlexPod Select for Hadoop Reference Architectures

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

FlexPod Select with Cloudera Distribution including Apache Hadoop (CDH)

Cloudera Distribution for Hadoop offers a single, integrated enterprise-class solution, which lets you efficiently query all of your data - structured and unstructured - and have a view beyond data sitting in relational databases. Cloudera's platform runs in real time, so you can work at the speed of thought as you build rapidly on deep insights, create competitive advantage and become truly data-driven.

FlexPod Select with Hortonworks Data Platform (HDP)

Hortonworks Data Platform is a 100% pure open source distribution of Apache Hadoop. As a data management platform, it is ideal for organizations that want to combine the power and cost-effectiveness of Hadoop with the advanced services and reliability required for enterprises. HDP operates alongside existing data center infrastructure, complementing existing ETL and data warehousing, and providing processing at a larger scale more cost effectively than ever before.
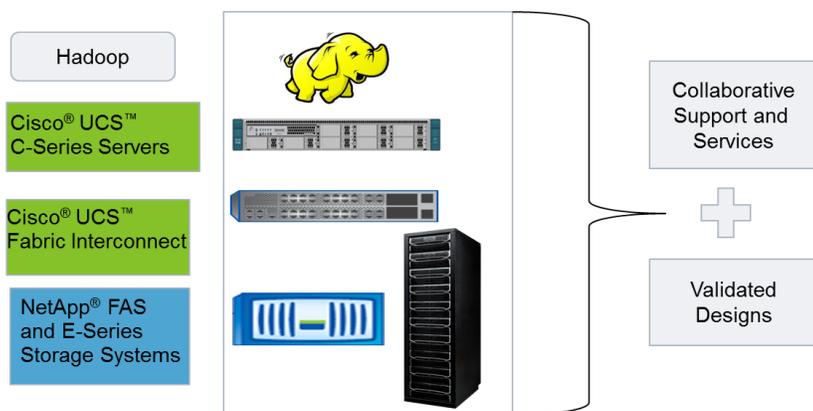
Getting Started

To learn how FlexPod Select can help set up and run a Hadoop cluster, contact your local NetApp or Cisco representative or partner. Cloudera reference design can be found here, and Hortonworks reference design can be found here.

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 on the full range of interoperable technologies involved.

Figure 1. FlexPod Select validates Hadoop with enterprise class infrastructure

www.cisco.com   www.netapp.com