

DataStax Enterprise Analytics

Think Big! Think Differently!

Situation

Intelligence and automated decision making have become mainstream in technology today. Financial industries use it for fraud detection, account activity reconciliation and even making decisions around credit lending. Automotive industries use it for Just in Time (JIT) inventory management, supply chain management and even controlling self-driving vehicles.

Complication

The challenges around using data for decision making is typically that the amount of data required to support decision making is huge. Large amounts of historical and trending data to compare against. Amounts of data that are not easily moved around or even accessible for decision making purposes in real-time, unless you have an architecture that can co-locate the ingestion of data alongside the intelligence required for decision making, such as the DataStax Enterprise Analytics engine.

Implication

The implication of processing data in this nature is that you need to be able to ingest and manipulate large amounts of data while making that data available to real-time analytical processes as effectively and efficiently as possible. Being able to mitigate the need for network hops, latency and contention for ingestion are key to success.

Position

DataStax Enterprise (DSE) Analytics provides a tightly integrated solution between Apache Cassandra and Apache Spark, providing the real-time analytical capabilities alongside huge data ingestion and manipulation capabilities in a simple to provision, comprehensive technology stack that powers modern businesses. Providing analytical capabilities co-located with the data and being "data aware" is transformational in an elegant solution that provides operational simplicity through integration while delivering real-time performance capabilities being co-located with the data. Having these processes co-located with the data mitigates network hops and communication bottlenecks by intelligently processing data using data locality techniques.

DataStax Enterprise and Apache Cassandra stores collections of data in "partitions" which are subdivided datasets to be distributed across a cluster of nodes. Leveraging distributed computing concepts enables the breakdown of the problem. The problem can be any number of things including data persistence, access patterns and workload isolation allowing DSE to scale not only horizontally but vertically as well to address potentially data dense workloads in a timely manner. Breaking down the problem across multiple compute resources gives DSE the ability to essentially increase the IO bandwidth to process greater amounts of data in a shorter amount of time, bring these processes into the realm of real-time analytics.

Opening Action

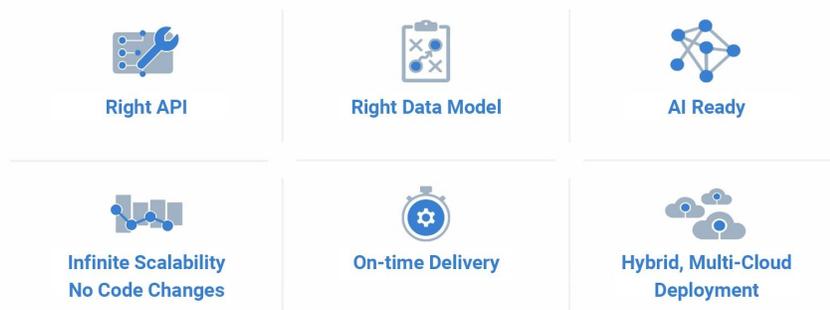
You can validate the possibilities of performing real-time analytics against co-located data yourself by simply downloading DataStax Enterprise, loading up some data and performing simple data science operations. To visualize your data as you're developing your analytics jobs, DataStax Enterprise provides DataStax Studio which is a web-based notebook-style IDE that supports processing of data that's stored in DataStax Enterprise and elsewhere. DataStax provides an abundance of examples within the platform that you can leverage to perform quick POCs or even learn by example.

Benefit

Analytics engines can be very utilitarian. DSE Analytics is tooled for acquiring data from a wide variety of data sources, having the ability to transform and manipulate data on the fly, while also delivering their primary functionality of performing intensive mathematical operations.

More prominently, analytics is a large part of a recently growing trend in Artificial Intelligence and Machine Learning, making DSE Analytics a great platform for delivering real time artificial intelligence to business problems. Business problems such as recommendation engines, JIT logistics solutions and fraud detection benefit from ML/AI. The ability to define feature stores, model repositories, training sets and scoring repositories in DataStax Enterprise allows you to provide a comprehensive real time AI platform constructed of the technologies that enable these systems. DataStax focuses on making it easy for developers to create the business outcomes through flexibility to write once and deploy globally.

Developer Outcomes

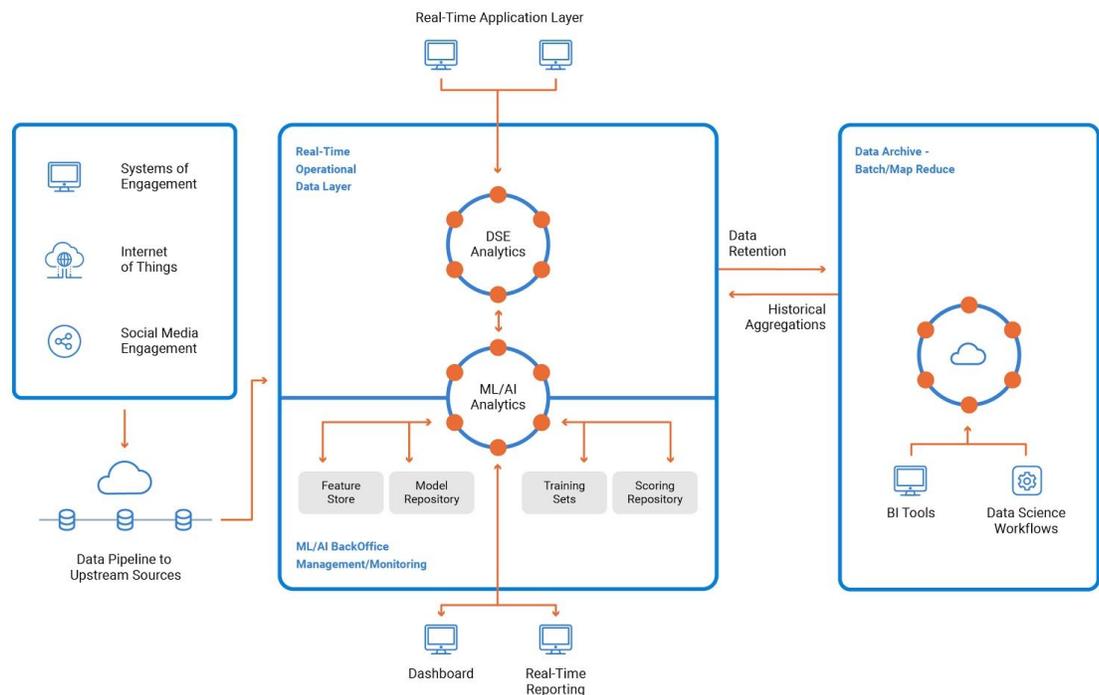


Main Idea #1: Smarter, Faster, More Accurate Artificial Intelligence

We see customers requiring the need to process ever increasing amounts of data in shorter periods of time. “In the face of unprecedented market shifts, data and analytics leaders require an ever-increasing velocity and scale of analysis in terms of processing and access to accelerate innovation and forge new paths to a post-COVID-19 world,” said Rita Sallam, Distinguished VP Analyst, during her presentation at virtual Gartner IT Symposium/Xpo™ 2020.

It's this acceleration of innovation where DataStax customers are well positioned. DSE provides a comprehensive operational data platform providing the architectural differentiators such as Search, Analytics and Graph, co-located with the data, to deliver low-latency data ingestion and analytics in real-time. AI and Machine Learning are critical, realigning supply and the supply chain to new demand patterns, delivering differentiators to the business.

Main Idea #2: Data and Analytics World Collide



Innovation has driven the need for new ways to process data and make decisions around that data. Businesses who leverage technology for decision making purposes tend to have greater differentiators in their markets. "The proliferation of augmented capabilities within analytics, business intelligence, and data science and machine learning products is making once-distinct markets collide. To harness the energy released, data and analytics leaders must anticipate big changes in products, investments and practices." says Carlie Idoine, Sr. Director Analyst, Gartner. DataStax Enterprise and its integrated analytics engine enabled by Apache Spark provides users to implement a comprehensive solution stack with elegance and intelligence.

Main Idea #3: Taking it to the next level

Extended considerations for your architectures can be enhanced by leveraging subject expertise of the various cloud providers. Putting data near the end-users, or placing it in a cloud to take advantage of features that exist in those clouds give you opportunities never before available to a single data management platform. The fact that DSE can run anywhere and everywhere, on-prem or in any cloud, allows you to negotiate and commoditize those cloud providers service offerings. Not only that, but DSE is a full solution stack, mitigating the complexity and costs of "services creep" into your architecture and total cost of ownership.

Position

Not only does DataStax Enterprise provide high availability and fault tolerance to the data, but it also provides a data analytics environment that delivers simplicity, efficiency and reduced total cost of ownership in comparison to outcomes of the Frankenstein architectures that are put together as requirements are defined. Security posturing is a key attribute for Enterprises looking to modernize while leveraging adjacent technology sets. Organizations that seize the opportunities of the newly catalyzed market could dramatically accelerate their analytics maturation.

Closing Action

Relationships define the foundation of data and analytics value.

Think Big. Think Differently.

Learn how you can architect DSE and DSE Analytics into your operational data store to take advantage of the benefits of performing analytics co-located with the data to improve latency on logical operations leading to transformational outcomes. Visit academy.datastax.com to learn more.

References

- ➔ “Gartner Top 10 Trends in Data and Analytics for 2020”, gartner.com, October 19, 2020
Contributor: Laurence Goasduff
<https://www.gartner.com/smarterwithgartner/gartner-top-10-trends-in-data-and-analytics-for-2020/>
- ➔ “Worlds Collide as Augmented Analytics Draws Analytics, BI and Data Science Together”, gartner.com, March 10, 2020
Analyst(s): Carlie Idoine
<https://www.gartner.com/en/documents/3981930/worlds-collide-as-augmented-analytics-draws-analytics-bi>

© 2020 DataStax, All Rights Reserved. DataStax, Titan, and TitanDB are registered trademarks of DataStax, Inc. and its subsidiaries in the United States and/or other countries.

Apache, Apache Cassandra, and Cassandra are either registered trademarks or trademarks of the Apache Software Foundation or its subsidiaries in Canada, the United States, and/or other countries.