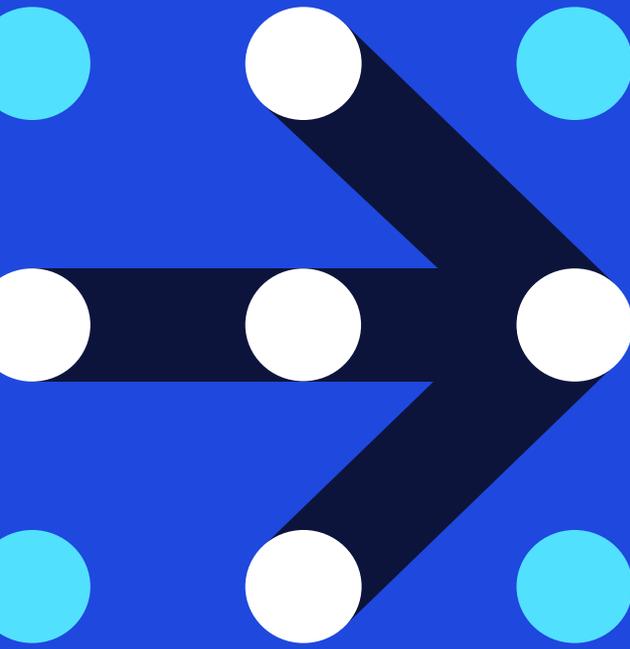


# Getting Started with NoSQL and Apache Cassandra

Accelerating the Transition from Relational to NoSQL



**Apache Cassandra™ is one of the top open source NoSQL databases used by Fortune 500 companies today, and is purpose-built for use cases with scale out and high-availability requirements. DataStax was founded in 2010 as the primary developer of and support organization for Apache Cassandra.**

DataStax Enterprise (DSE) was first released in 2011 and is the most advanced, enterprise-proven distribution of Cassandra available. Its masterless architecture is designed to handle big data workloads across multiple nodes with no single point of failure.



**Cassandra's peer-to-peer architecture guarantees no single point of failure to collect and store streaming data from infrastructure instruments and enables us to write and read from any node with minimal latency.**

Praveen Kumar, Sr. Manager Of Emerging Technologies & Platform, Equinix

[Read How](#)

**DataStax Academy  
gives you instant  
access to product  
downloads, learning  
tools, and much more.**

**Sign Up**

## **Model. Code. Go.**

If you're a data architect, data scientist, or application developer familiar with relational databases, this guide provides an opportunity to learn how to sharpen your skills, with advice from the foremost experts in distributed database technology, so you can migrate to Cassandra and build the next generation of powerful applications.

We'll cover the key differences between Cassandra and relational databases, how to translate your relational database skills to Cassandra, and how to get started quickly. We'll also tell you about resources that will help you on your journey, including:



**The DataStax Academy**



**DataStax Documentation**



**Learning paths and samples**

---

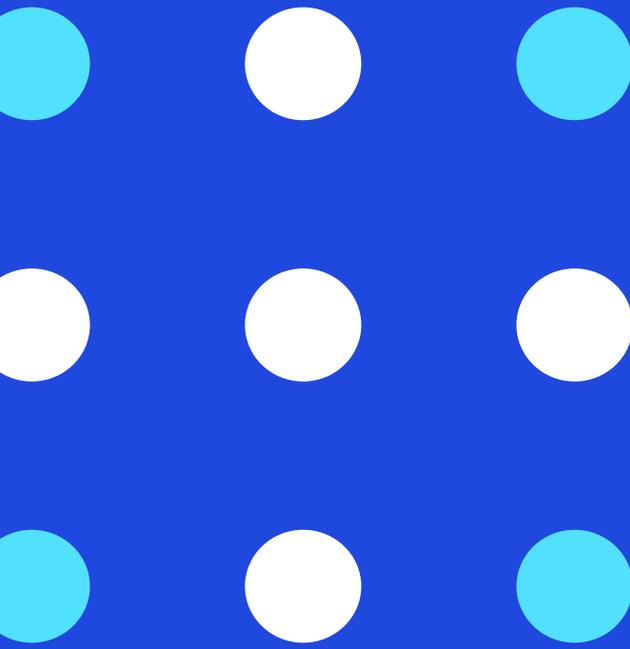
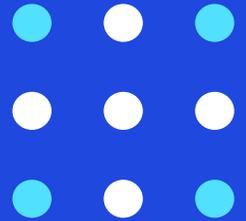
# Relational and Cassandra. Compared.



**We realized that we just couldn't scale effectively with a traditional MySQL solution. In fact, we couldn't even continue to offer the same level of service to our customers, simply because the speed of access to our data would have deteriorated as the amount of data grew. That's when we decided to move forward with Cassandra.**

Harry Robertson, Tech Lead, Ooyala

[Read More](#)

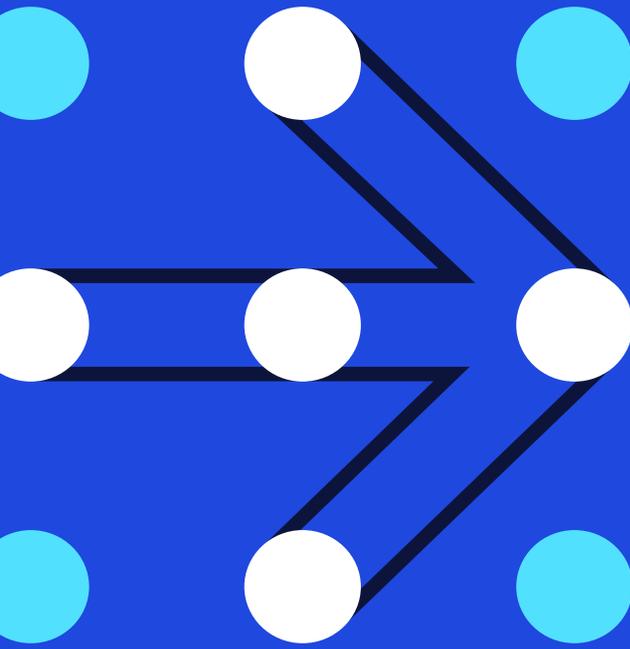
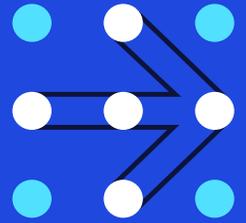


## Why Move to DataStax Distributions of Apache Cassandra?

Apache Cassandra provides incredible resiliency and high performance for distributed data management; it's a distributed, elastically scalable, highly available and fault-tolerant platform with tunable consistency. DataStax distributions expand on Cassandra's benefits with advanced security, performance, multi-model, and operational management capabilities.

Applications that are time-series based, have high data volumes like IoT use cases, or those that are real-time transaction-based—such as fraud detection and customer recommendations—are ideal candidates, as well as those that require high security and the utmost resiliency like eCommerce apps.

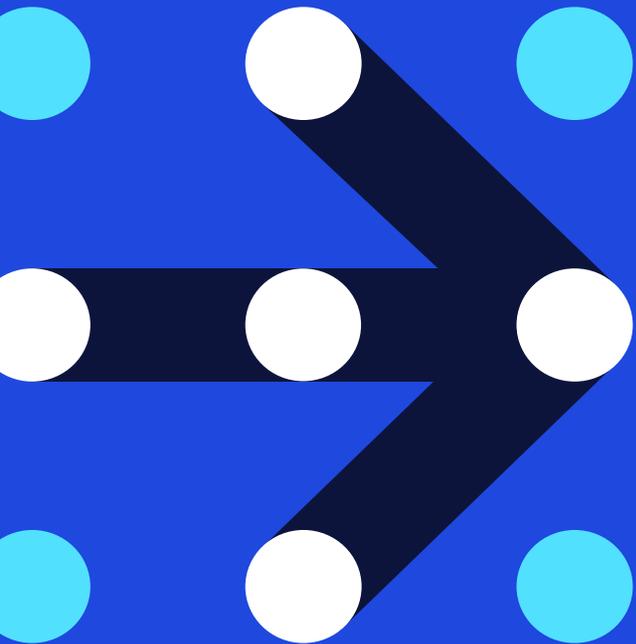
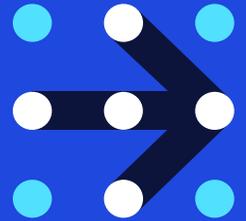
# Relational and Cassandra. Compared.



Key architectural differences between Cassandra and relational databases are as follows:

	Relational	Cassandra
<b>Data</b>	Structured	Unstructured
<b>Schema</b>	Fixed	Flexible
<b>Table</b>	An array of arrays	A list of nested key-value pairs
<b>Outermost Container</b>	Database	Keyspace
<b>Entities</b>	Tables	Tables and columns
<b>Row</b>	An individual record	Unit of replication
<b>Column</b>	Attributes of a relation	Unit of storage
<b>Relationships</b>	Foreign keys, joins, etc.	Collections

## Relational and Cassandra. Compared.



### Tapping into Graph Database Technology

DataStax also provides a graph database option that leverages all the benefits of Cassandra called DataStax Graph. It's an ideal way to build data models that represent the complex relationships between people, products, interactions, and transactions—with high performance essential for applications that glean insights from connected data like fraud detection, supply chain optimization, social network analysis, and customer-facing recommendation engines.

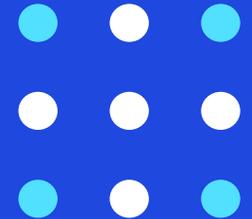
Data storage in a graph database can be compared to a pre-joined relational database, with built-in data relationships, so foreign keys are unnecessary. Data is retrieved by traversing the graph, so time-consuming and error-prone JOIN operations are not required.



[Learn how it compares with relational databases, and ways to migrate.](#)

[Learn Now](#)

# Getting Started



### DataStax has simplified the getting started process for DataStax distributions (DataStax Enterprise and DataStax Distribution of Apache Cassandra).

You can choose to run them on-premises or in any kind of cloud, including managed, public, private, hybrid, or multi-cloud. The DataStax Distribution of Apache Cassandra provides a production-ready version of Apache Cassandra and grants access to the DataStax Bulk, DataStax Apache Kafka Connector, Production Docker Image, developer tools, and a range of services and support.

With DataStax Enterprise, you get even faster performance, as well as greatly simplified database operations and management features with OpsCenter, advanced developer tools and security features for more operational simplicity, and multi-model development and mixed workloads with options like DataStax Graph, DataStax Analytics, and DataStax Search.



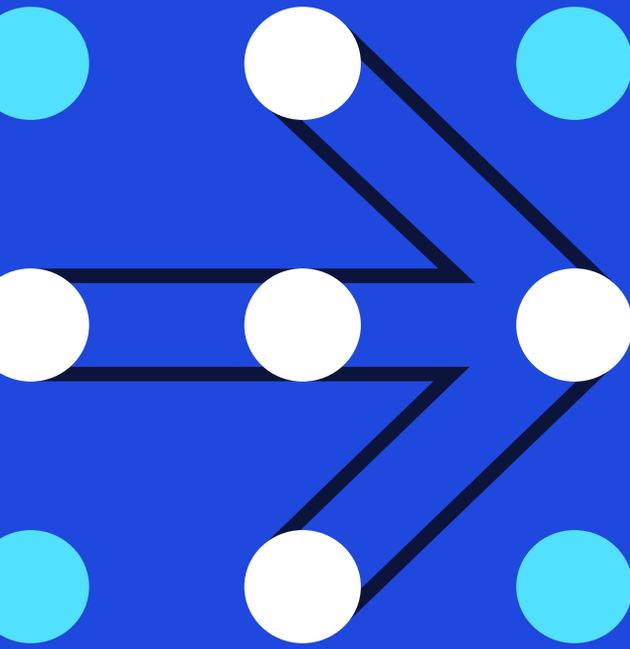
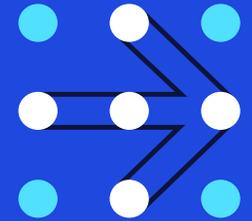
It's simple to start running DataStax Enterprise or the DataStax Distribution of Apache Cassandra, on-premises or in the cloud.

[Get Started](#)



Want to fast-track learning Apache Cassandra? Take the free self-paced training and master Cassandra's internal architecture with hands-on exercises.

[Get Started](#)



## What are the basic building blocks of Cassandra compared to RDBMS? Here are the highlights:

### Keyspaces

In Cassandra, a Keyspace is like the schema concept in relational database management systems. It's the top-level object and there is only one keyspace per application. Keyspaces contain tables, materialized views, and user-defined types, functions, and aggregates, and they control the replication for the objects they contain at each datacenter in the cluster.



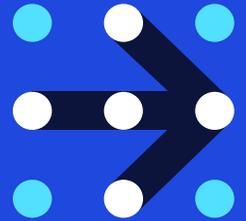
Did you know that CQL (Cassandra Query Language) provides full DDL, DML, and DCL that makes it easier to carry over core concepts from SQL?

[Learn More](#)



Keyspaces

[Learn More](#)



Managing  
tables

[Learn More](#)



Making  
schema  
changes

[Learn More](#)

## Tables

The next structural unit is called the table, where data is stored in tables containing rows of columns. Tables can be created, dropped, and altered at runtime without blocking updates and queries. Just like a relational database, Cassandra tables require a primary key. The difference is the first element in a primary key is called a partition key. The partition key has a special use in Apache Cassandra beyond showing the uniqueness of the record in the database, it's other purpose is one that is critical in distributed systems: determining data locality.

## Schema Changes

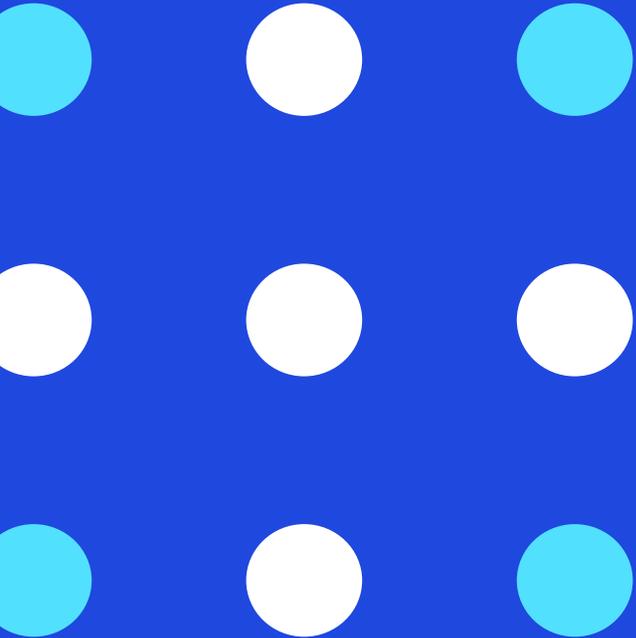
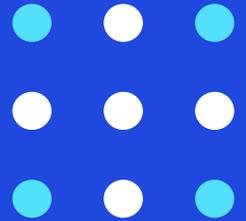
In Apache Cassandra, schema changes naturally take time to propagate to all nodes in the cluster depending on the size, network capacity, and load on the cluster. As a result, best practices suggest making these changes one at a time and ensuring they have fully spread throughout the cluster before moving onto subsequent alterations.



You can use CQL to **CREATE/ALTER/DROP/TRUNCATE** to manage keyspaces, types, functions, aggregates, tables, and indexes.

[Learn More](#)

# Introducing CQL (SQL Skills Welcome)



**We touched on CQL earlier—it’s a fast way to start working with Cassandra, especially when you’re familiar with SQL, so you can quickly start creating or altering keyspaces and tables, making changes to data, and performing queries.**

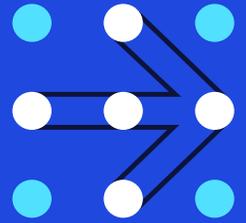
If you’re familiar with SELECT, INSERT, UPDATE, and DELETE in SQL, you’re ready for CQL and can refer to our [CQL Quick Reference Guide](#). We’ve also highlighted some of the key similarities and differences below:



[Learn all the CQL commands](#)

[Learn Now](#)

# Introducing CQL



## Operation

## CQL Examples

### SELECT

```
SELECT col_0, col_1 FROM my_table;
```

Retrieving all rows in a table looks the same when comparing SQL and CQL

```
SELECT col_0, col_1 FROM my_table WHERE ...;
```

In SQL, any column can be included in the WHERE clause, though in CQL only columns that are strictly declared in the primary key can be used as a restricting column. Also, each query must have a partition key defined at a minimum.

### INSERT UPDATE

```
INSERT INTO my_table (col_0, col_1) VALUES (val_0, val_1);
```

```
UPDATE my_table SET col_0=val_0 WHERE ...;
```

DSE and Cassandra are best-in-class at high throughput writes. Bear in mind that in CQL, both INSERT and UPDATE must include the partition key.

### DELETE

```
DELETE FROM my_table WHERE ...;
```

In SQL, there is only the option to remove the entire row(s) using the DELETE syntax. In CQL, you can delete specific columns using

```
DELETE my_col FROM my_table WHERE ...;
```



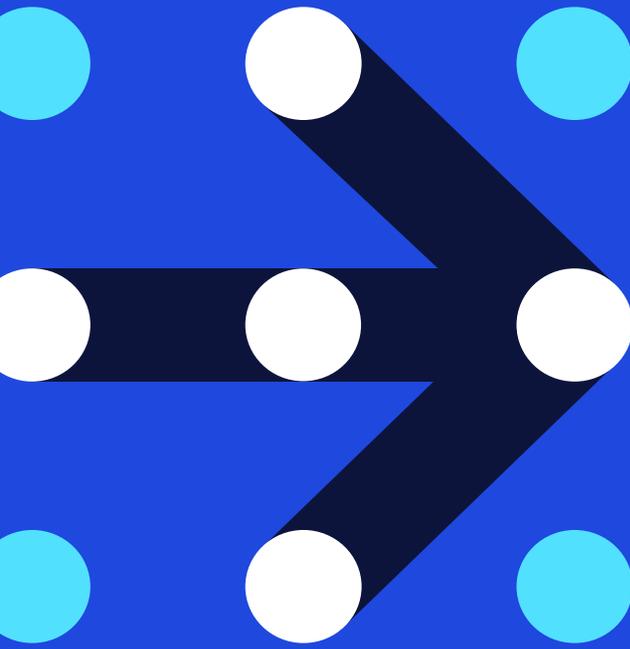
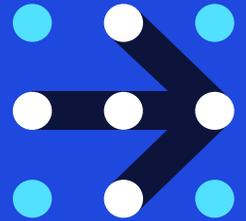
Did you know that in Cassandra, data isn't deleted in the same way it is in an RDBMS?

Apache Cassandra is designed for high write throughput and avoids reads-before-writes.

It uses SSTables, which are immutable once written. So, a delete is an update and updates are actually inserts (into new SSTables). Want to manage data effectively?

[Learn How](#)

# Introducing CQL



Just like how you control permissions and resources of the entities in SQL, you can do the same in CQL:

**\_CREATE ROLE**  
**\_ALTER ROLE**  
**\_DROP ROLE**  
**\_LIST ROLES**  
**\_GRANT**  
**\_REVOKE**

**\_RESTRICT**  
**\_UNRESTRICT**  
**\_RESTRICT ROWS**  
**\_UNRESTRICT ROWS**  
**\_LIST PERMISSIONS**



While native integration always provides the best performance, you can connect your Business Intelligence and other tools using DataStax ODBC/JDBC drivers.

[See How](#)

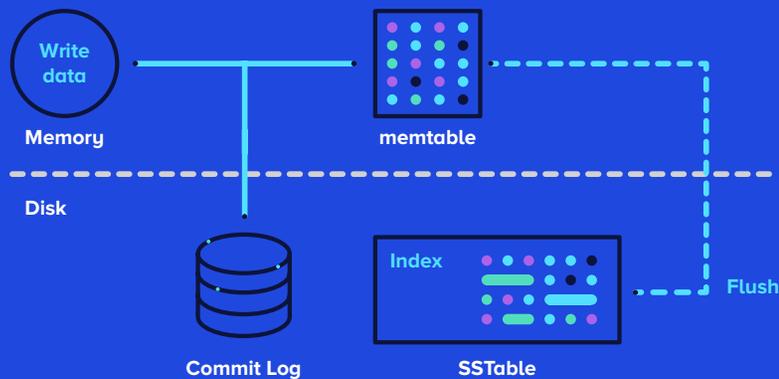
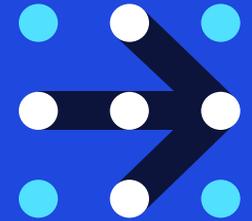
# Balancing Transactional Integrity



**Penn Mutual started out with a traditional RDBMS approach for the persistence layer of their Core Service but soon realized that it could not meet their requirements for application performance or scalability. They turned to DataStax Enterprise instead.**

[Learn More](#)

# Balancing Transactional Integrity



As a non-relational database, Apache Cassandra and DataStax Enterprise does not support joins or foreign keys and consequently does not offer consistency in the ACID sense. But it does offer atomic, isolated, and durable transactions with eventual and tunable consistency that allows the user to decide how strong or eventual they want each transaction's consistency to be.

- **Atomicity.** A write and a delete operation is atomic at the partition level. Insertions or updates of two or more rows in the same partition are treated as one write operation.
- **Isolation.** Write and delete operations are performed with full row-level isolation, so a row within a single partition on a single node is only visible to the client performing the operation.
- **Durability.** Writes are durable. All writes to a replica node are recorded both in memory and in a commit log on disk before they are acknowledged as a success.



Get all the details about DataStax and Apache Cassandra database internals.

Get Details

# Ways to Migrate Your Workloads

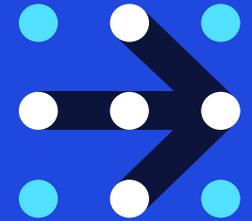


**We pulled in one year of data from Oracle and once we got it into DataStax Enterprise, built on the best distribution of Apache Cassandra, it was smooth ride and it was processing at a very high rate.**

Mukram Aziz, Sr. Manager of Data Services, Capital One

[Read How](#)

# Ways to Migrate Your Workloads



How do I ingest from an existing relational database (RDBMS) to an Apache Cassandra or DataStax Enterprise cluster?

[Learn How](#)

There are many ways to migrate data from your relational databases into Apache Cassandra and DataStax Enterprise. Here's where you can start.

---

## DataStax Bulk Loader (dsbulk)

Load and unload CSV or JSON data in and out of the DSE database. DataStax Bulk Loader efficiently and reliably loads small or large amounts of data, supporting developer and production environments.

---

## CQL COPY FROM

Imports data from a comma-separated values (CSV) file or a delimited text file into an existing table, mainly for datasets that have less than 2 million rows.

---

## sstableloader

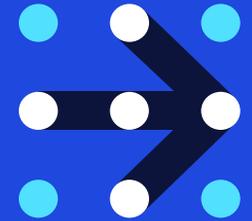
Bulk loads large volumes of external data into a cluster by streaming a set of SSTable data files to a live cluster.

---

Several Extract-Transform-Load (ETL) tools like Talend, Informatica, and StreamSets also support Apache Cassandra and DataStax Enterprise, providing sophisticated data transformation logic, point-and-click interfaces, scheduling, and more, to manage data movement.

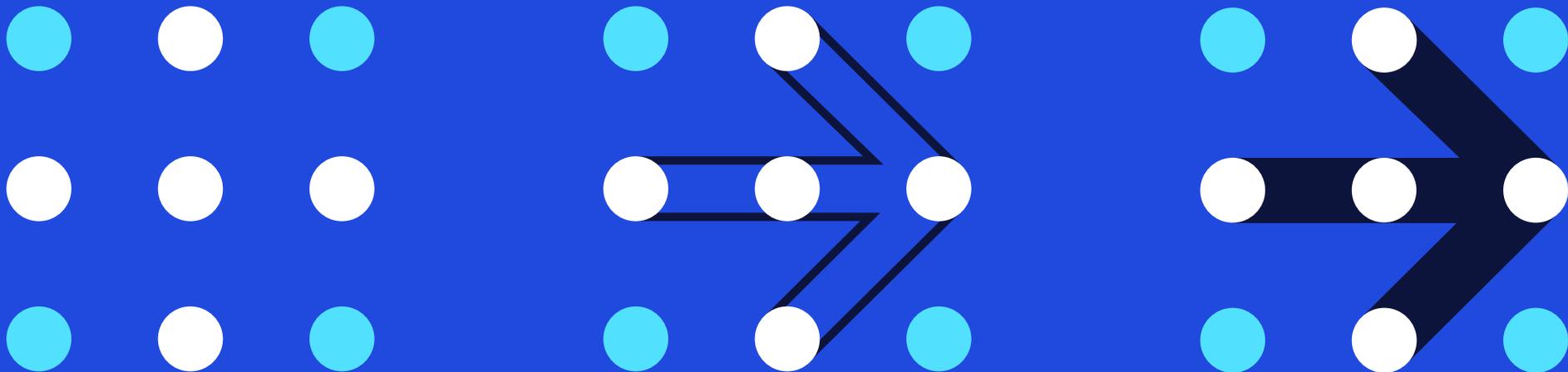
# Next Steps

## Next Steps



Take a Learning Path to gain an expert understanding of Apache Cassandra and DataStax Enterprise principles related to your role. Each Learning Path is composed of a sequence of recommended courses for your role, curated by our curriculum engineers. When you complete your path, you will receive a printable Certificate of Completion. You can follow your progress along the path in the chart below and switch to a different path at any time.

	 <b>Administrator</b>	 <b>Architect</b>	 <b>Developer</b>	 <b>Graph Specialist</b>
<b>DS101: Introduction to Apache Cassandra™</b>	Included	Included	Included	Included
<b>DS201: DataStax Enterprise 6 Foundations of Apache Cassandra™</b>	Included	Included	Included	Included
<b>DS210: DataStax Enterprise 6 Operations with Apache Cassandra™</b>	Included	Included	N/A	N/A
<b>DS220: DataStax Enterprise 6 Practical Application Data Modeling with Apache Cassandra™</b>	N/A	Included	Included	Included
<b>DS330: DataStax Enterprise 6 Graph</b>	N/A	N/A	N/A	Included



## Conclusion

Moving to Apache Cassandra with DataStax is faster and easier than ever before. That means that you can use the same technology that's powering the world's leading enterprises—like Capital One, Cisco, Comcast, Delta Airlines, eBay, Macy's, McDonald's, Safeway, Sony, and Walmart—and be up and running in no time.

As the world's leading Cassandra experts, DataStax provides the community with online training, certification, and full documentation; DataStax Examples provide sample code for developers to reference and shorten the path to getting started. You can also access DataStax expertise directly through professional services and support options that can ensure the right level development or deployment help whenever you need it in your journey.

Get started today at <https://www.datastax.com/get-started>.

## About DataStax

DataStax delivers the only active everywhere hybrid cloud database built on Apache Cassandra™: DataStax Enterprise and DataStax Distribution of Apache Cassandra, a production-certified, 100% open source compatible distribution of Cassandra with expert support. The foundation for contextual, always-on, real-time, distributed applications at scale, DataStax makes it easy for enterprises to seamlessly build and deploy modern applications in hybrid cloud. DataStax also offers DataStax Managed Services, a fully managed, white-glove service with guaranteed uptime, end-to-end security, and 24x7x365 lights-out management provided by experts at handling enterprise applications at cloud scale. More than 400 of the world's leading brands like Capital One, Cisco, Comcast, Delta Airlines, eBay, Macy's, McDonald's, Safeway, Sony, and Walmart use DataStax to build modern applications that can work across any cloud. For more information, visit [www.DataStax.com](http://www.DataStax.com) and follow us on Twitter @DataStax.

© 2019 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.

