DATASTAX

FEATURE IN FOCUS

DataStax Enterprise – Analytics

To effectively compete in today's Internet economy, modern enterprises need to utilize the data from their operational database systems in ways that provide the fastest possible time-to-insight so they can quickly make decisions that benefit their customer and business. This translates into being able to smartly run analytic tasks on data contained in their operational/transactional online applications as well as being able to integrate that data easily with historical information contained in data warehouses.

DataStax Enterprise supplies built-in analytic functionality so operational Cassandra data can easily be analyzed and used to make important business decisions, along with external connectivity to Hadoop data warehouses / lakes that allows for simplified integration of hot transactional and cold historical data.

DataStax Enterprise delivers constant uptime and linear scale performance for online applications needing transactional, analytical, search, and in-memory workload support in a single platform.

Transactional Analytics

DataStax Enterprise supports what is sometimes referred to as "trans-lytics" or HTAP (hybrid transactional analytical processing). You can run transactional and analytical workloads together on the same database cluster on the same data without resource contention. This allows for quick analysis of hot data needed to influence customers via recommendations and navigation suggestions in web and mobile applications.

Streaming and Near/Real-Time Analytics

With its Apache Spark™ integration, DSE supports use cases needing streaming and near/real-time analytics on incoming transactional data for the fastest possible data analysis. Spark analytics can be run directly on Cassandra data, across multiple data centers and the cloud.

Batch Analytics

DSE provides built-in MapReduce, Hive, Pig, and Mahout support for running longer batch analytics on Cassandra data. Batch analytics in DSE supplies an easy way to perform analytic tasks on transactional data that don't require immediate response times as do real-time analytics tasks.

Scalable Analytics

Additional capacity can be added online through node additions so analytic workloads can easily scale to meet incoming data and customer demands.

Always-On Analytics

DSE analytics is perfect for applications that need to run analytics on data in way that is always available and never goes down. DSE's always-on architecture, built on Cassandra, ensures 100% uptime for analytics operations.

Multi-Data Center and Cloud Analytics

DSE analytics run across multiple data centers and cloud availability zones, which allows analytics to be processed on OLTP data in different geographical regions delivering fast response times back to users in those locations.

Workload Isolation and Management

Analytic workloads performed on OLTP data should not impact OLTP operations; in other words, there should be a way to support both OLTP and analytic workloads with isolation between the two so no competition exists for either compute

or data resources. DSE delivers just this capability as nodes in a database cluster can be specified to run streaming / real-time or batch analytics on OLTP data. Cassandra's powerful replication automatically copies and moves data among nodes so there is no need to extract data from transactional databases and load them into another analytic database. Everything is contained within one database cluster.

Hadoop Integration

Because there are situations where operational and historical data must be combined for decision making purposes, DataStax Enterprise supports integration with Hadoop data warehouses/lakes such as those offered by Cloudera and HortonWorks. This supports use cases where both hot OLTP and cold historical data must be combined together at the same time. Note that DSE does not take the place of Hadoop data warehouse/lake deployments, nor can Hadoop operate in a transactional/operational manner to support Web, mobile, and IoT applications like Cassandra. DSE is a modern operational database, whereas Hadoop is progressive data warehouse framework. The difference between the two is much like the lines drawn in the RDBMS world between online and data warehouse databases.

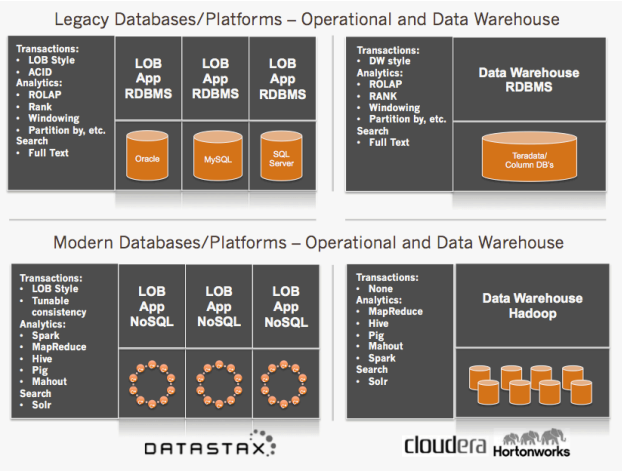

Figure 1 — Contrasting legacy vs. modern data platforms for both operational and data warehousing.

Figure 2 – DSE Ring with multiple analytics nodes.

Multi-Tempo Analytics Support

DSE delivers true "multi-tempo" analytics capabilities as it supports streaming, real/near-time, batch, and external Hadoop analytics all in the same database cluster with no contention for either compute or data resources. This negates the need for multiple analytics systems to support an application.

Visual Management and Monitoring

Analytics functionality can easily be visually provisioned, managed, and monitored with DataStax OpsCenter.

ODBC and JDBC Support

DataStax provides free ODBC and JDBC drivers for DSE analytics, which allows all popular BI and similar tools to integrate with and analyze data stored in DSE.

Further Reading

The integrated analytics capabilities of DataStax Enterprise provide the full spectrum of analytics support needed by today's Web, mobile, and IoT applications. For more resources and downloads of DataStax Enterprise, visit www.datastax.com today.

3975 Freedom Circle
Santa Clara, CA 95054
(650) 389-6000