Cloudera Ships Impala 1.0: Industry's First Production-Ready SQL-on-Hadoop Solution

Surpasses Industry Expectations With GA Release of Its Open Source, Real-Time Query Engine for Apache Hadoop

PALO ALTO, CA -- (Marketwired) -- 04/30/13 -- , the category leader that sets the standard for in the enterprise, today announced the general availability of Cloudera Impala, its open source, interactive SQL query engine for analyzing data stored in Hadoop clusters in real time. Cloudera was first-to-market with its offering, releasing Impala to open source as a public beta offering in October 2012. Since that time, it has worked closely with customers and open source users, rigorously testing and refining the platform in real world applications to deliver today's production-hardened and customer validated release, designed from the ground-up for enterprise workloads. The company noted that adoption of the platform has been strong: over 40 enterprise customers and open source users are using Impala today, including 37signals, Expedia, Six3 Systems, Stripe, and Trion Worlds. With its 1.0 release, Impala extends Cloudera's unified Platform for Big Data, which is designed specifically to bring different computation frameworks and applications to a single pool of data, using a common set of system resources.

"At Ovum, we believe that for Hadoop to cross over to the enterprise, it must become a first class citizen with IT, the business and the data center," said Tony Baer, principal analyst, Software and Enterprise Solutions at Ovum. "A large part of making Hadoop a first-class citizen in the enterprise is making it accessible to the large base of SQL developers and applications that already exist. With Impala, Cloudera has decisively planted the stake in bringing the worlds of Hadoop and enterprise SQL together. And it has done so in a way that addresses the expectations for performance that are taken for granted in the enterprise SQL world."

Cloudera Impala was recently recognized as the most popular and well-established real-time data processing and query solution for the Hadoop market by GigaOm Research(1), who predicts that the company's purpose-built system "will ultimately lead the SQL-on-Hadoop market."

"Cloudera's Impala is perhaps the most widely known SQL-on-Hadoop solution," said Joseph Turian, PhD and research analyst at GigaOm Research. "Cloudera has chosen to build its system from the ground up. This will allow it to optimize every part of the solution. It believes that by avoiding legacy, it can actually make a better architecture that is superior, both for end users and the ops staff."

Cloudera Impala is the first SQL-on-Hadoop solution of its kind and represents a major advancement in the evolution of Cloudera's Platform for Big Data. The company invested more than two years of intensive research and development to build Impala from the ground up, delivering the industry's first massively parallel processing (MPP) query engine that's native to Hadoop.

With Impala, users can query data stored in HDFS and HBase directly. The framework supports all standard file and data formats available, so users can choose the format that best suits their use case, including the latest in analytics-focused columnar formats like Parquet, and can promote data sharing and reuse across all computing workloads -- from batch to interactive SQL -- all from a single dataset.

This unique approach eliminates the need to migrate datasets into specialized systems or proprietary formats for analytics purposes and reduces system redundancy and latency that would exist in a legacy data warehouse environment. The Impala framework is optimized for use with CDH, Cloudera's 100-percent open
source distribution of Hadoop and related applications.

"At Stripe, we have a constant need to quickly ingest and detect patterns in data coming from banks and our own systems," said Colin Marc, developer at Stripe. "Impala is an excellent tool for that and its ability to perform speed-of-thought exploratory queries has been useful, both for analytics and development."

Cloudera Enterprise RTQ is an optional subscription module that adds technical support and management automation to Impala for Cloudera Enterprise customers. It is the first data management solution that moves Apache Hadoop decisively "beyond batch," enabling users to handle real-time workloads that previously required ongoing investment in expensive, dedicated enterprise data warehouse (EDW) solutions. Powered by Impala, Cloudera Enterprise with RTQ offers a single, massively scalable system that dramatically improves the economics and performance of large-scale enterprise data management, enabling petascale processing and interaction with that data in real time to deliver "speed-of-thought" insights.

"Six3 Systems is a recognized and proven industry leader in cyber security; our business is all about comparing current activity with observed historical norms, identifying non-obvious patterns in data, correlation of large/fast moving disparate data sources, and automating threat detection. The larger the data sets that our algorithms can run on, the greater the cyber security threat awareness that can be provided to decision makers, making Hadoop a great fit," said Wayne Wheeles, senior network forensics analytic/enrichment developer at Six3 Systems. "Impala integrates fully with our existing technologies, infrastructure and analytics providing a smooth transition into real-time, interactive data querying. With its innovative 'beyond batch' capabilities, we are now able to ask more sophisticated questions and gain actionable intelligence more quickly and efficiently, eliminating traditional data analysis bottlenecks and complexity."

"The ability to query data at the speed of thought is becoming a must-have in today's fast-paced gaming industry," said David Green, director of data services at Trion Worlds. "We're deploying Cloudera Impala to empower our support organization to access and analyze issues that customers experience on the fly, while they're connected. The ability to address customer challenges in real-time will drive a happier and more loyal customer base that is crucial to our business success."

Cloudera Impala has been widely embraced by Cloudera's partner ecosystem, with numerous companies certifying their solutions for integration with the platform, including Alteryx, Capgemini, IBM Cognos, Karmasphere, MicroStrategy, Pentaho, QlikView, SAP, Splunk, and Tableau.

"Our successful collaboration with Cloudera empowers organizations to unlock valuable business insights hidden in large, complex data sets in compelling new ways," said Paul Zolfaghari, president at MicroStrategy Incorporated. "We are very excited about Cloudera's continuing innovation in the SQL-on-Hadoop market. In our independent testing of Cloudera Impala, we experienced a massive performance increase in the accessibility of data stored in Hadoop. Through our platform integration with Impala, customers can now perform sophisticated point and click analytics on data stored in Hadoop directly from MicroStrategy applications."

"We have seen wonderful improvements in query performance when using Impala which will make Hadoop more valuable to our customers as they adopt it broadly and give more people interactive access," said Dan Jewett, Vice president of product management at Tableau Software. "We are very pleased with our experiences working with Impala and the Cloudera team, as one of the first partners to integrate with Impala."

"Impala represents a major advance for Cloudera and the Hadoop ecosystem as a whole. We've invested years of research and development and devoted a team comprised of the world's top engineering talent to execute it. We are immensely proud to be releasing a fully tested and production-hardened Impala to general availability today, and to be shattering industry forecasts for its delivery timetable," said Mike Olson, CEO at Cloudera. "Cloudera was first to recognize that Apache Hadoop would be a catalyst for business transformation in the 21st century. We have worked tirelessly to support the rapid development of the
platform to form a viable and open enterprise solution, with a rich and vibrant ecosystem to support it. We will continue to be a primary driver behind the evolution of a 100-percent open source Hadoop platform by setting a high bar that pushes the boundaries of what's possible to exceed the high expectations of our enterprise customers."

Cloudera Impala 1.0 is now available. Download it for free:

Cloudera Enterprise RTQ, powered by Impala is immediately available as an optional, supplemental module with any Cloudera Enterprise subscription. For more information, visit:

Cloudera Impala is an Apache-licensed open source project. The platform is open to community contributions and the source code is available for free download on GitHub. For more information, or to join Cloudera and other open source contributors in the development of the Impala platform, visit:

Platform documentation and project release guides, with instructions for how to configure, secure, and use Impala, can be downloaded for free at:

To learn more about what Cloudera's partners are saying about Impala 1.0, visit:

Founded in 2008, Cloudera pioneered the business case for Hadoop with CDH, the world's most comprehensive, tested and widely deployed distribution of Apache Hadoop in both commercial and non-commercial environments. Now, the company is redefining with its Cloudera Enterprise, empowering enterprises to Ask Bigger Questions™ and gain rich, actionable insights from all their data, to quickly and easily derive real business value that translates into competitive advantage. As the top contributor to the community and a leading educator of data professionals, with tens of thousands of nodes under management and hundreds of customers across diverse markets, Cloudera is the category leader that has set the standard for Hadoop in the enterprise.

Read the blog:
Follow on Twitter:
Visit on Facebook:

(1) GigaOm Research Sector RoadMap: SQL-on-Hadoop Platforms in 2013 by Joseph Turian, Ph.D.