Hortonworks' Big Data Play Reaches Beyond Hadoop

By [Virginia Backaitis](#) | *Mar 2, 2016*

Rob Bearden

Hortonworks has a new vision that not only leverages data at rest in Hadoop, data streaming in Spark and data in motion in Apache NiFi, but also data at the tipping point where data itself sets the roadmap.

During a series of announcements yesterday at an event called "The Future of Data," Hortonworks CEO Rob Bearden delivered game-changing news that will enable the next generation of data architecture.

Fundamental to that vision are Hortonworks Data Platform (HDP) for data at rest, and Hortonworks Data Flow (HDF), an Internet of Anything Engine (IoE) for data in motion. The two come together as Hortonworks Connected Data Platforms.

What's promising here is that insights gleaned from data at rest could potentially influence data in motion as — or even before — a decision needs to be made.

Fired-Up About Spark

Hortonworks is serious about Spark: Its HDP 2.4 is the first to come with support for Apache Spark 1.6.

"It's an entirely new way to manage data in motion and data at rest," Matt Morgan, Hortonworks' vice president of product and alliance marketing told CMSWire. "The integration enables a new generation of applications that can manage data-at-rest at scale while capturing information at the jagged edge."

Hortonworks CTO Scott Gnau and Hewlett Packard Enterprise (HPE) CTO Martin Fink together announced that their companies will be working in concert to enable a new class of analytic workloads that benefit from large pools of shared memory. HPE rewrote the shuffling engine of Spark in C++ to get better memory utilization, improved performance and usage for broader scalability, which will help enable new large-scale use cases.

Fink also took time to explain why Hortonworks, with its commitment to 100 percent open source, is the right Hadoop vendor/partner for HPE. The message underscored that an article Friday in the Wall Street Journal, which suggested Hortonworks was now writing proprietary software, was [misleading](#).

Addressing Its Customers

Tech companies have different types of enterprise customers. Some want the latest and greatest as soon as it is enterprise-ready, while others would rather upgrade in their own time. The same can be said for business partners.

Kudos to Hortonworks for recognizing this and adjusting its release cadences. From now on, core Apache Hadoop components (HDFS, MapReduce and YARN) and Apache Zookeeper will be updated annually and aligned with the [ODPi consortium](#).

Extended Services (including Spark, Hive, HBase, Ambari and more), which run on top of the Core, will be logically grouped together and released continually throughout the year to match the pace of innovation occurring within each project team in the community.

As part of this rapid distribution model, Hortonworks announced the general availability of Apache Spark 1.6, Apache Ambari 2.2 and SmartSense 1.2 in HDP 2.4, which is available immediately. Options for continuous and for express upgrades are also being offered.

The significance of these options cannot be overstated.

"This is a defining moment for how we deliver advancements to our customers," said Tim Hall, vice president of product management at Hortonworks. "We can give customers all the latest innovations in the moment without sacrificing a stable and reliable core. This will change the way people consume Hadoop."

Bigger Insights from Hortonworks DataFlow

Apache NiFi, the community driven project behind DataFlow is quickly winning the interest of IoT engineers.

"It's growing at a record pace," said Matt Morgan, vice president of product and alliance marketing, Hortonworks. One-hundred-and-thirty plugins are now available, including Apache Kafka, Apache Storm, Couchbase, Microsoft Azure Event Hub and Splunk processors, which enable an easy point and click user experience for any and all data.

Hortonworks has also hardened DataFlow with Kerberos for centralized authentication management across applications.

Like Tableau, But for Spark

Visualization tools are a must in the age of advanced analytics; Hortonworks President Herb Cunitz introduced a preview version of Apache Zeppelin during the event. "It's like Tableau for Spark," he said.

Impetus Technologies, a new Hortonworks partner, is now offering StreamAnalytix a tool for designing and monitoring pipelines for streaming data. It focuses on making complex event processing solutions like Storm and Spark more efficient with less coding. Together, HDF and StreamAnalytix accelerate business value from big data by eliminating complex and time-consuming manual processes and coding.

What It Means

Constellation Research analyst Holger Mueller came away from the event as well as subsequent meetings with Hortonworks executives and customers with a fairly simple, but insightful conclusion, "It's clear they (Hortonworks) are committed to become the next data platform for the enterprise. They are sure to have a hold on data at rest, now have to fight for data in motion, which they are not yet known for. And gaining a second leg isn't easy."

Still, Mueller sees data in motion as a "must have" which signifies that Hortonworks' strategy is solid and making execution paramount.

That's something that Bearden and company seem to be willing to bet the farm on. Just look at the promises they've made to Wall Street.