



Market Insight Report Reprint

Databricks details machine learning, data science and analytics strategy to target coders and noncoders

November 12 2021

by **Krishna Roy**

The company has evolved since it was founded in 2013 to build a business around the Apache Spark data-processing engine. Over the past couple of years, Databricks has focused in earnest on supporting machine learning, data science and analytics use cases for technical personnel with coding expertise.

451 Research

S&P Global

Market Intelligence

This report, licensed to Databricks, developed and as provided by S&P Global Market Intelligence (S&P), was published as part of S&P's syndicated market insight subscription service. It shall be owned in its entirety by S&P. This report is solely intended for use by the recipient and may not be reproduced or re-posted, in whole or in part, by the recipient without express permission from S&P.

Introduction

When 451 Research previously caught up with Databricks in June, the company announced enhancements related to machine learning, data management and data sharing. As we noted at the time, machine learning has always been a key functionality thanks to the underlying ML capabilities of the Apache Spark data-processing engine (as well as the MLflow open source machine learning platform), which Databricks continues to build a business around. However, the vendor has brought machine learning, data science and analytics to the fore in recent years using packaged offerings.

This year's debut of Databricks Machine Learning to provide a managed service in the cloud for experiment tracking, model training, and feature development and management, as well as feature and model serving, exemplifies this strategy. Furthermore, Databricks has inked its first two acquisitions, according to 451 Research's Knowledgebase, to cement its machine learning, data science and analytics approach, which places code generation at the heart of it – regardless of whether users adopt a code or graphical UI-based strategy for these projects.

THE TAKE

Databricks is looking to expand its addressable audience beyond its classic users – including data engineers and data scientists who are comfortable and familiar with code-based offerings – to include the larger pool of nonexperts that require machine learning, data science and analytics and aren't coders, yet need the underlying code from these projects for regulatory compliance, explainability and other corporate mandates. That's a differentiation as Databricks is one of the few vendors that has adopted this approach thus far. However, we expect more firms to follow suit as low-/no-code strategies, which have been embraced in other tech sectors, take hold in the data science and analytics realm.

That said, Databricks is amassing experience as a tuck-in acquirer, which it could put to further use to shore up its competitive standing. The company certainly has the funding in place to do so following an eye-popping \$1.6bn series H round in August, which increased its post-money valuation to \$38bn. Furthermore, Databricks is on the glide path to an IPO, as noted here, which is another milestone that would set the vendor apart from its data science and analytics rivals, which are mainly private entities.

Details

Databricks has shed further light on its strategy to support machine learning, data science and analytics use cases. The vendor notes that it has found that most organizations have different expertise levels when it comes to these disciplines, which it is aiming to address with a code-based and graphical UI-based approach. This strategy involves serving up code-first functionality for coders, such as data scientists and machine learning engineers, as well as drag-and-drop graphical UI capabilities for citizen data scientists and other nonexperts lacking coding skills. Equally important, Databricks offers access to the code generated under the hood for coders and noncoders to support its operationalization into production systems and other situations where code documentation is required.

Having embraced machine learning, data science and analytics using a mix of open source components and homegrown development, including Databricks AutoML, the company has added another string to its bow of late – M&A. Databricks SQL, which is designed for expert analysts that write queries in the SQL language to deploy for ad-hoc analytics, visual analysis and dashboard use cases, was obtained when it reached for Redash in June 2020.

Databricks' second purchase was in October, when it nabbed Germany-based startup 8080 Labs to get its its low-/no-code data science and analytics efforts underway. 8080 Labs was the company behind bamboolib, which is an open source graphical UI for the Pandas data analysis and manipulation tool written for the Python programming language, which is popular for data science – bamboolib also generates code under the covers by essentially exporting Python code in a process akin to recording macros in an Excel spreadsheet. The offering is available as a paid-for version with additional enterprise functionality so users don't have to share work, for example. The free bamboolib Community Edition will remain available on an unsupported basis until integration occurs.

The vendor plans to integrate bamboolib into the Databricks Lakehouse Platform, which is designed to combine the low-cost storage and agility benefits of a data lake with the data structure and management features of the data warehouse to support multiple workloads, including machine learning, analytics and data science. The goal of the integration is to serve data scientists and other experts who already use bamboolib for data exploration and data prep, as well as expand into targeting nonexperts by drawing on the user-friendliness it provides via a graphical UI. Furthermore, bamboolib's graphical UI will become the front end for Databricks AutoML and Databricks SQL to provide noncoders with access to their functionality through drag-and-drop capabilities – again with the option to access the underlying code, if required.

Databricks AutoML is currently designed for machine learning experts to quickly generate baseline machine learning models – aka, models that are relatively simple to set up and have a reasonable chance of producing good results – as well as notebooks. It is offered as part of Databricks Machine Learning, which, as we previously noted, involved pulling together existing capabilities into a dedicated offering targeted at data engineers and data scientists, as well as introducing AutoML and a feature store to provide a repository for all features associated with machine learning models. Additionally, Databricks Machine Learning features a cloud-native proprietary notebook – similar to Google Docs in its collaborative capabilities – as well as hooks into RStudio, Jupyter Notebooks and languages popular among data scientists and data engineers, including Scala and Python.

CONTACTS

The Americas

+1 877 863 1306

market.intelligence@spglobal.com

Europe, Middle East & Africa

+44 20 7176 1234

market.intelligence@spglobal.com

Asia-Pacific

+852 2533 3565

market.intelligence@spglobal.com

www.spglobal.com/marketintelligence

Copyright © 2021 by S&P Global Market Intelligence, a division of S&P Global Inc. All rights reserved.

These materials have been prepared solely for information purposes based upon information generally available to the public and from sources believed to be reliable. No content (including index data, ratings, credit-related analyses and data, research, model, software or other application or output therefrom) or any part thereof (Content) may be modified, reverse engineered, reproduced or distributed in any form by any means, or stored in a database or retrieval system, without the prior written permission of S&P Global Market Intelligence or its affiliates (collectively, S&P Global). The Content shall not be used for any unlawful or unauthorized purposes. S&P Global and any third-party providers, (collectively S&P Global Parties) do not guarantee the accuracy, completeness, timeliness or availability of the Content. S&P Global Parties are not responsible for any errors or omissions, regardless of the cause, for the results obtained from the use of the Content. THE CONTENT IS PROVIDED ON "AS IS" BASIS. S&P GLOBAL PARTIES DISCLAIM ANY AND ALL EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, ANY WARRANTIES OF MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE OR USE, FREEDOM FROM BUGS, SOFTWARE ERRORS OR DEFECTS, THAT THE CONTENT'S FUNCTIONING WILL BE UNINTERRUPTED OR THAT THE CONTENT WILL OPERATE WITH ANY SOFTWARE OR HARDWARE CONFIGURATION. In no event shall S&P Global Parties be liable to any party for any direct, indirect, incidental, exemplary, compensatory, punitive, special or consequential damages, costs, expenses, legal fees, or losses (including, without limitation, lost income or lost profits and opportunity costs or losses caused by negligence) in connection with any use of the Content even if advised of the possibility of such damages.

S&P Global Market Intelligence's opinions, quotes and credit-related and other analyses are statements of opinion as of the date they are expressed and not statements of fact or recommendations to purchase, hold, or sell any securities or to make any investment decisions, and do not address the suitability of any security. S&P Global Market Intelligence may provide index data. Direct investment in an index is not possible. Exposure to an asset class represented by an index is available through investable instruments based on that index. S&P Global Market Intelligence assumes no obligation to update the Content following publication in any form or format. The Content should not be relied on and is not a substitute for the skill, judgment and experience of the user, its management, employees, advisors and/or clients when making investment and other business decisions. S&P Global Market Intelligence does not endorse companies, technologies, products, services, or solutions.

S&P Global keeps certain activities of its divisions separate from each other in order to preserve the independence and objectivity of their respective activities. As a result, certain divisions of S&P Global may have information that is not available to other S&P Global divisions. S&P Global has established policies and procedures to maintain the confidentiality of certain non-public information received in connection with each analytical process.

S&P Global may receive compensation for its ratings and certain analyses, normally from issuers or underwriters of securities or from obligors. S&P Global reserves the right to disseminate its opinions and analyses. S&P Global's public ratings and analyses are made available on its websites, www.standardandpoors.com (free of charge) and www.ratingsdirect.com (subscription), and may be distributed through other means, including via S&P Global publications and third-party redistributors. Additional information about our ratings fees is available at www.standardandpoors.com/usratingsfees.