How to Migrate From DataStax Enterprise to Instaclustr Managed Apache Cassandra®

There are a number of reasons you may be looking to migrate from DataStax Enterprise (DSE) to open source Apache Cassandra®. Perhaps you want to cut costs or eliminate vendor lock-in, or you feel excited about participating in the vibrant open source community. No matter your reason for making the change, we're confident you'll find Apache Cassandra to be a powerful replacement for DataStax Enterprise and that it will provide you with maximal value over the long term.

This document is a guide for migrating from DataStax Enterprise to Instaclustr Managed Cassandra. It's intended for solutions architects, engineers, IT Directors, or others on the front lines of planning and executing the migration. Throughout this material, we will go over different considerations to bear in mind for planning an effective migration with zero downtime.

Instaclustr
Now part of **Spot by NetApp**

Questions to Ask Prior to Migration

Before you migrate, there are several questions you should explore about your current architecture—and your desired future state—that will meaningfully influence how you approach the migration. Make sure you consider the following:

■ **What DataStax Enterprise version are you on?**

The DataStax Enterprise software version you are on will impact your migration pattern options. Datastax has progressively made it more difficult for customers to freely migrate off of their platform.

From DataStax Enterprise 6.0 onward, DataStax has changed file formats. The SSTable files used to store information in the database have been altered so that it's no longer compatible with open source Apache Cassandra. Because of this change, DataStax Enterprise users on 6.0 or later only have the option to do an out-of-place migration and cannot perform an in-place migration. Despite these challenges, it is still possible to migrate from 6.0 onward to Cassandra without any downtime.

But if you haven't moved to 6.0 yet, now is the time to make your move.

■ **What DataStax Enterprise proprietary features are you *actually* using?**

If you're a DataStax Enterprise customer, you may think you have no choice but to stay with DataStax because you are locked in with proprietary features that you require (or at least value). But fortunately, these features can be replaced with open source alternatives or by features present in the Instaclustr Console, all without any license fees or long-term vendor lock-in.

Below are proprietary features of DataStax Enterprise and alternatives you can use instead:

DataStax OpsCenter

Many DataStax Enterprise customers use DataStax OpsCenter for management and monitoring of their cluster. OpsCenter allows users to:

- Add and expand clusters
- Configure nodes
- Monitor performance metrics
- Monitor cluster health
- Perform basic cluster repair
- Backup clusters

When you migrate to Instaclustr's Managed Apache Cassandra service, you can handle core management tasks such as adding and expanding clusters and configuring nodes from directly within the Instaclustr Console or using the Provisioning API.

If you're using Prometheus or Grafana within your tech stack, you can also use these tools to replace some of the monitoring features of OpsCenter if you prefer that over the Instaclustr Console. Instaclustr's Monitoring API can easily be configured to port metrics from your Instaclustr Managed Cassandra instance to Prometheus. These metrics can then be queried with PromQL and viewed within a web UI in Prometheus. You can also pair Grafana with Prometheus for improved visualization of those time-series metrics.

DataStax Enterprise customers may also be accustomed to OpsCenter for its repair service which synchronizes data across nodes and repairs any corrupted data. For Instaclustr Managed Cassandra service customers, this process is fully supported within our Managed Platform and with our expert Technical Operations team.

Another feature you may be concerned about replacing is the ability to manage backups in OpsCenter. OpsCenter allows for both automatic backups and manually backing up a DataStax Enterprise cluster. This is once again a process that is handled automatically in the background for Instaclustr Managed Cassandra customers.

DSE Search

DataStax Enterprise Search can be used to query data in Cassandra and generate reports. This functionality can easily be replaced with any number of open source tools. OpenSearch, the replacement for the Open Distro of Elasticsearch, is the most obvious replacement here. OpenSearch offers powerful search and analytics functions and can be integrated with Cassandra. Instaclustr also offers a Managed OpenSearch solution that can easily be integrated with your Managed Cassandra instance.

Another alternative to DSE Search is the Cassandra-Lucene Index, the Java library that OpenSearch was built from. The Cassandra-Lucene Index is available as an add-on to Cassandra as part of the Instaclustr Managed Platform. Lucene acts as an index attached to your Cassandra clusters, and is more like a bolt-on product when compared with OpenSearch, which is a completely separate technology from Cassandra. Lucene works well for those who want to work just from within Cassandra rather than adding another technology, but it can be difficult to configure, manage, and run. OpenSearch is a better fit for those willing to add another solution to their stack and for those who want to run more advanced queries than are available with Lucene. OpenSearch lets you scale search capacity separately from core Cassandra, and you get access to the standard APIs that come with it.

DSE Analytics

DSE Analytics offers an analytics toolset for processing streams of data and generating reports. Built off of open source Apache Spark, it unsurprisingly can be replaced with Spark itself. Spark is a fast and general cluster computing system for big data, providing an optimized engine that supports general computation graphs for data analysis. Instaclustr Managed Cassandra customers can easily add Managed Spark to their deployment with a click of a button within the Instaclustr Console at the time of cluster provisioning.

Common Concerns Around Migration

Downtime

One major concern DataStax customers may have is, will my application have downtime during the migration? Understandably, downtime is not an option for many organizations, especially those with revenue-generating applications. Fortunately both in-place and out-of-place migrations (which we'll discuss in detail later in this piece) to Cassandra can be achieved without any downtime.

Data Loss

You may also be concerned about losing critical data during the migration process. Just as with downtime, migration from DataStax Enterprise to Instaclustr Managed Cassandra can be achieved without any data loss so long as it is properly executed.

Performance

One common selling point DataStax will use for DataStax Enterprise is that they offer performance enhancements over what is available in the open source space. But in fact, we've seen customers experience performance improvements moving from DSE to Cassandra.

Top 10 Bank Gets Performance Boost Migrating to Cassandra

Instaclustr helped one of the top 10 U.S. banks migrate from DataStax Enterprise to open source Apache Cassandra. The team moved three clusters and converted approximately 200 DataStax instances across different environments and geographic regions. This was accomplished as a rolling migration without any data loss or downtime. Once the migration was complete, the bank saw a 5% performance improvement even without any changes to their data model or architecture.

Migration Patterns

One of the things you'll need to decide on when planning your migration is if you'll perform an in-place or out-of-place migration. In order to move to Instaclustr Managed Cassandra, first, you must move off of DataStax Enterprise and onto an Apache Cassandra cluster via either an in-place or out-of-place migration. Then, we work with you to help migrate your Cassandra cluster onto our managed service.

For an in-place migration to open source Cassandra, you migrate one node at a time within your cluster, checking each node for any issues before taking action on the next.

For an out-of-place migration, you set up a second cluster running Apache Cassandra. Then, you can have your application write to your DSE cluster and your Cassandra cluster simultaneously, until eventually the data on both clusters is the same and you can retire your DSE cluster. Or, alternatively, you can provision a Cassandra cluster and copy your data over in full from DSE to Cassandra, connect your Cassandra cluster to your application, and deprovision your DSE instance.
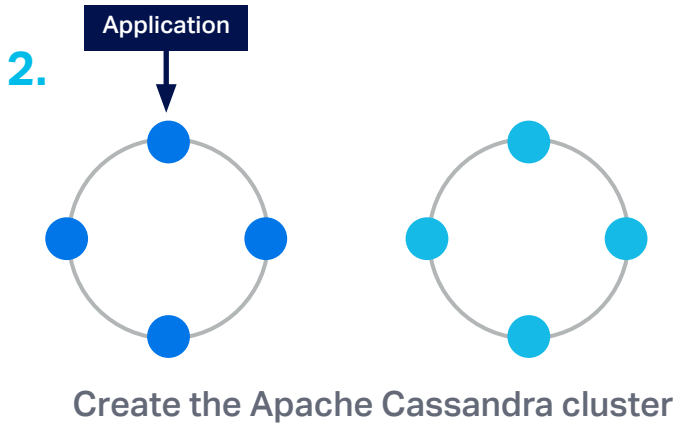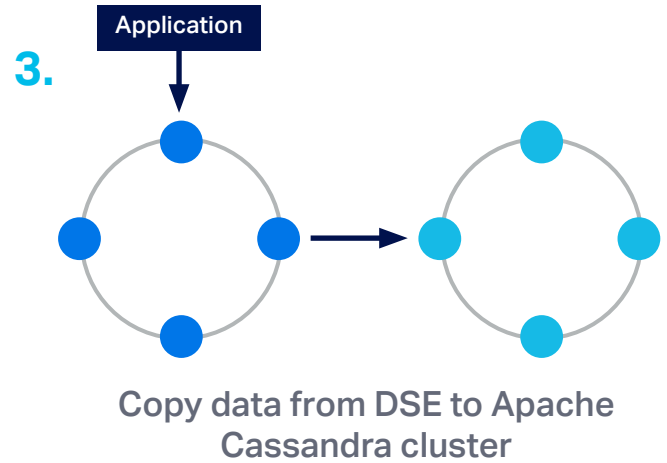
IN-PLACE

1.
Application

2.
Application

3.
Application

4.
Application
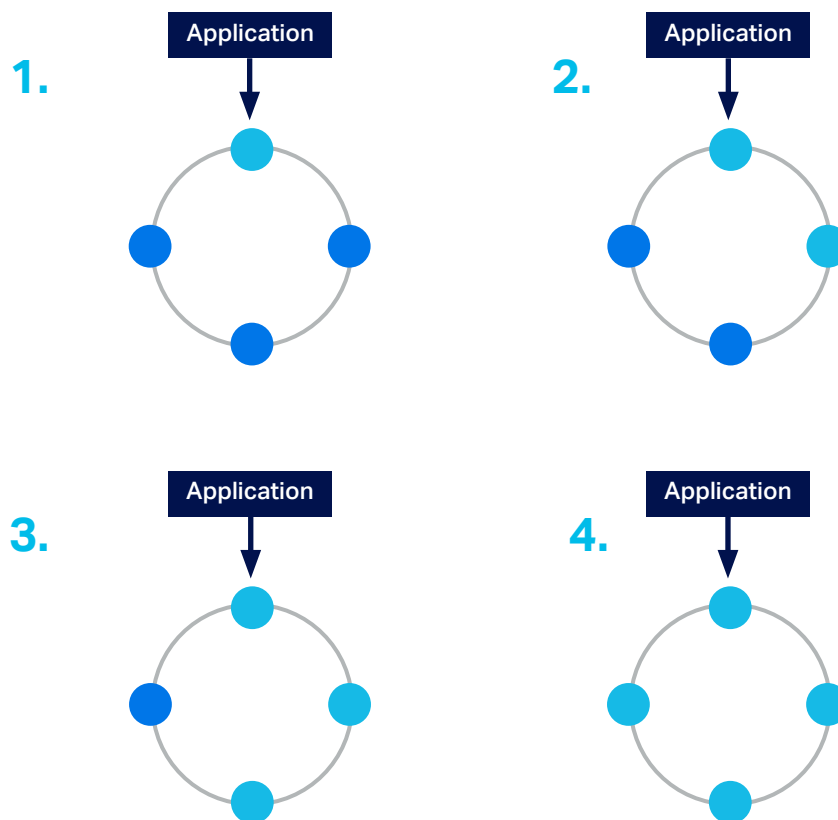
Transition each cluster one by one

OUT-OF-PLACE

1.
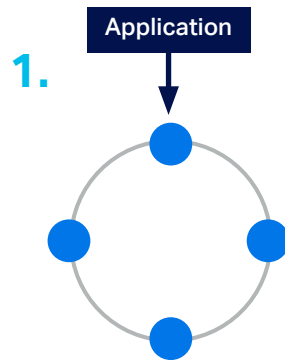
Application

Your original DSE cluster

2.

Application

Create the Apache Cassandra cluster

3.

Application

Copy data from DSE to Apache Cassandra cluster

4.
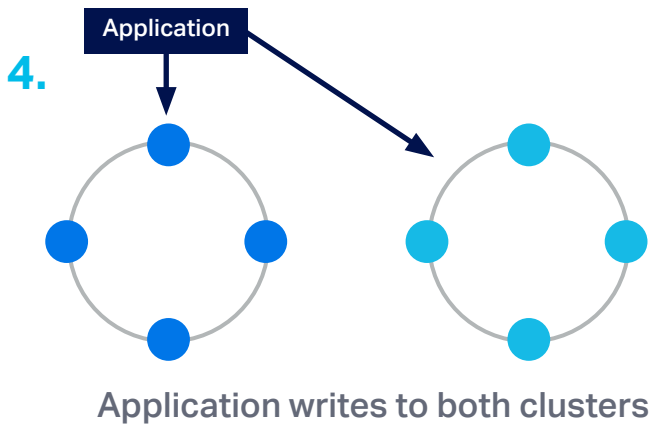
Application

Application writes to both clusters

5.

Application
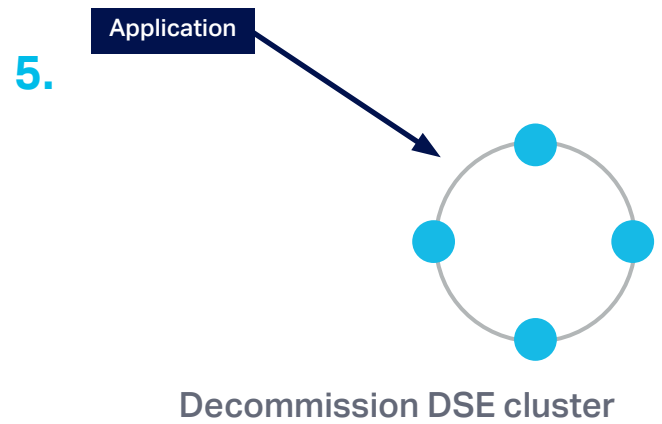
Decommission DSE cluster

A benefit of performing an in-place migration is that you don't have to provision additional hosting infrastructure to run one cluster with DSE and another cluster with Cassandra. But a benefit of an out-of-place migration is that by running the two clusters at once, you can compare them side-by-side before retiring the DSE cluster.

The following section will talk through the mechanics of both types of migration patterns in greater detail.

Migration Steps

The following are the steps you will perform to migrate from DataStax Enterprise to Apache Cassandra via an in-place migration.

- **Gap Analysis:** As discussed above, review proprietary features included with DataStax Enterprise that you will be replacing with alternative, open source solutions.

- **Remove Dependencies on Proprietary Features:** Remove any dependencies your application has on DataStax proprietary features.

- **Install Cassandra:** Download and install Apache Cassandra.

- **Switch Nodes to Cassandra:** To prevent an application outage, perform this process for one node at a time: shut down the DSE node, run the migration script, and start the node running Cassandra. Depending on the version, this process can be as short as 90 seconds per node.[2]

- **System Check:** Review the health and performance of your Cassandra cluster to ensure it is functioning properly.

- **Remove DSE:** Deprovision your DataStax Enterprise cluster and related hosting infrastructure.

For an out-of-place migration pattern, steps one, two, and three in the process are the same as the above. Then, you execute the following steps:

- **Provision Out-of-Place Infrastructure:** Establish your new cloud or on-prem infrastructure that will host your Cassandra cluster.

- **Synchronize Data:** Copy your data over in full from your DSE cluster to your Cassandra cluster. Or, start porting data from your application to both your DSE cluster and your Cassandra cluster until they hold equivalent data.

- **Connect Application to Cassandra:** If you don't already have your application connected to Cassandra from the previous step, now is the time to configure your application to use the Cassandra cluster rather than the DSE cluster in your data layer.

[2] Instaclustr has developed tools to make this step of the process safer, faster, and more resilient.

- **Remove DSE:** Deprovision your DataStax Enterprise cluster and related hosting infrastructure.

After you've migrated to your open source Cassandra instance, our expert Consulting team can help you work through the following additional steps to get onto our managed service:

- **Install Data Center:** Our Consulting team will create a second data center and attach it to your cluster. This data center will be running as part of Instaclustr's managed service.

- **Synchronize Data:** We synchronize the data between your Cassandra cluster and the cluster running in our data center.

- **Disconnect Old Cluster:** Disconnect the managed service cluster from the original, self-managed cluster and deprovision it later.

Don't Wait to Migrate

If you're a current DataStax Enterprise customer, you can expect to become increasingly locked-in over time, reducing your choices for your data architecture and making it harder to migrate to open source solutions like Cassandra. If you think migrating to Cassandra could be a fit for you, don't delay. Reach out to our expert team to assess if migration is a good fit for you.

About Instaclustr

Instaclustr helps organizations deliver applications at scale through its managed platform for open source technologies such as **Apache Cassandra®**, **Apache Kafka®**, **Apache Spark™**, **Redis™**, **OpenSearch®**, **PostgreSQL®**, and **Cadence®**.

Instaclustr combines a complete data infrastructure environment with hands-on technology expertise to ensure ongoing performance and optimization. By removing the infrastructure complexity, we enable companies to focus internal development and operational resources on building cutting edge customer-facing applications at lower cost. Instaclustr customers include some of the largest and most innovative Fortune 500 companies.