Bloor

InBrief

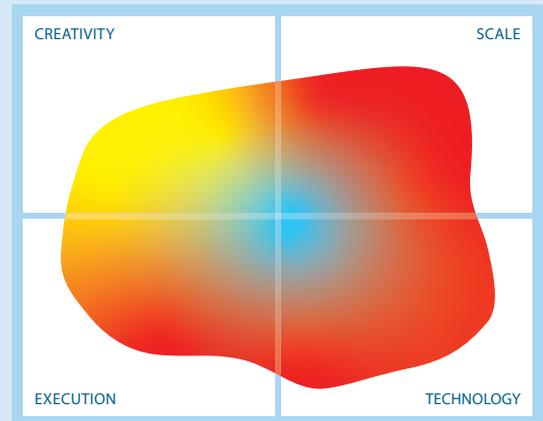**Philip Howard –** Research Director, Information Management

DATASTAX

DataStax Enterprise

The company

DataStax is a database vendor that was founded in 2010. Its primary offering is DataStax Enterprise (DSE), the leading database built using Apache Cassandra. In 2015, the company acquired Aurelius, the chief developers of the Titan graph database, and have since leveraged their expertise to develop DSE Graph, a property graph database add-on for DSE. DSE is available on-premises, in-cloud, or as part of a hybrid solution, and is additionally deployable as a Docker container. The company has customers using Kubernetes, but this is not formally supported at present (it is planned for later in 2019). DataStax is headquartered in Santa Clara, CA, and has additional US offices in San Francisco, Austin, and Atlanta, as well as international offices in the UK, France, Germany, Ireland, Japan and Australia. As of this writing, DataStax has more than 600 employees, ~250 of which are dedicated engineering or product staff.

> " Search and analytics were some of the key capabilities we were looking for and with DataStax Enterprise, we got a unified platform that provides all these and more all in the same cluster. This was a significant reason why we chose DataStax Enterprise to power our app. "
>
> **You Are My Guide**

Figure 1 – Showing how DataStax is built on top of Cassandra

The image in this Mutable Quadrant is derived from 13 high level metrics, the more the image covers a section the better. Execution metrics relate to the company, Technology to the product, Creativity to both technical and business innovation and Scale covers the potential business and market impact.

What is it?

DSE is a distributed NoSQL database, using CQL (Cassandra Query Language), that is oriented towards (though not exclusive to) cloud and hybrid-cloud architectures. It is built on top of Cassandra, as illustrated in *Figure 1*. It boasts numerous capabilities above and beyond what Cassandra alone offers, including native search and analytics, auto-management functionality, and significant increases to speed and performance.

As can be seen in this diagram, DSE provides multi-model capabilities and, unlike some other multi-model products you can leverage all of the models, not just within a single database instance but also within a single query. For example, the optimiser can automatically invoke Spark or search (Solr) from a Gremlin (graph) query. This has the advantage that if you are a Gremlin or CQL developer you don't need to know or understand Spark (or Solr). One possible limitation is with respect to document model implementations where DataStax requires that a schema is defined.

Note that from the perspective of supporting hybrid processing environments DataStax takes the view that this should not only encompass analytic and transactional processing but also search.
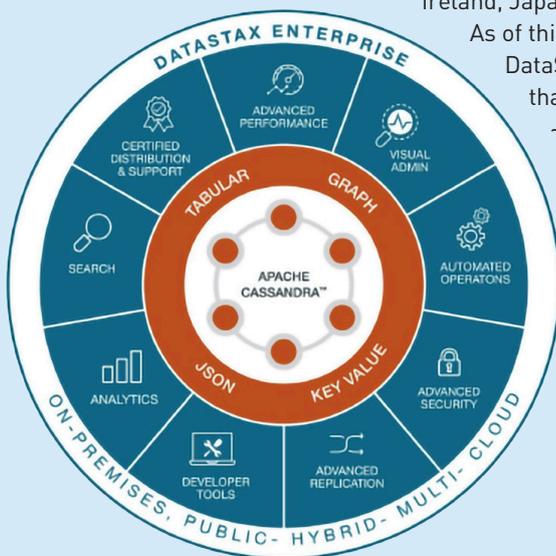
Architecture	★★★★		Flexibility	★★★★
Concurrent analytics	★★★★		In-process analytics	★★★★
Criticality	★★★★		Performance	★★★★
Ease of use	★★★★		Scale	★★★★

> **"** The key benefit of using DSE is the co-location of data and technology with Cassandra and Solr for search and Cassandra with Spark for analytics. This results in the real-time nodes having access to data instantly and not requiring time-consuming or costly ETL processes to move data between systems, because all the data is transparently replicated in the cluster. **"**
>
> **Macquarie**

How does it work?

Architecturally, the most notable feature of DSE is that it uses a master-less architecture in which all nodes are the same, with the result that there is no single point of failure. This particularly suits environments where you want to deploy across multiple clouds or in hybrid on-premises and cloud deployments. It also suits the way that DataStax supports workload management, which is illustrated in *Figure 2*.

As can be seen, you can support any workload within a node, you can specify that a particular node has a specific task or you can have clusters – (elastically) scalable individually - dedicated to a particular task, or you can mix and match these.

From a transactional standpoint the database supports the atomicity, isolation and durability of ACID guarantees but tuneable consistency.

Figure 2 – How DataStax supports workload management

The latter is enabled by choosing to use either asynchronous or synchronous replication. The former provides eventual consistency and the latter immediate consistency but with the trade-off of reduced performance.

As far as analytics and search are concerned the company offers specific enterprise components known as DSE Analytics and DSE Search, which work in conjunction with both DSE itself and DSE Graph. As mentioned, DSE Analytics is integrated with Spark and the company claims that DSE Analytics is

significantly faster than open source Spark. The product also supports Python and it has customers using both R and TensorFlow though these are not formally supported as yet. PMML (predictive modelling mark-up language) is not supported. It is worth also noting that DSE Graph in and of itself boasts some significant differentiators. This includes its dual processing engines, allowing you to easily switch between transactional and analytical processing, and DataStax Studio, a particularly impressive example of a visual development environment for graph.

Finally, it is worth commenting on DSE's Kafka integration, which enables data to be streamed into the DSE environment. This is currently only a one-way process, but the company plans to support export to Kafka in a future release.

Why should you care?

Cassandra initially made its name as a NoSQL database because it was designed from the outset to support key enterprise requirements such as constant availability, resilience, and disaster recovery, as well as scalability. Many other NoSQL databases did not start from this position and only added mission-critical capabilities – if they did – later. We prefer the approach taken by the developers of Cassandra. Moreover, in DSE there are substantial additional elements that go beyond Cassandra itself, some of which are at the feature level and some of which, such as the multi-model support, and the search and analytics capabilities, are more substantial.

The Bottom Line

DSE is almost unique in supporting both graph and conventional analytics alongside transactional processing and search. No other company we have spoken to sees hybrid processing as a three-way (transactions, analytics and search) environment, and we think DataStax's approach makes a lot of sense.
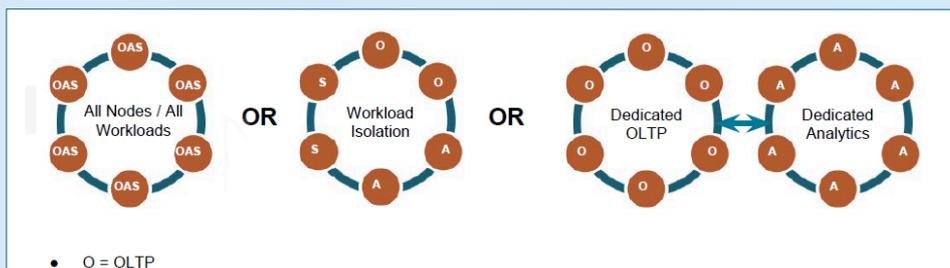
FOR FURTHER INFORMATION AND RESEARCH CLICK HERE