

Deliver Always-On, Real-Time Insights at Scale

with DataStax
Enterprise Analytics

AIU	HJI	WWE	PLQ	EER	ORT	OPY
1,822 (-35)	20,369 (+590)	890 (-20)	6,350 (-290)	10,985 (+580)	26,800 (-15)	6,800 (+15)
MBC	LJH	MJB	PON	NFR	UON	OMJ
3,605 (+210)	9,542 (-128)	2,609 (+35)	7,654 (+169)	6,522 (+122)	20,000 (-54)	3,652 (+182)
TRV	OMN	MMJ	IT	KLM	15,000	5,285 (-130)
3,204 (63)	5,211 (+158)	7,100 (-60)	7,150 (-150)	782 (+74)	1,901 (+101)	5,285 (-130)
MBB	WFF	HJM	QLC	LSD	10,000	12,630 (+330)
3,320 (-130)	712 (+12)	134 (+5)	2,022 (-18)	631 (+40)	5,000	12,630 (+330)



AIU	1,822	12,349,000
EJK	3,680	238,681,000
HPL	1,062	85,678,000
KEE	485	8,369,800
NAH	8,569	189,301,000
QOP	6,602	10,000,000
TIK	890	
WIG	6,280	
AHD	2,436	

CONTENTS

Meeting the Needs of the Right-Now Customer	3
Introducing DSE Analytics	3
Faster Performance than Open Source Apache Cassandra™ and Apache Spark	4
DataStax Enterprise File System (DSEFS)	4
AlwaysOn SQL	4
Spark SQL	4
DSE Analytics Solo	5
Workloads	5
Streaming Analytics	6
Business Intelligence (BI)	6
ETL	6
Use Cases	7
Predictive Analytics	7
Real-Time Analytics — At Scale	7
Advanced Threat Detection	8
Customer Stories	8
IHS Markit	8
Clear Capital	9
DSE: Inject Freshest Data and Insights Into Decision-Making	9
About DataStax	10

MEETING THE NEEDS OF THE RIGHT-NOW CUSTOMER

Today's hyper-connected customers expect highly personalized and accurate information "Right-Now".

This has made analytics the lifeblood of all modern enterprises.

Be it for applications like customer 360, fraud detection, personalization and recommendations, or forecasting and modeling, Right-Now Enterprises need data that is fresh and always available to be able to glean instant insights for business decisions that improve customer experience, increase operational efficiencies, and decrease risk.

Retailers, for example, need analytics to better understand customer behavior both online and in-store so that they can create individualized, timely offers and drive smart inventory decisions. Banks and financial services companies use analytics to improve fraud detection and trading decisions, and enhance customer intimacy. Manufacturing firms collect data from their machines and use analytics to anticipate machinery failures.

Running a business without being able to quickly and easily analyze all the business-critical data is simply not an option for enterprises in the Right-Now economy. However, the exploding growth in data volume, variety, and velocity has rendered traditional systems unfit for managing and processing data at cloud scale.

Traditional systems also require using an extract, transform, and load (ETL) process to collect data from various sources, transform it into a suitable format, and then load it into the database for storage and analysis, all of which requires time and hence makes the data useless for real-time analytics. As per a [study conducted by IDC](#), two-thirds of data moved via ETL was at least five days old by the time it was available for analytics.

As a result, using these traditional systems has resulted in limited and delayed insights, slow query performance, and increased complexities for enterprises.

Enterprises need a modern architecture that allows them to:

- Cost-effectively store and analyze terabytes or petabytes of data and provides the flexibility to do so across any type of data
- Analyze streaming data as quickly as it comes in
- Run analytics jobs directly against the freshest data in the database without the costly and time-consuming ETL process

This paper explains how DataStax Enterprise (DSE) Analytics enables enterprises to secure insights from streaming and historical data at cloud scale, while being able to run analytics in a production environment across multiple data centers or cloud regions.

INTRODUCING DSE ANALYTICS

An integral part of DataStax Enterprise (DSE), an always-on, distributed cloud database, DSE Analytics integrates real-time and batch operational analytics on transactional data without any data or resource contention, allowing you to process streaming data and historical data and improve your business intelligence (BI) reporting.

Built on a production-certified version of Apache Spark™ along with integrated search and graph capabilities, DSE Analytics provides highly available, production-ready analytics that enables enterprises to securely build instantly responsive, contextual, always-on applications and generate ad-hoc reports.

DSE Analytics gives you the flexibility to easily process data, both structured and unstructured, to or from existing databases, including Apache Hadoop™, HDFS, Amazon S3, Oracle, MySQL, and IBM DB2, so you can glean insights from disparate sources and all data types. DSE Analytics also gives you access to many advanced analytics functionalities via integration with Spark's MLlib and SparkR, making data science and machine learning workflows easy and scalable on the most recent data without time-consuming ETL operations.

With DSE Analytics you get:

Faster Performance than Open Source Apache Cassandra™ and Apache Spark

DSE Analytics enables faster performance thanks to Continuous Paging. You can retrieve the data from the DSE database three times faster using the DSE Spark connector than using open-source Spark to read from open-source Apache Cassandra.

DataStax Enterprise File System (DSEFS)

The DataStax Enterprise File System (DSEFS) is a continuously available, scalable, HDFS-compatible distributed file system with no single point of failure. It also has no messy external dependencies, such as Apache Zookeeper™, which introduce other fragilities. Designed for use cases that require leveraging a distributed file system for data ingestion, data staging, and state management for Spark Streaming applications (such as checkpointing or write-ahead logging), DSEFS is similar to HDFS, but without the deployment complexity or single points of failure that come with HDFS.

AlwaysOn SQL

Although Spark has an SQL engine, and DSE inherits that benefit, Spark does not provide an enterprise-class ODBC/JDBC service that can put your SQL applications into production. That's where DSE AlwaysOn SQL comes in.

DSE AlwaysOn SQL is an easy-to-use, production-ready ODBC/JDBC service that provides SQL access to data in the DSE database and DSEFS, allowing you to confidently put ODBC/JDBC applications into production. It offers high availability, which means the service restarts automatically in the event of any failures, cached datasets are refreshed automatically, and client applications connect seamlessly to the service without having to know where in the data center the service is running — it's all handled for you.

DSE AlwaysOn SQL also provides authentication and authorization, including role-based management of SQL users via the same users managed within DSE Advanced Security, ensuring only the right users access the right data and removing the need for any extra, error-prone security setup steps.

Whether it's routine data cleansing and transformations, data loading, investigative queries, or BI, AlwaysOn SQL allows non-stop access to your DSE data via the familiar SQL language. By providing ODBC and JDBC interfaces, existing SQL tools, such as BI tools, can be immediately leveraged against your DSE data.

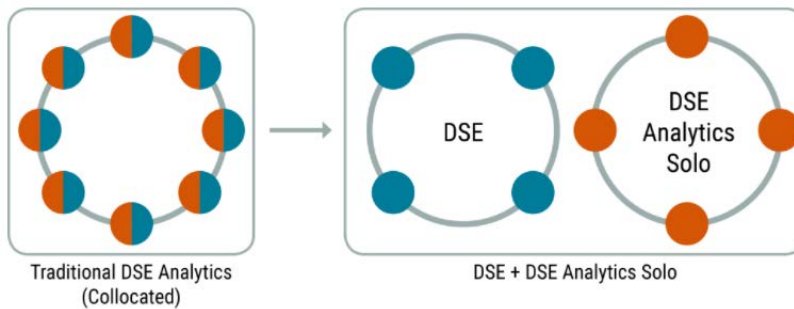
Spark SQL

Spark SQL, a unified relational query language for traversing distributed data, allows you to perform relational queries over data stored in DSE clusters and executed using Spark.

With [DataStax Studio](#), DSE gives you access to an intelligent, visual query builder that simplifies and accelerates the creation of Spark SQL queries and the review of results. This powerful query builder allows you to identify potential typos through syntax highlighting, and provides context-aware suggestions and validation so you can write queries without constantly referring to documentation or schemas and avoid errors before query execution.

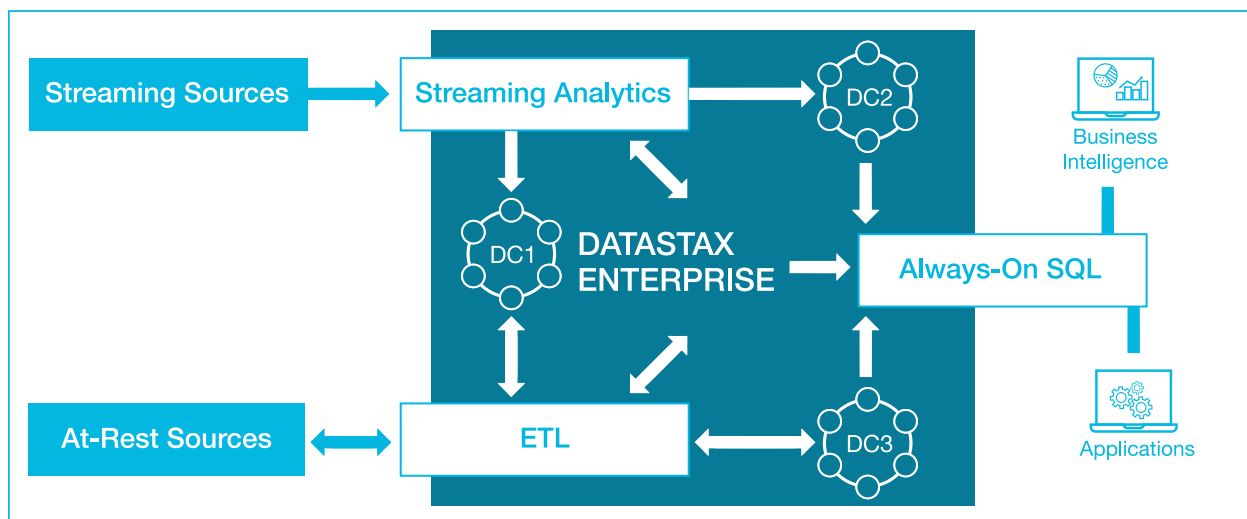
DSE Analytics Solo

DSE Analytics Solo, a DSE add-on, allows you to create a standalone cluster dedicated to operational analytics, effectively separating compute from storage and ensuring predictable performance. It delivers all of the powerful features of [DSE Analytics](#), including highly available, scalable, secure Apache Spark deployment and a fault-tolerant Spark resource manager with secured communications. DSE Analytics Solo gives you the flexibility to quickly and cost-effectively add or remove analytic processing nodes as the use case requires and manage those nodes via the same [DSE OpsCenter](#) management suite that manages the database nodes.



WORKLOADS

DSE Analytics is designed to support different workloads, ensuring enterprises get the right data at the right time for improved decision making.



These workloads include:

Streaming Analytics

As more and more data gets generated from a variety of connected devices, sensors, and web and enterprise applications, enterprises need to be able to analyze this continuous stream of data as quickly as it comes in to extract actionable insights in real time and build applications that provide real-time context.

With its Spark integration, DSE Analytics supports use cases that require real-time analytics on streaming and time-series data for the fastest possible data analysis. Unlike some other analytic solutions that require ETL, DSE Analytics allows analytics to be run directly on the data in the DSE database, eliminating the pitfalls related to data consistency.

DSE Analytics also supports the new Structured Streaming API in Spark via a Structured Streaming sink, enabling simple, efficient, and robust streaming of data into DSE from Apache Kafka™, file systems, or other sources.

And thanks to its unique masterless, distributed architecture, linear scalability, and flexible data model, DSE also allows you to handle the challenges associated with streaming analytics, including processing high volumes of structured and unstructured data coming from multiple, disparate data sources with seconds or milliseconds latency analytics requirements.

Using the streaming analytics capabilities of DSE Analytics, you can actively monitor data and take preemptive action - for example to send relevant and timely offers to customers or to provide alert on an imminent equipment failure or to detect and prevent a fraudulent transaction.

Business Intelligence (BI)

Today's business world moves at the speed of light, requiring enterprises to have easy-to-use applications that allow them to analyze data quickly and without the help of data experts. So, having a BI application has become a standard requirement for enterprises.

DSE Analytics gives the power of self-service BI to decision makers, enabling them to create visual, interactive dashboards and ad-hoc reports that provide accurate insights to significantly improve decision making in real time.

DSE's ODBC and JDBC drivers allow you to connect to a database cluster with your favorite BI tools (like Tableau and Microsoft Excel), access data stored on DSE nodes, and create join and rollup queries on DSE data.

DSE Analytics also enables users to explore and run complex analytic queries against terabytes and petabytes of data using standard SQL, while providing the high availability, fine-grained security, and governance necessary to ensure a reliable, enterprise-wide usage of BI applications using SQL.

ETL

Modern data layers frequently require multiple technologies to effectively support Right-Now Applications.

No data system should operate in a silo, as there will always be data of interest in other systems, or other use cases outside of your business that will benefit from your high-value data. While ETL operations are something to be avoided, if possible, they can be a necessary component of data flows in support of multiple lines of business.

DataStax recognizes this fact — that DSE is not alone in the data center — by also supporting necessary data flow ETL components via DSE Analytics. DSE Analytics addresses the needs of ETL operations both simple and complex by leveraging the same underlying parallel processing engine that powers all of Spark.

These more batch-like operations can tap into the powerful Spark engine in a variety of programming languages, including SQL, Scala, and Python, enabling data engineers to work in the language of their choice. And since DSE Analytics includes a secure, highly available resource manager, production ETL flows can be confidently deployed and efficiently executed in a production-ready environment using in-house tools and expertise.

USE CASES

DSE Analytics powers a wide variety of business-critical use cases across different verticals. Some of these key use cases are:

Predictive Analytics

Which equipment needs maintenance or replacement to avoid breakdown? Which products need to be re-stocked to avoid losing a customer to the competition? Which product should you recommend for purchase right now? Today's enterprises are constantly trying to anticipate and stay ahead of these business and customer needs and are turning to predictive analytics to boost their bottom line and competitive advantage.

DSE Analytics, with its streaming and batch analytics capabilities and BI support, allows enterprises to harness growing volumes of real-time and historical data to predict business and customer behavior and adapt quickly to meet their needs.

DSE Analytics also gives you also access to many advanced analytics functionalities via its built-in integration with Spark ML and MLlib, allowing you to run parallel machine learning functions on massive data sets for pattern detection.

A telecommunications firm, for example, can use predictive analytics to notify its customers when its data plan is running low and follow that up with a personalized plan renewal offer — all in real time.

Anticipating customers' needs and providing them with timely, relevant solutions is an incredibly powerful way to drive customer loyalty and prevent churn.

Real-Time Analytics — At Scale

In today's hyper-connected world, where customers expect instant responsiveness, the faster companies can decipher actionable insights, the faster they'll be able to act on them to deliver delightful customer experiences and outsmart the competition. Enterprises no longer have the luxury of relying on old, non real-time-updated data for decision making.

DSE Analytics provides a scalable, highly available solution that empowers the Right-Now Enterprises with fresh, always available data from multiple data sources for accurate, in-the-moment insights at cloud scale. Its high throughput and low latency lets you easily handle user and workload surges, ensure data access for all users, and provide an instantly responsive end-user experience.

Having access to real-time insights can bring immense benefits to a variety of applications across different verticals. For example, DSE's real-time analytics and search capabilities enable the Intercontinental Exchange to gain quick visibility into 'tick' data, pricing trends, projected pricing, and other market data pertinent to a specific stock and/or bond and provide reliable up-to-the-second financial information to traders, brokers, and investors, which empowers them to make better trading and investment decisions.

Advanced Threat Detection

Cyber threats evolve quickly. Enterprises, including financial services organizations and retailers, are increasingly dealing with security breaches that result in the loss of sensitive data of millions of customers. As data continues to increase in volume and complexity, it's becoming more and more difficult for traditional analytic tooling and infrastructures to keep up.

DSE Analytics enables enterprises to store, process, and analyze millions of transactions in real time, while enabling machine learning and ad hoc queries at scale.

DSE Analytics lets you run analytics on streaming data and quickly identify patterns and anomalies to take actions such as decline a fraudulent transaction, block an IP from a DOS attack, or shut down a piece of equipment, in real time.

DSE also comes with DSE Advanced Security, which includes features such as unified authentication, authorization, and proxy execution. These enterprise-grade security features ensure that only an authenticated user can query the database tables and that users only have access to tables and data they have permissions to read, thereby keeping your data protected.

CUSTOMER STORIES

IHS Markit

500 Million Components: IHS Markit Converts Data to Value with DSE

The UK-based IHS Markit provides critical information, analytics, and solutions to Fortune 500 enterprises across business lines, improving their operational efficiency and providing deep insights that lead to well-informed decisions.

IHS Markit manages a digital catalog of more than 500 million electronic parts and components. Customers can use this database to monitor component status; receive part changes and end-of-life notifications; quickly and accurately identify alternate component sources; optimize spend; meet regulatory compliance requirements; and more.

To improve the catalog's power, speed, and dependability, IHS Markit chose DSE for its massively linear scalability and masterless architecture, with DSE winning out over competing technologies such as Apache Hadoop® with Apache HBase™.

With DSE Analytics, IHS Markit can now build dashboards for visualizations and real-time business analysis. And with its integrated machine learning capabilities, DSE Analytics also allows IHS Markit to automate content classification and entry, a feature that lets the company get market data updates in way less time — in many cases weeks ahead of what could be done prior to DSE.

DSE Analytics also allows IHS Markit to eliminate the legacy ETL processes, enabling the company to not only drive faster updates to market, but, more importantly, radically reduce the overall maintenance costs and complexity of back-end systems.

Additionally, thanks to DSE Search, IHS Markit easily optimizes queries across a diverse set of complex attributes to generate fast and reliable results.

Backed by DSE, IHS Markit's digital catalog now helps customers find the parts and insights they need right when they need it, while allowing IHS Markit to improve its bottom line and employee productivity.

Clear Capital

Delivering around-the-clock, accurate appraisals with DSE

Clear Capital is the leading provider of real estate asset valuation and collateral risk assessment solutions for large financial services companies. Clear Capital's customers, which include some of the largest US banks and investment firms, rely on Clear Capital to make market-level decisions with only the most recent and accurate data.

As the largest source of valuation data for residential and commercial properties in the United States, Clear Capital recognized the need for a database that could support a high volume of small transactions, linear scalability, real-time performance, and continuous availability to power their cutting-edge valuation review management software-as-a-service platform and deliver highly accurate and recent valuation data quickly to financial institutions and banks.

DSE was the only solution that met all of Clear Capital's needs. In fact, with DSE, Clear Capital could take down an entire availability zone within their Amazon environment with zero service interruptions. Clear Capital also chose DSE for its integrated search and analytics capabilities.

An important component of Clear Capital's appraisal system is its geospatial search and analysis capability, which allows the company to gain valuable market insights based on aggregate statistics, feeding into the accuracy of the valuations delivered. DSE's integrated search and analytics capabilities power Clear Capital's geospatial search and allow it to deliver real-time insights to customers that rely on the most recent data to support their market-level decisions.

DSE: INJECT FRESHEST DATA AND INSIGHTS INTO DECISION-MAKING

Deep insights come from being able to leverage your data using highly available, real-time analytics. These insights, when unlocked, can be used to make decisions that help enterprises provide meaningful, interactive experiences and innovate at a speed not previously possible.

DSE Analytics, with its streaming and batch analytics capabilities, empowers enterprises to run from the simplest to the most complex analytic queries across massive datasets with incredible speed, availability, and reliability — resulting in fresh in-the-moment insights that build customer loyalty and boost your bottom line.

For information about DataStax Enterprise please visit <https://www.datastax.com/products/datastax-enterprise>.

ABOUT DATASTAX

DataStax powers the Right-Now Enterprise with the always-on, distributed cloud database built on Apache Cassandra™ and designed for hybrid cloud. The foundation for real-time applications at massive scale, our flagship product, DataStax Enterprise, makes it possible for companies to exceed expectations through consumer and enterprise applications that provide responsive and meaningful engagement to each customer wherever they go. Our product also gives businesses full data autonomy, allowing them to retain control and strategic ownership of their most valuable asset in a hybrid cloud world. DataStax helps more than 400 of the world's leading brands like Capital One, Cisco, Comcast, eBay, McDonald's, Microsoft, Safeway, Sony, UBS, and Walmart transform their businesses through right-now applications focused on enterprise optimization and customer experience. For more information, visit DataStax.com and follow us on @DataStax.

DataStax is a registered trademark of DataStax, Inc. and its subsidiaries in the United States and/or other countries. Apache Cassandra is a trademark of the Apache Software Foundation or its subsidiaries in Canada, the United States, and/or other countries.





Deliver Always-On, Real-Time Insights at Scale

with DataStax
Enterprise Analytics