IBM Launches Industry's First Development Environment for Apache Spark – Delivered in the Cloud for Rapid Adoption

IBM Cloud becomes hub to help Data Scientists analyze big data quickly and simply using Apache Spark

SAN FRANCISCO, CA - 07 Jun 2016: IBM ([NYSE:IBM](#)) today announced the first cloud-based development environment for near real-time, high performance analytics, giving data scientists the ability to access and ingest data and deliver insight-driven models to developers. Available on the IBM Cloud Bluemix platform, the Data Science Experience provides 250 curated data sets, open source tools and a collaborative workspace to help data scientists uncover and share meaningful insights with developers, making it easier to rapidly develop applications that are infused with intelligence.

Building on its $300 million investment in developing Apache Spark as a type of "analytics operating system," IBM created the Data Science Experience to extend the speed and agility of Spark to more than two million members of the R community through new contributions to SparkR, SparkSQL and Apache SparkML. As a result, data scientists who work in R will have faster access to more data, and in turn, more insights delivered from the IBM Cloud.

The Data Science Experience's open and collaborative environment allows data scientists to accelerate and simplify data ingestion, curation and analysis by bringing together the content, data, models, and open source resources from IBM and others including H2O, RStudio, Jupyter Notebooks on Apache Spark in a single security-rich managed environment.

"With Apache Spark, we see an opportunity to significantly transform the role of the data scientist by providing access to curated data sets, open source tools and a collaborative platform to accelerate innovation," said Bob Picciano, Senior Vice President, IBM Analytics. "IBM's Digital Science Experience is the killer enterprise app for Apache Spark, and gives data scientists new opportunities to deliver insight-driven models to developers, and opens the door for unprecedented innovation from the open source community."

IBM is already helping organizations across industries use data science applications built on Apache Spark to get new business insight, drive growth, and improve efficiency. Some examples include:

- Bernhardt Furniture: Using IBM Spark, IBM Bluemix, and mobile technologies, the Bernhardt IT team designed a virtual showroom app for iPad devices that gives the sales team immediate access to the latest product information. Real-time analysis of traffic patterns and product trends allows Bernhardt to now make rapid adjustments to product placement, pricing and availability status. The new tablet-oriented sales ordering process has established Bernhardt Furniture as both a fashion-forward and technology-leading company.

- USA Cycling: USA Cycling Women's Team Pursuit is using IBM Spark, Watson IoT, mobile and cloud to derive instantaneous insights leading to game-changing training strategies and racing tactics. The team can now get advanced analysis of rider data, calculate dynamic race positioning, and determine the grouping of riders over the race track.

- SETI: IBM, NASA and the SETI Institute are working together to analyze more than six terabytes of complex deep space radio signals to hunt for patterns that might identify the presence of intelligent extraterrestrial life. With IBM Analytics on Apache Spark, SETI has been able to embark on a new Stellar Pair Eavesdropping campaign which enables the organization to look for potential communications between planets that might be orbiting in double star systems. More than half of all stars are, in fact, these types of planets. By extracting new features from millions of observations, researchers are able to use machine learning techniques to classify signals and sharpen their focus for subsequent deep analysis on clusters of signals which are anomalous or outliers.

IBM continues to collaborate with leading data science organizations including Galvanize, H2O.ai, LightBend and RStudio to promote an integrated and unified data science ecosystem. Additionally, IBM is joining the R Consortium to help accelerate data science's readiness for the enterprise.

IBM is leading the way in the growing Analytics ecosystem having contributed to related projects including Apache Toree, EclairJS, Apache Quarks, Apache Mesos, Apache Tachyon now called Alluxio, and major contributions to Apache Spark sub-projects SparkSQL, SparkR, MLLib, and PySpark with over 3,000 total contributions in the last year.

In addition, IBM has built Spark into the core of its platforms including Watson, Commerce, Analytics, Systems, Cloud as well as more than 30 offerings including IBM BigInsights for Apache Hadoop, IBM Analytics on Apache Spark, Spark with Power Systems, Watson Analytics, SPSS Modeler and IBM Stream Computing. IBM also open-sourced its breakthrough SystemML machine learning technology to advance Spark's machine learning capabilities in 2015.

"Just as IBM played a critical role in the development of Computer Science, we can see many similarities today. Computer Science went mainstream with the introduction of the PC," said Picciano. "With Data Science, the major roadblock is having access to large data sets and having the ability to work with so much data. With today's announcement, clients can have both."

For more information on the IBM Data Scientist Experience and IBM Spark solutions, see datascience.ibm.com.

---

https://uk.newsroom.ibm.com/2016-Jun-07-IBM-Launches-Industrys-First-Development-Environment-for-Apache-Spark-Delivered-in-the-Cloud-for-Rapid-Adoption