



Yahoo! Advances Hadoop From Science to the World's Largest Internet Deployment to Mainstream Business Use

Chief Product Officer Blake Irving to Discuss the Business-Critical Role of Hadoop at Yahoo!, Powering Consumer and Advertiser Experiences With Advanced Cloud Computing

SANTA CLARA, Calif., Jun 29, 2010 (BUSINESS WIRE) -- Yahoo! Inc. (Nasdaq:YHOO), the leading developer of Apache Hadoop, today announced significant enhancements to the open source software, accelerating the potential for enterprise-wide adoption by mainstream businesses. Hadoop is the open source technology at the epicenter of big data and cloud computing, helping companies get value from their data and better manage their businesses.

As Internet usage continues to grow, data proliferates, making it challenging for businesses of all sizes to manage data in secure and useful ways. Yahoo!, one of the largest online companies in the world, initially used Hadoop for applied science projects and has built it into an enterprise-class platform being used across its business to develop increasingly personalized consumer experiences, built on relevance and trust. Hadoop plays a key role in Yahoo!'s popular global home page, Yahoo! Search, Yahoo! Mail, and many more.

"Hadoop is where science meets big data - it's the technical underpinning that powers our innovative consumer and advertiser products on the world's most-advanced digital canvas," said Blake Irving, Executive Vice President and Chief Product Officer at Yahoo!. "Yahoo!'s cloud and Hadoop make it possible for Yahoo! to rapidly personalize our content and advertising, and deliver highly relevant experiences, while maintaining the trust of our 600 million users."

Improving Hadoop With Security, Ease of Use, and Reliability for the Enterprise

Today, at the Yahoo!-hosted Third Annual [Hadoop Summit](#), Yahoo! announced the beta release of Hadoop with Security and Oozie, Yahoo!'s workflow engine for Hadoop. The enterprise will benefit from these open source releases because they include better controls for managing business-sensitive data and enabling complex processes to be delivered via Hadoop. Hadoop with Security and Oozie are interoperable and have been tested and deployed at Yahoo! on tens of thousands of servers.

Today's Contributions Include:

- **Hadoop with Security:** a set of significant security updates to Hadoop, enabling strong authentication.
 - Integrates Hadoop with Kerberos, a mature, open source authentication standard, enabling more secure collaboration and sharing of authenticated data.
 - Enables multi-tenancy, or the use of hardware by multiple internal parties, providing authenticated secure access and processing of sensitive data.
- **Oozie, Yahoo!'s workflow engine for Hadoop:** an open-source workflow management and coordination engine to manage jobs running on Hadoop, including Hadoop Distributed File System, Pig and MapReduce.
 - Designed for Yahoo!'s rigorous use cases that require managing complex work processes and ETL (extract, transform, load) at global scale
 - Integrates with Hadoop with Security
 - Tested and deployed across Yahoo!

"Businesses across all sectors are looking for ways to leverage the vast quantities of data they are accumulating, and Apache Hadoop is an efficient solution for processing data at scale. Hadoop has matured and is now becoming an enterprise-ready cloud computing technology with the addition of Kerberos authentication," said Melanie Posey, research director at IDC Research. "Now organizations of various sizes can leverage Yahoo!'s Hadoop investment and deployments to run it on their own systems and build out their own Hadoop deployments without starting from scratch on internal science experiments."

Science and Research on Hadoop and the Cloud

Yahoo! Labs has been at the forefront of using and developing a variety of open source cloud software and has also been an early adopter of Hadoop. Since 2005, Yahoo! Labs has used Hadoop to conduct science at true Internet scale, leveraging Yahoo!'s global network to unearth insights into consumer behavior, social systems, economics, machine learning, and a host of scientific disciplines critical to the development of the Web.

Several projects developed by scientists in Yahoo! Labs have migrated to production in the Yahoo! Cloud and a few have been open sourced to the community. Examples include Pig, a programming language for performing procedural data processing tasks on top of Hadoop, and Zookeeper, a service for managing performance across distributed computing environments that recently won the best paper award at USENIX ATC '10.

In addition, Yahoo! continues to partner closely with the global academic and scientific community as both a founding member of the Open Cirrus(TM) Testbed, which is advancing cloud computing research at an international scale, and the Open Cloud Consortium, a testbed for systems research on large-scale data clouds. Top research universities such as Carnegie Mellon, the University of California at Berkeley, Cornell University, and the University of Massachusetts at Amherst use Hadoop on Yahoo!'s [M45 supercomputer](#) for a broad range of computer science research.

About Hadoop at Yahoo!

Yahoo! is the leading contributor to and user of Apache Hadoop, currently running the world's largest Hadoop implementation. Hadoop was initiated as an open source project by a team of Yahoo! employees in 2005 to address the critical business need to leverage the company's exponentially increasing volumes of data in a scalable way. In five years, the company has taken Hadoop from a 20-server prototype in Yahoo! Labs to a 35,000-server deployment running in production across Yahoo!'s global network. With the combination of global platforms like the Yahoo! cloud, and the enrichment and analysis of data at scale with Hadoop, Yahoo! is able to deliver highly relevant content and optimize the quality of the online experiences across the Yahoo! network.

For more information on Hadoop and Yahoo!, visit Yahoo.com and search for: "Hadoop and Yahoo!."

Distribution

The Yahoo! Distribution of Hadoop with Security (beta) and Oozie, the workflow engine for Hadoop are available through the Yahoo! Developer Network at <http://developer.yahoo.com/hadoop/>.

About the Hadoop Summit

To foster the community collaboration around Hadoop, Yahoo! is hosting the third annual Hadoop Summit today in Santa Clara, California. The event brings together leaders in the community, including Amazon Web Services, Facebook, IBM, and Cloudera, along with hundreds of members from the Hadoop developer and user communities. More information on the Hadoop Summit is available at <http://developer.yahoo.com/events/hadoopsummit2010/>.

About Yahoo!

Yahoo! attracts hundreds of millions of users every month through its innovative technology and engaging content and services, making it one of the most visited Internet destinations and a world-class online media company. Yahoo!'s vision is to be the center of people's online lives by delivering personally relevant, meaningful Internet experiences. Yahoo! is headquartered in Sunnyvale, California. For more information, visit <http://pressroom.yahoo.com> or the company's blog, Yodel Anecdotal (<http://yodel.yahoo.com>).

Yahoo! is the trademark and/or registered trademark of Yahoo! Inc.

All other names are trademarks and/or registered trademarks of their respective owners.

SOURCE: Yahoo! Inc.

Yahoo!

Karen Mahon, 408-394-6140

karenm@yahoo-inc.com

or

OutCast Communications

Meghan Hughes, 415-963-2552

meghan@outcastpr.com

Copyright Business Wire 2010