



cloudera®

Dell Apache Hadoop Solutions

Leverage enterprise-optimized hardware, software and services with the Dell | Cloudera Solution.

Streamlined to solve business problems

When you move forward with the Dell™ | Cloudera® Solution, you have the confidence that comes with Cloudera's distribution including Apache® Hadoop® (CDH), one of the world's leading distributions of Hadoop in commercial and non-commercial environments.

CDH delivers a streamlined path for putting Apache Hadoop to work solving business problems in production. Ideal for enterprises seeking a stable, tested Hadoop solution without proprietary vendor lock-in, CDH is the bridge between the insights of organizations using Hadoop in production and the continuous stream of innovations from the Apache community.

Fully documented and ready to go, CDH consists of 100-percent open source Apache Hadoop plus the comprehensive set of open source software components needed to use Hadoop, which can be activated on an as-needed basis. CDH is thoroughly tested and certified to integrate with a wide range of operating systems and hardware, databases and data warehouses, and business intelligence and extract, transform and load (ETL) systems. This broad compatibility helps your organization take advantage of Hadoop while leveraging your existing tools and resources.

Simplicity and certainty

Due to the explosive growth in data volumes and complexity, companies of all sizes are adopting the open source Apache Hadoop platform to store, process, and derive value from all of their data. There are many advantages to Hadoop, particularly in the areas of flexibility and economics. But just like any open source technology, implementation into a production environment presents a unique set of challenges and questions.

How can you guarantee system stability? Which components of the open source platform do you need and how can you ensure that they will integrate correctly? How well is the technology documented? Who do you call for support? The uncertainty caused by questions like these drives some companies to invest in proprietary vendor solutions—despite the higher price tag and risk of vendor lock-in.

Cloudera created its distribution of Hadoop for this very purpose—to remove the uncertainty and barriers that sometimes dissuade enterprises from deploying open source Hadoop in production.

Everything you need

The Dell | Cloudera Solution gives you everything you need to tackle your big-data challenges—including software, hardware, services, reference architecture and deployment guide. Choose from PowerEdge™ C6105, R720, R720xd or C8000 series servers, Force10 or PowerConnect™ switches—components proven in cloud and big-data deployments. Select Cloudera Enterprise or CDH distributions of Apache Hadoop, then speed deployment and ease ongoing operation of your big data cluster with Dell's open source

Crowbar software framework. At any time, tap into the worldwide network of experts in the online Hadoop and Crowbar communities.

Dell Services makes getting started easy. Options include, but are not limited to, custom solution design, hardware and software deployment, ongoing support and training. This gives you the assurance that your solution is backed by expert hardware and software support that can be tailored to your specific needs.

Streamlined path to success with Hadoop

- Gain all the components needed for use in production; implement components on an as-needed basis.
- Remove uncertainty with a fully documented distribution.
- Protect existing investments and the integration work is done for you.

Stable and reliable Hadoop without vendor lock-in

- Leverage 100% pure Apache Hadoop, hardened for production environments.
- Increase your confidence with a distribution thoroughly tested by expert engineers.
- Scale across tens of thousands of nodes in enterprise environments.

Community driven

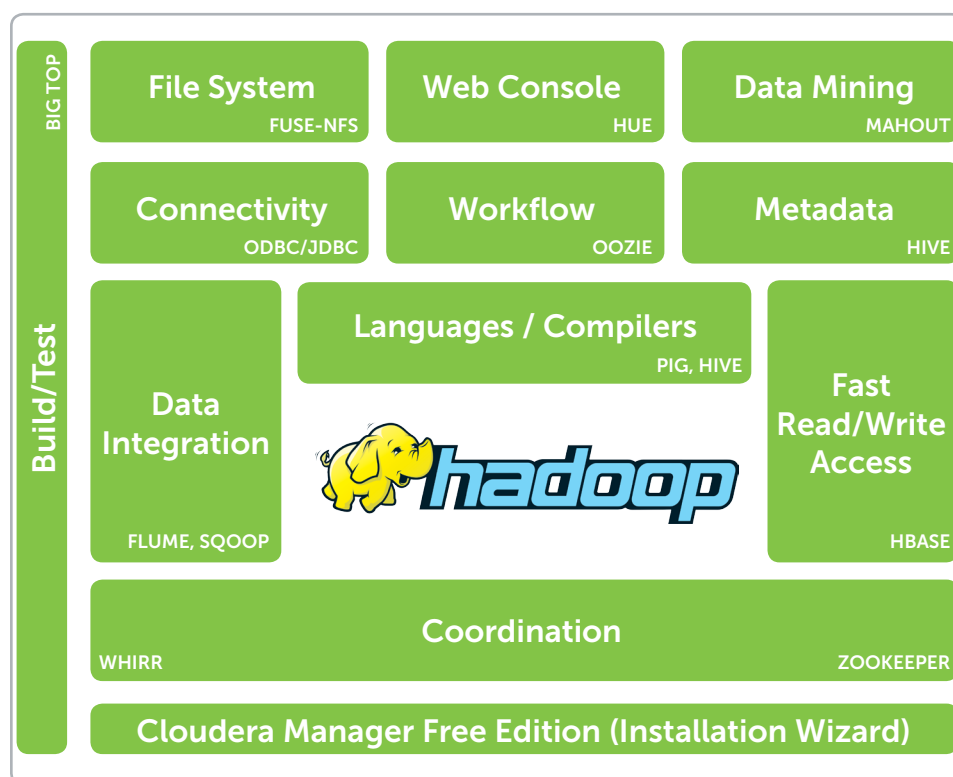
- Incorporate only open source components—no forks or proprietary underpinnings.
- Gain access to a continuous stream of innovations from the Hadoop community.
- Leverage a proven distribution that is free to download.

What the Dell Cloudera Solution delivers	
A complete, packaged set of all components needed to use Apache Hadoop in production	CDH contains everything you need for a successful implementation, and you can deploy the different components as you need them. All of the integration work is done for you and the entire solution is fully documented. By taking the guesswork out of building out your Hadoop stack, CDH gives you a streamlined path to success in solving real business problems.
100% pure Apache Hadoop, hardened for production environments	With CDH, Cloudera maintains a vigorous commitment to upholding the open source standard. From the core file system to the user interface, CDH contains absolutely no code forks or proprietary underpinnings and is free to download. The entire distribution is quality tested by Cloudera and proven at scale across tens of thousands of nodes and varying workloads in enterprise environments. CDH is stable, reliable Hadoop without vendor lock-in.
The Cloudera Ready certification framework	CDH has been tested and certified to integrate with a wide range of hardware, operating systems, and data analysis tools—lowering risk, easing deployment, and protecting your investments in your existing data analysis and business intelligence solutions.

CDH components	
Apache Hadoop	Reliable, scalable distributed storage and computing
DataFu	Library with user defined functions for data mining and statistics tasks
Apache Flume	Distributed service for collecting and aggregating log and event data
Apache HBase	Hadoop database for random, real-time read/write access
Apache Hive	Browser-based desktop interface for interacting with Hadoop
Hue	High-level language for expressing data analysis programs
Apache Mahout™	Tool to build scalable machine learning libraries
Apache MRv1	MapReduce software framework for processing large data sets
Apache Oozie	Server-based workflow engine for Hadoop activities
Apache Pig	SQL-like language and metadata repository
Apache Sqoop	Transfers bulk data between Apache Hadoop and structured datastores
Apache Whirr	Library for running Hadoop in the cloud
Apache Zookeeper	Highly reliable distributed coordination service

CDH 4.1.2 distribution details	
Component versions	
Apache Hadoop	v2.0.0 +552
DataFu	v0.0.4+9
Apache Flume	v1.2.0 +122
Apache HBase	v0.92.1+160
Apache Hive	0.9.0+155
Hue	v2.1.0+221
Apache Mahout™	v0.7+8
Apache MRv1	v1-0.20.2 +1265
Apache Oozie	v3.2.0+126
Apache Pig	v0.10.0+58
Apache Sqoop	v1.4.1+54
Apache Whirr	v0.8.0+18
Apache Zookeeper	v3.4.3+28
Supported operating system versions	
Red Hat	RHEL 6, RHEL 5
CentOS	CentOS 6
Ubuntu	Lucid, Precise
SUSE	SUSE 11 SP1
Supported build infrastructure and cloud platforms	
Build Infrastructure	Apache Maven
Cloud Platforms	Rackspace, Amazon EC2, Softlayer

CDH components



© 2012 Dell Inc. All rights reserved. Dell, the DELL logo, the DELL badge and PowerEdge are trademarks of Dell Inc. Other trademarks and trade names may be used in this document to refer to either the entities claiming the marks and names or their products. Dell disclaims proprietary interest in the marks and names of others. This document is for informational purposes only. Dell reserves the right to make changes without further notice to the products herein. The content provided is as-is and without expressed or implied warranties of any kind.

To learn more, visit: www.Dell.com/Hadoop

