Conquer the challenges of big data

Organizations that are gathering insights from vast data volumes and varied data types find that managing large volumes of unstructured data exceeds the capacity and capabilities of traditional data intelligence systems.

These systems were specifically designed for structured data types from sources such as relational databases. Gathering data intelligence and developing perspectives from extremely large amounts of data requires a scalable system that can process multi-structured data volumes quickly and responsively and that can easily scale to manage growing data volumes.

With the explosive growth in data volumes and complexity, organizations of all sizes are turning to the open source Apache Hadoop platform to store, process and generate value from their data stores. Hadoop solutions are not just about being able to capture data but also about being able to work with the many new and different varieties of unstructured data - social media data, sensor data, machine generated data and more. There are many advantages to using Hadoop, particularly in scalability, flexibility and economics. But as with any open source technology, it presents a unique set of challenges when deployed into production.

Cloudera created its enterprise distribution of Hadoop for this very purpose; to remove the uncertainty and barriers that may dissuade an organization from deploying open source Hadoop into production processes.

Get the benefits of a complete solution

- Accelerate big data cluster deployments with the Dell | Cloudera Apache Hadoop 5.4 Solution, powered by Intel
- Perform interactive analysis on any data stored in Hadoop HDFS and HBASE with Cloudera Impala
- Interact with your data via SQL and leading business intelligence solutions with Impala
- Configure, manage and monitor your Hadoop Cluster with Cloudera Manager
- Rely on the Dell Global Support Network and the rich knowledge base of Cloudera Enterprise

Cloudera Enterprise, with Apache Hadoop at the core, is:

Unified—one integrated system, bringing diverse users and application workloads to one pool of data on common infrastructure; no data movement required

Secure—perimeter security, authentication, granular authorization and data protection

Governed—enterprise-grade data auditing, data lineage and data discovery

Managed—native high-availability, fault-tolerance and self-healing storage, automated backup and disaster recovery, and advanced system and data management

Open—Apache-licensed open source to help ensure your data and applications remain yours, and an open platform to connect with all of your existing investments in technology and skills
Leverage enterprise-optimized hardware, software and services

When you deploy Apache™ Hadoop™ technology within your enterprise, your first priority is to quickly store and manage your data and then begin to derive the business value from all of it, instead of worrying about managing an unproven software framework.

These are among the benefits of the Dell | Cloudera Apache Hadoop 5.4 Solution, accelerated by Intel. This end-to-end package delivers the core elements of Hadoop, including scalable storage and distributed computing, within a turnkey solution based on Cloudera Enterprise software and Dell hardware with Intel processors.

Hadoop for the enterprise

Designed specifically for mission-critical environments, Cloudera Enterprise includes CDH, the world’s most popular open source Hadoop-based platform, as well as advanced system and data management tools, open source software components that help with Hadoop usability, and dedicated support from Hadoop experts.

Better still, Cloudera Enterprise is thoroughly tested and certified by Dell to support a wide range of big data use cases—and to give you complete confidence in your Hadoop environment.

Key features

The Dell | Cloudera Solution is based on the Cloudera CDH 5.4 Enterprise or Community distribution of Hadoop.

Changes in Cloudera 5.4

<table>
<thead>
<tr>
<th>Updates to key areas of focus</th>
<th>Data Governance</th>
<th>Updates with audit and lineage completeness, addition of xPlain.io and a redesigned metadata search UI (xPlain: Continuous data model optimization and identifying troublesome queries)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Open Standards</td>
<td>Includes Spark 1.3, HBase 1.0, Hadoop 2.6, Solr 4.10.3, Kafka 11, Impala 2.2, Hive 1.1</td>
</tr>
<tr>
<td></td>
<td>Platform Support</td>
<td>More cloud support and more deployment options</td>
</tr>
<tr>
<td></td>
<td>Core Scalability</td>
<td>Scale and performance improvements, high stress testing, Navigator back-end improvements for performance and scalability, Hue responsiveness improvements - &gt; 10x gain in response times, Search scalability work at Petabyte scale customers, Spark shuffle service - &gt; reliability of Spark jobs</td>
</tr>
</tbody>
</table>

| Updates to Cloud platform support for | OpenStack, Microsoft Azure, Amazon AWS |

| Enterprise Security Requirement updates | Core Platform updates | Improvements to Solr security for PCI certification, Updated support and authentication in Kerberos |
|                                       | Manageability improvements | Navigator Key Trustee installed and monitored by Cloudera Manager |
|                                       | Simplified setups | Backup and Disaster Recovery (BDR): Automated replication of ACLs, BDR: Automated replication of HDFS encryption zones |

| Core Storage updates to | HDFS | Hierarchical storage management updates |
|                         | HBase | Update to Hbase 1.0 |

| Hive-on Spark preview: | 5x improvements in speed over MapReduce |
| Impala | Deep focus to updates on stability |

| Spark – includes updates for production readiness | Spark Streaming low latency solution with continued improvements in metrics, monitoring and debugging with improved cluster utilization of Spark |
| Kafka | Rich monitoring of Kafka from within Cloudera Manager and support for MirrorMaker |
Increase uptime with Cloudera Support
With Cloudera Enterprise, you have the most experienced technical support team beside you every step of the way. Cloudera Support features the industry’s only predictive and proactive support capabilities, focused on securing more uptime, rapid issue resolution and prevention, better performance for your mission-critical applications, and faster delivery of the all features you care about.

Services & Support
The Dell | Cloudera Apache Hadoop 5.4 Solution, accelerated by Intel, comprises Cloudera’s Distribution of Hadoop, Dell PowerEdge series servers, Dell Networking, and joint service and support. This combined solution delivers an integrated, highly optimized solution for massive distributed processing against a single file system namespace.

Hadoop Deployment & Consulting
Solution Deployment – A turnkey set of services that include on-site implementation of all solution hardware, configuration of servers and network switches, and installation and configuration of Cloudera Hadoop. Deployment services are performed by trained Dell technicians in accordance with the Dell | Cloudera Hadoop Solution 5.4 Reference Architecture and best practices.

Big Data Consulting – Dell Services can optimize your Hadoop solution implementation by unlocking vital data resources through extensive expertise and proven delivery. From high level strategy and architecture design to implementation, management and support, we can help drive the success of your big data business intelligence program.

Gain the value of a partnered solution
The Dell | Cloudera Apache Hadoop Solution, accelerated by Intel, gives you everything you need to tackle your big-data challenges—including software, hardware, services, reference architecture and deployment guides. The Dell | Cloudera Apache Hadoop 5.4 Solution includes Dell PowerEdge R730XD, Dell Networking and optional Dell Services, which are tested, validated and proven in big data deployments. And at any time, tap into the worldwide network of experts within the online Hadoop community.

Learn more about Dell Hadoop solutions at:
Dell.com/Hadoop or DellBigData.com.

©2015 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.

SB240_Dell_Cloudera_Enterprise_Apache_Hadoop_Distribution_2.2.15

Intel and the Intel logo are trademarks of Intel Corporation in the U.S. and/or other countries.