

IBM s'engage dans Spark, le projet Open Source le plus important de la décennie

IBM rejoint la communauté Spark, et envisage de former plus d'un million de Data Scientists

Paris - 15 juin 2015: IBM (NYSE: [IBM](#)) annonce aujourd'hui son engagement en faveur d'Apache Spark, le plus important projet Open Source de la décennie. Au centre de cet engagement, IBM envisage d'embarquer Spark au sein même de ses plateformes Analytics et Commerce. Il sera en outre disponible en mode SaaS sur sa célèbre plateforme Bluemix. A cette occasion, ce sont plus de 3500 chercheurs et développeurs d'IBM qui travailleront sur des projets liés à Spark, dans plus d'une douzaine de laboratoires à travers le monde. Dans le même temps, plus d'un million de Data Scientists et d'Analystes de Données seront formés sur celui-ci.

Afin de favoriser le développement de l'écosystème de Spark, IBM met en place les mesures suivantes :

- L'intégration de Spark au coeur de ses plateformes d'Analytics et de Commerce.
- Les solutions Cloud d'IBM Watson pour la santé mettront à profit Spark comme un atout majeur pour sa plateforme de recherche et de conseils. L'objectif étant de délivrer rapidement des informations de qualité pour les prestataires de soins et les chercheurs, en leur donnant accès à de nouvelles analyses autour des données de santé.
- IBM va mettre à disposition en Open Source sa technologie de Machine Learning IBM SystemML et va également collaborer avec Databricks pour faire évoluer les capacités de Spark.
- IBM offrira Spark en mode SaaS sur sa plateforme BlueMix afin de permettre aux développeurs d'exploiter les données de manière plus efficace.
- IBM ouvrira également un Spark Technology Center à San Francisco dédié à la Data Science et à la communauté des développeurs afin de favoriser l'innovation au sein des applications intelligentes.
- IBM va former plus d'un million de Data Scientists et d'Analystes de Données sur Spark, à travers des partenariats avec AMPLab, DataCamp, MetiStream, Galvanize et Big Data University MOOC.

IBM Announces Major Commitment to Advance Apache® Spark™, Calling It the Most Significant Open Source Project of the Next Decade

IBM Joins Spark Community, Plans to Educate More Than 1 Million Data Scientists

ARMONK, NY – June 15, 2015: IBM (NYSE:IBM) today announced a major commitment to Apache® Spark™, the most important new open source project in a decade that is being defined by data. At the core of this commitment, IBM plans to embed Spark into its industry-leading Analytics and Commerce platforms, and to offer Spark as a cloud service on Bluemix. IBM will also put more than 3,500 IBM researchers and developers to

work on Spark-related projects at more than a dozen labs worldwide; donate its breakthrough IBM SystemML machine learning technology to the Spark open source ecosystem; and educate more than one million data scientists and data engineers on Spark.

As data and analytics are embedded into the fabric of business and society – from popular apps to the Internet of Things (IoT) – Spark brings essential advances to large-scale data processing. First, it dramatically improves the performance of data dependent apps. Second, it radically simplifies the process of developing intelligent apps, which are fueled by data.

To further accelerate open source innovation for the Spark ecosystem, IBM is taking the following actions:

- IBM will build Spark into the core of the company's analytics and commerce platforms.
- IBM's Watson Health Cloud will leverage Spark as a key underpinning for its insight platform, helping to deliver faster time to value for medical providers and researchers as they access new analytics around population health data.
- IBM will open source its breakthrough IBM SystemML machine learning technology and collaborate with Databricks to advance Spark's machine learning capabilities.
- IBM will offer Spark as a cloud service on IBM Bluemix to make it possible for app developers to quickly load data, model it, and derive the predictive artifact to use in their app.
- IBM will commit more than 3,500 researchers and developers to work on Spark-related projects at more than a dozen labs worldwide, and open a Spark Technology Center in San Francisco for the Data Science and Developer community to foster design-led innovation in intelligent applications.
- IBM will educate more than 1 million data scientists and data engineers on Spark through extensive partnerships with AMPLab, DataCamp, MetiStream, Galvanize and Big Data University MOOC.

"IBM has been a decades long leader in open source innovation. We believe strongly in the power of open source as the basis to build value for clients, and are fully committed to Spark as a foundational technology platform for accelerating innovation and driving analytics across every business in a fundamental way," said Beth Smith, General Manager, Analytics Platform, IBM Analytics. "Our clients will benefit as we help them embrace Spark to advance their own data strategies to drive business transformation and competitive differentiation."

Spark Drives Business Transformation for IBM Clients

Spark has grown quickly in popularity among developers and data scientists as an essential platform for helping organizations more easily integrate Big Data into applications, and is quickly gaining momentum with IBM clients looking to transform business decision-making:

- Real-time transportation planning software from Optibus is changing the way public transport is organized. "Spark, together with IBM, provides a highly scalable platform for Optibus, making it easy for us to expand our software as a service offering into new markets, and helps us simplify deployment, maintenance and application development for transportation companies worldwide," said Amos Haggiag, Optibus CTO and Co-Founder.
- Findability Sciences, a global consulting and contextual data technology solutions company, is using IBM Analytics and Spark to help clients tap into the power of Big Data. "Apache Spark with IBM BigInsights has given us tremendous capacity for our implementations for small and medium businesses, where MapReduce was not efficient. With Spark, the performance has improved multi fold. We're now able to process streaming data from

IoT devices and offer analytics for data in motion for things like traffic, commuters and parking," said Anand Mahurkar, CEO of Findability Sciences.

- Independence Blue Cross (IBC) is the largest health insurer in the Philadelphia area, serving more than two million people in the region and seven million nationwide. It's using Spark to help drive product innovation and develop new services. "Apache Spark is quickly maturing into a power tool for development of machine-learning analytic applications. It allows our IBC researchers and academic partners to work together more seamlessly, which means we can get new claims and benefits apps up and out to customers much faster," said Darwin Leung, Director of Informatics at Independence Blue Cross.

- IBM, NASA, and the SETI Institute are collaborating to analyze terabytes of complex deep space radio signals using Spark's machine learning capabilities in a hunt for patterns that might betray the presence of intelligent extraterrestrial life. "With Spark as a Service on Bluemix, we'll be able to work with IBM to develop promising new ways to analyze signal data as we hunt for evidence of intelligence elsewhere in the cosmos. This is an exciting example of synergy in the service of science," said Dr. Seth Shostak, Senior Astronomer and Director of the Center for SETI Research.

IBM is one of four founding members of the UC Berkeley AMPLab, where Spark was first invented in 2009, and as a result participates in multi-day research retreats, provides advice and real-world insight, and interacts closely with AMPLab researchers on projects of mutual interest. "As a sponsor of the AMPLab, IBM contributes to the greater Spark community and provides guidance for the continued evolution and improvement of the Berkeley Data Analytics Stack, the open source platform of which Spark is a key component," said Professor Michael Franklin, Director of the UC Berkeley AMPLab.

Spark is agile, fast and easy to use. And because it is open source, it is improved continuously by a worldwide community. Over the course of the next few months, IBM scientists and engineers will work with the Apache Spark open community to rapidly accelerate access to advanced machine learning capabilities and help drive speed-to-innovation in the development of smart business apps. By contributing SystemML, IBM will help data scientists iterate faster to address the changing needs of business and to enable a growing ecosystem of app developers to apply deep intelligence into every thing. For more information on IBM SystemML, visit: http://researcher.ibm.com/researcher/view_group_pubs.php?grp=3174

For more information on IBM Analytics and Spark, visit www.ibm.com/spark

###

Apache®, Apache Spark™, Spark™, are trademarks of the Apache Software Foundation in the United States and/or other countries.
