



Impact Report

DataStax expands Spark support as Cassandra matures as a distributed data platform

Analyst: [Matt Aslett](#) 19 Dec, 2014

DataStax recently updated its DataStax Enterprise offering, based on the Apache Cassandra distributed NoSQL database, to version 4.6, including support for Apache Spark 1.1 and Apache Spark Streaming for in-memory analytics and stream processing. As such, DataStax Enterprise is becoming a multipurpose data-processing platform, one that presenters at the recent Cassandra Summit Europe proved is being used to support critical applications in a number of industries such as gaming, utilities, social networking, the Internet of Things and financial services.

The 451 Take

There was an impressive array of enterprise use cases on display at Cassandra Summit Europe, including the Web-based properties that were early Cassandra adopters and – perhaps more importantly – established enterprises that are seeking alternatives to the incumbent relational providers as they seek to define their future data-processing platforms. With the likes of ING proving (at least to itself) that when appropriately deployed, Cassandra's masterless architecture can be used to support active-active clustering while also maintaining consistency, expect increasing numbers of enterprises to be attracted to DataStax and Cassandra given the combination of commercial support, open source community-led development, and distributed architecture.

Context

The Apache Cassandra distributed database was originally created by Facebook to power its inbox search. It has come a long way since then, having been released under the Apache open source license in 2008 and become a top-level Apache Software Foundation project in February 2010. Early adopters included Web properties such as Digg, Rackspace, Twitter and eBay while perhaps the best-known user today is Netflix, which relies on Cassandra to deliver its subscription video-on-demand services.

DataStax was founded in 2010 by Apache Cassandra project chair Jonathan Ellis and fellow former Rackspace employee Matt Pfeil to provide commercial support and services for Cassandra and has similarly come a long way since then. Now led by CEO Billy Bosworth (Ellis and Pfeil remain with the company as CTO and chief customer officer, respectively), the company has [raised \\$190m in funding](#), has 400 or so employees and boasts a similar number of subscription customers for its DataStax Enterprise distribution, which offers additional capabilities above and beyond Cassandra – not least integrated search, analytics and in-memory processing capabilities.

Customers

The recent Cassandra Summit Europe event provided more evidence of how the core capabilities of Cassandra and DataStax Enterprise are driving adoption in a variety of industries and not just for 'lightweight' customer-facing applications such as search but also mission-critical run-the-business apps. As noted, Web properties such as Netflix and eBay already depend on DataStax/Cassandra and attendees heard how Russian social networking site Ok.ru also relies on Cassandra, having migrated its systems from Microsoft SQL Server as part of a move toward commodity hardware, open source and multi-datacenter high availability.

These criteria were among the list of requirements that cropped up again and again at Cassandra Summit Europe as users discussed their reasons for selecting Cassandra. Similarly, DemonWare, which provides online software and services to support game development parent company Activision Blizzard, turned to Cassandra to replace its existing MySQL-based architecture as it sought to deal with the success of titles such as *Call of Duty* and *Skylanders*.

Cassandra's peer-to-peer architecture was particularly important for DemonWare as the former architecture was feeling the strain of players wanting to maintain their profiles and keep track of progress as they moved between games and platforms, which the existing approach of sharding for different games and platforms could not handle.

It's not just Web properties that are adopting Cassandra, however. Another example came from French shipping and delivery firm Chronopost, which

has adopted Cassandra in place of Oracle Database in order to meet its requirements of multi-datacenter availability and the ability to scale to meet demand at peak times. UK-based energy provider British Gas was also at the event to explain how Cassandra and DataStax are playing a key role in its Connected Homes initiative, providing the underlying data processing to support its Hive-connected thermostat as well as Smart Energy reports and the nascent Connected Boiler program, which will ultimately be used to deliver predictive boiler maintenance to the company's heating customers.

Hive is currently used by 100,000 British Gas customers to control their central heating from their cell phones and is an emerging Internet of Things use case. Another was presented by water pressure management firm i2O, which has deployed Cassandra (replacing Microsoft SQL Server) as part of its combination of hardware devices and a cloud-based software platform that is used to monitor and control water pressure within a delivery network and has been deployed to save an estimated 235 million liters of water per day by preventing leakage.

In addition to commodity hardware, open source and multi-datacenter high availability, the aforementioned Ok.ru was also seeking SQL support and ACID transactions, not things that Cassandra is known for as a NoSQL database. However, Ok.ru found that in its case the SQL-like Cassandra Query Language (CQL) served its purpose and claims to have been able to solve the challenge of supporting ACID transactions by implementing its own C*One storage engine.

Ok.ru isn't the only company deploying Cassandra in transactional environments. Credit Suisse briefly outlined how it has been using the NoSQL database for five years as part of a strategy to create a data-processing platform for the next 20 years, including supporting its Hippo risk management platform. Additionally, Dutch financial services firm ING explained how it has proven that Cassandra can fulfill its needs of a scalable, consistent data-processing platform that is active-active across multiple datacenters without a single point of failure.

While ING is not at this point in production (the presenters had only first started to explore the potential use of Cassandra by attending the same event a year before), it has proven to the satisfaction of the CTO that when appropriately deployed, Cassandra's masterless architecture and CQL can be deployed to support active-active clustering while also maintaining consistency.

Strategy

As noted, Cassandra's core masterless architecture and open source development were highlighted as key adoption drivers. However, the existence of DataStax as a commercial support provider was also referenced by a number of users. The company may have started out providing commercial support services but is now best described as an advanced distribution provider with its DataStax Enterprise offering additional capabilities such as the full version of its OpsCenter management console, enterprise search capabilities based on Apache Solr, the option to run Cassandra in-memory, and commercial support for the Apache Spark in-memory data-processing engine.

DataStax used Cassandra Summit Europe to announce the release of DataStax Enterprise 4.6, which updated Spark support to version 1.1, including support for the Spark Streaming micro-batch stream-processing technology in addition to existing support for Spark SQL (and its predecessor, Shark). DataStax Enterprise 4.6 also added integration with LDAP and Active Directory.

Balancing the commercial and community responsibilities will be key to the long-term success of both DataStax and Cassandra, and it is interesting to note that while the company is the primary contributor to Cassandra, the development team led by CTO and Apache Cassandra project chair Jonathan Ellis is quite separate from the DataStax Enterprise development, testing and support team, which is led by EVP of engineering Martin van Rysyck.

Competition

There are multiple databases that might be considered by potential Cassandra and DataStax Enterprise adopters but the most repeatedly mentioned by Cassandra Summit Europe was Basho's Riak, which we would expect given that it has a similar masterless architecture and underlying key value data model. We would also anticipate would-be users to be potentially evaluating the likes of Aerospike and Amazon Web Services' DynamoDB, as well as possibly MongoDB, Couchbase, MarkLogic, Apache HBase, FoundationDB and Aerospike.

All of the above are NoSQL databases. However, as the examples from Cassandra Summit Europe proved, Cassandra and DataStax Enterprise are also being used to replace applications previously deployed on existing relational, SQL, databases. Oracle is the most likely to be the displaced provider thanks to the popularity of its Oracle Database and MySQL offerings, while Microsoft, IBM and SAP are also among the incumbent giants. Depending on the level of transactional requirements, adopters might also look at the new breed of NewSQL relational database vendors such as NuoDB, MemSQL, VoltDB and TransLattice.

SWOT Analysis

Strengths

The adoption of DataStax and Cassandra has been driven by the combination of commercial support and open source community-led development, as well as a distributed architecture.

Opportunities

We believe the long-term move toward distributed architecture will encourage many enterprises to reconsider their database requirements, with Cassandra providing an increasingly mature alternative.

Weaknesses

Cassandra is disruptive to the database establishment and can be viewed with suspicion by administrators and support staff that have been raised with and rely on relational databases.

Threats

The company faces no shortage of competitors, including rival NoSQL firms, emerging NewSQL experts and the incumbent database giants.

This report falls under the following categories. Click on a link below to find similar documents.

Company: [DataStax](#)

Other Companies: [Activision Blizzard](#), [Aerospike](#), [Amazon](#), [Apache Software Foundation](#), [Amazon Web Services](#), [Basho](#), [British Gas](#), [Chronopost](#), [Couchbase](#), [Credit Suisse](#), [Digg](#), [eBay](#), [Facebook](#), [FoundationDB](#), [Hippo](#), [i2O](#), [IBM](#), [ING Groep](#), [MarkLogic](#), [MemSQL](#), [Microsoft](#), [MongoDB](#), [Netflix](#), [NuoDB](#), [Ok.ru](#), [Oracle](#), [Rackspace](#), [SAP](#), [TransLattice](#), [Twitter](#), [VoltDB](#)

Analyst(s): [Matt Aslett](#)

Sector(s):

[Information management / Data management / Non-relational databases](#)

[Information management / Data management / Relational databases](#)

Copyright © 2000-2014 The 451 Group. All Rights Reserved.