



Delivering on Spark: Databricks Offers Simba-developed ODBC Connectivity, SQL Capability

Case Study: Databricks Chooses Simba for ODBC Connectivity in the Cloud

Databricks CEO Ion Stoica, CTO Matei Zaharia, and team spun Databricks out from the U.C. Berkeley AMPLab research center in 2013. Their mission was ambitious: Revolutionize what data analysts could do with Big Data. At AMPLab, they had collaborated to create the Apache Spark open-source data-processing engine for Hadoop. At Databricks, they have continued to strive to deliver unprecedented performance gains with Spark over the traditional Hadoop data access paradigm. The Databricks team—including customer engagement lead Arsalan Tavakoli-Shiraji—recognized the need to incorporate direct SQL-to-Spark connectivity, and chose Simba Technologies to provide it.

The Databricks story begins in Northern California: While at the University of California at Berkeley's AMPLab data-analytics research center, then-PhD student Matei Zaharia and professor Ion Stoica decided that they could create a faster data-processing engine to overcome what they saw as performance limitations in the Hadoop data-access model. They built Spark, the open-source cluster-computing framework for Hadoop. They introduced Spark to the world in 2009, open-sourced it in 2010, and turned it over to the Apache Software Foundation in 2013.

Later that year, the two—now re-titled as corporate CTO and CEO, respectively—joined with a cohort of their AMPLab colleagues to found Databricks and offer a commercial distribution for Spark, as well as related services. Along the way, they attracted interest, investment, and even a few board members from influential Silicon Valley venture capital firms NEA and Andreessen Horowitz. Databricks' mission was funded, unambiguous, and ambitious: Revolutionize how data analysts could extract more value from Big Data.

Tableau, Power BI, Excel, Microstrategy, Lumira, Sisense, & Other BI Tools

SQL Queries

ODBC Connectivity

databricks Spark

The Databricks Cloud stack

"When you look at the marketplace... [when] you talk to any of the people in the space, Simba is the answer if you want to provide that connectivity."

—Arsalan Tavakoli-Shiraji, Customer Engagement Lead, Databricks

The Challenge: SQL on Spark

Spark offers massive data-processing performance improvement (up to 100x) over MapReduce. The Databricks suite of offerings has grown to include the Databricks Cloud platform supported with Spark SQL (which supplanted Shark).

But as they assessed the data analytics market space, Stoica, Zaharia, and team realized that Spark—no matter the performance—would succeed only if analysts could

access Spark data using their preferred BI tools and methods.

"The second you try to tell people that they need to change their tools to use their underlying data store, you've lost," observes Databricks' Customer Engagement Lead Arsalan Tavakoli-Shiraji. "It was a no-brainer for us: If we want to get adoption going forward, we need to support these tools. It's just table stakes."

"Tableau Software said 'We want to release a Mac Spark driver in short order.' So we called Simba. Even Tableau was surprised they could deliver both the Windows and Mac drivers at launch...The fact that we could make something like that happen quickly...was terrific."
—Arsalan Tavakoli-Shiraji, Databricks

The Solution: Simba ODBC Connectivity

That realization meant ensuring Spark ODBC connectivity to BI tools like Tableau, Excel, Microstrategy, Qlik, and others. More importantly, those target analysts needed to be able to get at Spark data using the data-query approach with which they were most familiar: SQL. And that led Databricks to Simba Technologies.

"When you look at the marketplace...[when] you talk to any of the people in the space, Simba is the answer if you want to provide that connectivity," explains Tavakoli-Shiraji.

Databricks selected Simba Technologies to provide its ODBC 3.8 driver for Databricks Spark SQL, its powerful SQL-query engine for Apache Spark. Databricks now includes the driver free of charge in its Databricks Cloud platform. For Databricks, the Simba ODBC driver enables Databricks customers to query their Spark data with their preferred BI tool (like Tableau, Excel, Qlik, MicroStrategy, or others) on Windows, Linux, and Mac.

"One of the top most-requested features for Databricks Cloud is 'Plug in my BI tools to work with this,'" says Tavakoli-Shiraji. "What's the simplest way to get people going? It's ODBC connectivity [from Simba]. It's there, it's stable, it works."

What's Next: Extensibility

A development quick-turn turned into a valuable "big deal" for Databricks, and reinforced the company's commitment to Simba Technologies.

Requirements:

- ODBC Connectivity
- Direct SQL-to-Spark Capability
- Extensibility
- Windows, Linux, and Mac platform support

Solution:

- Simba Technologies ODBC Driver for Spark SQL and Databricks Cloud

"Tableau Software said to us, 'We want to release a Mac Spark driver in short order,' explains Tavakoli-Shiraji. "So we called Simba." The Simba dev team quickly extended the Databricks Spark ODBC Driver to Tableau for the Mac. And partner reaction was immediate (and positive). "Even Tableau was surprised they could deliver both Windows and Mac drivers at launch," continues Tavakoli-Shiraji. "The fact that we could make something like that happen quickly...was terrific."

Tavakoli-Shiraji concludes that Databricks Cloud users value its connectivity and compatibility with BI tools: "We've knocked down the barriers for less-technical users to be able to work with their data, even if they don't know about all the complexities that are being abstracted under the covers."

The Simba-developed Databricks ODBC Driver is available from Databricks. Learn more at <http://www.databricks.com>.