

DataStax stays the open core course as it reconnects with the Apache Cassandra community

Publication Date: 18 Dec 2018 | Product code: INT002-000204

Tony Baer



Ovum view

Summary

With its latest open source and commercial releases, DataStax is reaffirming its relationships with the open source community as it firms up its open core product direction. It has made two releases over the past month: its latest open source distribution of Cassandra and the latest update to its commercial DataStax Enterprise offering. DataStax is walking the fine line that is typical of open source database platforms for differentiating and protecting its intellectual property. The recent experiences of open source database companies MongoDB and Redis prove that there is no one-size-fits-all solution to this challenge. With DataStax Enterprise 6.7, the company is doubling down on open core with continued support of the Cassandra open source project and forging ahead with adding its own graph, analytics support, and Docker integration. The key challenge at this point is extrinsic to DataStax: the Apache open source community needs to mobilize, filling the roles formerly led by DataStax employees as the project focuses on its own path.

Reconnecting the paths

The Apache Cassandra open source wide column NoSQL database has been differentiated by its ability to support extreme scale and performance; it is typically deployed as an operational database. Apache Cassandra has garnered significant third-party support and adoption; for instance, Pegasystems has built its AI-enhanced customer experience management system around Apache Cassandra as the data lake, because of the platform's flexibility, scale, and performance.

Like MongoDB and Redis, Cassandra was developed by creators who went on to form their own companies to commercialize support and develop enterprise platforms. But unlike the other platforms, where the project (and company) founders retained control of the open source project, Cassandra remains a community project governed by the rules of the Apache Foundation and carries an Apache software license. As such, the challenges for DataStax differed from those of MongoDB and Redis, which recently developed new open source licenses designed to limit the ability of third parties to monetize the platforms without contributing to the open source communities. The context is different for Apache Cassandra and DataStax the company. Cassandra is a community project that is governed not by a single vendor, but by the community, and with the Apache license, there are no commercial restrictions on third parties for monetizing it.

Several years ago, DataStax began charting a different path from the Apache community, putting greater emphasis on developing its own platform. It acquired Aurelius, which provided the technology that is allowing DataStax to extend its enterprise platform with graph database support, and DataScale, which led to the introduction of DataStax Managed Cloud. The fruits of those efforts include the latest DataStax Enterprise (SE) 6.7 release, which is discussed in more detail below.

Given the reality that the market for databases for the hybrid cloud is getting increasingly competitive, Ovum believes that DataStax and the Apache Cassandra community need each other. Having previously ramped down development activity in Cassandra, DataStax is shifting gears and putting more skin back in the game. For instance, it is bringing back its conference for the Cassandra user community conference, which will be called DataStax Accelerate, next spring. And it is ramping up

contributions to the project, but not reclaiming leadership, to keep the project solidly in community hands.

That set the stage for DataStax's release last month of its DataStax Distribution of Apache Cassandra. The DataStax distro will add several features beyond the core open source stack, including drivers, bulk load utilities, and business day commercial support. So why is DataStax releasing a commercial distribution of the pure Apache platform? It fills a gap, providing a commercially supported offering for customers that want the open source version rather than the full DataStax platform.

The core Apache Cassandra platform has matured to the point where it is now well-rounded – the basic features required for an operational database are in place. For the community, it is a matter of filling out the project leadership, forming consensus in the absence of a single driving force, and ensuring that new features are vetted and tested on a timely schedule. Whereas DataStax formerly dominated contributions, today they are coming from a more diverse group that includes Uber, Netflix, Facebook, and Apple.

We expect that with more diverse leadership, consensus on direction won't be as efficient as it was when a single party was the primary driver. DataStax is still playing a role, especially with documentation and training. And in the upcoming Apache Cassandra 4.0, for which code was frozen back in September, it is leading the effort to replace Apache Thrift (a lightweight language for writing remote procedure calls to databases) with the more SQL-like CQL as the query language. Nonetheless, both sides still have much more work to do to present a united front in this increasingly competitive part of the database market.

DataStax 6.7

As the follow-up to DSE 6.0, the 6.7 release is a jump in the numbering scheme so that all the components are now versioned in common. The highlights of the new release include better interoperability with Spark for faster analytics, enhanced geospatial indexing, Kafka integration, and incremental upgrades in cloud backup and restore and performance diagnostics.

In more detail, DSE is adding arms-length support for Spark integration. That is done by using dedicated specific nodes in a Cassandra cluster in conjunction with the DataStax Enterprise File System (DSEFS). Each DSEFS-enabled node can handle up to 40TB when run for analytics. As an operational database, DataStax is hardly alone in adding Spark connectors; the fact that it can handle multiple terabytes of data makes it an attractive analytics data source. Spark connectors are becoming checkbox items for platforms like Splice Machine, MongoDB, and Couchbase. Kafka and Docker integration are key to making DSE a target for ingesting streaming data, while Docker support provides a pathway for customers to support cloud applications. The next logical step for DSE – Kubernetes support – will await a future release.

Other upgrades are more modest. Enhanced geospatial indexing is a first step toward exposing DSE data to location-based use cases such as supply chain, logistics, and IoT; the next step will be supporting geospatial as native data types. For performance monitoring, new agentless data collectors and a console have been introduced; we expect that the next dot-zero release will include more granular performance management capabilities that will, hopefully, incorporate machine learning for not only exposing bottlenecks, but helping resolve or nip them in the bud.

This release, which builds atop the previous release's emphasis on manageability and extensibility with connectivity, is an important stepping stone to what we expect will be in the next major upgrade: enhancing DataStax's embrace of hybrid cloud architecture to provide maximum flexibility for deployment, and edge capabilities that we would like to see for bolstering IoT applications.

Appendix

Further reading

"DataStax Enterprise 6 extends Apache Cassandra," INT002-000091 (April 2018)

On the Radar: DataStax, IT014-002864 (December 2013)

"DataStax adds JSON and graph computing to its Cassandra distribution," IT0014-003141 (July 2016)

Open Source and Big Data, IT0014-003193 (February 2017)

Author

Tony Baer, Principal Analyst, Information Management

tony.baer@ovum.com

Ovum Consulting

We hope that this analysis will help you make informed and imaginative business decisions. If you have further requirements, Ovum's consulting team may be able to help you. For more information about Ovum's consulting capabilities, please contact us directly at consulting@ovum.com.

Copyright notice and disclaimer

The contents of this product are protected by international copyright laws, database rights and other intellectual property rights. The owner of these rights is Informa Telecoms and Media Limited, our affiliates or other third party licensors. All product and company names and logos contained within or appearing on this product are the trademarks, service marks or trading names of their respective owners, including Informa Telecoms and Media Limited. This product may not be copied, reproduced, distributed or transmitted in any form or by any means without the prior permission of Informa Telecoms and Media Limited.

Whilst reasonable efforts have been made to ensure that the information and content of this product was correct as at the date of first publication, neither Informa Telecoms and Media Limited nor any person engaged or employed by Informa Telecoms and Media Limited accepts any liability for any errors, omissions or other inaccuracies. Readers should independently verify any facts and figures as no liability can be accepted in this regard – readers assume full responsibility and risk accordingly for their use of such information and content.

Any views and/or opinions expressed in this product by individual authors or contributors are their personal views and/or opinions and do not necessarily reflect the views and/or opinions of Informa Telecoms and Media Limited.

CONTACT US

ovum.informa.com

askananalyst@ovum.com

INTERNATIONAL OFFICES

Beijing

Boston

Chicago

Dubai

Hong Kong

Hyderabad

Johannesburg

London

Melbourne

New York

Paris

San Francisco

Sao Paulo

Shanghai

Singapore

Sydney

Tokyo

