REPORT REPRINT

DataStax continues to invest in enterprise version as the gap widens with Apache Cassandra

JAMES CURTIS

04 MAY 2018

The company's latest release, which focuses on performance and availability, among other updates, provides ample evidence that the gap between open source Apache Cassandra and DataStax's enterprise version is growing ever wider.

451 Research®

 |

In our most recent coverage of DataStax, we described a company transitioning from serving as the primary supporter of Apache Cassandra to one channeling much of its development efforts into its own enterprise version, known as DataStax Enterprise (DSE). Its most recent version of that offering is the first major release since the company publicly announced its shift in focus. With this latest release, DataStax has added faster reads/writes, along with enhancements in availability and reduced administrative efforts, among other updates.

THE 451 TAKE

DSE 6 represents DataStax's first major release since the company declared that it was going to invest more in its enterprise version than supporting open source Apache Cassandra. And based on version 6's added functionality, the gap is certainly widening between DSE and Cassandra. Incorporating the thread-per-core architecture, data loader and NodeSync are examples of differentiation. While the Cassandra community's primary contributor has reduced its commits to the Apache Cassandra project, the move for DataStax appears so far to be going well as it has posted strong revenue. The competition, however, remains stiff for DataStax as several NoSQL vendors are likewise targeting similar use cases and updates from traditional relational database providers are being added as well.

CONTEXT

DataStax is a well-known player in the broader database market. Early on, the company was known for serving as the commercial supporter of Apache Cassandra, a wide-column database considered part of the NoSQL family of databases. However, last year, DataStax shifted its focus from pouring development efforts into the open source version of Cassandra to instead concentrating its efforts on its enterprise version. While Apache Cassandra does – and will continue to – form the core of the company's enterprise offering (DataStax Enterprise), the disparity between open source Cassandra and DSE will only get wider as time goes on. In other words, the majority of enterprise functionality that would previously have been contributed back to the community will only be available from DataStax. In fact, that is already taking shape with the recent release of DSE 6, described in more detail later in this report.

The company is privately held and while it has not previously disclosed financial metrics, it has reported that the shift in focus along with other strategic pivots, such as changing its go-to-market emphasis, have gone well. Management reports that DataStax has surpassed the $100m revenue mark and is on track to be cash-flow positive. Funding-wise, the company has raised $189m to date and has 480 total employees. DataStax does not disclose customer count and part of the reasoning for that is that it is looking to expand its partner or OEM-based program to work with SIs to bundle its products as part of their services. While that strategy is growing, its primary roadmap is direct to large enterprises.

The company points out two general uses cases that it focuses on. One is customer experience (CX), which includes customer 360, personalization and recommendations, loyalty, and consumer fraud detection. The other is what DataStax calls enterprise optimization, which includes use cases such as e-commerce, inventory and asset management, logistics, compliance, supply chain management, and identity management.

PRODUCTS

Our prior coverage of DataStax in September 2017 addressed its change in strategy as well as the release of DataStax Enterprise 5.1, which was made generally available in April 2017. That being the case, the company's latest DSE 6 marks a year since the prior release.

As such, DataStax has added several updates to DSE 6 that are worth highlighting. Performance updates, for instance, have been an area of emphasis. The company has implemented a thread-per-core architecture over Apache Cassandra's staged, event-driven architecture that provides twice the read and write performance as the previous version. Further, it has introduced a bulk loader feature that enables fast loading of data into and out of DSE to other data targets. Similarly, DataStax claims a 4x performance boost over previous versions.

Driving greater availability, or sometimes referred to as node or server repair operations, the company has unveiled its NodeSync functionality. NodeSync kicks in when a node goes down. Previously, syncing nodes was done manually or with scripting, but with NodeSync, the process is automated and transparent to the user. Security improvements have also been introduced, including separation of duties, as well as the ability to audit by role.

For analytical operations, DataStax has made improvements to the DSE Spark Connector. The company was responsible for the development of the initial Spark connector for Cassandra but for DSE 6, it has added some optimizations and enhancements that include providing 'always on' functionality such that if a node goes down, the query will continue to run.

Graph is also an ongoing effort for DataStax since its acquisition of Aurelius in 2015. At a high level, the company is working to make graph more natively integrated within DSE in terms of how data is loaded and queried. While that vision is not fully realized now, it did improve the DSE Graph engine such that it can more efficiently traverse graph data without as much user input.

Finally, and more specifically related to the administrative side, DataStax has automated how it delivers and carries out system patches. Previously a manual exercise, the new functionality downloads the patches and applies them in a rolling restart fashion with no expected downtime. Further, the company's DataStax Studio product supports SparkSQL, thus enabling more advanced analytical queries to be executed employing the SQL query language.

COMPETITION

While Apache Cassandra continues to sit at the core of DataStax Enterprise, the open source version of Cassandra continues to be the company's foremost competitor. DSE is considered a multi-model database, but organizations leveraging open source Cassandra would need to utilize other open source tools or build their own functionality to match the capabilities in DSE.

Given DataStax's use case focus, we can identify a handful of other NoSQL database vendors that would compete with DataStax, both cloud-only providers as well as firms offering on-premises and cloud products. For instance, DataStax notes that it encounters Microsoft's Azure Cosmos DB and AWS's DynamoDB, particularly as it relates to DataStax's Managed Cloud service, which is available on AWS and Azure.

Other NoSQL vendors with on-premises and cloud offerings include Couchbase, which is also multi-model. The company has been quite active in positioning its database as an 'engagement database' that likewise targets web and mobile scenarios with a CX focus. Aerospike similarly positions itself as a system of engagement based on a key-value data model and is differentiated by its SSD, flash-based architecture. Redis Labs, known for its in-memory capabilities based on a key-value model, drives a flexible platform offering with its Modules strategy, where graph is available as a module. MongoDB is often cited as a primary rival to DataStax and offers multiple data models, although it is more based on the document model and may not be a true fit in terms of use cases that both companies are chasing.

A few other NoSQL providers with multiple data models that are potential competitors to DataStax include ArangoDB, MapR-DB, OrientDB and InterSystems with Cache, a non-relational database. And then there is ScyllaDB, which has been quite active in positioning itself as a drop-in replacement for Apache Cassandra.

Traditional relational database specialists continue to vie with DataStax as well. Specifically, the company notes that it sees Oracle Database, Microsoft SQL Server, PostgreSQL and MySQL, among others. DataStax points out that its masterless architecture is often a point of differentiation. Also worth mentioning is Instaclustr, which offers a managed cloud service for Apache Cassandra and recently increased its commitment to the open source project.

SWOT ANALYSIS

STRENGTHS

While Apache Cassandra is at the heart of DataStax Enterprise, the company has developed and continues to add to its collection of tools and capabilities that are not just built on top of but are also natively integrated with the Cassandra database.

WEAKNESSES

There is perhaps still some ongoing confusion surrounding the company's relationship with Cassandra, although that appears to be dissipating as DataStax releases updates to its proprietary enterprise offering.

OPPORTUNITIES

Distributed digital transformation projects and applications – those that require a mix of workloads, scalability and availability – are a good fit for DataStax. These efforts can come from either net new applications or from enterprise applications that may have been initially powered by relational databases.

THREATS

While the gap between open source Cassandra and DataStax is widening, there are other entities that support and are building tools on top of Cassandra that overlap with DSE. In addition, there are several NoSQL database vendors that are focusing on similar use cases.