Oracle® Cloud

Migrating Oracle Data Hub Cloud Service Clusters to Oracle Cloud Infrastructure

Release 19.2.1
F18496-01
May 2019

Oracle Cloud Migrating Oracle Data Hub Cloud Service Clusters to Oracle Cloud Infrastructure, Release 19.2.1

F18496-01

Primary Author: Ashwin Agarwal

Contents

1

Learn About Migrating to Oracle Cloud Infrastructure

These topics help you learn about the benefits to migrating your existing Oracle Data Hub Cloud Service clusters to Oracle Cloud Infrastructure, and also provide an overview of the migration process and tools.

Topics:

Why Migrate to Oracle Cloud Infrastructure

Oracle encourages you to migrate your existing cloud resources to Oracle Cloud Infrastructure regions. You can gain several advantages by doing so.

In Oracle Cloud, you provision resources in specific regions, which are localized to geographic locations. Certain regions support the Oracle Cloud Infrastructure platform.

Oracle Cloud Infrastructure is Oracle's modern cloud platform that's based on the latest cloud technologies and standards. It provides more consistent performance and better features at lower costs. Oracle continues to invest in Oracle Cloud Infrastructure, including the addition of new regions, services, and features. See Data Regions for Platform and Infrastructure Services.

You can benefit from these additional administrative features when you migrate your cloud resources to Oracle Cloud Infrastructure:

- Organize cloud resources into a hierarchy of logical compartments.
- Create fine-grained access policies for each compartment.

To learn more, see Upgrade Your Classic Services to Oracle Cloud Infrastructure.

About the Migration Scope

A Data Hub cluster can be migrated to one of the following:

- **DataStax Distribution of Apache Cassandra (DDAC)**: DDAC is a certified version of Apache Cassandra that goes through the same rigorous quality assurance and certification testing process as DataStax Enterprise. It includes business day, 8x5 support or 24x7 worldwide support from DataStax.
- **DataStax Enterprise (DSE)**: DSE is more featured than Apache Cassandra or Data Hub. Integrated within each node of DSE is powerful indexing, search, analytics, and graph functionality, provided by combining Apache Cassandra with Apache Solr, Apache Spark, and DSE Graph.

> **Note:**
>
> You must purchase licenses for DDAC or DSE from DataStax to run them on Oracle Cloud Infrastructure.

About Oracle Cloud Infrastructure

Get familiar with basic Oracle Cloud Infrastructure security, network, and storage concepts.

Cloud resources in Oracle Cloud Infrastructure are created in logical compartments. You also create fine-grained policies to control access to the resources within a compartment.

You create instances within an Oracle Cloud Infrastructure region. You also specify an availability domain (AD), if supported in the selected region.

A virtual cloud network (VCN) is comprised of one or more subnets, and an instance is assigned to a specific subnet. Oracle Cloud Infrastructure does not allow you to reserve specific IP addresses for platform services.

A subnet's security lists permit and block traffic to and from specific IP addresses and ports.

Instances can communicate with resources outside of Oracle Cloud by using Oracle Cloud Infrastructure FastConnect, which provides a fast, dedicated connection to your on-premises network. Alternatively, use an IPSec VPN.

A bucket in Oracle Cloud Infrastructure Object Storage can be used to store files and share them with multiple instances. A user's generated authentication token (auth token) is required to access the bucket.

To learn more, see Key Concepts and Terminology in the Oracle Cloud Infrastructure documentation.

2

Prepare to Migrate Oracle Data Hub Cloud Service to Oracle Cloud Infrastructure

You can perform a live migration of your existing Oracle Data Hub Cloud Service cluster to *DataStax Distribution of Apache Cassandra* or *DataStax Enterprise* on Oracle Cloud Infrastructure IaaS by deploying new nodes and syncing the data. At a high level, this is accomplished by upgrading the existing nodes in Data Hub, joining new nodes in Oracle Cloud Infrastructure to the Data Hub cluster as a new data center, and by running a rebuild command or loading snapshot files.

Topics

Options to Deploy DDAC/DSE on Oracle Cloud Infrastructure

- The Oracle Cloud Infrastructure team along with the DataStax Solutions team has built Terraform deployment templates to automate deployment of DDAC and DSE on Oracle Cloud Infrastructure. These templates deploy all required infrastructure resources (virtual cloud network, subnet, security list, compute instances, storage) as well as DDAC/DSE. Please reach out to the Oracle Sales team to gain access to the Terraform template and its documentation.
- Both DDAC and DSE are available for download from DataStax's website. The customer is required to obtain a license from DataStax when installing the DDAC or DSE. Follow DataStax's documentation for installing and running of both solutions.

Size your Oracle Cloud Infrastructure Instances

The number and size (in terms of cores, RAM, and block storage) of your new nodes in Oracle Cloud Infrastructure should match those in Data Hub, choosing core number first.

- For example, if you deployed your Data Hub cluster using an OC6 (8 OCPU/60 GB RAM) shape you would deploy the same number of VM.Standard2.8 (8 OCPU/120 GB RAM) instances. A full list of Oracle Cloud Infrastructure shapes can be found here.

- The block storage volume for each node should be sized at 200% of the data contained on each Data Hub node to allow for growth and compaction operations. Disk sizing is discussed in detail here.

Check Data Hub Node Configuration

The `cassandra.yaml` file (default location: `/etc/cassandra/`) on all nodes should set the `endpoint_snitch` property to `GossipingPropertyFileSnitch`.

Upgrade your Data Hub Nodes

For both migrations, all existing Data Hub nodes need to be upgraded to either DDAC or DSE. This is a mandatory step because some operations required cannot be performed on a cluster with mixed versions. Instructions and a compatibility matrix are available here.

Prepare your Network

The new cluster nodes will be deployed into an Oracle Cloud Infrastructure VCN. These new nodes should either be deployed in a:

- Private subnet of an Oracle Cloud Infrastructure VCN that has been connected with your Oracle Cloud Infrastructure Classic network by following these instructions. Also in the VCN's security list port 7000 (port 7001 if using inter-node SSL) need to be open between the Data Hub andOracle Cloud Infrastructure CIDR blocks as described here.
- Public subnet. On each Oracle Cloud Infrastructure node set the `broadcast_address` property in `cassandra.yaml` to the node's public IP. The provided terraform templates will configure this property if the nodes are deployed to a public subnet. It should be noted that after the migration you can manually set this to the node's private IP. Also in the VCN's security list port 7000 (port 7001 if using inter-node SSL) need to be open between the Data Hub and Oracle Cloud Infrastructure CIDR blocks as described here.

3

Migrate an Oracle Data Hub Cloud Service Cluster to Oracle Cloud Infrastructure

These topics lists the steps to migrate an Oracle Data Hub Cloud Service cluster to Oracle Cloud Infrastructure.

Topics

Deploy the DataStax Cluster

Deploy nodes in Oracle Cloud Infrastructure as a stand-alone DDAC or DSE cluster by using the DataStax Oracle Cloud Infrastructure terraform templates that you receive from your sales professional. Deploying a stand-alone cluster first allows you to check your deployment and configuration, or perform any performance testing desired without touching your production cluster.

Stop Critical Services

Stop the *DDAC/DSE* services on all Oracle Cloud Infrastructure nodes by running either of the following commands:

```
sudo service dse stop
```

or

```
sudo service ddac stop
```

Add OCI Nodes as a New Data Center

Add the Oracle Cloud Infrastructure nodes as a *new datacenter* by following these instructions but **stop before running** the `nodetool rebuild` command. Since you're reusing nodes (i.e. this was a stand-alone cluster) the prerequisites listed are required.

Trigger Migration of Data

At this point you can migrate data.
If there is < 1 TB of data per node, it is quickest to use the `nodetool rebuild` command in the instructions linked above by SSH-ing onto each Oracle Cloud

Infrastructure node. Note, this puts a load on the Data Hub nodes. We recommend that you run this on one node to gauge load and rebuild time. If the load on the existing cluster is reasonable and the time taken is not too long, gradually increase the number of nodes being rebuilt.

If there is > 1 TB of data per node or `nodetool rebuild` command was prohibitively slow, it is recommended to:

1. Install the Oracle Cloud Infrastructure CLI on all Data Hub and Oracle Cloud Infrastructure nodes.
2. Snapshot the data for each keyspace that needs to be migrated.
3. Create a bucket in Oracle Cloud Infrastructure Object Storage.
4. Copy the snapshots from the Data Hub nodes to the Oracle Cloud Infrastructure nodes using the Oracle Cloud Infrastructure CLI per the details provided below:
   - From each Data Hub node upload each snapshot to Oracle Cloud Infrastructure object storage by running a command like the one below. Values in <> are either determined by the snapshot command run above or the bucket name you choose.

     ```
     oci os object bulk-upload --bucket-name <bucket> \
     --src-dir /var/lib/cassandra/data/<keyspace_name>/<table_name-UUID>/
     snapshots/<snapshot_name> \
     --object-prefix <node_hostname>/<keyspace_name>/<table_name-UUID>/
     <snapshot_name>
     ```

   - On each Oracle Cloud Infrastructure node you can download all snapshots from the corresponding Data Hub node by running:

     ```
     oci os object bulk-download --bucket-name <bucket> \
     --download-dir <snapshot_tmp_dir> \
     --prefix <node_hostname>
     ```

5. Finally, follow these instructions to restore the snapshot on the new nodes.

> **Note:**
>
> These operations need to be performed quicker than `gc_grace_seconds` (default 10 days) at the risk of deleted data reappearing.

4

Complete the Post-Migration Tasks

After successfully migrating your Oracle Data Hub Cloud Service clusters from Oracle Cloud Infrastructure Classic to Oracle Cloud Infrastructure, test your applications thoroughly, and then perform cleanup and other optional tasks.

Topics

Decommission your Data Hub Nodes

Follow these instructions.

Terminate your Data Hub Nodes

At this point all data has been migrated, all clients are using only the new nodes/data center, and the Data Hub nodes are not part of the cluster. The Data Hub resources can now be terminated.