The Apache Software Foundation Announces Apache™ Spark™ as a Top-Level Project

Thursday 27 February, 2014

Forest Hill, MD –27 February 2014– The Apache Software Foundation (ASF), the all-volunteer developers, stewards, and incubators of more than 170 Open Source projects and initiatives, announced today that Apache Spark has graduated from the Apache Incubator to become a Top-Level Project (TLP), signifying that the project's community and products have been well-governed under the ASF's meritocratic process and principles.

Apache Spark is an Open Source cluster computing framework for fast and flexible large-scale data analysis. Dubbed a "Hadoop Swiss Army knife" by The Register, Spark is recognized for its remarkable speed and ease of use, running programs up to 100x faster than Apache Hadoop MapReduce in memory, and with APIs that allow developers to quickly write applications in Java, Python, or Scala.

"It's great to see Apache become Spark's permanent home," said Matei Zaharia, Vice President of Apache Spark. "Spark has quickly become one of the most active projects in the Hadoop ecosystem, with dozens of organizations contributing, and we look forward to working closely with the rest of the Apache community."

Initially created in 2009 at the University of California at Berkeley's AMPLab (the research center also responsible for the original development of Apache Mesos), the Spark distributed computing framework for advanced analytics in Apache Hadoop can easily be used standalone or on Hadoop YARN, EC2 or Mesos. Integrated with Apache Hadoop, Spark is well suited for machine learning, interactive queries, and stream processing, and can read from HDFS, HBase, Cassandra, as well as any Hadoop data source.

"This is a major milestone for the students and researchers in the AMPLab," said Mike Franklin, Director of the AMPLab at UC Berkeley. "Spark demonstrates the real impact that research can have and validates the support AMPLab has received from our White House-announced NSF Expeditions in Computing Award and our 20+ industrial sponsors and collaborators."

"Through our work on Spark at both AMPLab and Databricks, we've focused on making it much easier for organizations to get insights from big data," said Ion Stoica, CEO at Databricks and Professor at UC Berkeley. "We're doing this together with a fantastic open source community. We look forward to continue working with the community to accelerate the development and adoption of Apache Spark."

Since entering the Apache Incubator in June 2013, Apache Spark bolstered its community through code contributions by more than 120 developers from 25 organizations. Apache Spark is in use at an array of global corporations that include Alibaba, Cloudera, Databricks, IBM, Intel, and Yahoo, among others.

Andrew Feng, Distinguished Architect at Yahoo, said "Yahoo has played a leading role in evolving Hadoop and related big-data technologies, including Spark. While Apache Hadoop serves as the foundation of our big-data platform, Spark is an attractive technology for iterative applications such as machine learning. Yahoo has made significant contributions to the development of Spark and we congratulate Spark on becoming an Apache top-level project."

"I'm really proud of the community aspect that has become infectious in Apache Spark and that really grew out of the energy in the project starting in the AMP Lab and through its movement to the ASF," said Chris Mattmann, Apache Spark Incubator Mentor at the ASF, and Chief Architect, Instrument and Science Data Systems Section at NASA JPL. "Matei, Patrick, Reynold, and many of the leaders of the project have really done a tremendous job and I'm excited to see the next generation of Hadoop-style systems have a home at the ASF."

"We have some very exciting features coming in the next months, so stay tuned for even more powerful versions of Spark," added Zaharia.

Availability and Oversight
As with all Apache products, Apache Spark software is released under the Apache License v2.0, and is overseen by a self-selected team of active contributors to the project. A Project Management Committee (PMC) guides the Project's day-to-day operations, including community development and product releases. For documentation and ways to become involved with Apache Spark, visit

Related Sectors:

Computing & Telecoms ::

Related Keywords:

Apache ::

Scan Me:

http://spark.apache.org/

About The Apache Software Foundation (ASF)
Established in 1999, the all-volunteer Foundation oversees more than one hundred and seventy leading Open Source projects, including Apache HTTP Server --the world's most popular Web server software. Through the ASF's meritocratic process known as "The Apache Way," more than 400 individual Members and 3,500 Committers successfully collaborate to develop freely available enterprise-grade software, benefiting millions of users worldwide: thousands of software solutions are distributed under the Apache License; and the community actively participates in ASF mailing lists, mentoring initiatives, and ApacheCon, the Foundation's official user conference, trainings, and expo. The ASF is a US 501(c)(3) charitable organization, funded by individual donations and corporate sponsors including Budget Direct, Citrix, Cloudera, Comcast, Facebook, Google, Hortonworks, HP, Huawei, IBM, InMotion Hosting, Matt Mullenweg, Microsoft, Pivotal, Produban, WANdisco, and Yahoo.
For more information, visit http://www.apache.org/ or follow @TheASF on Twitter.

# # #

Company Contact:

The Apache Software Foundation

T. 16179218656
E. press@apache.org
W. https://www.apache.org

View Online

Newsroom: Visit our Newsroom for all the latest stories:
https://www.apache.pressat.co.uk