

# Hortonworks focuses on Hadoop stability and reliability with HDP 1.0

**Analyst:** Matt Aslett

10 Jul, 2012

Hortonworks recently introduced version 1.0 of its Hortonworks Data Platform (HDP) distribution of Apache Hadoop. Based on the stable Apache Hadoop 1.0 codebase, HDP 1.0 highlights the company's belief that a focus on stability and reliability is required to encourage greater mainstream adoption of the open source data-processing framework.

## The 451 Take

The delivery of HDP 1.0 is a good way to mark the one-year anniversary of Hortonworks' formation and is a good reflection of the company's perspective on the current state of Hadoop adoption – highlighting its belief that a stable and reliable platform, rather than innovative new features, will be important to encourage more widespread adoption by the early majority. That this is in direct contrast to the attitude taken by Hortonworks' closest rival is all the more interesting, and the adoption rates of their respective distributions will provide an indication of mainstream attitudes toward Hadoop.

## Context

Despite widespread interest in the Apache Hadoop data-processing framework and early-stage adoption by Web properties, it remains an open question as to how soon mainstream enterprises will convert Hadoop development and test projects into production deployments. While other Hadoop supporters are more bullish, Hortonworks – formed in mid-2011 by many of the main engineers that made up Yahoo's core Hadoop architecture team – believes that in order for Hadoop to cross the chasm into mainstream adoption, there needs to be a greater focus on Hadoop's

reliability and stability. It is in this context that the company's recently released HDP 1.0 distribution must be viewed. Based on the stable Apache Hadoop 1.0 code base, rather than the more experimental but more feature-rich Apache Hadoop 2.0, it is designed to provide a platform that will encourage adoption by the 'early majority' (HDP 2.0, based on Apache Hadoop 2.0, is in beta testing).

In a nod to the importance of the high-availability (HA) advances in Apache Hadoop 2.0, HDP 1.0 includes an HA architecture developed with VMware and designed to be compatible with Hadoop 2.0, but is otherwise designed to stay as close as possible to the Apache Hadoop 1.0 code line. In addition to the core Hadoop Distributed File System (HDFS) and MapReduce projects, HDP 1.0 also includes Pig and Hive for analysis, the HBase NoSQL data store, as well as Oozie for workflow and scheduling. Additional components include Talend's Open Studio for Big Data data-integration software, HCatalog for common metadata services, and the Hortonworks Management Center – a Hadoop management and monitoring console based on the Apache Ambari project – all of which are open source.

The Hortonworks Management Center provides dashboard views of cluster operation data, as well as server and storage utilization and performance levels, and features open APIs to plug into systems and configuration management tools such as Ganglia, Nagios and Puppet. It also supports simplified cluster provisioning via automated compatibility checks and testing. Talend released Open Studio for Big Data under the Apache license to improve integration with Hadoop, and introduced support for Oozie and HCatalog, a project abstracted from the Hive metadata services to provide consistent metadata and data models across all components in the Hadoop stack. HCatalog also has a REST-based API that enables integration with other data-based applications and services.

Hortonworks reports that some 800 companies registered for the private beta of HDP 1.0, with strong interest coming from the telecom, financial services, oil and gas, and Web sectors. The company highlights Hadoop's role as a data 'refinery' for processing large data sets for analysis, potentially via native analytic tools, but more likely via integration with traditional BI and analytic platforms. Hortonworks already provides Hadoop support and training, but with the introduction of HDP 1.0 has also added support subscriptions to its portfolio. The company offers three production support subscriptions, priced per cluster, and a developer subscription, priced per developer. Hortonworks now has more than 85 employees, compared with the 22 it started with a year ago.

## **Competition**

The core competitors for Hortonworks are Cloudera, IBM, EMC Greenplum and MapR Technologies,

as well as DataStax. Of these, Cloudera is the closest rival and it is interesting to see Hortonworks focusing on the 100% open source nature of its distribution, in comparison with Cloudera Enterprise, which combines subscription-based support for the open source Cloudera Distribution including Apache Hadoop (CDH) with the Cloudera Manager administration software. Hortonworks reports that it is seeing interest from customers and prospects driven specifically by its 100% open source nature, but we also see Cloudera Enterprise specifically driving customers and prospects to Cloudera – particularly as it is more feature-rich than Hortonworks Management Center.

Cloudera is also more aggressive in terms of encouraging adoption of new Hadoop features – its recently introduced CDH 4 distribution is based on the Apache Hadoop 2.0 code base, which is more experimental, but does include advanced features including HA, security and extensibility. Therefore, it will be interesting to see how adoption of HDP 1.0 compares with CDH 4 to get an indication of how eager the early majority is to adopt Apache Hadoop 2.0. MapR and EMC are partners and are differentiated, among other things, by their replacement of HDFS with Direct Access NFS, as well as the addition of Distributed NameNode HA and JobTracker HA. DataStax similarly replaces HDFS with CassandraFS, an interface to the Apache Cassandra database, in its DataStax Enterprise offering. IBM also offers additional complementary functionality in its InfoSphere BigInsights Enterprise but recently announced that it will support alternative Hadoop distributions, starting with Cloudera, as part of its 'big data' platform.

## **SWOT Analysis**

### **Strengths**

Hortonworks has undoubted expertise in deploying and managing Hadoop clusters, and employs some of the major contributors to Hadoop and its associated projects.

### **Opportunities**

There is significant enterprise interest in Hadoop, and we have yet to see adoption among mainstream enterprises.

### **Weaknesses**

The company is working very hard to catch up, but is behind other distributors in terms of commercial partnerships and customers.

### **Threats**

Hortonworks has partnered with some of the industry's largest data management incumbents, such as Microsoft, Informatica and Teradata, but will have to compete with others such as Oracle and IBM.

Reproduced by permission of The 451 Group; © 2011. This report was originally published within 451 Research's Market Insight Service. For additional information on 451 Research or to apply for trial access, go to: [www.451research.com](http://www.451research.com)