



News Release

Think Big, a Teradata Company Expands Capabilities for Building Data Lakes with Apache Spark

2016-04-13

Spark deployment challenges prompt rising demand for Teradata's big data services across the world

Teradata (NYSE: **TDC**), the **big data analytics** and marketing applications company, today announced that Think Big, a global Teradata consulting practice with leadership expertise in deploying Apache Spark™ and other big data technologies, is expanding its data lake and managed service offerings using Apache Spark. Spark is an open source cluster computing platform used for product recommendations, predictive analytics, sensor data analysis, graph analytics and more.

Today, customers can use a data lake with Apache Spark in the cloud, on general "commodity built" Hadoop environments, or with Teradata's Hadoop Appliance, the most powerful, ready-to-run enterprise platform, preconfigured and optimized to run enterprise-class big data workloads.

While interest in Spark continues to increase, many companies struggle to keep up with the rapid pace of change and frequency of releases of the open source platform. Think Big has successfully incorporated Spark in its frameworks for building enterprise-quality data lakes and analytical applications.

"Many organizations are experimenting with Apache Spark, in hopes of leveraging its strengths with streaming data, query, and analytics – often in conjunction with a data lake," said Philip Russom, Ph.D., director of data management research, The Data Warehousing Institute (TDWI). "But users soon realize that Spark is not easy to use and that data lakes take more planning and design than they thought. Users in this situation need to turn to outside help in the form of consultants and managed service providers who have a track record of success with Apache Spark and data lakes across a diverse clientele. Think Big has such experience."

Think Big is building replicable service packages for Spark deployment including adding Spark as an execution engine for its Data Lake and Managed Services offers. Through its training branch--Think Big Academy—the consultancy is also launching a series of new Spark training offers for corporate clients. Led by experienced instructors, these classes help train managers, developers, and administrators on using Spark and its various modules including machine learning, graph, streaming and query.

Also, Think Big's Data Science team will open source routines for distributed **K-Modes** clustering with Spark's Python application programming interface (API). These routines improve clustering of categorical data for customer segmentation and churn analysis. This code will be available with other Think Big open source efforts on Think Big's **GitHub** page.

"Our Think Big consulting practice is expanding quickly from the Americas across Europe and China because demand is exploding for the expertise, experience and methods to help companies get a data lake using Spark and Hadoop right, the first time," said Ron Bodkin, president of Think Big. "The deployment of Spark should be part of an information and analytics strategy. We know from experience what use cases are relevant, what the right questions are, and where to watch for deployment landmines. We understand business user expectations as well as technology requirements. We can help generate tangible business value, and our Spark customers are already doing so in domains ranging from omni-channel consumer personalization to real-time failure detection in high-tech manufacturing."

Long before big data buzz became trendy, Think Big was already the world's first and leading pure-play big data services firm, implementing analytic solutions based on emerging technologies. Today, Think Big provides managed services for Hadoop in the areas of platform and application support with well-defined processes, robust tools, and experienced big data experts to affordably manage, monitor, and maintain the Hadoop platform. Initiating each engagement with a well-tested transition process, Think Big assesses and improves a client's production support, development, and sustainment teams – for efficient, effective deployment.

Relevant News Links

- Think Big SPARK enablement services: For details visit the Think Big web page
- Teradata positioned as a Leader in the 2016 Gartner Magic Quadrant for Data Warehouse and Data Management Solutions for Analytics – **Get the new report here**

About Teradata

At Teradata, we believe that people thrive when empowered with trusted information. We offer the most complete

cloud analytics and data platform for AI. By delivering harmonized data and trusted AI, we enable more confident decision-making, unlock faster innovation, and drive the impactful business results organizations need most. See how at **Teradata.com**.

Media Contact

Jennifer Donahue

Teradata

Jennifer.Donahue@Teradata.com