IBM Expands Access and Value of z Systems Mainframe Data with Apache Spark

- New z/OS Platform for Apache Spark gives data scientists and developers real-time, simple and more secure access to mainframe data - Partner solutions help accelerate insights from structured and unstructured data to support clients' cognitive business transformations

ARMONK, N.Y. - 30 Mar 2016: IBM (NYSE: IBM) is making it easier and faster for organizations to access and analyze data in-place on the IBM z Systems mainframe with the new z/OS Platform for Apache Spark. This is creating new opportunities for data scientists and developers to apply advanced analytics to the system's rich data sets for real-time insights.

IBM z/OS Platform for Apache Spark enables Spark, an open-source analytics framework, to run natively on the z/OS mainframe operating system. The new offering, available now, enables data scientists to analyze data in place on the system origin, without the need to extract, transform and load (ETL), by breaking the tie between the analytics library and underlying file system.

In the cognitive era, where data is the new natural resource and computer systems are able to understand, reason and learn, businesses must be able to develop and capitalize on insights before they are no longer relevant. With this offering, which includes accelerators from z Systems business partners, organizations can more easily take advantage of z Systems data and capabilities to understand market changes and individualized client needs and make business adjustments in real-time, speeding time to value.

z Systems handles critical data and transactions for many of the world's major banks, insurers, retailers and transport companies. It features the industry's fastest commercial microprocessor and the ability to perform in-transaction analytics, scoring predictive models within a transaction in 2 milliseconds or less. Organizations can now leverage these capabilities, applying advanced in-memory analytics through Spark without moving data off the mainframe, saving time and money and limiting risk.

"As businesses of all sizes transform into real-time digital organizations, they must be able to get a clear picture of all their enterprise data without the excessive time and risk of ETL," said Rod Smith, IBM Fellow, Emerging Internet Technologies. "With Apache Spark enabled natively on IBM platforms – now including z Systems – customers can perform analytics alongside the transactional systems that house key data, while drawing contextual insights from other data sources, enabling them to engage with customers and generate revenue in real time."

IBM z/OS Platform for Apache Spark includes Spark open source capabilities consisting of the Apache Spark core, Spark SQL, Spark Streaming, Machine Learning Library (MLlib) and Graphx, combined with the industry's only mainframe-resident Spark data abstraction solution. The new platform helps enterprises derive insights more efficiently and securely through:

- **Streamlined development** – Developers and data scientists can use their existing expertise with programming languages such as Scala, Python, R and SQL to reduce time to value for actionable insights.

- **Simplified data access –** Optimized data abstraction services remove complexity, providing seamless access to enterprise data in traditional formats such as IMS, VSAM, DB2 z/OS, PDSE or SMF with familiar tools via Apache Spark APIs.

- **In-place data analytics** – Apache Spark uses an in-memory approach for processing data to deliver results quickly. The platform includes data abstraction and integration services that enable z/OS analytics applications to leverage standard Spark APIs. This allows the organization to analyze data in-place, avoiding costly processing and security considerations associated with ETL.

- **Open source capabilities –** The platform offers an Apache Spark distribution of the open source, in-memory processing engine that is designed for big data.

IBM is also working with three partners, DataFactZ, Rocket Software and Zementis, to create customized solutions using IBM z/OS Platform for Apache Spark:

- [DataFactZ](#) is a new partner that is working with IBM to develop Spark analytics based on Spark SQL and MLlib for data and transactions processed on the mainframe.

- [Rocket Software](#) has been a long-standing IBM partner and this collaboration extends to z/OS Apache Spark. For example, the new Rocket Launchpad solution will allow clients to try the platform using data on z/OS.

- [Zementis](#) is complementing its in-transaction predictive analytics offering for z/OS with a standards-based execution engine for Apache Spark. The solution allows users to deploy and execute advanced predictive models that can help them anticipate end users' needs, compute risk or detect fraud in real-time at the point of greatest impact, while processing a transaction.

The new z/OS Platform for Apache Spark and partner solutions will allow data scientists and data wranglers, who are charged with gathering data from different sources, to use the formats and tools they prefer to collect and analyze data.

IBM announced last year a [commitment to Spark](#) that included putting more than 3,500 IBM researchers and developers to work on projects related to the framework. As part of the commitment to advancing open source technologies for analytics on the mainframe, z Systems has also established a new [GitHub organization](#) for developers to collaborate and build tools around z/OS on Spark. For example, a combination of [Project Jupyter](#) and any NoSQL database can provide a flexible and extendible data processing and analytics solution.

This approach can help make modern open source tools more accessible by enabling developers to choose their tools and languages, providing new visual aids that monitor analytics results across disparate data environments and enabling modern data processing techniques and skills.

The IBM z/OS Platform for Apache Spark is now available for developers working with z/OS to download. To learn more about the IBM z Systems portfolio, visit http://www.ibm.com/systems/z/or the IBM Systems blog.

QUOTES

DataFactZ:

"With the proliferation of and access to data comes the proliferation of threats and vulnerabilities," said Krishna Kallakuri, CEO of DataFactZ. "Our solution for z/OS leverages Spark's MLlib to detect and identify real-time, POS transactions on the mainframe as fraudulent or not so customers can take better advantage of their data without increasing risks."

Rocket Software:

"Running Apache Spark natively on z/OS enables organizations to move their analytics closer to their data, instead of the other way around," said Bryan Smith Vice President of R&D and Chief Technology Officer for Rocket Software. "Customers interested in getting started with Apache Spark on z/OS can sign up for Rocket Launchpad, an engagement model designed to help organizations develop creative solutions to solve their most challenging data problems."

Zementis

"Apache Spark enables organizations to apply data science more efficiently. As cognitive capabilities become a key differentiator for smarter business applications, establishing a common platform and a common process to operationalize machine learning and advanced predictive analytics based on the Predictive Model Markup Language (PMML) industry standard is essential." said Dr. Michael Zeller, CEO of Zementis.

---

https://uk.newsroom.ibm.com/2016-Mar-30-IBM-Expands-Access-and-Value-of-z-Systems-Mainframe-Data-with-Apache-Spark