Cloudera Migration Assistant 4.1.1

End-to-end migration walkthroughs for Cloudera Migration Assistant

Date published: 2026-06-18
Date modified: 2026-07-29

CLOUDERA

https://docs.cloudera.com/

Legal Notice

Contents

End-to-end migration walkthroughs..... 4

Migrating HDFS data to HDFS end to end.....4

Planning Ozone storage layout end to end [Technical Preview].....15

Migrating HDFS data to Ozone end to end.....26

Migrating Airflow DAGs from to Git end to end [Technical Preview]..... 40

Discovering Ranger policies end to end [Technical Preview].....53

End-to-end migration walkthroughs

Complete how-to tasks that walk from Cloudera Migration Assistant deployment through workload execution for each supported migration path on CDP.

These walkthroughs are for operators who want one narrative from master and agent deployment through migration execution. Each topic reuses procedures from the Administrator and Migration guides so install, security, and workload steps are not duplicated here.

What you can do with Cloudera Migration Assistant

After you deploy the master and agents, you register source and target clusters, run scans to inventory data and services, and group assets into collections for migration plans. Typical data paths on CDP include HDFS to HDFS replication with Cloudera Replication Manager, planning and deploying an Ozone storage layout on the target, and migrating HDFS data into Ozone. You can also follow technical preview walkthroughs for Cloudera Data Engineering Airflow DAGs to Git and for Ranger policy discovery. Code Migrator is not available on CDP in release 4.1.

For install, TLS, databases, and platform configuration, use the *Administrator* guide. For individual tasks (registering a cluster, running a scan, creating a collection), use the *Migration* guide. Use the table below when you want a single end-to-end sequence.

Choose a walkthrough

Migration path	End-to-end topic
HDFS to HDFS	*Migrating HDFS data to HDFS end to end*
Ozone storage layout (technical preview)	*Planning Ozone storage layout end to end*
HDFS to Ozone	*Migrating HDFS data to Ozone end to end*
Cloudera Data Engineering Airflow DAGs to Cloudera Workflow Orchestrator Git (technical preview)	*Migrating Airflow DAGs from to Git end to end*
Ranger policy discovery (technical preview)	*Discovering Ranger policies end to end*

Supported platforms and version requirements are listed in *Supported platforms and migration paths*.

Screenshot backlog

End-to-end workflow topics may reference screenshots that are not yet captured in this release. Stage builds show placeholder figures where noted in individual workflows.

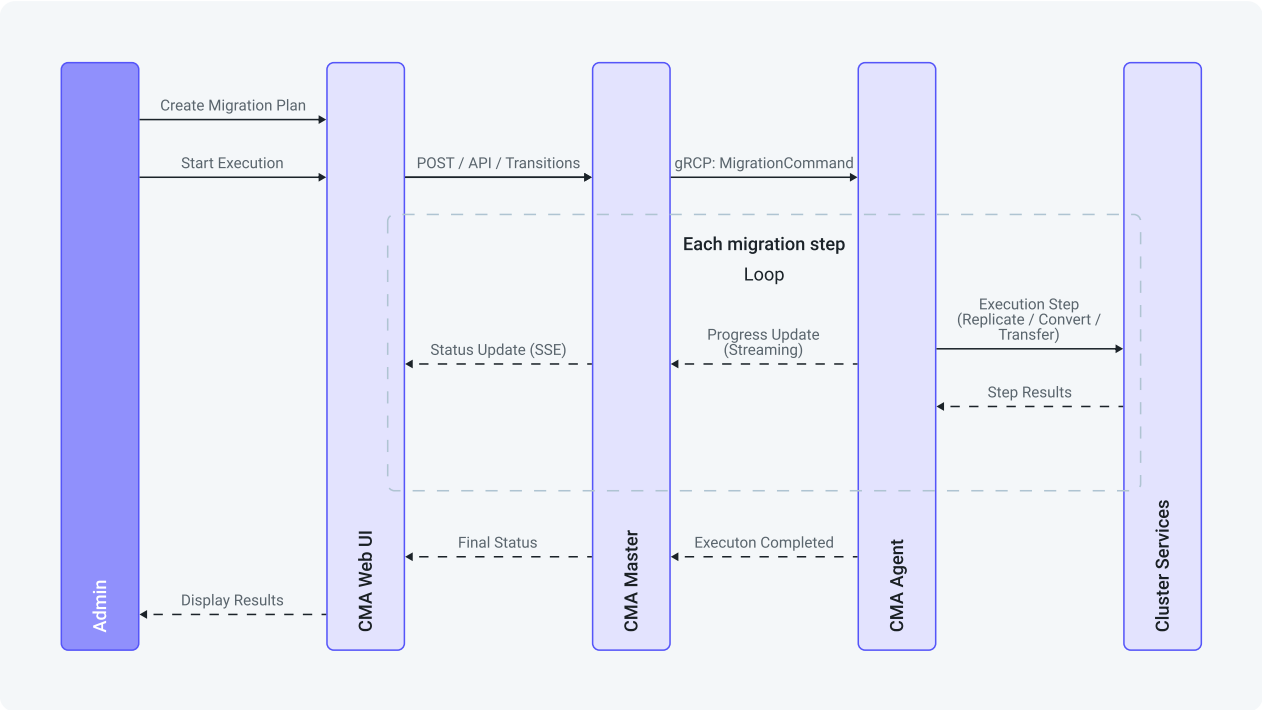
Migrating HDFS data to HDFS end to end

Deploy Cloudera Migration Assistant, register clusters, and run an HDFS to HDFS migration using Cloudera Replication Manager replication.

About this task

This end-to-end workflow combines deployment, platform preparation, discovery, and workload steps. Detailed reference procedures are reused from the Administrator and Migration guides.

Figure 1: Migration flow



Before you begin

Verify supported platforms and migration paths in *Supported platforms and migration paths* before you start.

Procedure

Phase 1 — Deploy the CMA Master and CMA Agents

Procedure

1. Download both CSD files to the /opt/cloudera/csd/ directory on the Server host.

   ```
   wget -P /opt/cloudera/csd/ https://archive.cloudera.com/cma/VERSION/csd/
   CMA_MASTER-VERSION-BUILD.jar
   ```

   ```
   wget -P /opt/cloudera/csd/ https://archive.cloudera.com/cma/VERSION/csd/
   CMA_AGENT-VERSION-BUILD.jar
   ```

   automatically detects the CSD files.
2. Change the ownership of the CSD files.

   ```
   chown cloudera-scm:cloudera-scm /opt/cloudera/csd/CMA_MA
   STER-VERSION-BUILD.jar /opt/cloudera/csd/CMA_AGENT-VERSION-BUILD.jar
   ```
3. Restart for the changes to take effect.

   ```
   systemctl restart cloudera-scm-server
   ```
4. Log into .
5. Restart the Cloudera Management Service.
6. Go to Hosts Parcels .
7. Click Parcel Repositories & Network Settings.

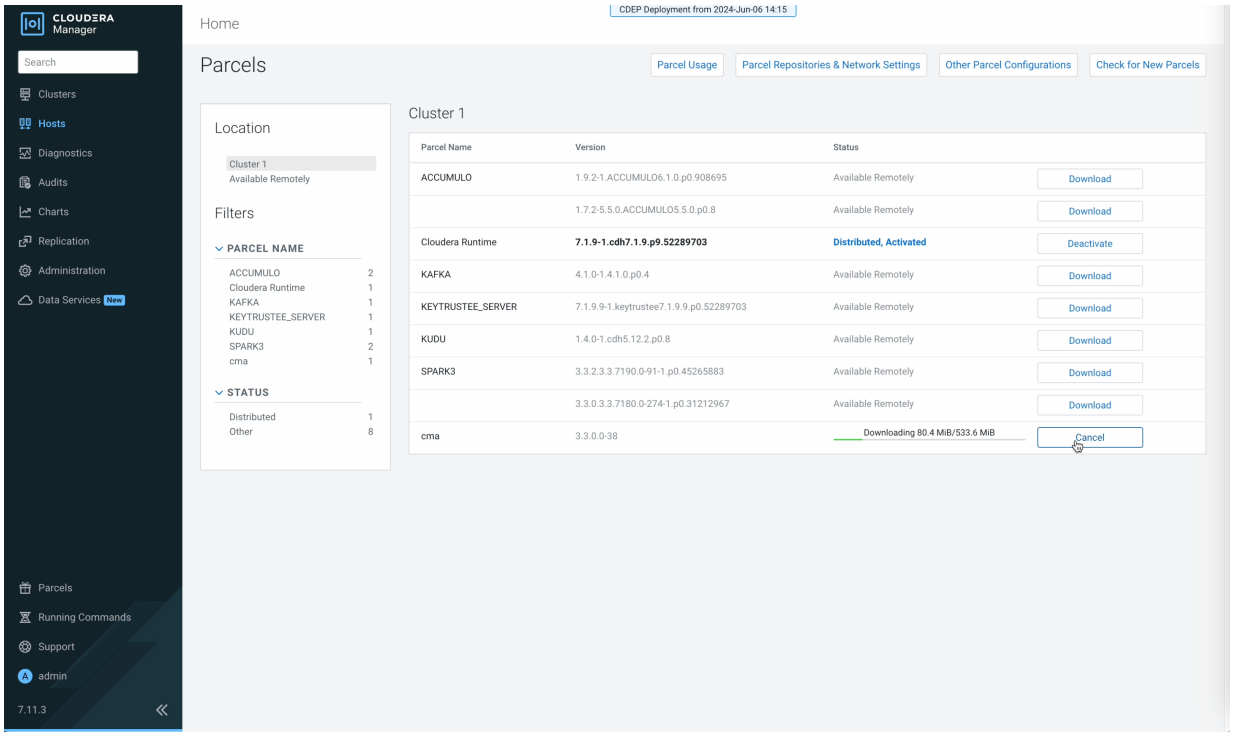
8. Add the Remote Parcel Repository URL for .

```
https://archive.cloudera.com/cma/VERSION/parcels/
```

Note: Make sure that the Remote Parcel Repository URL uses an HTTPS link. To install a different version of the parcel, change the URL as needed.

9. Click Save & Verify Configuration to commit the change.
10. Click Close.
    You are redirected to the Parcels page.
11. Search for , and click Download to download the parcel to the local repository.
12. After download completes, click Distribute to distribute the parcel to all clusters.

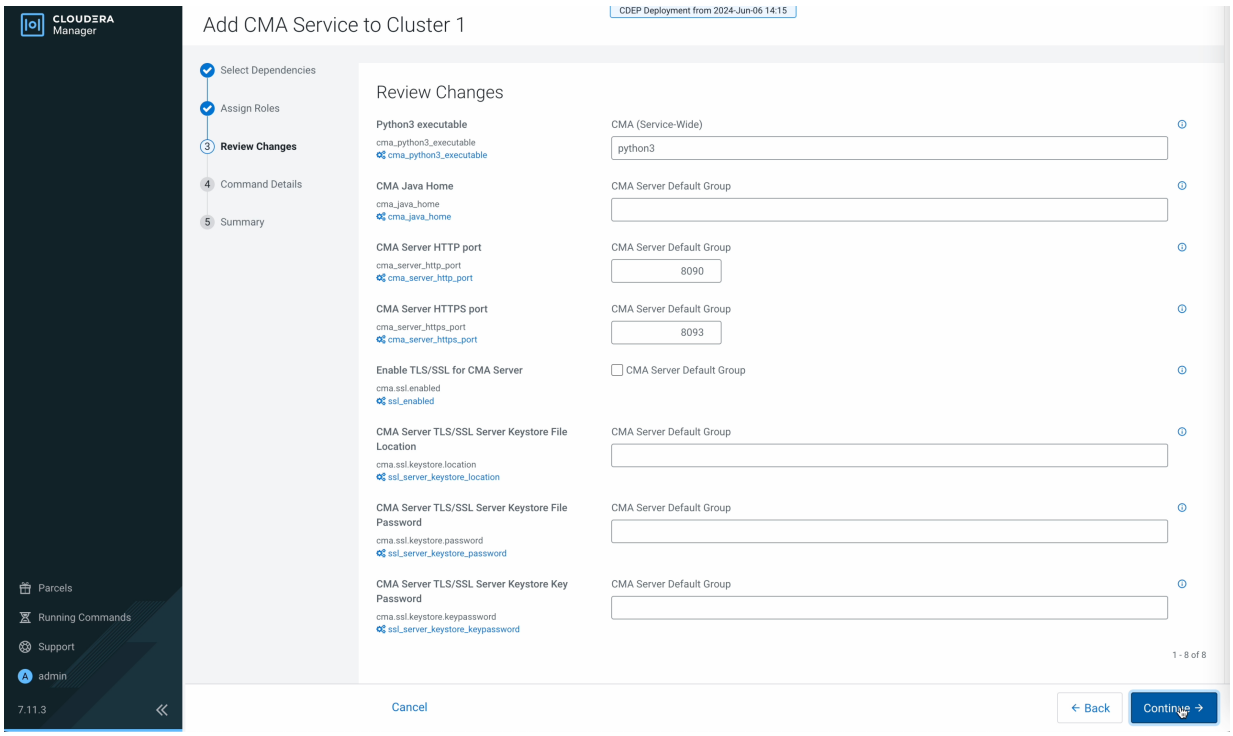
Figure 2: Distribute the parcel



13. Click Activate to activate the parcel.
14. Click OK when confirmation is required.
15. Click Clusters in the left navigation pane.
16. Select Add Service from the drop-down menu to the right of your cluster.
17. From the list, select the service type to add, then click Continue. The single parcel provides two service types: CMA_MASTER (central control node) and CMA_AGENT (cluster-side executor).

    The **Add Service wizard** opens.

Figure 3: Add service

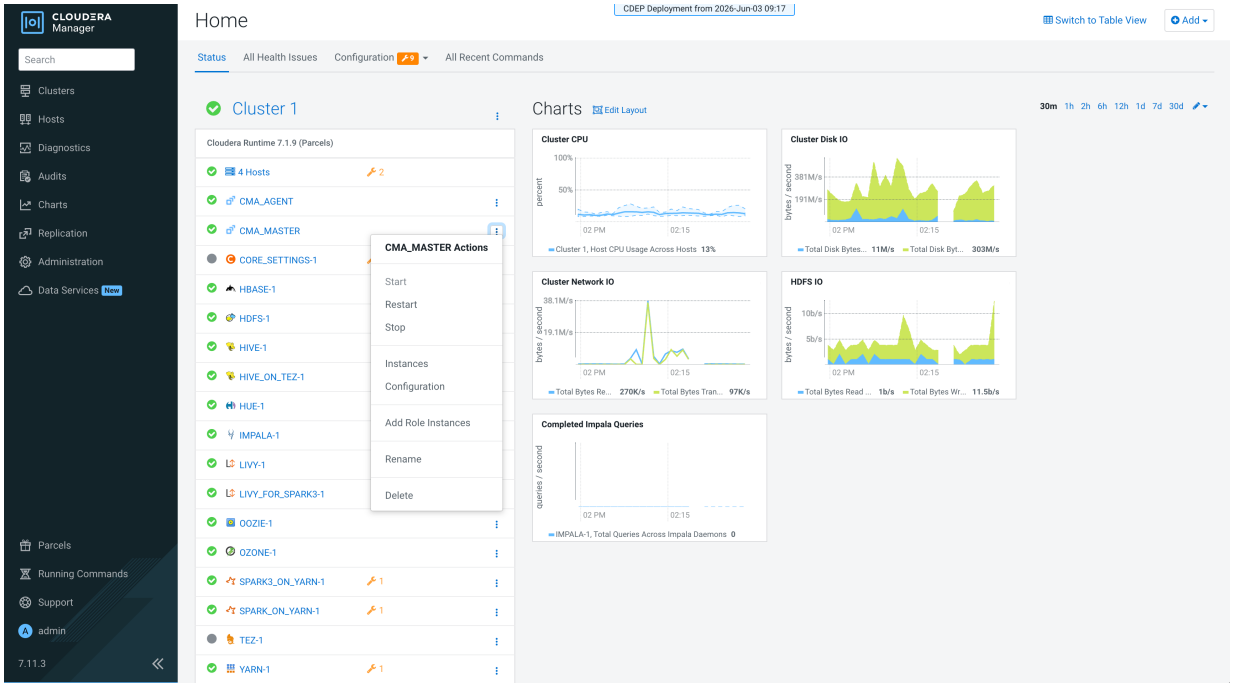


18. Assign the service roles to the hosts where Python 3.11 and JDK 17+ are installed, and click Continue.

19. Review service configurations and click Continue. At minimum, set OAuth2 client credentials (cma_client_secret, cma_agent_client_secret). If the is on a different cluster from the , also set Gateway URL (cma_gateway_url).

The first run of the service starts. When the command finishes, the service is added to the cluster.

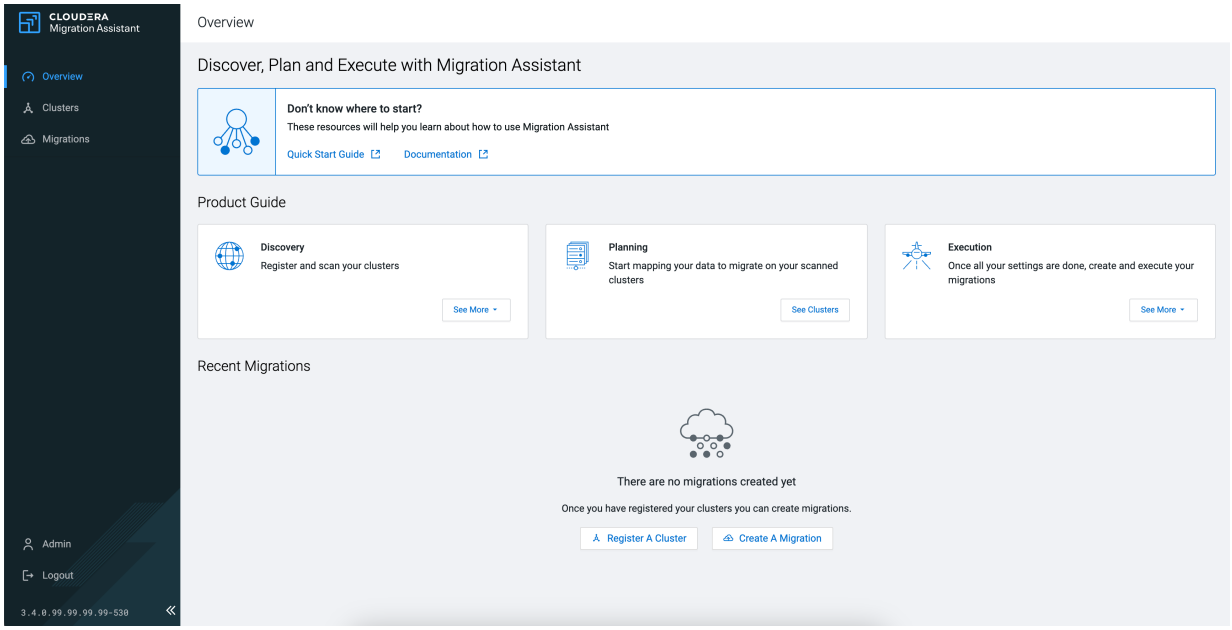
Figure 4: service started



20. Repeat steps 16–19 to add the other service type if needed (for example, add CMA_AGENT after CMA_MAST ER, or vice versa).

21. Go back to the cluster homepage, open the service page, and click the Server UI tab to open the UI.

Figure 5: landing page



22. Set the service dependency to none in the service configuration in Cloudera Manager because Cloudera Manager cannot discover the Master automatically across clusters.

23. Set the **Gateway URL** to cma_gateway_url in the service configuration.

```
curl -sk https://<master-host>:8093/api/agent/config
```

Procedure

Phase 2 — Configure Cloudera Replication Manager on the target cluster

Procedure

1. Enable the HDFS to Ozone flag on the Server host.

For HDFS to Ozone migrations, requires the CMF_FF_API_H2O_REPLICATION feature flag to be enabled. Without this flag, the API endpoints for HDFS to Ozone replication are not available.

Note: This flag is only required for HDFS to Ozone migrations. It is not required for HDFS-to-HDFS, Hive, or HBase replication.

a) Append the following line to /etc/default/cloudera-scm-server on the Server host:

```
export CMF_FF_API_H2O_REPLICATION=true
```

b) Restart the Server.

```
sudo systemctl restart cloudera-scm-server
```

c) Wait for to become available again on port 7180 before proceeding with the setup.

2. Register the instance in the UI to allow to locate and communicate with the replication service.

a) In the Web UI, go to Settings.

b) Go to the App Access section.

c) Configure your connection fields using one of the following validation methods:

- If has already detected a service on one of the registered clusters, select it from the Suggestions drop-down list to auto-fill the form.
- Manually enter your target replication parameters in the following fields:
  - App URL — The full URL of the REST API, including protocol, host, and port, for example, https://rm-host.example.com:9011.
  - App username — The username for authenticating with the API.
  - App password — The password for authenticating with the API.

d) Click Save and Push to Agents.

Procedure

Phase 3 — Register source and target clusters

Procedure

1. In on the source cluster, select Add Service and select CMA_AGENT.
2. Configure the required properties.

Table 1: configuration properties

Property	Description
cma_gateway_url	URL of the Gateway, for example, https://master-host:8093. **Required when the is on a different cluster than the .** The agent uses this URL to discover the Master host, gRPC port, and Auth Server URI automatically through the discovery endpoint (GET /api/agent/config). When connecting from a different cluster, set the service dependency to none in .
cm_username	administrator user name.
cm_password	administrator password.
cma_user_home	Home directory for the cma-agent user, for example, /var/lib/cma-agent.

3. Set cma_gateway_url when the is on a different cluster than the .
4. Start the service.
5. Wait for the agent to connect to the and finish the registration automatically.

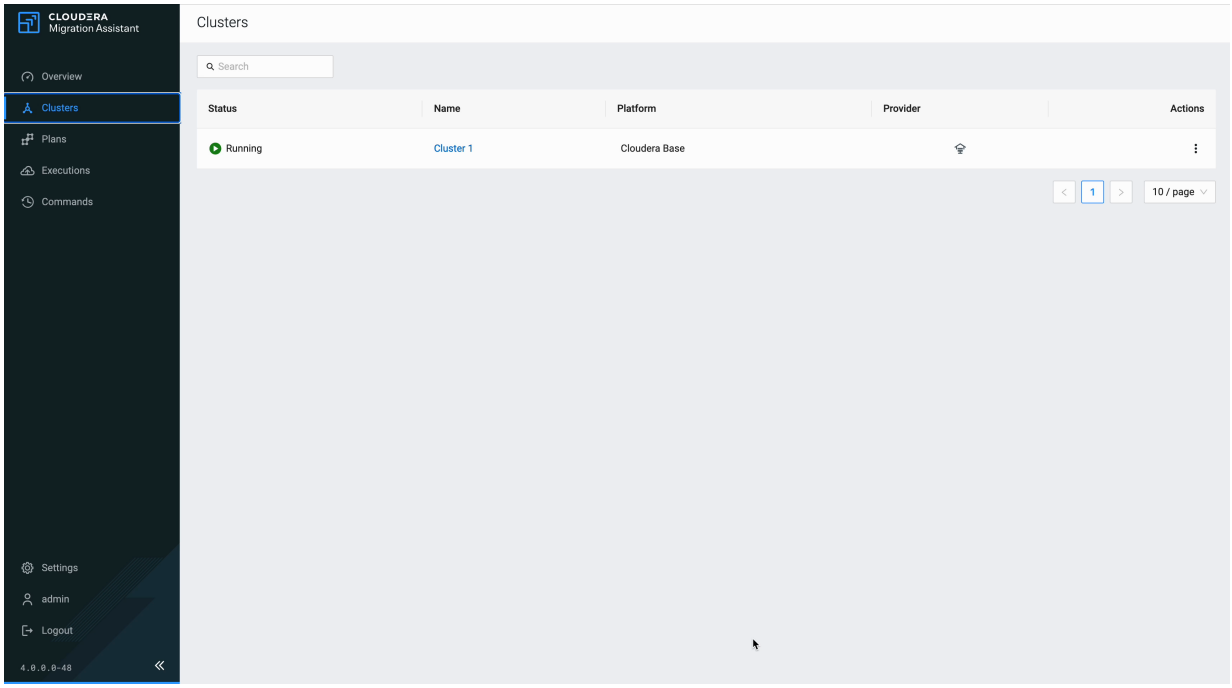
Procedure

Phase 4 — Scan clusters

Procedure

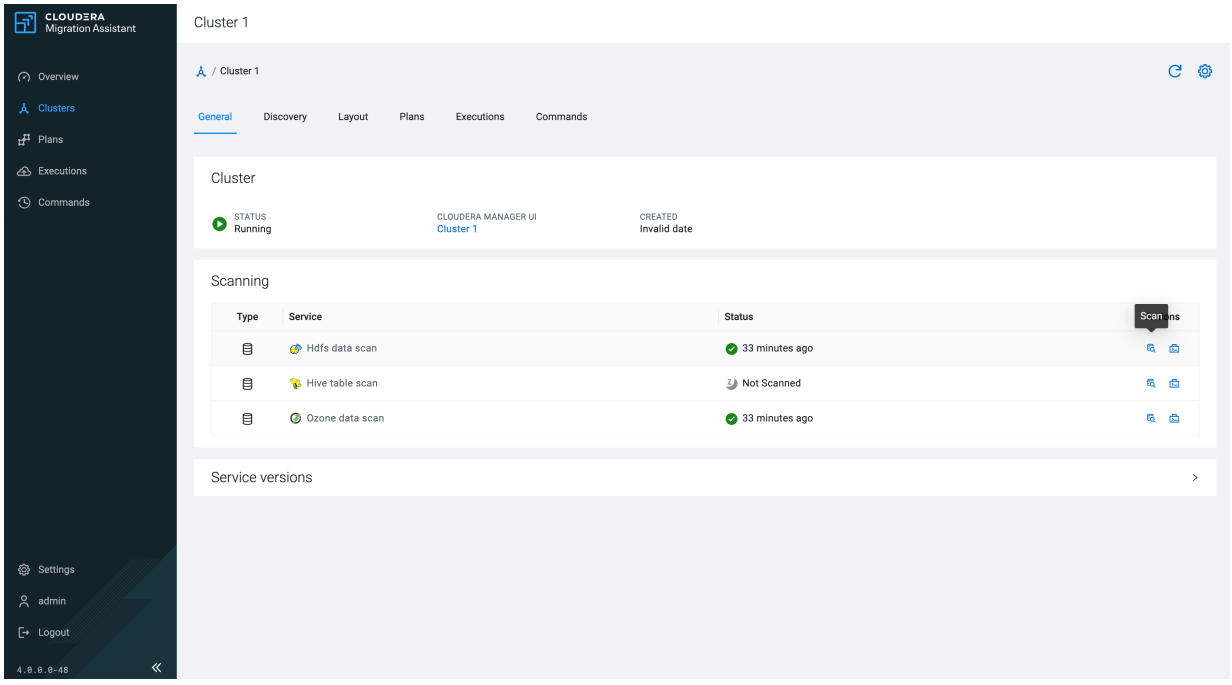
1. Navigate to the Clusters page and click the cluster you want to scan.

Figure 6: Clusters page



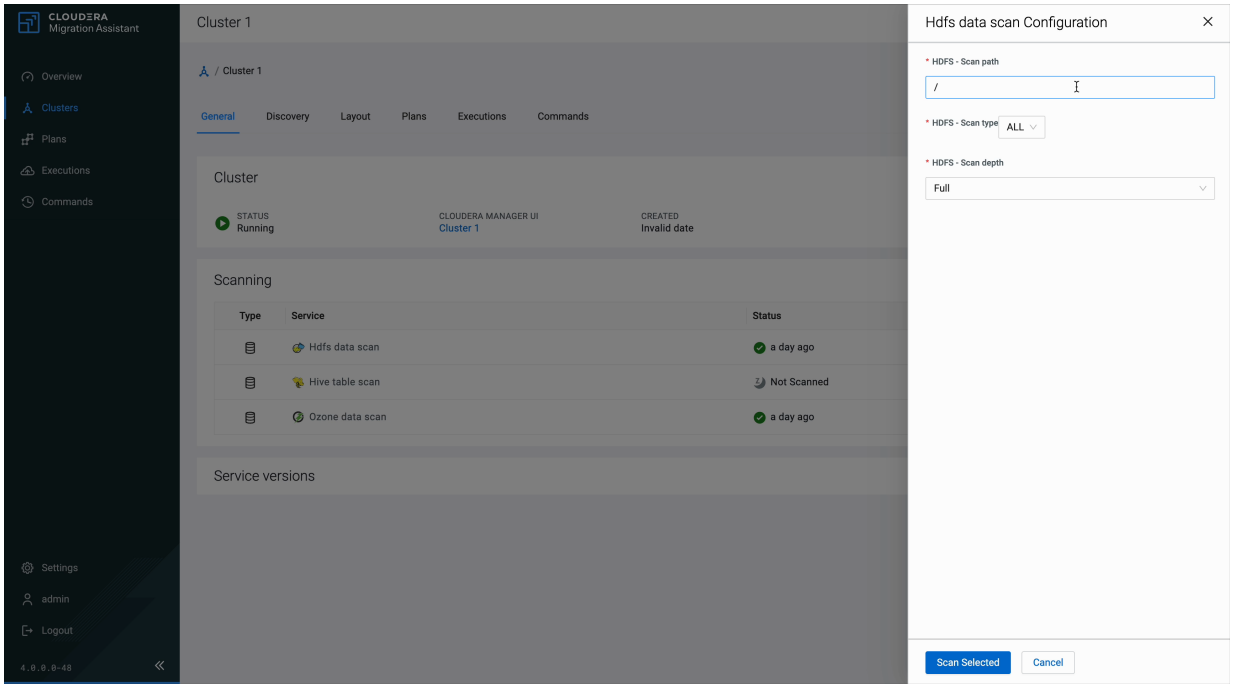
2. On the cluster detail page, find the Scanning section, which lists the available scan types for your cluster. Click the configuration icon on the right side of the scan type you want to trigger.

Figure 7: cluster scanning



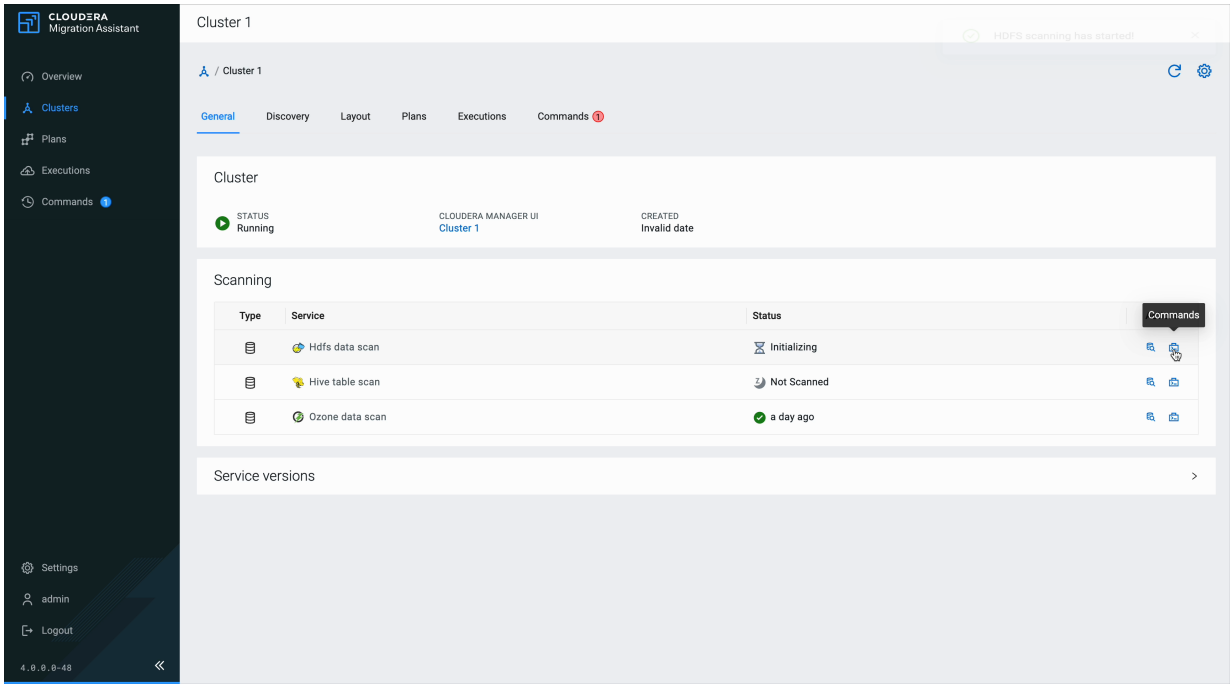
3. In the scan configuration drawer, configure the scan parameters (for example, scan path, scan type, scan depth) and click Scan Selected.

Figure 8: scan configuration



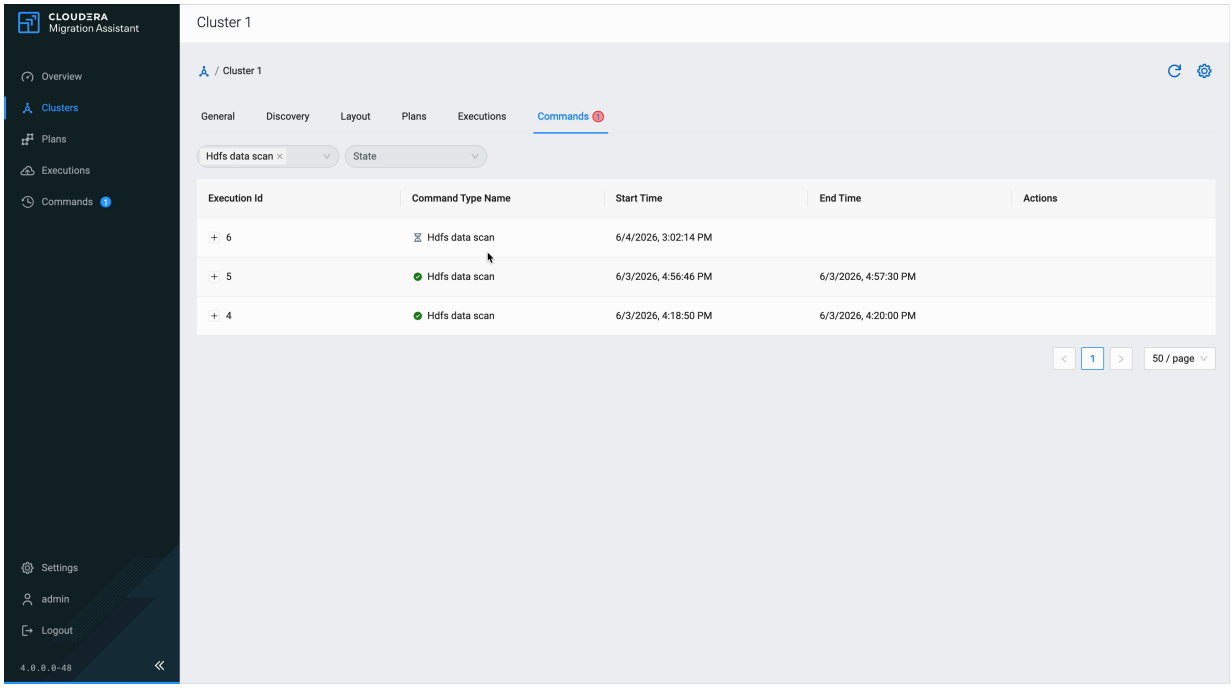
4. The scan starts and the status changes to Initializing. Track scan progress by clicking the Commands tab, which shows a badge with the number of active commands.

Figure 9: scan in progress



5. On the Commands tab, expand a command to view its Output Log, Error Log, and Configuration.

Figure 10: scan commands



Note: Scan results are also available in the data directory on the source cluster at <CMA_ROOT_DIR>/data/sources/ClusterName/Source_ID/cluster-scan/ on the host.

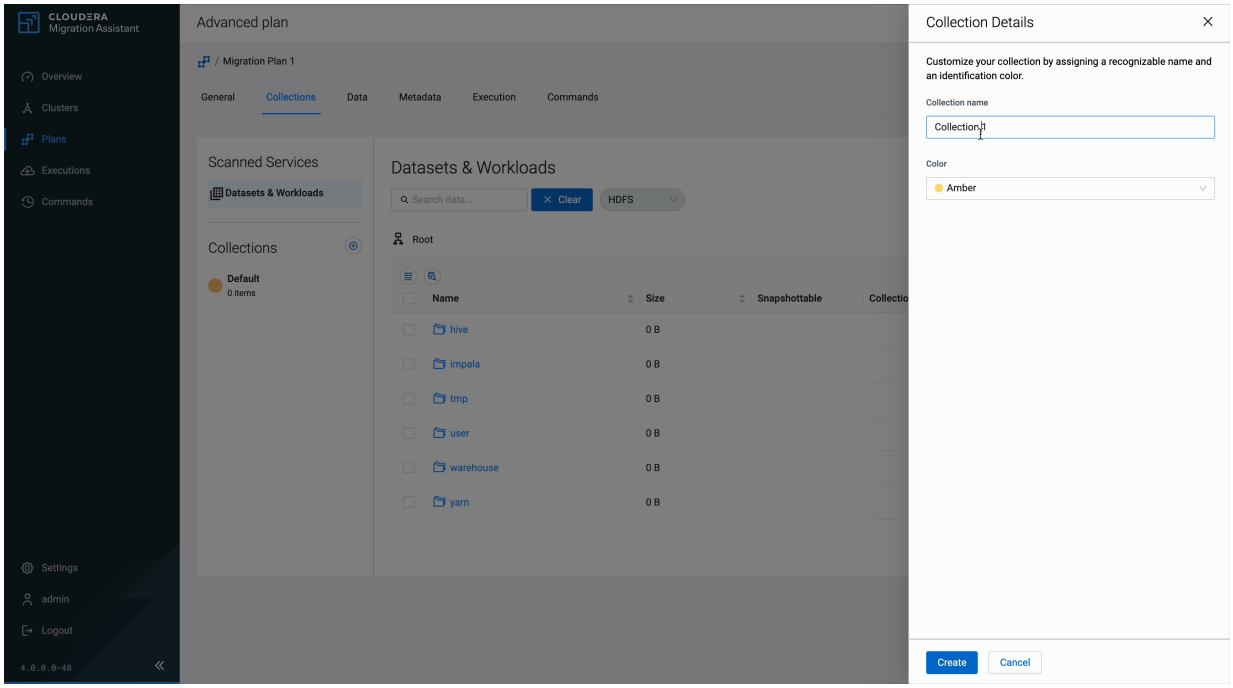
Procedure

Phase 5 — Create collections (optional)

Procedure

1. Click the + icon next to Collections in the sidebar.
2. In the **Collection Details** drawer, enter a Collection name and select a Color for identification.

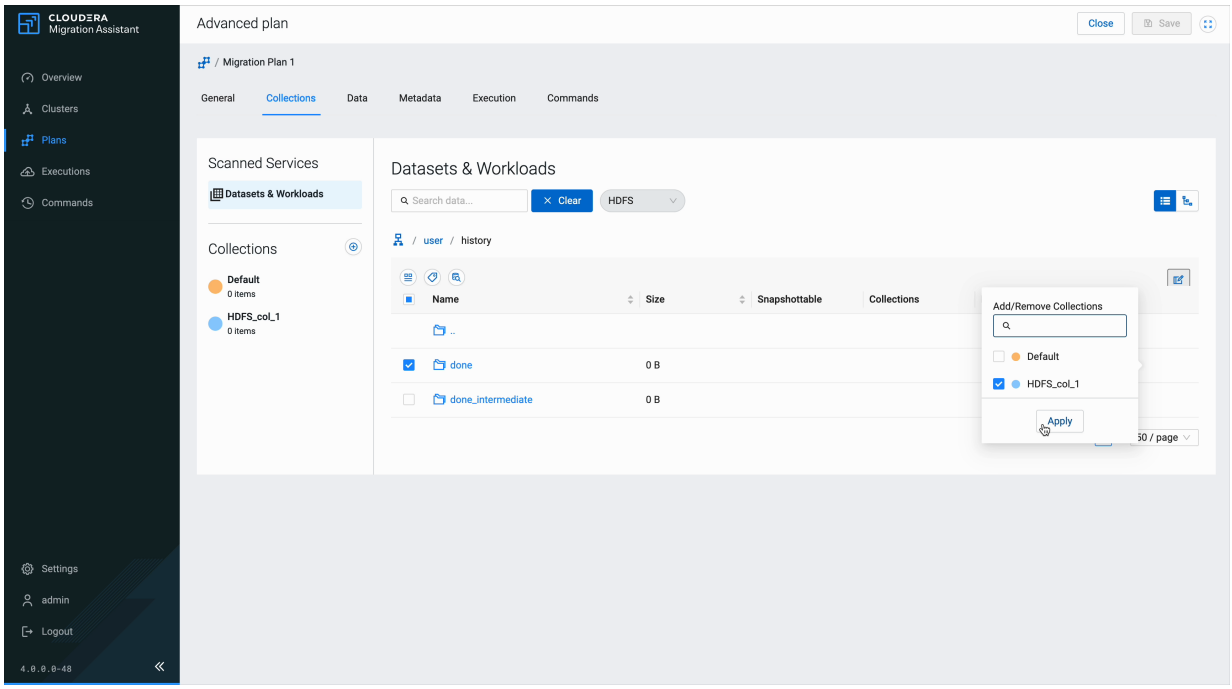
Figure 11: create collection



3. Click Create.

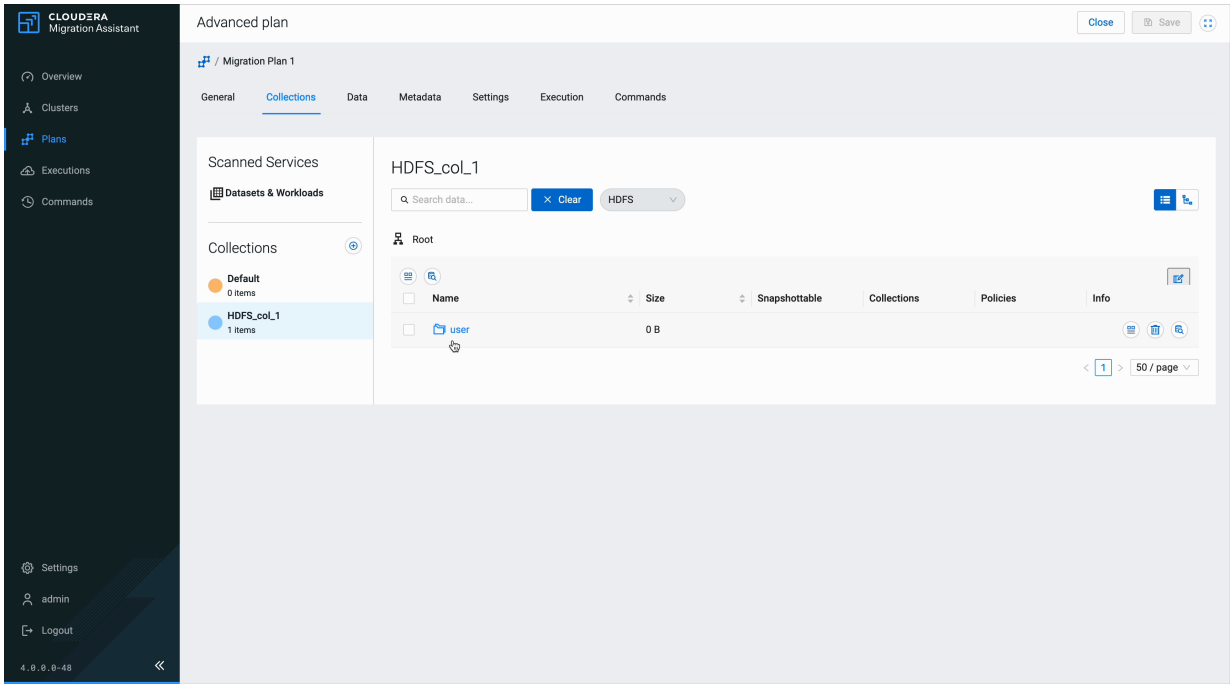
   The new collection appears in the sidebar with 0 items.
4. Browse the data tree to find the items you want to migrate. Click folder names to navigate deeper into the hierarchy.
5. Select items using the checkboxes next to each item.
6. Click the collection icon that appears on the selected item row.
7. In the **Add/Remove Collections** popup, check the collections you want to assign the item to.
8. Click Apply.

Figure 12: assign to collection



9. Click a collection name in the sidebar to view its contents. The data browser switches to show only the items assigned to that collection.

Figure 13: view collection



Procedure

Phase 6 — Run the migration workload

Procedure

Step 1 — Scan the source cluster (HDFS)

1. Open Clusters, then select the source cluster.
2. Click Start Scanning.
3. Select Hdfs data scan.
4. Constrain the scan path, for example, /tmp or /user, to shorten runs on massive namespaces.
5. Click Scan selected and watch the progress drawer until completion.

   After success, Discovery lists HDFS trees with sizing and ownership snapshots.

Step 2 — Scan the target cluster (HDFS)

6. From **Clusters**, open the target cluster.
7. Click Start Scanning.
8. Select Hdfs data scan.
9. Click Scan selected and wait until the status turns green.

   HDFS hierarchies populate inside Discovery afterward.

Step 3 — Create a migration plan

10. Go to Plans.
11. Click Create Plan.
12. Set Name, select the source HDFS cluster, and the target HDFS cluster.
13. Complete the wizard (Create) and land on the plan detail surface.

Step 4 — Map HDFS directories to HDFS paths on the Data tab

The Data tab displays source HDFS on the left and target HDFS on the right.

14. Go to Plan Data .
15. On the target side, browse to the parent directory that should receive migrated data.
16. On the source side, go to the parent directory that contains data you want to migrate.
17. Hover the directory and click Copy. persists a structural mapping.
18. Switch to the **Review** view any time to view a list of your current structure mappings.
19. Repeat the mapping workflow for additional directories whenever you broaden scope.

    You can map each directory to separate target prefixes or combine them under shared parents. The layout must meet the requirements for your tenancy model.

Step 5 — Create and run an execution

20. Open Execution.
21. Select Create Execution.
22. Complete the steps in the configuration wizard (General, Scope, Review), and then click Finish.
23. Open the execution dashboard and trigger Run All.

    During execution the following stages occur:

    - **Policy creation** – registers policies mapping HDFS prefixes to destination HDFS paths.
    - **Data replication** – invokes DistCp-style transfers governed by settings.
    - **Verification** – tracks step health until completion (COMPLETED or FAILED statuses per step).

Step 6 — Verify the migration

24. Confirm that the execution steps display COMPLETED.
25. Inspect HDFS paths inside Discovery on the target cluster.
26. Validate migrated paths using the HDFS CLI on the target cluster.

```
hdfs dfs -ls /path/to/migrated/data
```

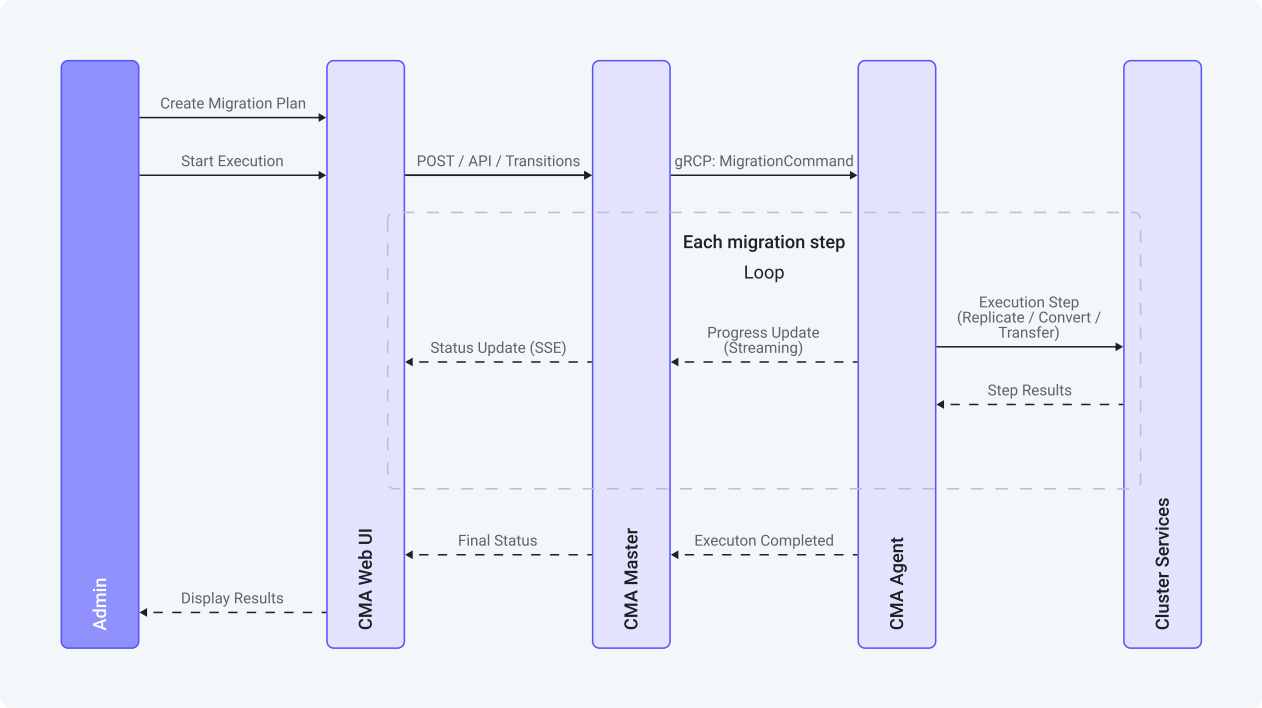
Planning Ozone storage layout end to end [Technical Preview]

Deploy Cloudera Migration Assistant, scan Ozone on the target cluster, and plan and deploy an Ozone storage layout before data migration.

About this task

This end-to-end workflow prepares the target cluster Ozone hierarchy using the **Storage Layout Editor**. It does not copy HDFS data. For the full HDFS to Ozone path, see *Migrating HDFS data to Ozone end to end*. Detailed steps are reused from the Administrator guide through content references.

Figure 14: Migration flow



Before you begin

Verify supported platforms in *Supported platforms and migration paths*. Ozone layout provisioning is technical preview; see *Managing Ozone storage*.

Procedure

Phase 1 — Deploy the CMA Master and CMA Agents

Procedure

1. Download both CSD files to the /opt/cloudera/csd/ directory on the Server host.

   ```
   wget -P /opt/cloudera/csd/ https://archive.cloudera.com/cma/VERSION/csd/
   CMA_MASTER-VERSION-BUILD.jar
   ```

   ```
   wget -P /opt/cloudera/csd/ https://archive.cloudera.com/cma/VERSION/csd/
   CMA_AGENT-VERSION-BUILD.jar
   ```

   automatically detects the CSD files.
2. Change the ownership of the CSD files.

   ```
   chown cloudera-scm:cloudera-scm /opt/cloudera/csd/CMA_MA
   STER-VERSION-BUILD.jar /opt/cloudera/csd/CMA_AGENT-VERSION-BUILD.jar
   ```
3. Restart for the changes to take effect.

   ```
   systemctl restart cloudera-scm-server
   ```
4. Log into .
5. Restart the Cloudera Management Service.
6. Go to Hosts Parcels .
7. Click Parcel Repositories & Network Settings.

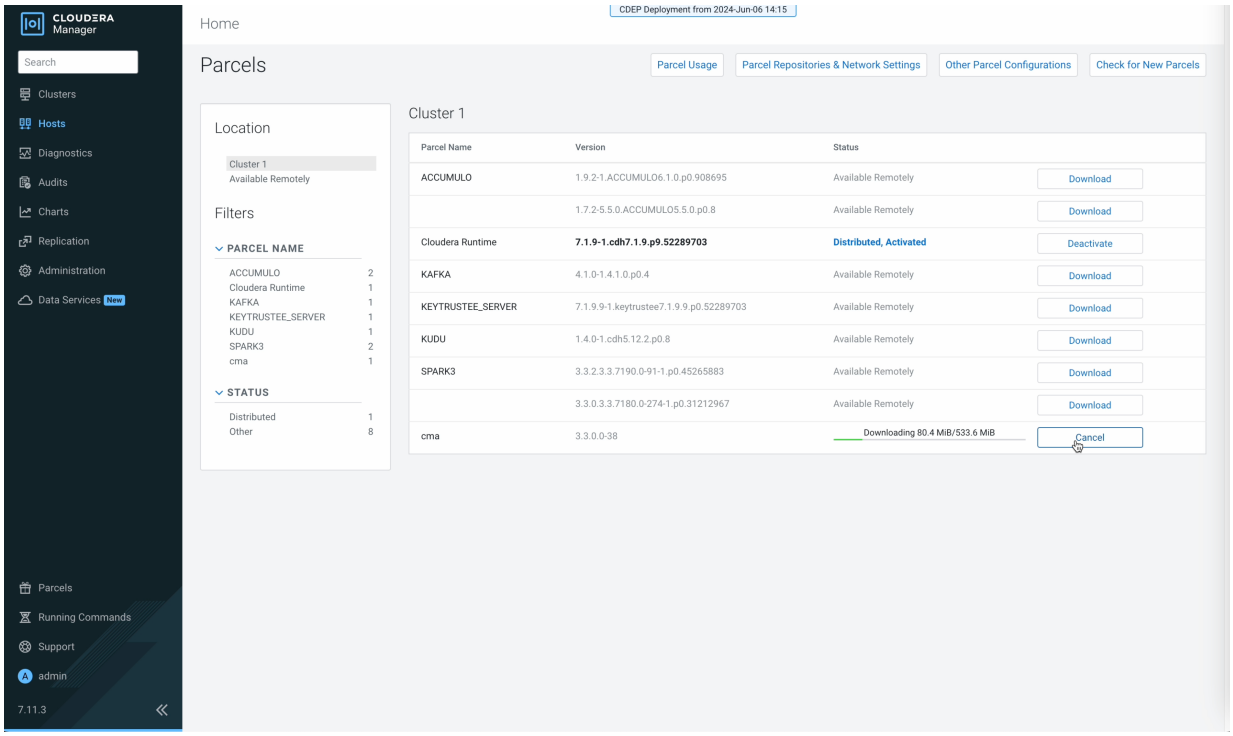
8. Add the Remote Parcel Repository URL for .

```
https://archive.cloudera.com/cma/VERSION/parcels/
```

Note: Make sure that the Remote Parcel Repository URL uses an HTTPS link. To install a different version of the parcel, change the URL as needed.

9. Click Save & Verify Configuration to commit the change.
10. Click Close.
    You are redirected to the Parcels page.
11. Search for , and click Download to download the parcel to the local repository.
12. After download completes, click Distribute to distribute the parcel to all clusters.

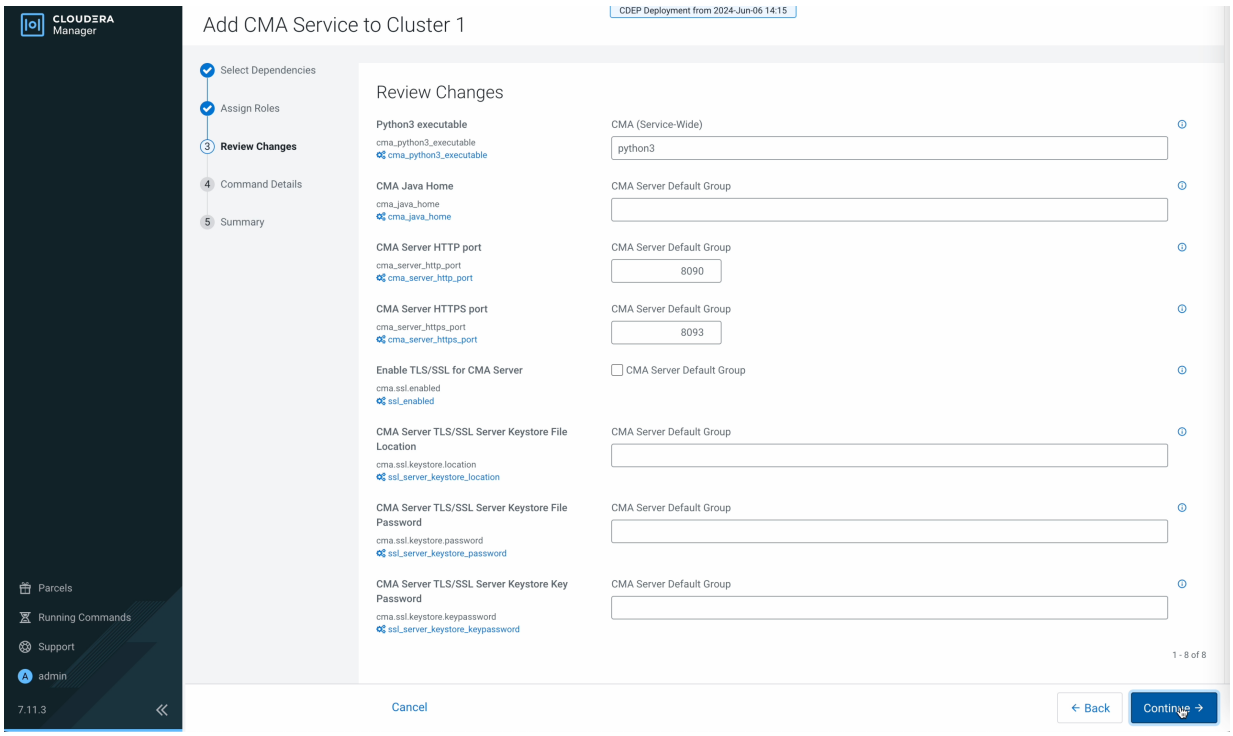
Figure 15: Distribute the parcel



13. Click Activate to activate the parcel.
14. Click OK when confirmation is required.
15. Click Clusters in the left navigation pane.
16. Select Add Service from the drop-down menu to the right of your cluster.
17. From the list, select the service type to add, then click Continue. The single parcel provides two service types: CMA_MASTER (central control node) and CMA_AGENT (cluster-side executor).

    The **Add Service wizard** opens.

Figure 16: Add service

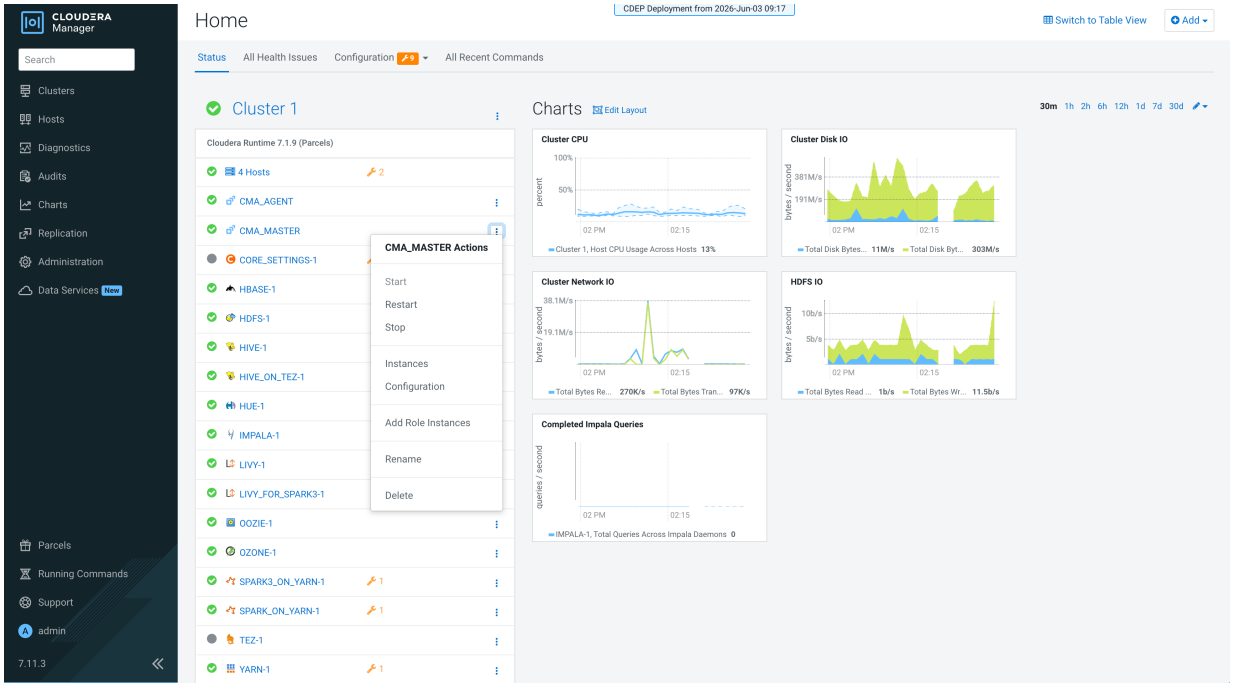


18. Assign the service roles to the hosts where Python 3.11 and JDK 17+ are installed, and click Continue.

19. Review service configurations and click Continue. At minimum, set OAuth2 client credentials (cma_client_secret, cma_agent_client_secret). If the is on a different cluster from the , also set Gateway URL (cma_gateway_url).

The first run of the service starts. When the command finishes, the service is added to the cluster.

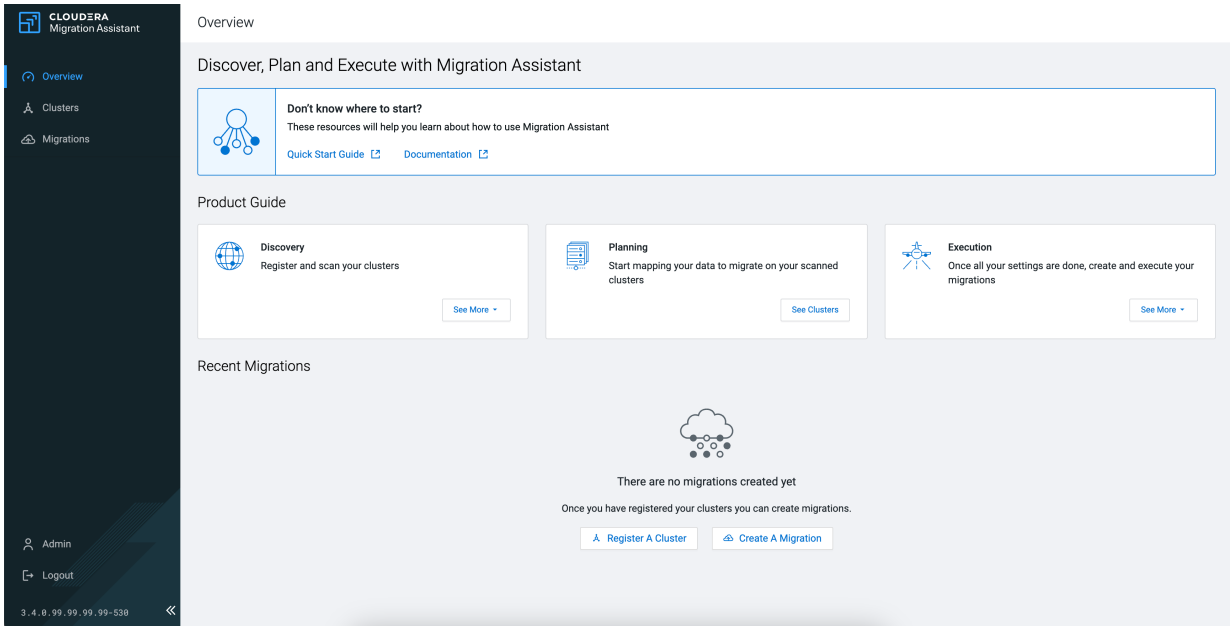
Figure 17: service started



20. Repeat steps 16–19 to add the other service type if needed (for example, add CMA_AGENT after CMA_MAST ER, or vice versa).

21. Go back to the cluster homepage, open the service page, and click the Server UI tab to open the UI.

Figure 18: landing page



22. Set the service dependency to none in the service configuration in Cloudera Manager because Cloudera Manager cannot discover the Master automatically across clusters.
23. Set the **Gateway URL** to cma_gateway_url in the service configuration.

```
curl -sk https://<master-host>:8093/api/agent/config
```

Procedure

Phase 2 — Configure Cloudera Replication Manager on the target cluster

Procedure

1. Enable the HDFS to Ozone flag on the Server host.

   For HDFS to Ozone migrations, requires the CMF_FF_API_H2O_REPLICATION feature flag to be enabled. Without this flag, the API endpoints for HDFS to Ozone replication are not available.

   Note: This flag is only required for HDFS to Ozone migrations. It is not required for HDFS-to-HDFS, Hive, or HBase replication.

   a) Append the following line to /etc/default/cloudera-scm-server on the Server host:

   ```
   export CMF_FF_API_H2O_REPLICATION=true
   ```

   b) Restart the Server.

   ```
   sudo systemctl restart cloudera-scm-server
   ```

   c) Wait for to become available again on port 7180 before proceeding with the setup.
2. Register the instance in the UI to allow to locate and communicate with the replication service.
   a) In the Web UI, go to Settings.
   b) Go to the App Access section.
   c) Configure your connection fields using one of the following validation methods:

- If has already detected a service on one of the registered clusters, select it from the Suggestions drop-down list to auto-fill the form.
- Manually enter your target replication parameters in the following fields:
  - App URL — The full URL of the REST API, including protocol, host, and port, for example, https://rm-host.example.com:9011.
  - App username — The username for authenticating with the API.
  - App password — The password for authenticating with the API.

d) Click Save and Push to Agents.

Procedure

Phase 3 — Register source and target clusters

Procedure

1. In on the source cluster, select Add Service and select CMA_AGENT.
2. Configure the required properties.

Table 2: configuration properties

Property	Description
cma_gateway_url	URL of the Gateway, for example, https://master-host:8093. **Required when the is on a different cluster than the .** The agent uses this URL to discover the Master host, gRPC port, and Auth Server URI automatically through the discovery endpoint (GET /api/agent/config). When connecting from a different cluster, set the service dependency to none in .
cm_username	administrator user name.
cm_password	administrator password.
cma_user_home	Home directory for the cma-agent user, for example, /var/lib/cma-agent.

3. Set cma_gateway_url when the is on a different cluster than the .
4. Start the service.
5. Wait for the agent to connect to the and finish the registration automatically.

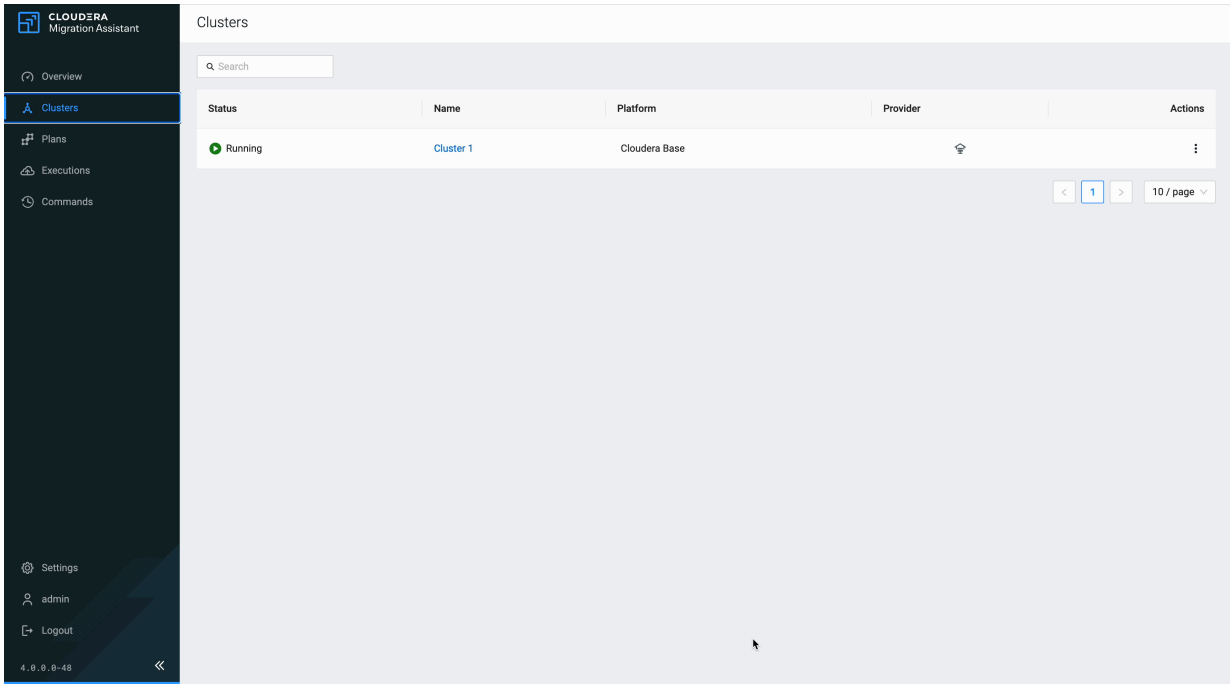
Procedure

Phase 4 — Scan clusters (include Ozone on the target)

Procedure

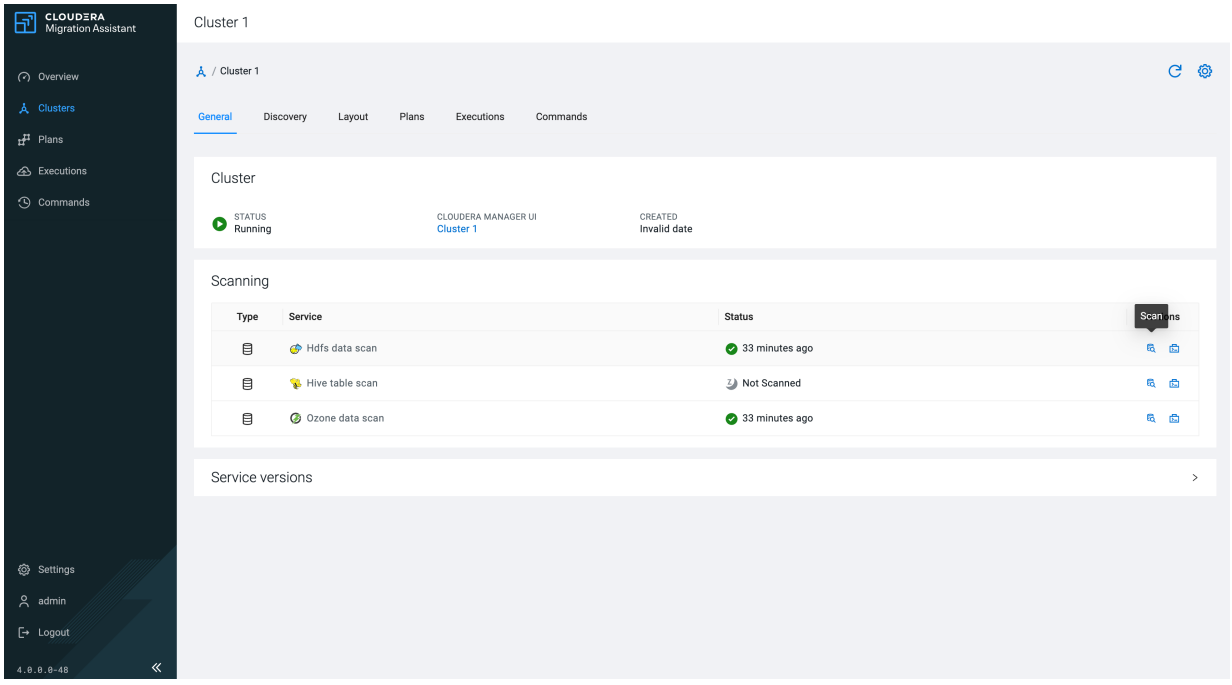
1. Navigate to the Clusters page and click the cluster you want to scan.

Figure 19: Clusters page



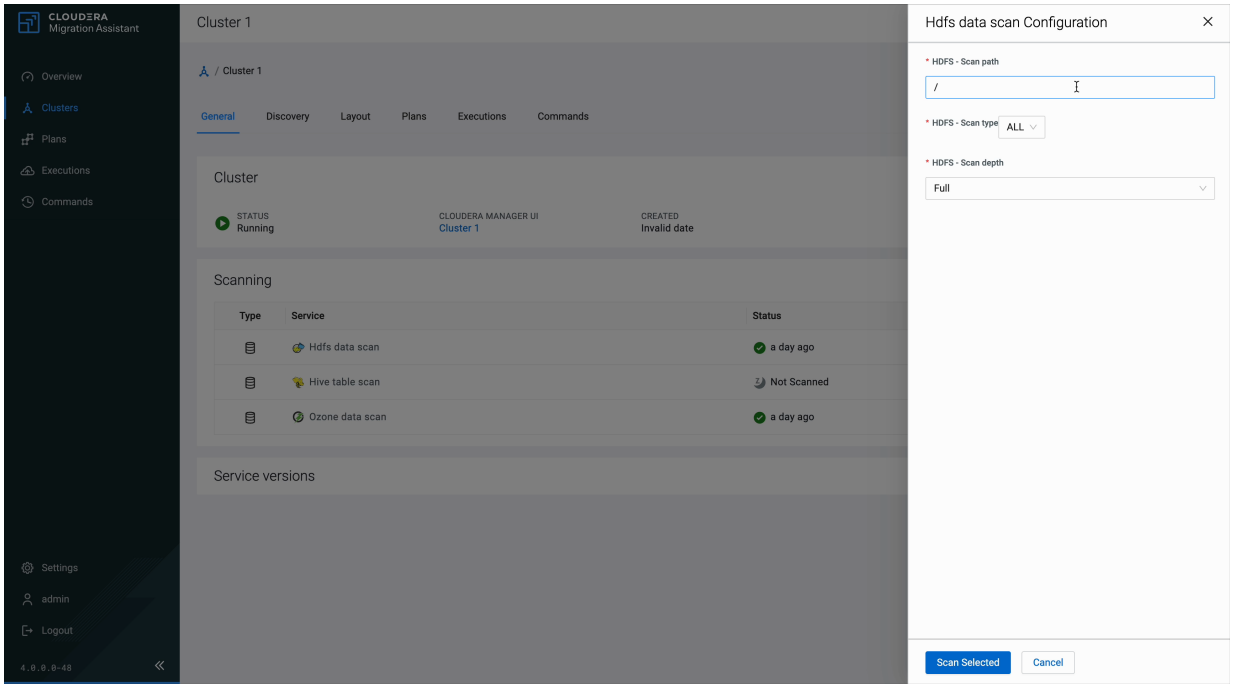
2. On the cluster detail page, find the Scanning section, which lists the available scan types for your cluster. Click the configuration icon on the right side of the scan type you want to trigger.

Figure 20: cluster scanning



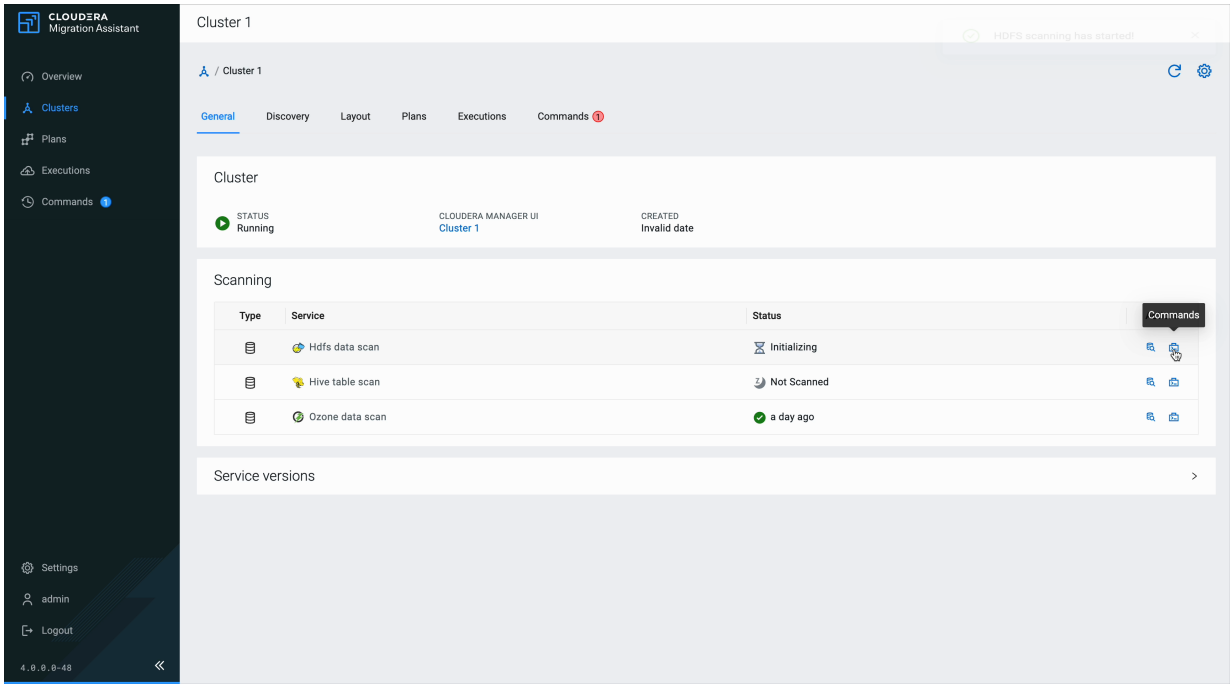
3. In the scan configuration drawer, configure the scan parameters (for example, scan path, scan type, scan depth) and click Scan Selected.

Figure 21: scan configuration



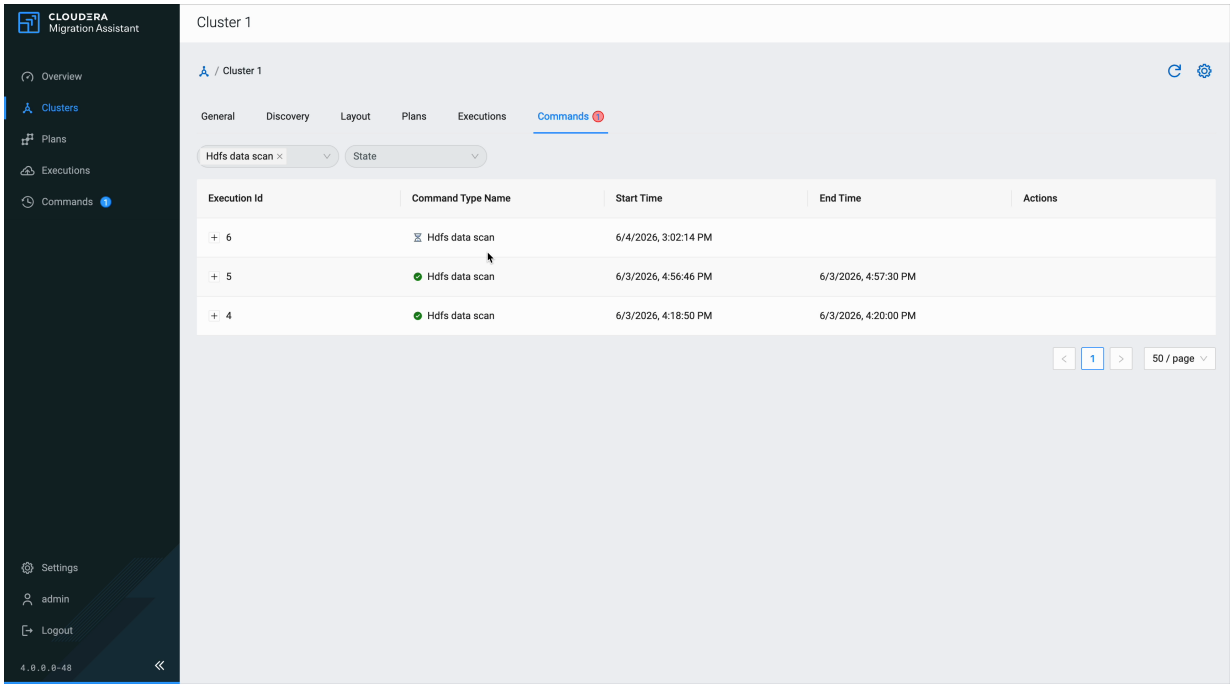
4. The scan starts and the status changes to Initializing. Track scan progress by clicking the Commands tab, which shows a badge with the number of active commands.

Figure 22: scan in progress



5. On the Commands tab, expand a command to view its Output Log, Error Log, and Configuration.

Figure 23: scan commands



Note: Scan results are also available in the data directory on the source cluster at <CMA_ROOT_DIR>/data/sources/ClusterName/Source_ID/cluster-scan/ on the host.

Procedure

Scan Ozone storage to discover existing volumes, buckets, and folders.

a) Click on the cluster you want to scan on the **Clusters** page.
b) Click Start Scanning to open **Scan Settings**.
c) Select Ozone scan.

Important:

The Ozone service must be available on the cluster before initiating the scan. checks service availability at scan time and reports an error if the Ozone service is not responding.

d) Click Scan selected.

You are redirected to the scanning progress page where you can monitor whether the scan completed successfully or encountered an error.

Procedure

Phase 5 — Plan and deploy the Ozone storage layout

Procedure

Known limitations

No input validation

does not validate layout definitions for volumes, buckets, or quotas.

No feasibility pre-check

does not verify whether the specified settings are feasible on the target cluster.

Limited migration testing

Ozone Layout Provisioning has not been tested as part of the end-to-end migration workflow.

Use this feature for evaluation purposes only. Behavior and APIs may change in future releases.

provides a **Storage Layout Editor** for designing and deploying Ozone storage structures.

Both the **Storage Layout Editor** and the **Ozone browser** use colors and text styles to indicate the status of volumes, buckets, and folders. The two views display different stages of the same lifecycle.

The **Storage Layout Editor** displays both unsaved (local) changes and changes that were saved to the database but not yet deployed to the cluster.

Table 3: Storage Layout Editor statuses

Status	Color	Style	Description
New or locally modified (unsaved)	Light red	Normal	Node was created or edited in the editor but not yet saved to the database.
Locally deleted (unsaved)	Light red	Strikethrough	Node was deleted in the editor but the deletion has not been saved yet.
Saved, pending CREATE	Red	Normal	Node was saved to the database and is queued for deployment to the cluster.
Saved, pending MODIFY	Default	Italic	Node was saved to the database and is queued for deployment to the cluster.
Saved, pending DELETE	Default	Strikethrough	Node was saved as a deletion and is queued for removal from the cluster.
Saved, no pending event	Default	Normal	Node is deployed and up to date.

The **Ozone browser** only displays nodes that have been saved to the database. Unsaved editor changes are not visible here. If you have unsaved changes in the editor, the browser shows the last saved state of those nodes until you save.

Table 4: Ozone browser statuses

Status	Color	Style	Description
Pending CREATE (saved, not yet deployed)	Red	Normal	Node was created and saved, awaiting deployment to the cluster.
Pending MODIFY (saved, not yet deployed)	Default	Italic	Node was modified and saved, awaiting deployment to the cluster.
Pending DELETE (saved, not yet deployed)	Default	Strikethrough	Node is queued for deletion. All children of a pending-delete node are also displayed with strikethrough.
No pending event	Default	Normal	Node is deployed and up to date.

Note: The strikethrough style in the browser propagates to all children of a node marked for deletion, so you can see the full impact of a pending delete operation before it is deployed.

1. Plan and deploy an Ozone storage layout using the **Storage Layout Editor**.
   a) Go to the Layout tab in the cluster view to access the **Storage Layout Editor**.

      Design the layout structure in the Plan tab and apply the layout to the cluster in the Deploy tab.

The Plan tab displays the following panels.

- Left panel (**Available Templates**) — Contains reusable volume structures you can drag into the editor.
- Right panel (**Editor**) — Is the working area where you build the target Ozone structure.

2. Add volumes and buckets manually.
   a) Select the cluster on the **Clusters** page.
   b) Go to the Layout tab.
   c) In the Plan tab, click the Add Volume button at the top of the editor panel.
   d) Provide the volume properties in the drawer that opens, then click Create.
   e) To add buckets to a volume, select the volume node and click Add Bucket.
   f) To add folders to a bucket, select the bucket node and click Add Folder.
   g) Click Save Plan in the toolbar to save all pending changes to the database.

The **Available Templates** panel on the left displays reusable volume structures. You can use the following template types:

- System templates — Built-in templates provided by . These cannot be deleted.
- User templates — Templates you create from volumes in the editor. These can be deleted.

3. Search and filter templates.

   Use the Search field at the top of the **Available Templates** panel to filter templates by name. Use the Type drop-down list to filter by template type.

4. Drag a template into the editor.
   a) In the **Available Templates** panel, click the template name to expand it and see its volumes, buckets, and folders.
   b) Drag the node you want (volume, bucket, or folder) by its drag handle into the editor panel on the right.

      You can drag any node from a template directly into the editor to add it to the layout. Volumes, buckets, and folders can all be dragged individually using the drag handle (the grip icon on the right side of each node). When you drag a node, the full subtree (all its children) moves with it.

5. Save a volume as a template.

   You can save any volume in the editor as a template to reuse it later. Saving a volume as a template stores the entire volume structure, including its buckets and folders, in the **Available Templates** panel.

   a) In the editor, hover over the volume node you want to save.
   b) Click Add to Template that is the download icon that is displayed on hover.
   c) Provide a name for the template.
   d) Click Save.

      The template is displayed in the **Available Templates** panel and is stored in the database.

6. Delete a user template.
   a) In the **Available Templates** panel, hover over the template name.
   b) Click the  icon next to the template name.
   c) Click Delete to confirm the deletion.

7. Import a layout from YAML or JSON.

   You can define a complete Ozone storage structure in a YAML or JSON file and import it into the editor.

   This example shows a YAML file defining volumes, buckets, and folders in a hierarchical structure.

```
volumes:
          - name: "my-volume"
            namespace-quota: 100
            space-quota: "10TB"
            user: "admin"
            buckets:
```

```
  - name: "my-bucket"
    create-snapshot: true
    namespace-quota: 1000
    space-quota: "1TB"
    user: "admin"
    layout: "FILE_SYSTEM_OPTIMIZED"
    replication-type: "RATIS"
    replication: "THREE"
    bucketkey: "my-encryption-key"
    enforcegdpr: false
    folders:
      - name: "data"
        folders:
          - name: "raw"
          - name: "processed"
      - name: "logs"
```

a) In the Plan tab, click Import Structure in the toolbar.
b) Upload the YAML or JSON file or paste the YAML or JSON content.
c) Click Upload.

validates the structure and loads the volumes, buckets, and folders into the editor. Validation errors are displayed if the YAML or JSON is malformed or contains invalid values.

d) Click Save Plan to save the imported structure to the database.

8. Deploy the layout to the cluster.

a) Click the Deploy tab in the **Layout** page.

The **Deployment Queue** table displays all saved volumes, buckets, and folders that have pending changes. The CREATE, MODIFY, or DELETE events are not yet applied to the cluster.

b) Review the pending changes in the **Volumes**, **Buckets**, and **Folders** tables.

Warning: Deployment is irreversible. After clicking Deploy to Ozone, rollback is not possible. Return to the Plan tab to make modifications before deploying.

c) Click Deploy to Ozone.

sends the deploy command to the on the cluster. Monitor progress in the Commands tab on the cluster view. The agent processes the changes in the following order:

a. Create, update, or delete volumes.
b. Create, update, or delete buckets.
c. Create, update, or delete folders.

Results

The target cluster has a saved and deployed Ozone storage layout. You can create collections and migration plans that reference these Ozone locations.

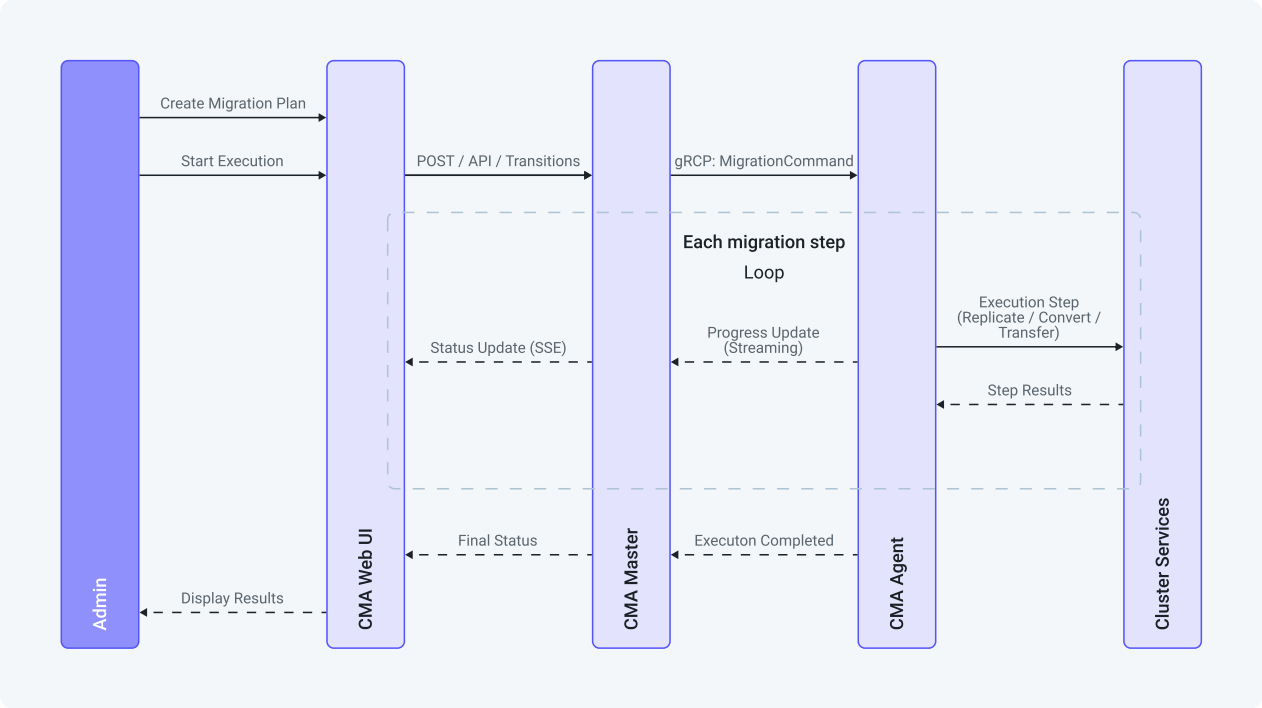
Migrating HDFS data to Ozone end to end

Deploy Cloudera Migration Assistant, prepare Ozone layout, and run an HDFS to Ozone migration.

About this task

This end-to-end workflow combines deployment, platform preparation, discovery, and workload steps. Detailed reference procedures are reused from the Administrator and Migration guides.

Figure 24: Migration flow



Before you begin

Verify supported platforms and migration paths in *Supported platforms and migration paths* before you start.

Procedure

Phase 1 — Deploy the CMA Master and CMA Agents

Procedure

1. Download both CSD files to the /opt/cloudera/csd/ directory on the Server host.

   ```
   wget -P /opt/cloudera/csd/ https://archive.cloudera.com/cma/VERSION/csd/
   CMA_MASTER-VERSION-BUILD.jar
   ```

   ```
   wget -P /opt/cloudera/csd/ https://archive.cloudera.com/cma/VERSION/csd/
   CMA_AGENT-VERSION-BUILD.jar
   ```

   automatically detects the CSD files.
2. Change the ownership of the CSD files.

   ```
   chown cloudera-scm:cloudera-scm /opt/cloudera/csd/CMA_MA
   STER-VERSION-BUILD.jar /opt/cloudera/csd/CMA_AGENT-VERSION-BUILD.jar
   ```
3. Restart for the changes to take effect.

   ```
   systemctl restart cloudera-scm-server
   ```
4. Log into .
5. Restart the Cloudera Management Service.
6. Go to Hosts Parcels .
7. Click Parcel Repositories & Network Settings.

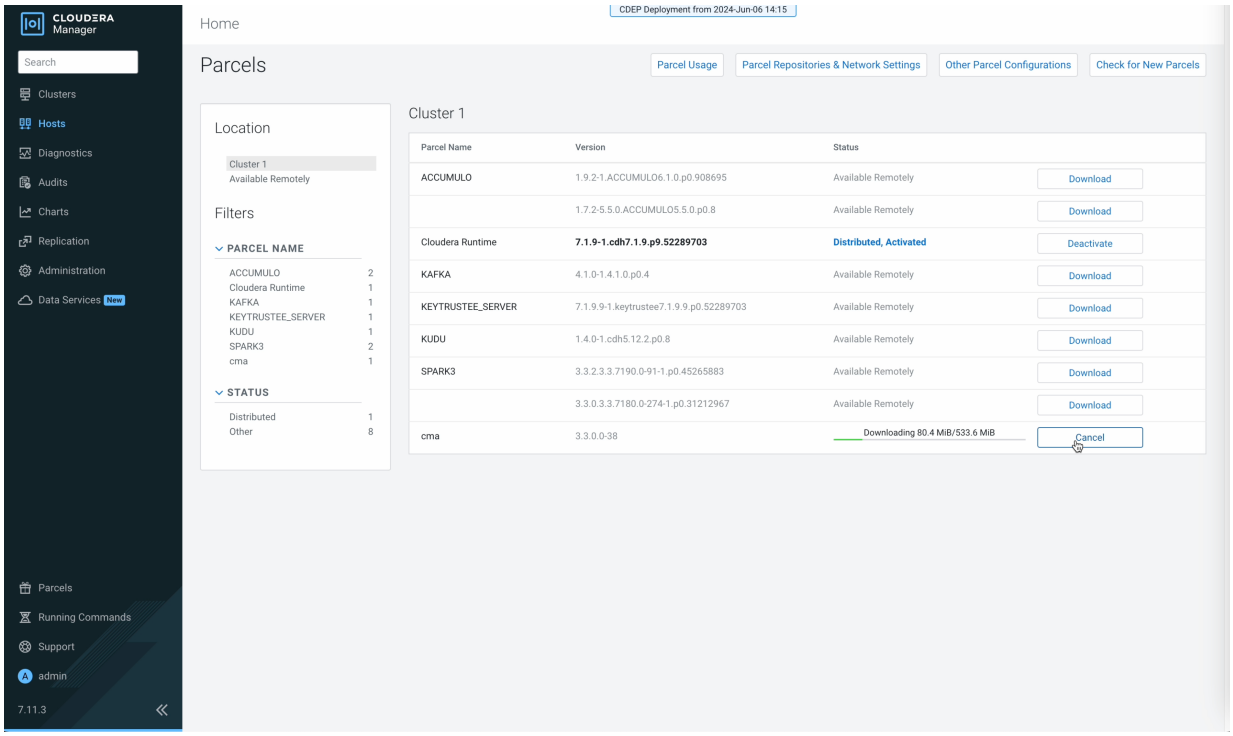
8. Add the Remote Parcel Repository URL for .

```
https://archive.cloudera.com/cma/VERSION/parcels/
```

Note: Make sure that the Remote Parcel Repository URL uses an HTTPS link. To install a different version of the parcel, change the URL as needed.

9. Click Save & Verify Configuration to commit the change.
10. Click Close.
    You are redirected to the Parcels page.
11. Search for , and click Download to download the parcel to the local repository.
12. After download completes, click Distribute to distribute the parcel to all clusters.

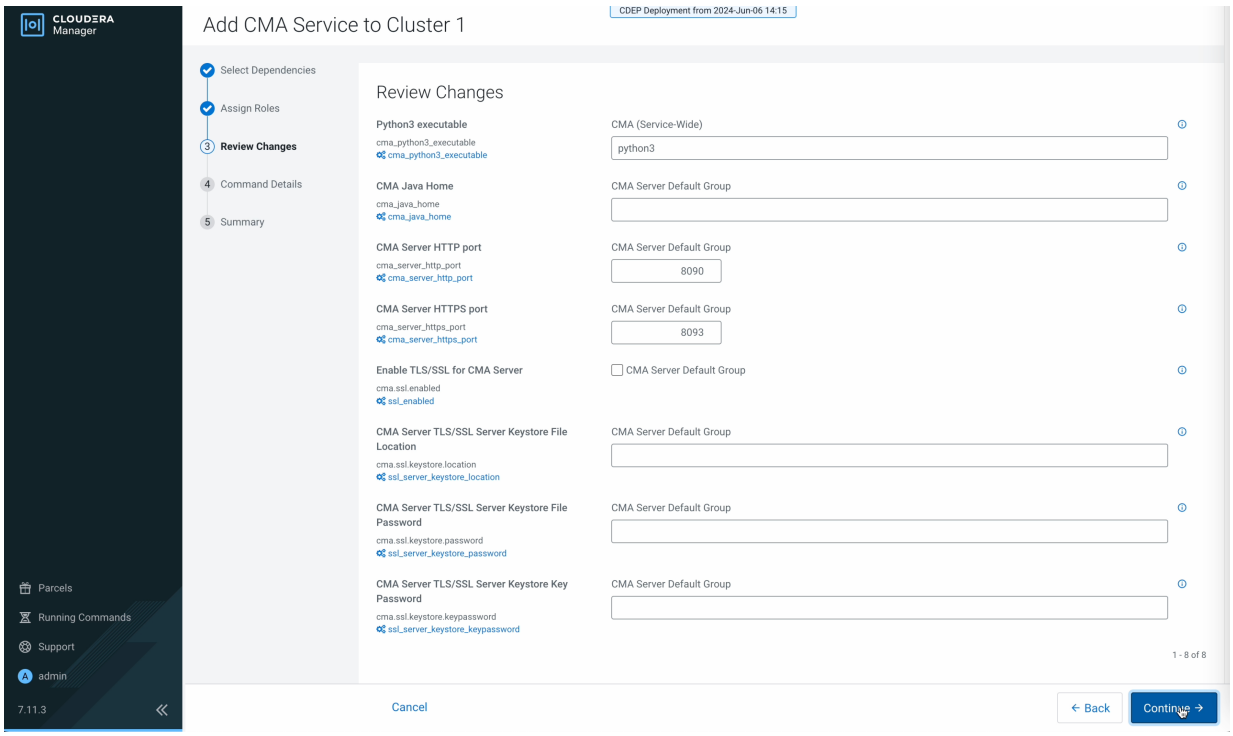
Figure 25: Distribute the parcel



13. Click Activate to activate the parcel.
14. Click OK when confirmation is required.
15. Click Clusters in the left navigation pane.
16. Select Add Service from the drop-down menu to the right of your cluster.
17. From the list, select the service type to add, then click Continue. The single parcel provides two service types: CMA_MASTER (central control node) and CMA_AGENT (cluster-side executor).

    The **Add Service wizard** opens.

Figure 26: Add service

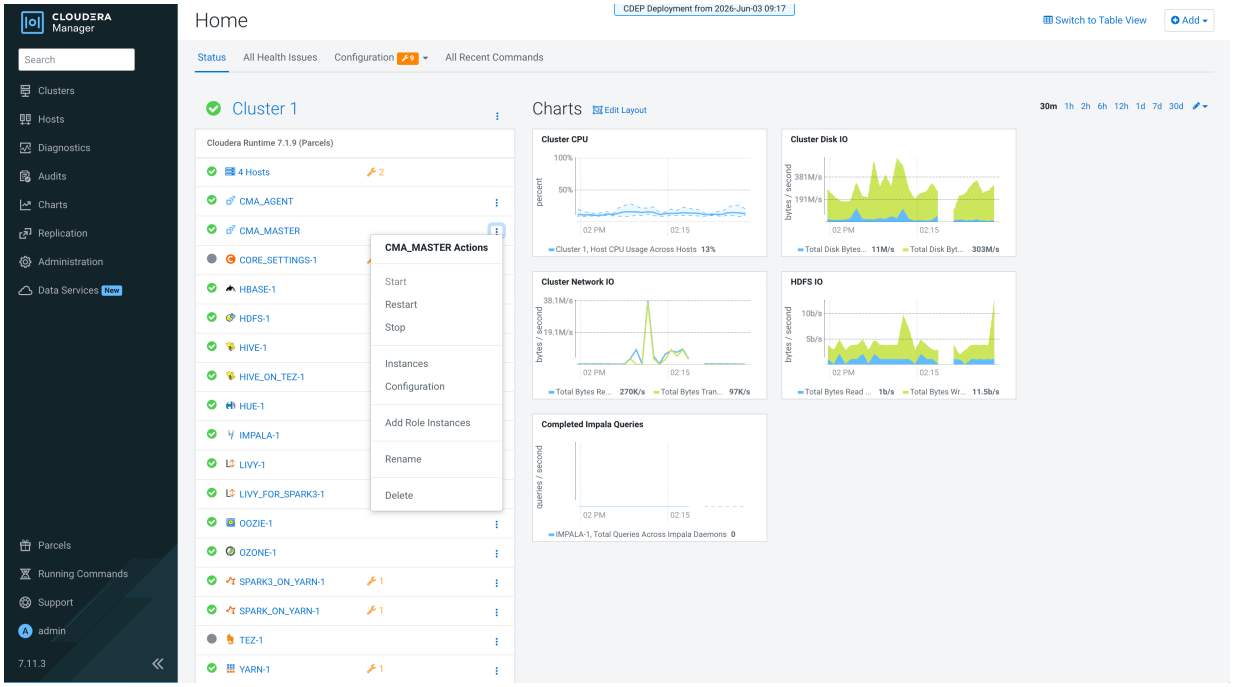


18. Assign the service roles to the hosts where Python 3.11 and JDK 17+ are installed, and click Continue.

19. Review service configurations and click Continue. At minimum, set OAuth2 client credentials (cma_client_secret, cma_agent_client_secret). If the is on a different cluster from the , also set Gateway URL (cma_gateway_url).

The first run of the service starts. When the command finishes, the service is added to the cluster.

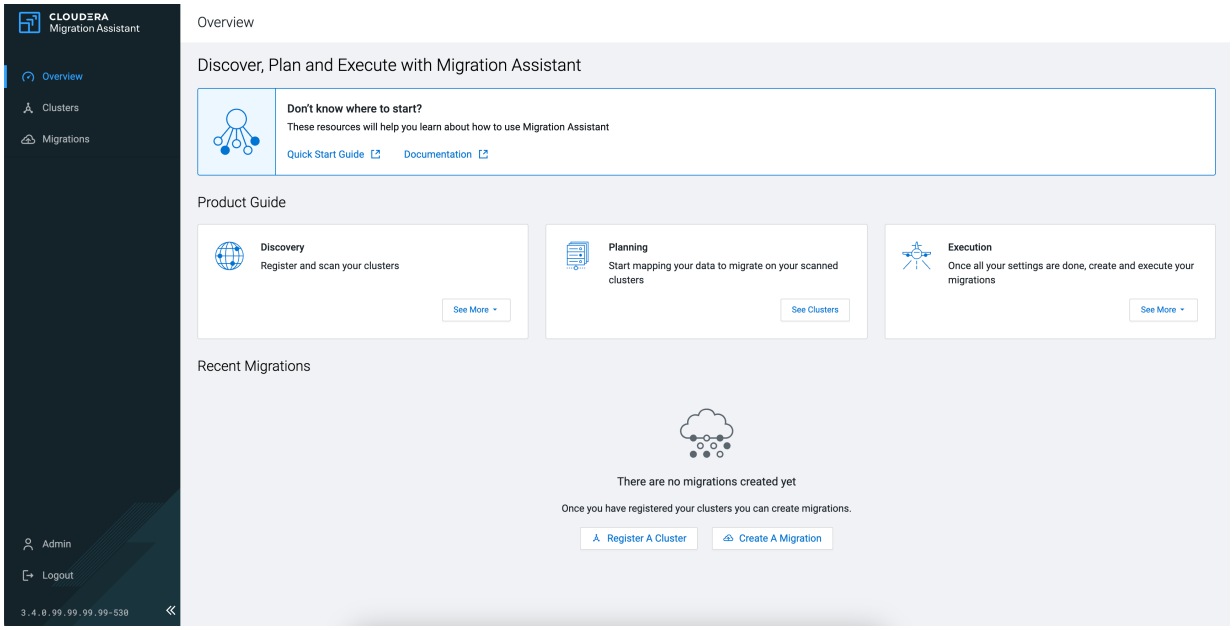
Figure 27: service started



20. Repeat steps 16–19 to add the other service type if needed (for example, add CMA_AGENT after CMA_MAST ER, or vice versa).

21. Go back to the cluster homepage, open the service page, and click the Server UI tab to open the UI.

Figure 28: landing page



22. Set the service dependency to none in the service configuration in Cloudera Manager because Cloudera Manager cannot discover the Master automatically across clusters.
23. Set the **Gateway URL** to cma_gateway_url in the service configuration.

```
curl -sk https://<master-host>:8093/api/agent/config
```

Procedure

Phase 2 — Configure Cloudera Replication Manager on the target cluster

Procedure

1. Enable the HDFS to Ozone flag on the Server host.

   For HDFS to Ozone migrations, requires the CMF_FF_API_H2O_REPLICATION feature flag to be enabled. Without this flag, the API endpoints for HDFS to Ozone replication are not available.

   Note: This flag is only required for HDFS to Ozone migrations. It is not required for HDFS-to-HDFS, Hive, or HBase replication.

   a) Append the following line to /etc/default/cloudera-scm-server on the Server host:

   ```
   export CMF_FF_API_H2O_REPLICATION=true
   ```

   b) Restart the Server.

   ```
   sudo systemctl restart cloudera-scm-server
   ```

   c) Wait for to become available again on port 7180 before proceeding with the setup.
2. Register the instance in the UI to allow to locate and communicate with the replication service.
   a) In the Web UI, go to Settings.
   b) Go to the App Access section.
   c) Configure your connection fields using one of the following validation methods:

- If has already detected a service on one of the registered clusters, select it from the Suggestions drop-down list to auto-fill the form.
- Manually enter your target replication parameters in the following fields:
  - App URL — The full URL of the REST API, including protocol, host, and port, for example, https://rm-host.example.com:9011.
  - App username — The username for authenticating with the API.
  - App password — The password for authenticating with the API.

d) Click Save and Push to Agents.

Procedure

Prepare Ozone storage on the target (before Phase 6)

Procedure

Known limitations

No input validation

does not validate layout definitions for volumes, buckets, or quotas.

No feasibility pre-check

does not verify whether the specified settings are feasible on the target cluster.

Limited migration testing

Ozone Layout Provisioning has not been tested as part of the end-to-end migration workflow.

Use this feature for evaluation purposes only. Behavior and APIs may change in future releases.

provides a **Storage Layout Editor** for designing and deploying Ozone storage structures.

Both the **Storage Layout Editor** and the **Ozone browser** use colors and text styles to indicate the status of volumes, buckets, and folders. The two views display different stages of the same lifecycle.

The **Storage Layout Editor** displays both unsaved (local) changes and changes that were saved to the database but not yet deployed to the cluster.

Table 5: Storage Layout Editor statuses

Status	Color	Style	Description
New or locally modified (unsaved)	Light red	Normal	Node was created or edited in the editor but not yet saved to the database.
Locally deleted (unsaved)	Light red	Strikethrough	Node was deleted in the editor but the deletion has not been saved yet.
Saved, pending CREATE	Red	Normal	Node was saved to the database and is queued for deployment to the cluster.
Saved, pending MODIFY	Default	Italic	Node was saved to the database and is queued for deployment to the cluster.
Saved, pending DELETE	Default	Strikethrough	Node was saved as a deletion and is queued for removal from the cluster.
Saved, no pending event	Default	Normal	Node is deployed and up to date.

The **Ozone browser** only displays nodes that have been saved to the database. Unsaved editor changes are not visible here. If you have unsaved changes in the editor, the browser shows the last saved state of those nodes until you save.

Table 6: Ozone browser statuses

Status	Color	Style	Description
Pending CREATE (saved, not yet deployed)	Red	Normal	Node was created and saved, awaiting deployment to the cluster.
Pending MODIFY (saved, not yet deployed)	Default	Italic	Node was modified and saved, awaiting deployment to the cluster.
Pending DELETE (saved, not yet deployed)	Default	Strikethrough	Node is queued for deletion. All children of a pending-delete node are also displayed with strikethrough.
No pending event	Default	Normal	Node is deployed and up to date.

Note: The strikethrough style in the browser propagates to all children of a node marked for deletion, so you can see the full impact of a pending delete operation before it is deployed.

1. Plan and deploy an Ozone storage layout using the **Storage Layout Editor**.
   a) Go to the Layout tab in the cluster view to access the **Storage Layout Editor**.

      Design the layout structure in the Plan tab and apply the layout to the cluster in the Deploy tab.

      The Plan tab displays the following panels.

      - Left panel (**Available Templates**) — Contains reusable volume structures you can drag into the editor.
      - Right panel (**Editor**) — Is the working area where you build the target Ozone structure.
2. Add volumes and buckets manually.
   a) Select the cluster on the **Clusters** page.
   b) Go to the Layout tab.
   c) In the Plan tab, click the Add Volume button at the top of the editor panel.
   d) Provide the volume properties in the drawer that opens, then click Create.
   e) To add buckets to a volume, select the volume node and click Add Bucket.
   f) To add folders to a bucket, select the bucket node and click Add Folder.
   g) Click Save Plan in the toolbar to save all pending changes to the database.

The **Available Templates** panel on the left displays reusable volume structures. You can use the following template types:

- System templates — Built-in templates provided by . These cannot be deleted.
- User templates — Templates you create from volumes in the editor. These can be deleted.

3. Search and filter templates.

   Use the Search field at the top of the **Available Templates** panel to filter templates by name. Use the Type drop-down list to filter by template type.
4. Drag a template into the editor.
   a) In the **Available Templates** panel, click the template name to expand it and see its volumes, buckets, and folders.
   b) Drag the node you want (volume, bucket, or folder) by its drag handle into the editor panel on the right.

      You can drag any node from a template directly into the editor to add it to the layout. Volumes, buckets, and folders can all be dragged individually using the drag handle (the grip icon on the right side of each node). When you drag a node, the full subtree (all its children) moves with it.

5. Save a volume as a template.

   You can save any volume in the editor as a template to reuse it later. Saving a volume as a template stores the entire volume structure, including its buckets and folders, in the **Available Templates** panel.

   a) In the editor, hover over the volume node you want to save.
   b) Click Add to Template that is the download icon that is displayed on hover.
   c) Provide a name for the template.
   d) Click Save.

      The template is displayed in the **Available Templates** panel and is stored in the database.

6. Delete a user template.

   a) In the **Available Templates** panel, hover over the template name.
   b) Click the  icon next to the template name.
   c) Click Delete to confirm the deletion.

7. Import a layout from YAML or JSON.

   You can define a complete Ozone storage structure in a YAML or JSON file and import it into the editor.

   This example shows a YAML file defining volumes, buckets, and folders in a hierarchical structure.

```
volumes:
          - name: "my-volume"
            namespace-quota: 100
            space-quota: "10TB"
            user: "admin"
            buckets:
              - name: "my-bucket"
                create-snapshot: true
                namespace-quota: 1000
                space-quota: "1TB"
                user: "admin"
                layout: "FILE_SYSTEM_OPTIMIZED"
                replication-type: "RATIS"
                replication: "THREE"
                bucketkey: "my-encryption-key"
                enforcegdpr: false
                folders:
                  - name: "data"
                    folders:
                      - name: "raw"
                      - name: "processed"
                  - name: "logs"
```

   a) In the Plan tab, click Import Structure in the toolbar.
   b) Upload the YAML or JSON file or paste the YAML or JSON content.
   c) Click Upload.

      validates the structure and loads the volumes, buckets, and folders into the editor. Validation errors are displayed if the YAML or JSON is malformed or contains invalid values.
   d) Click Save Plan to save the imported structure to the database.

8. Deploy the layout to the cluster.

   a) Click the Deploy tab in the **Layout** page.

      The **Deployment Queue** table displays all saved volumes, buckets, and folders that have pending changes. The CREATE, MODIFY, or DELETE events are not yet applied to the cluster.
   b) Review the pending changes in the **Volumes**, **Buckets**, and **Folders** tables.

Warning: Deployment is irreversible. After clicking Deploy to Ozone, rollback is not possible. Return to the Plan tab to make modifications before deploying.

c) Click Deploy to Ozone.

sends the deploy command to the on the cluster. Monitor progress in the Commands tab on the cluster view. The agent processes the changes in the following order:

a. Create, update, or delete volumes.
b. Create, update, or delete buckets.
c. Create, update, or delete folders.

Procedure

Phase 3 — Register source and target clusters

Procedure

1. In on the source cluster, select Add Service and select CMA_AGENT.
2. Configure the required properties.

Table 7: configuration properties

Property	Description
cma_gateway_url	URL of the Gateway, for example, https://master-host:8093. **Required when the is on a different cluster than the .** The agent uses this URL to discover the Master host, gRPC port, and Auth Server URI automatically through the discovery endpoint (GET /api/agent/config). When connecting from a different cluster, set the service dependency to none in .
cm_username	administrator user name.
cm_password	administrator password.
cma_user_home	Home directory for the cma-agent user, for example, /var/lib/cma-agent.

3. Set cma_gateway_url when the is on a different cluster than the .
4. Start the service.
5. Wait for the agent to connect to the and finish the registration automatically.

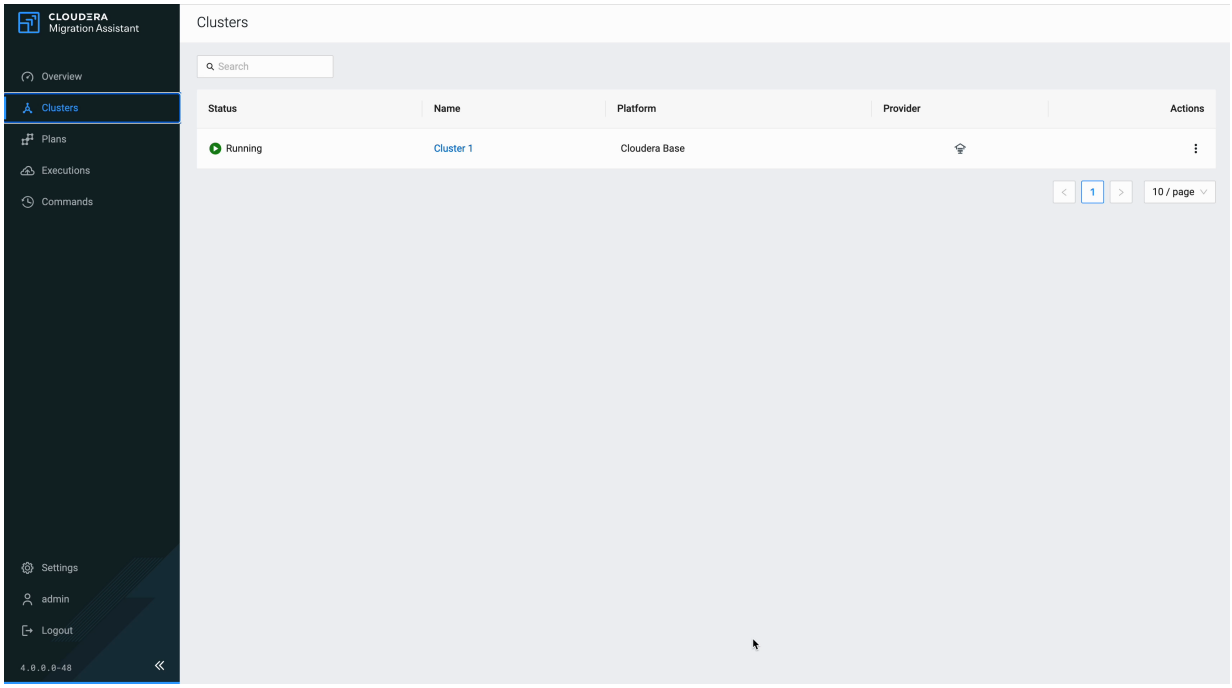
Procedure

Phase 4 — Scan clusters

Procedure

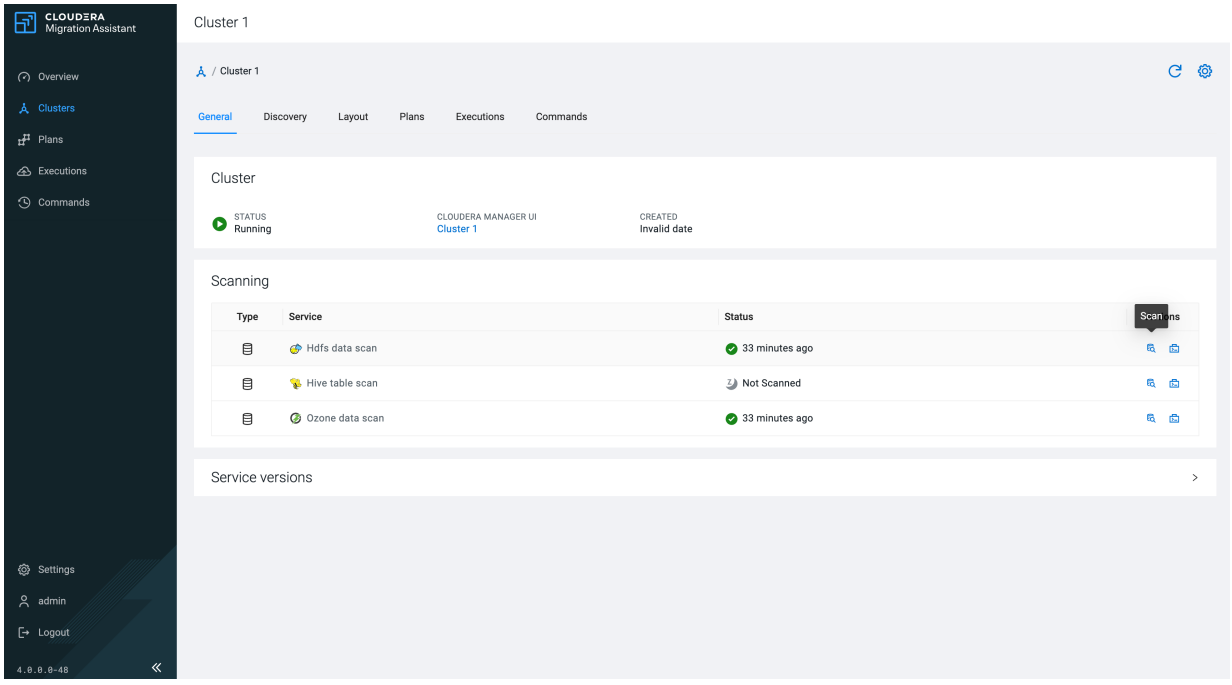
1. Navigate to the Clusters page and click the cluster you want to scan.

Figure 29: Clusters page



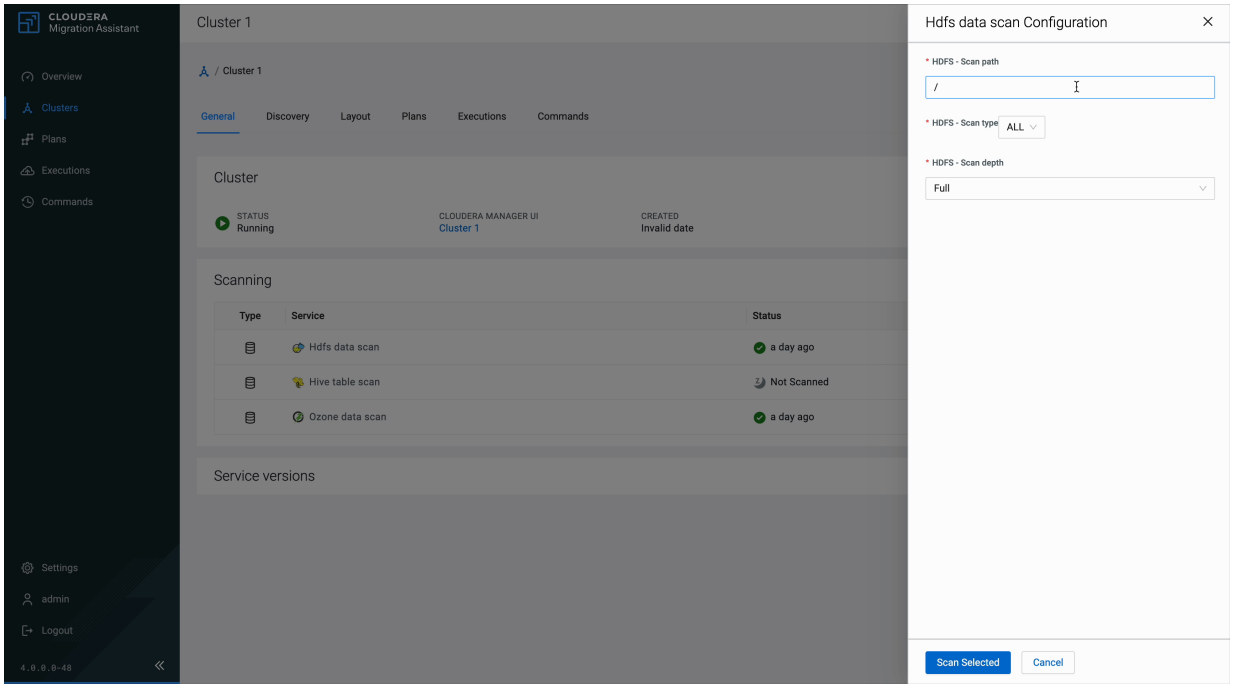
2. On the cluster detail page, find the Scanning section, which lists the available scan types for your cluster. Click the configuration icon on the right side of the scan type you want to trigger.

Figure 30: cluster scanning



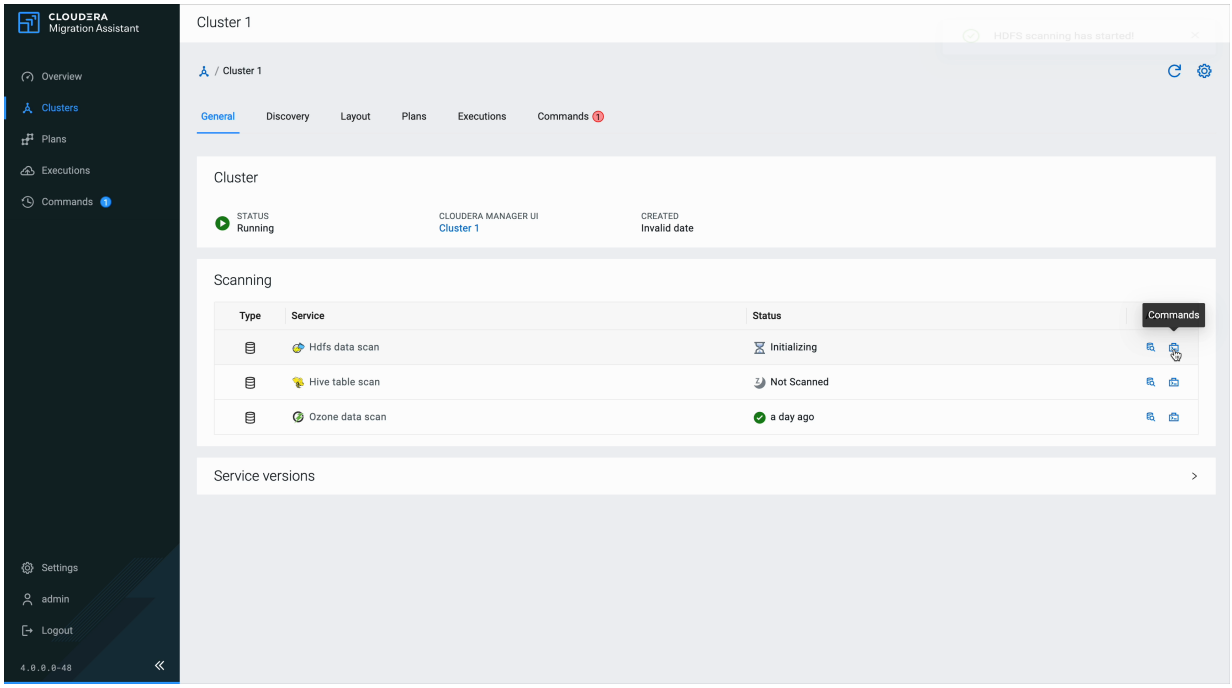
3. In the scan configuration drawer, configure the scan parameters (for example, scan path, scan type, scan depth) and click Scan Selected.

Figure 31: scan configuration



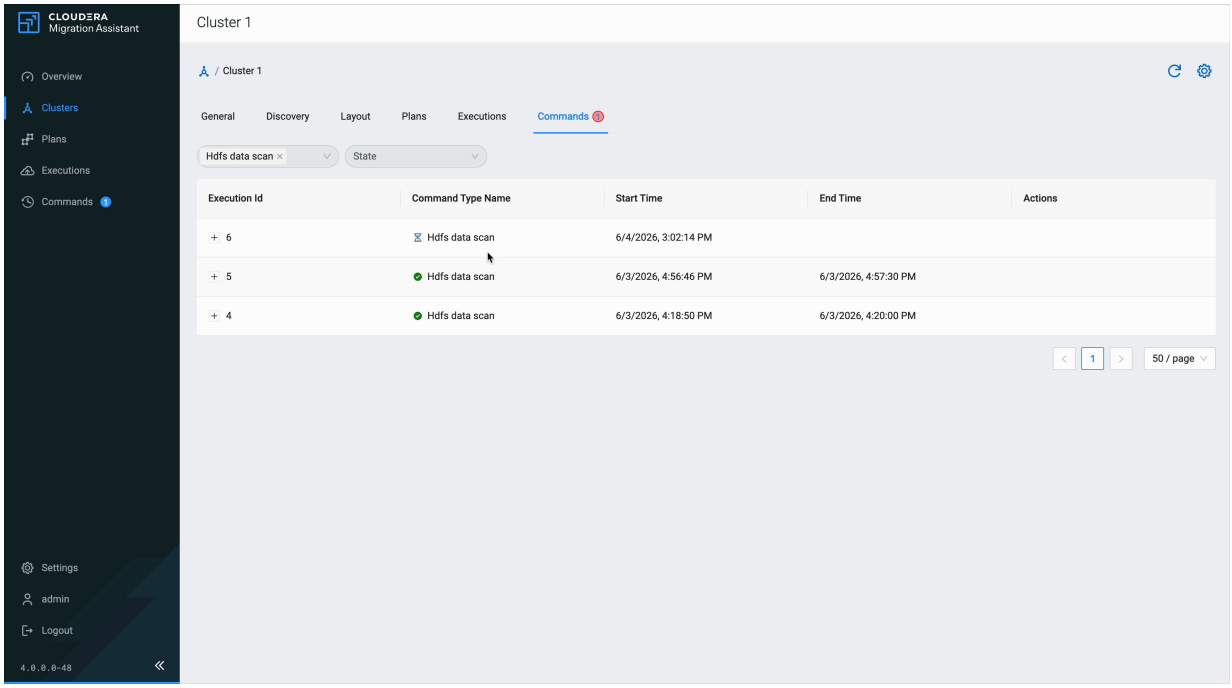
4. The scan starts and the status changes to Initializing. Track scan progress by clicking the Commands tab, which shows a badge with the number of active commands.

Figure 32: scan in progress



5. On the Commands tab, expand a command to view its Output Log, Error Log, and Configuration.

Figure 33: scan commands



Note: Scan results are also available in the data directory on the source cluster at <CMA_ROOT_DIR>/data/sources/ClusterName/Source_ID/cluster-scan/ on the host.

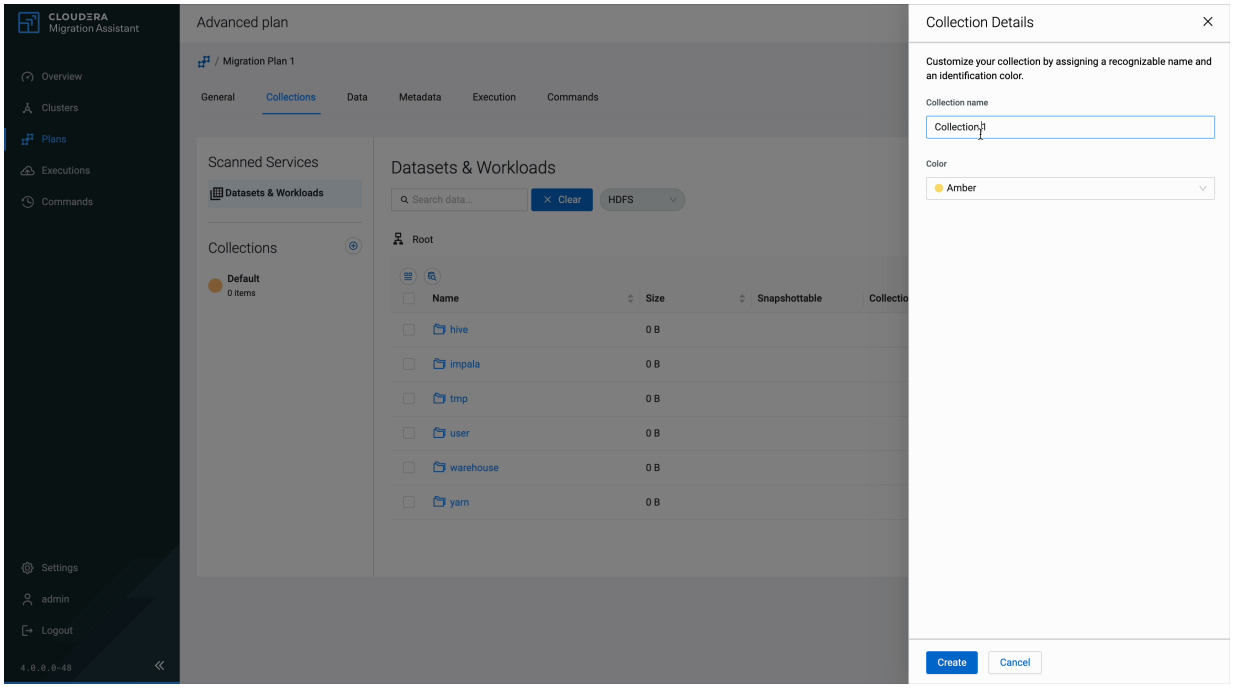
Procedure

Phase 5 — Create collections (optional)

Procedure

1. Click the + icon next to Collections in the sidebar.
2. In the **Collection Details** drawer, enter a Collection name and select a Color for identification.

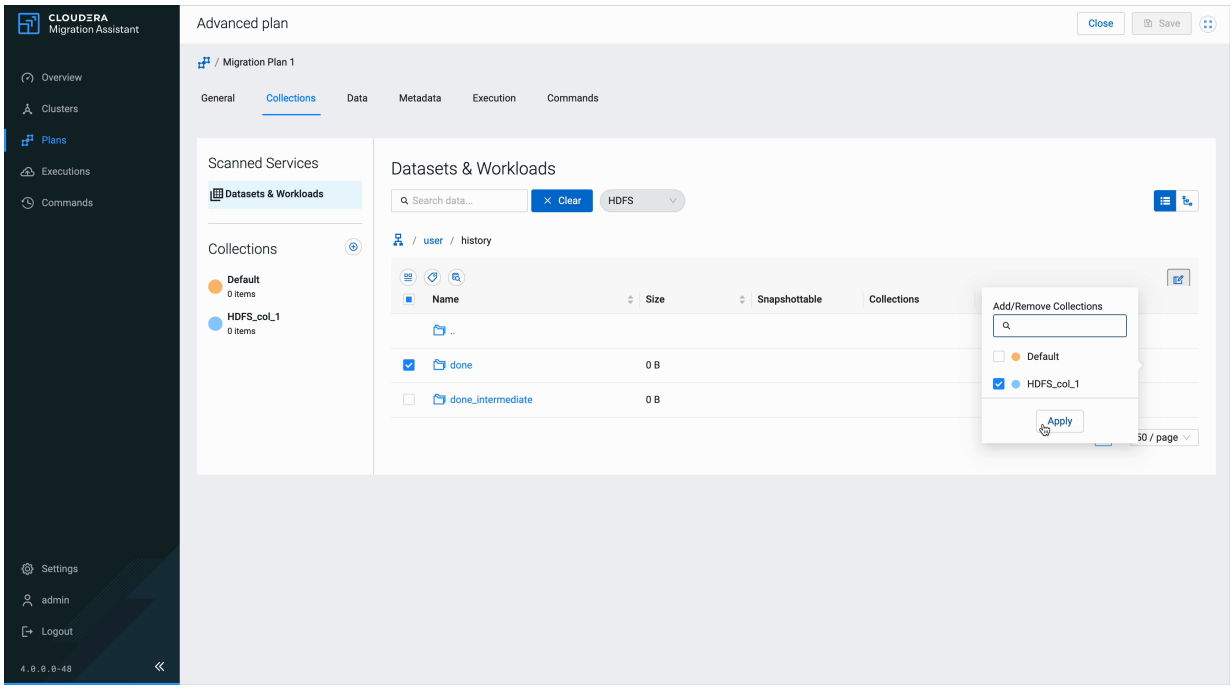
Figure 34: create collection



3. Click Create.

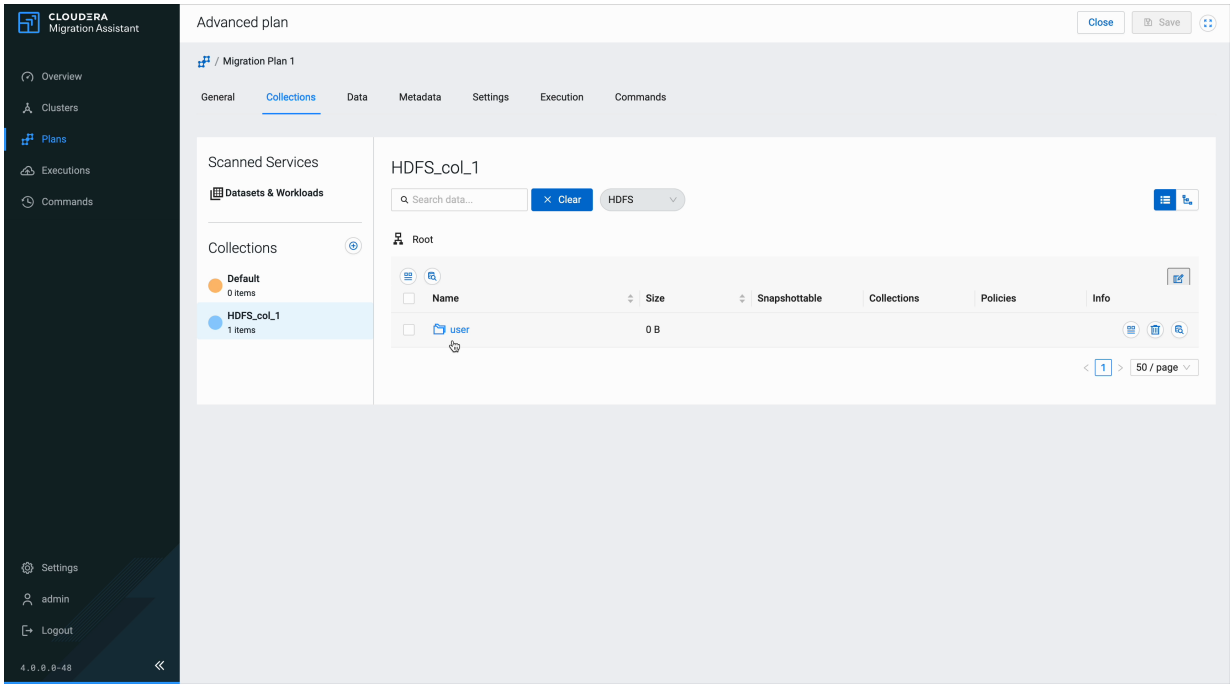
   The new collection appears in the sidebar with 0 items.
4. Browse the data tree to find the items you want to migrate. Click folder names to navigate deeper into the hierarchy.
5. Select items using the checkboxes next to each item.
6. Click the collection icon that appears on the selected item row.
7. In the **Add/Remove Collections** popup, check the collections you want to assign the item to.
8. Click Apply.

Figure 35: assign to collection



9. Click a collection name in the sidebar to view its contents. The data browser switches to show only the items assigned to that collection.

Figure 36: view collection



Procedure

Phase 6 — Run the migration workload

Procedure

Step 1 — Scan the source cluster (HDFS)

1. Open Clusters, then select the source cluster.
2. Click Start Scanning.
3. Select Hdfs data scan.
4. Constrain the scan path, for example, /tmp or /user, to shorten runs on massive namespaces.
5. Click Scan selected and watch the progress drawer until completion.

   After success, Discovery lists HDFS trees with sizing and ownership snapshots.

Step 2 — Scan the target cluster (Ozone)

6. Confirm Apache Ozone is healthy in before scanning.
7. From **Clusters**, open the target cluster.
8. Click Start Scanning.
9. Select Ozone data scan.
10. Click Scan selected and wait until the status turns green.

    Ozone hierarchies populate inside Discovery afterward.

Step 3 — Create a migration plan

11. Go to Plans.
12. Click Create Plan.
13. Set Name, select the source HDFS cluster, and the target Ozone cluster.
14. Complete the wizard (Create) and land on the plan detail surface.

Step 4 — Map HDFS directories to Ozone paths on the Data tab

The Data tab displays source HDFS on the left and target storage on the right.

15. Go to Plan Data .

16. Switch the target browser mode from HDFS to Ozone using the service picker.

17. Go to your target volume, bucket, and folder staging area.

18. On the source side, go to the parent directory that contains data you want to migrate.

19. Hover the directory and click Copy. persists a structural mapping.

20. Switch to the **Review** view any time to view a list of your current structure mappings.

21. When Ozone misses a folder UI-side, create it before copying structures.

a) In the target browser, go to the parent bucket or folder.
b) Click Create Folder.
c) Provide the folder name and confirm creation.

updates its metadata immediately and executions push the filesystem change during processing.

22. Repeat the mapping workflow for additional directories whenever you broaden scope.

You can map each directory to separate buckets or combine them into shared buckets. The layout must meet the requirements for your tenancy model.

Step 5 — Create and run an execution

23. Open Execution.

24. Select Create Execution.

25. Complete the steps in the configuration wizard (General, Scope, Review), and then click Finish.

26. Open the execution dashboard and trigger Run All.

During execution the following stages occur:

- **Policy creation** – registers policies mapping HDFS prefixes to destination Ozone keys.
- **Data replication** – invokes DistCp-style transfers governed by settings.
- **Verification** – tracks step health until completion (COMPLETED or FAILED statuses per step).

Step 6 — Verify the migration

27. Confirm that the execution steps display COMPLETED.

28. Inspect Ozone hierarchies inside Discovery.

29. Validate migrated objects using the Ozone CLI.

```
ozone fs -ls ofs://<service-id>/<volume>/<bucket>/<path>
```

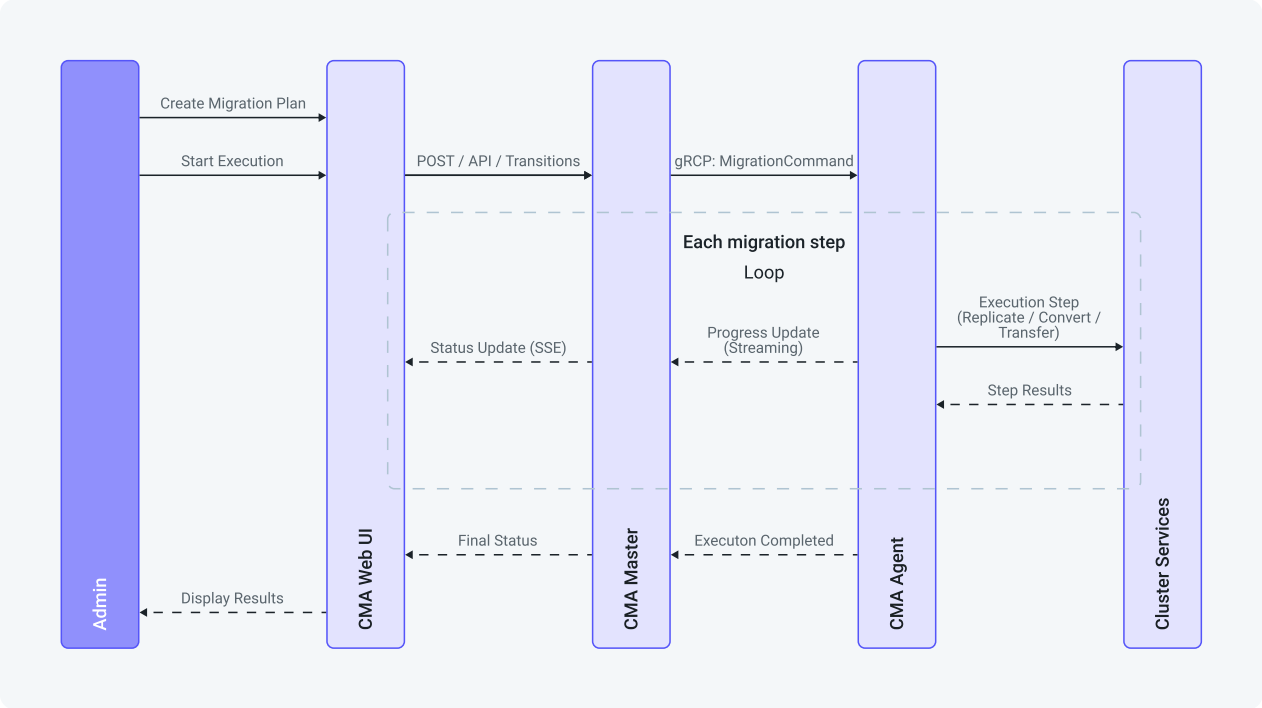
Migrating Airflow DAGs from to Git end to end [Technical Preview]

Deploy Cloudera Migration Assistant, configure Cloudera Data Engineering and Git sources, and migrate DAGs to a Cloudera Workflow Orchestrator Git repository.

About this task

This end-to-end workflow combines deployment, platform preparation, discovery, and workload steps. Detailed reference procedures are reused from the Administrator and Migration guides.

Figure 37: Migration flow



Before you begin

Verify supported platforms and migration paths in *Supported platforms and migration paths* before you start.

This workflow is technical preview. See *Setting up to Cloudera Workflow Orchestrator Git migration*.

Procedure

Phase 1 — Deploy the CMA Master and CMA Agents

Procedure

1. Download both CSD files to the /opt/cloudera/csd/ directory on the Server host.

   ```
   wget -P /opt/cloudera/csd/ https://archive.cloudera.com/cma/VERSION/csd/
   CMA_MASTER-VERSION-BUILD.jar
   ```

   ```
   wget -P /opt/cloudera/csd/ https://archive.cloudera.com/cma/VERSION/csd/
   CMA_AGENT-VERSION-BUILD.jar
   ```

   automatically detects the CSD files.
2. Change the ownership of the CSD files.

   ```
   chown cloudera-scm:cloudera-scm /opt/cloudera/csd/CMA_MA
   STER-VERSION-BUILD.jar /opt/cloudera/csd/CMA_AGENT-VERSION-BUILD.jar
   ```
3. Restart for the changes to take effect.

   ```
   systemctl restart cloudera-scm-server
   ```
4. Log into .
5. Restart the Cloudera Management Service.
6. Go to Hosts Parcels .

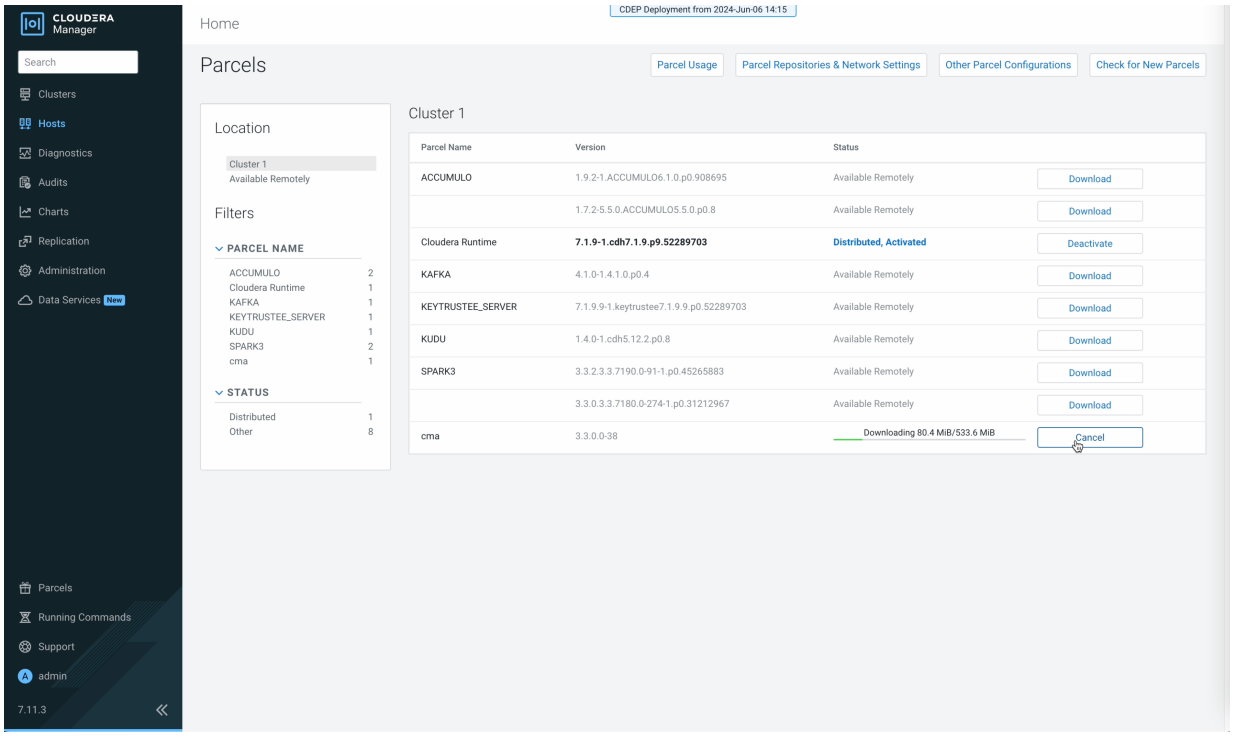
7. Click Parcel Repositories & Network Settings.
8. Add the Remote Parcel Repository URL for .

```
https://archive.cloudera.com/cma/VERSION/parcels/
```

Note: Make sure that the Remote Parcel Repository URL uses an HTTPS link. To install a different version of the parcel, change the URL as needed.

9. Click Save & Verify Configuration to commit the change.
10. Click Close.
    You are redirected to the Parcels page.
11. Search for , and click Download to download the parcel to the local repository.
12. After download completes, click Distribute to distribute the parcel to all clusters.

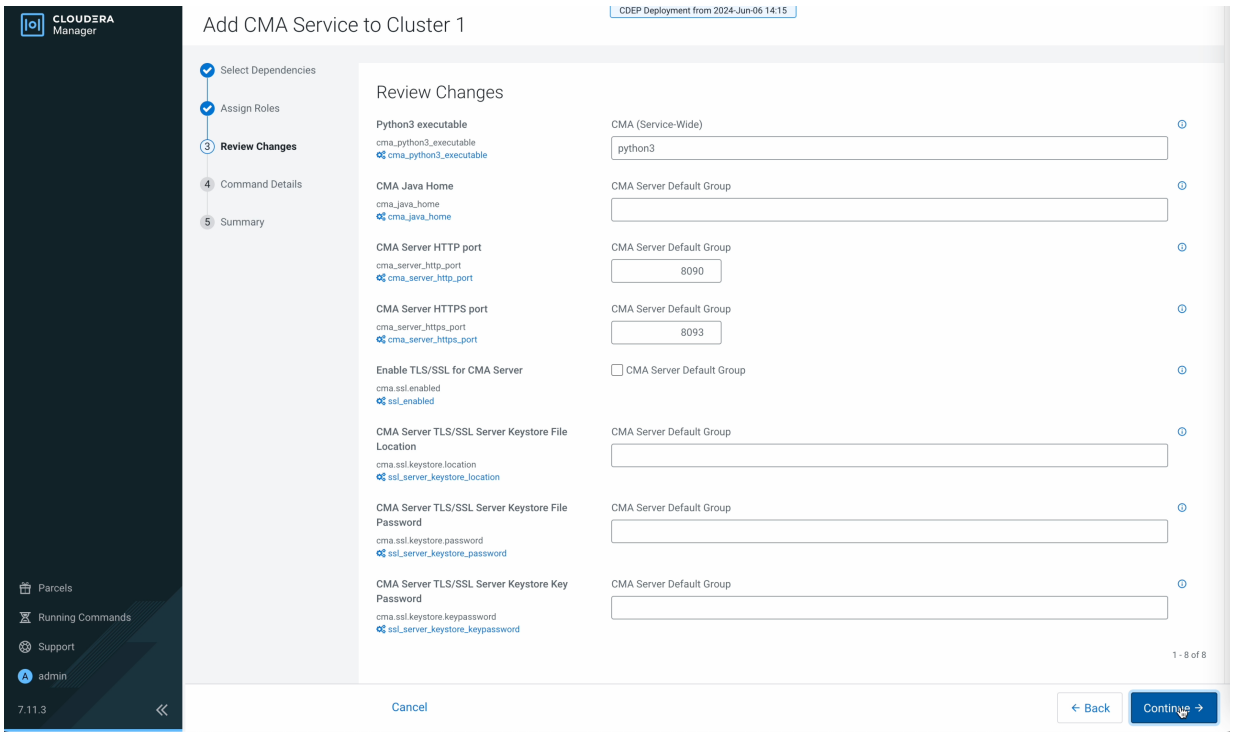
Figure 38: Distribute the parcel



13. Click Activate to activate the parcel.
14. Click OK when confirmation is required.
15. Click Clusters in the left navigation pane.
16. Select Add Service from the drop-down menu to the right of your cluster.
17. From the list, select the service type to add, then click Continue. The single parcel provides two service types: CMA_MASTER (central control node) and CMA_AGENT (cluster-side executor).

    The **Add Service wizard** opens.

Figure 39: Add service

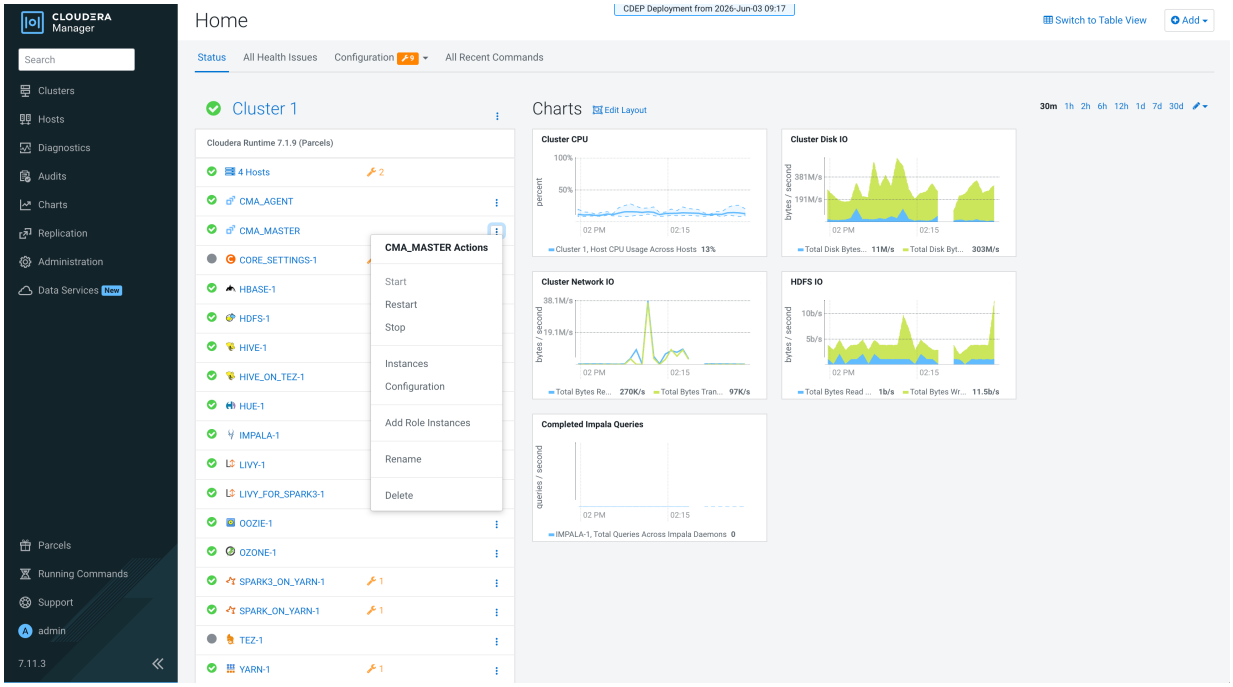


18. Assign the service roles to the hosts where Python 3.11 and JDK 17+ are installed, and click Continue.

19. Review service configurations and click Continue. At minimum, set OAuth2 client credentials (cma_client_secret, cma_agent_client_secret). If the is on a different cluster from the , also set Gateway URL (cma_gateway_url).

The first run of the service starts. When the command finishes, the service is added to the cluster.

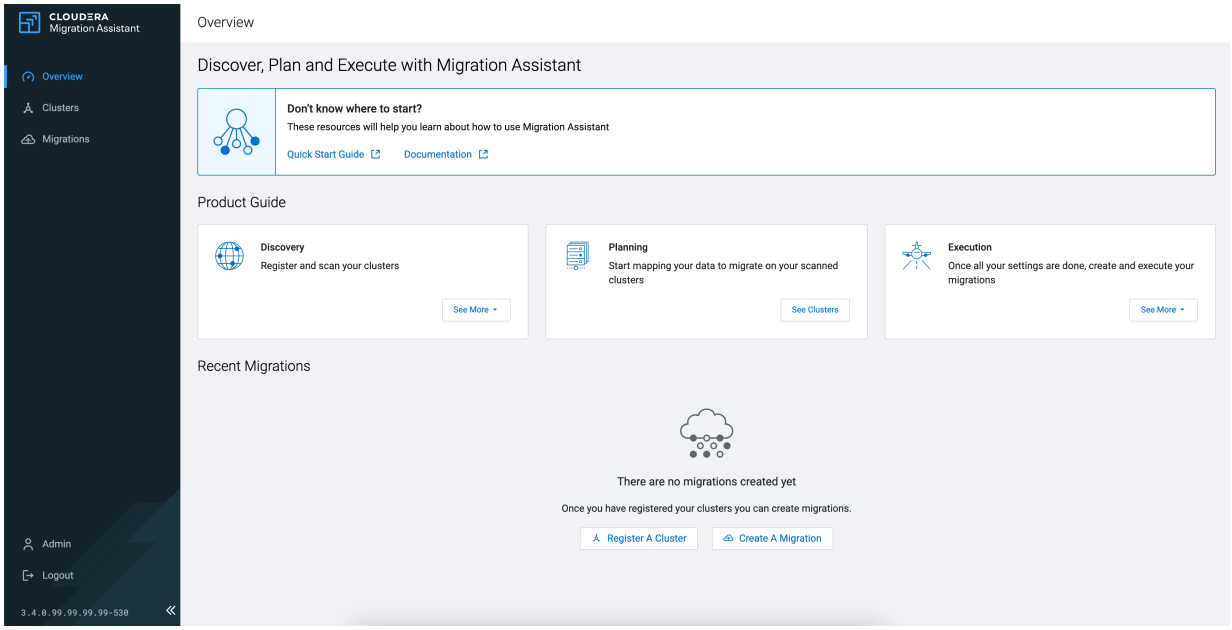
Figure 40: service started



20. Repeat steps 16–19 to add the other service type if needed (for example, add CMA_AGENT after CMA_MAST ER, or vice versa).

21. Go back to the cluster homepage, open the service page, and click the Server UI tab to open the UI.

Figure 41: landing page



22. Set the service dependency to none in the service configuration in Cloudera Manager because Cloudera Manager cannot discover the Master automatically across clusters.
23. Set the **Gateway URL** to cma_gateway_url in the service configuration.

```
curl -sk https://<master-host>:8093/api/agent/config
```

Procedure

Phase 2 — Configure Cloudera Replication Manager on the target cluster

Procedure

1. Enable the HDFS to Ozone flag on the Server host.

   For HDFS to Ozone migrations, requires the CMF_FF_API_H2O_REPLICATION feature flag to be enabled. Without this flag, the API endpoints for HDFS to Ozone replication are not available.

   Note: This flag is only required for HDFS to Ozone migrations. It is not required for HDFS-to-HDFS, Hive, or HBase replication.

   a) Append the following line to /etc/default/cloudera-scm-server on the Server host:

   ```
   export CMF_FF_API_H2O_REPLICATION=true
   ```

   b) Restart the Server.

   ```
   sudo systemctl restart cloudera-scm-server
   ```

   c) Wait for to become available again on port 7180 before proceeding with the setup.
2. Register the instance in the UI to allow to locate and communicate with the replication service.
   a) In the Web UI, go to Settings.
   b) Go to the App Access section.
   c) Configure your connection fields using one of the following validation methods:

- If has already detected a service on one of the registered clusters, select it from the Suggestions drop-down list to auto-fill the form.
- Manually enter your target replication parameters in the following fields:
  - App URL — The full URL of the REST API, including protocol, host, and port, for example, https://rm-host.example.com:9011.
  - App username — The username for authenticating with the API.
  - App password — The password for authenticating with the API.

d) Click Save and Push to Agents.

Procedure

Prepare Cloudera Data Engineering and Git sources (before Phase 2b)

Procedure

1. Enable the cma.cde-git-migration-enabled flag.
   a) In the UI, go to Settings.
   b) Click the Feature Flags tab.
   c) Locate the cma.cde-git-migration-enabled flag and set it to true.
   d) Click Save.
   e) Go to the service in .
   f) Click the Configuration tab.
   g) Search for CMA_AGENT_SERVER_role_env_safety_valve.
   h) Add the Cloudera Control Plane and Git repository environment variables listed in the tables in the next step.
   i) Click Save to save the advanced configuration snippet configuration.
   j) Restart the service.
   k) In the UI, go to Clusters.
   l) Confirm that two new clusters are displayed. If the clusters are not displayed within a few minutes, check the log for registration errors.
2. Configure the advanced configuration snippet in on the cluster where the is deployed:

   and Git credentials are passed to the through the CMA_AGENT_SERVER_role_env_safety_valve advanced configuration snippet. These environment variables are read at startup. The agent uses them to register the and Git clusters.

   The Cloudera Control Plane settings allow the to authenticate with the and the Knox gateway.

Table 8: Cloudera Control Plane environment variables

Variable	Description	Example
CMA_AGENT_CLUSTER_CLOUD_CONTROL_DATACENTER	API endpoint URL.	https://api.us-west-1.cdp.cloudera.com
CMA_AGENT_CLUSTER_CLOUD_CONTROL_ACCESS_KEY_ID	access key ID for Control Plane authentication. Obtained from the in User Management Access Keys .	a1b2c3d4-...
CMA_AGENT_CLUSTER_CLOUD_CONTROL_PRIVATE_KEY	private key corresponding to the access key ID.	<YOUR_CDP_PRIVATE_KEY>
CMA_AGENT_CLUSTER_CLOUD_CONTROL_USERNAME	username. Must be an **LDAP user**. Used by the to authenticate against the Knox gateway when fetching DAG metadata.	admin
CMA_AGENT_CLUSTER_CLOUD_CONTROL_PASSWORD	password corresponding to the username.	<CDE_PASSWORD>

Note: CMA_AGENT_CLUSTER_CLOUD_CONTROL_USERNAME and CMA_AGENT_CLUSTER_CLOUD_CONTROL_PASSWORD are credentials, not credentials. They are used only for Knox authentication against virtual cluster APIs.

The Git repository settings allow the Git Agent to scan the target repository and create Pull Requests or Merge Requests.

Table 9: Git repository environment variables

Variable	Description	Example
CMA_AGENT_CLUSTER_GIT_REPOURL	HTTPS URL of the target Git repository. SSH URLs are not supported.	https://github.com/your-org/your-repo.git
CMA_AGENT_CLUSTER_GIT_TOKEN	Personal access token for the Git repository. For GitHub, the token requires the repo scope. For GitLab, it requires api or read_repository + write_repository.	<YOUR_GIT_PAT>
CMA_AGENT_CLUSTER_GIT_BRANCH	Base branch name. creates a source branch named cma/<planName> from this branch and opens a Pull Request or Merge Request back to it. Defaults to main if not set.	main

3. Save and restart the .

   After the restarts, it registers both clusters with automatically. No manual steps are required in the UI.
4. Verify the cluster registration.

Table 10: Verification values for auto-registered clusters

Cluster Name	Platform Label	Purpose
<clusterName>-cde	Cloudera Data Services	source that is used for scanning DAGs
<clusterName>-git	Git	Git target that is used for scanning the repository and creating Pull Requests and Merge Requests

Procedure

Add per-user Cloudera Data Engineering and Git credentials in the UI. See *Managing user connections*.

Procedure

Phase 3 — Register source and target clusters

Procedure

1. In on the source cluster, select Add Service and select CMA_AGENT.
2. Configure the required properties.

Table 11: configuration properties

Property	Description
cma_gateway_url	URL of the Gateway, for example, https://master-host:8093. **Required when the is on a different cluster than the .** The agent uses this URL to discover the Master host, gRPC port, and Auth Server URI automatically through the discovery endpoint (GET /api/agent/config). When connecting from a different cluster, set the service dependency to none in .
cm_username	administrator user name.

Property	Description
cm_password	administrator password.
cma_user_home	Home directory for the cma-agent user, for example, /var/lib/cma-agent.

3. Set cma_gateway_url when the is on a different cluster than the .
4. Start the service.
5. Wait for the agent to connect to the and finish the registration automatically.

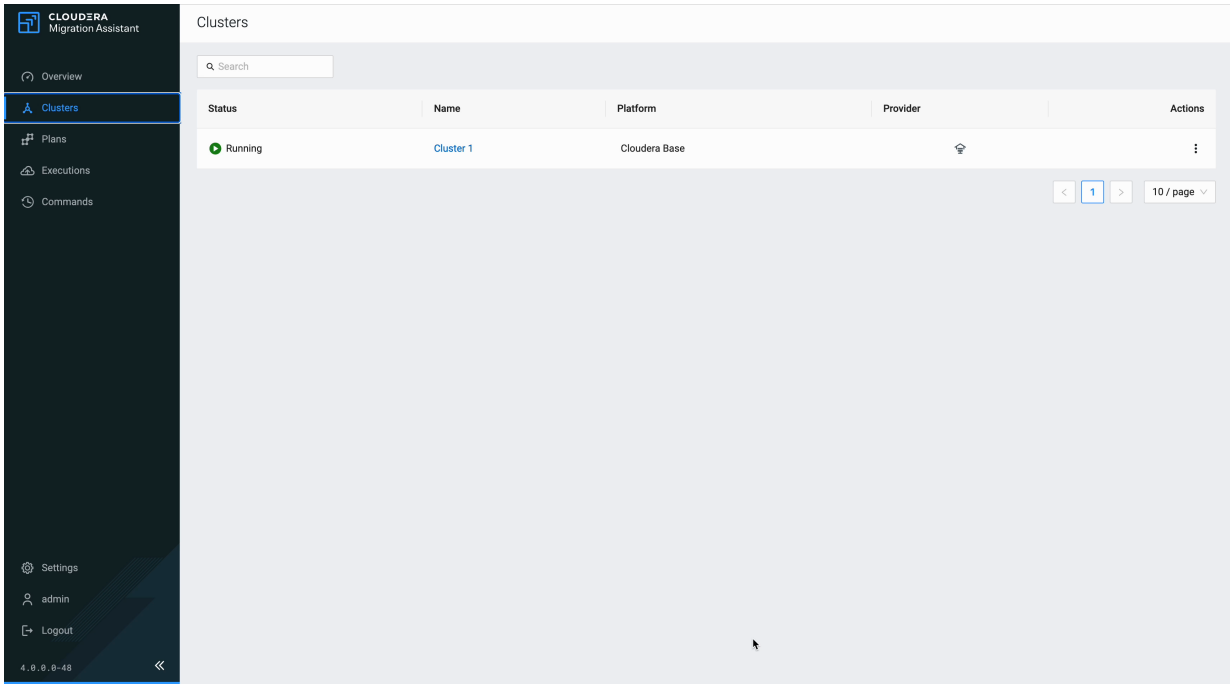
Procedure

Phase 4 — Scan clusters

Procedure

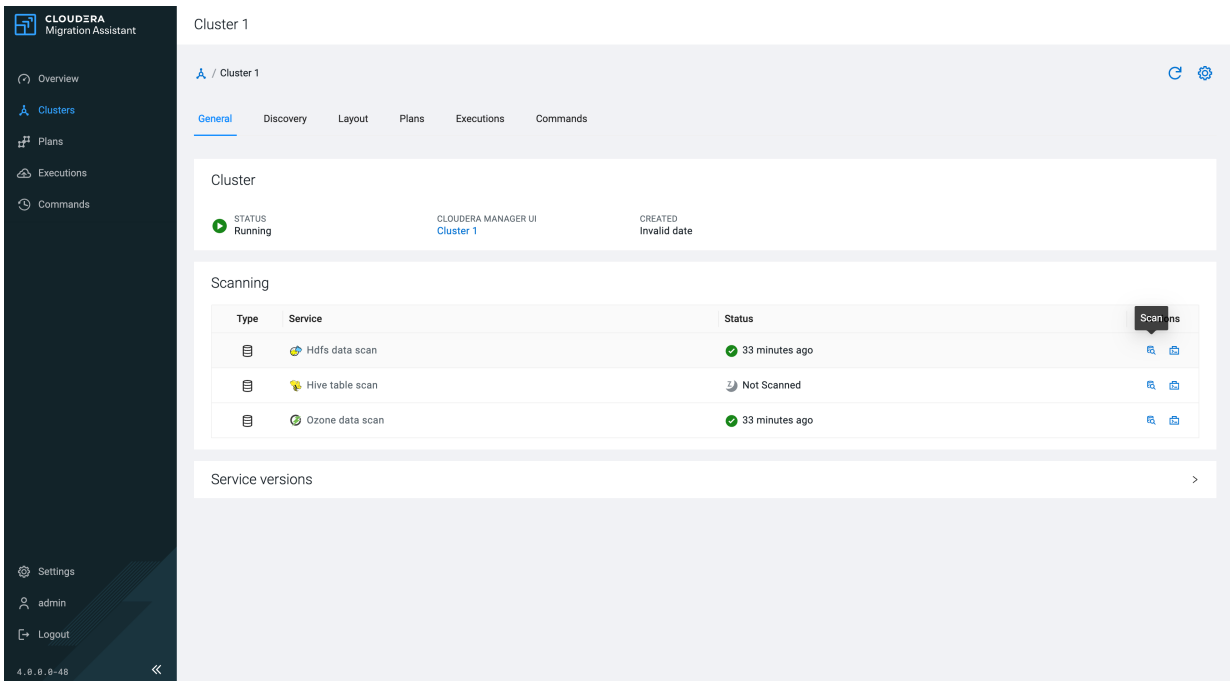
1. Navigate to the Clusters page and click the cluster you want to scan.

Figure 42: Clusters page



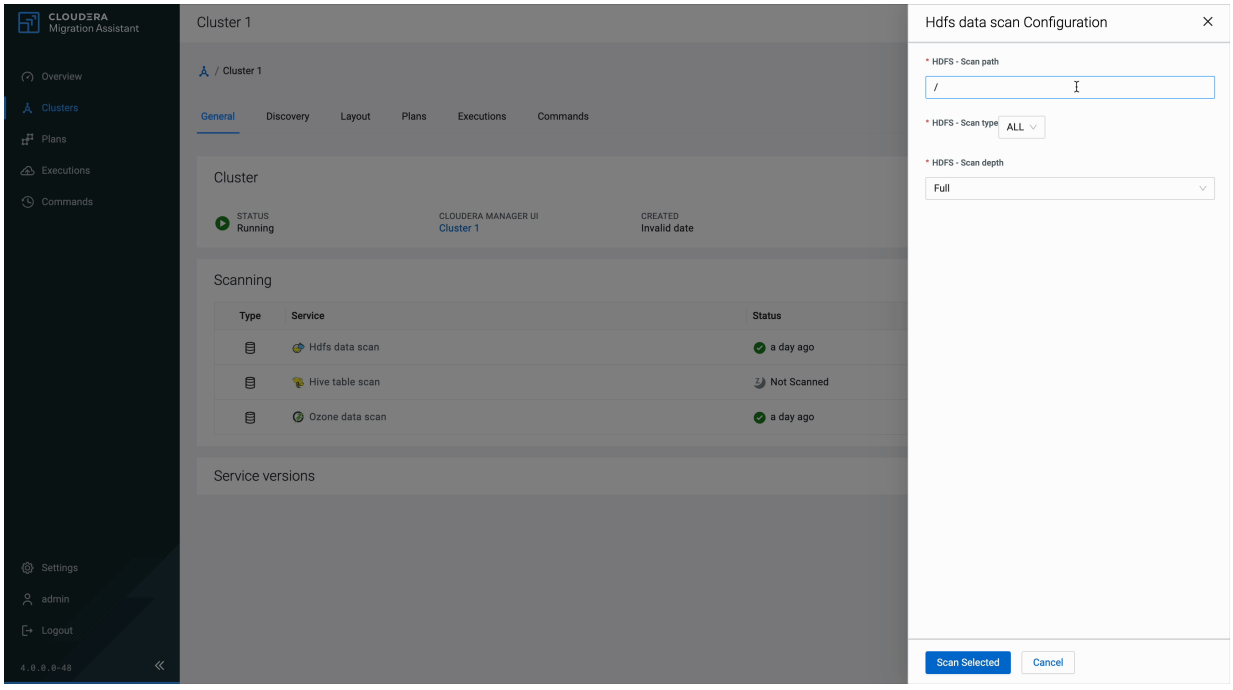
2. On the cluster detail page, find the Scanning section, which lists the available scan types for your cluster. Click the configuration icon on the right side of the scan type you want to trigger.

Figure 43: cluster scanning



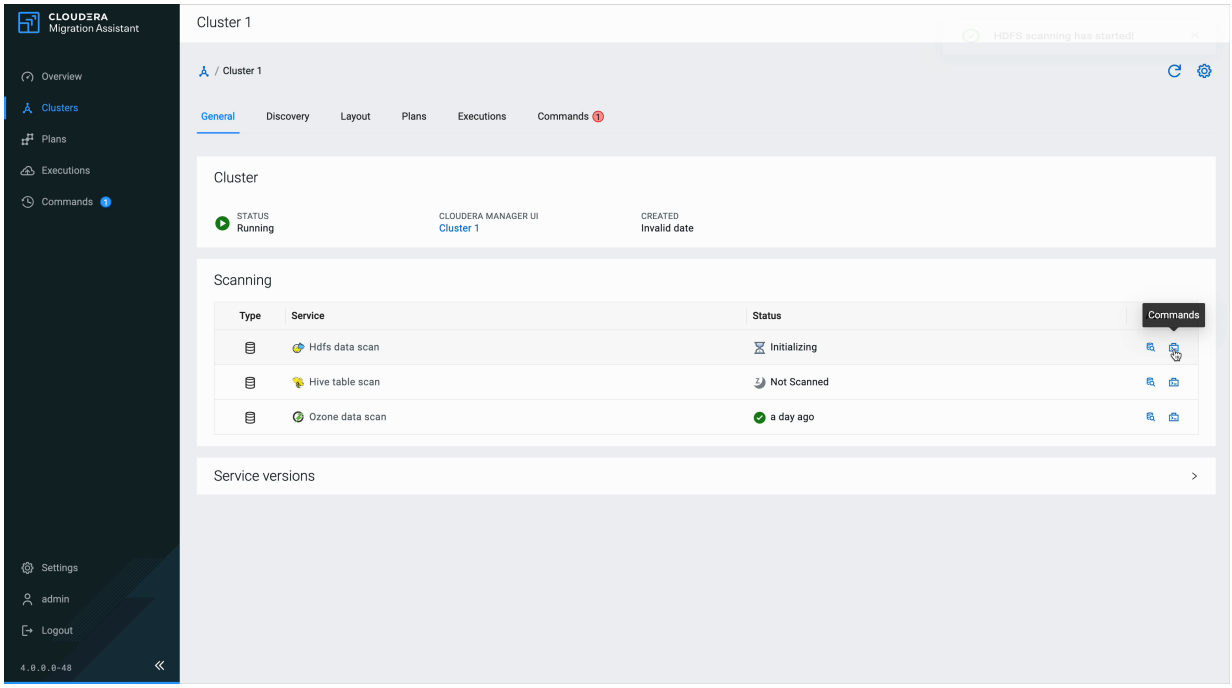
3. In the scan configuration drawer, configure the scan parameters (for example, scan path, scan type, scan depth) and click Scan Selected.

Figure 44: scan configuration



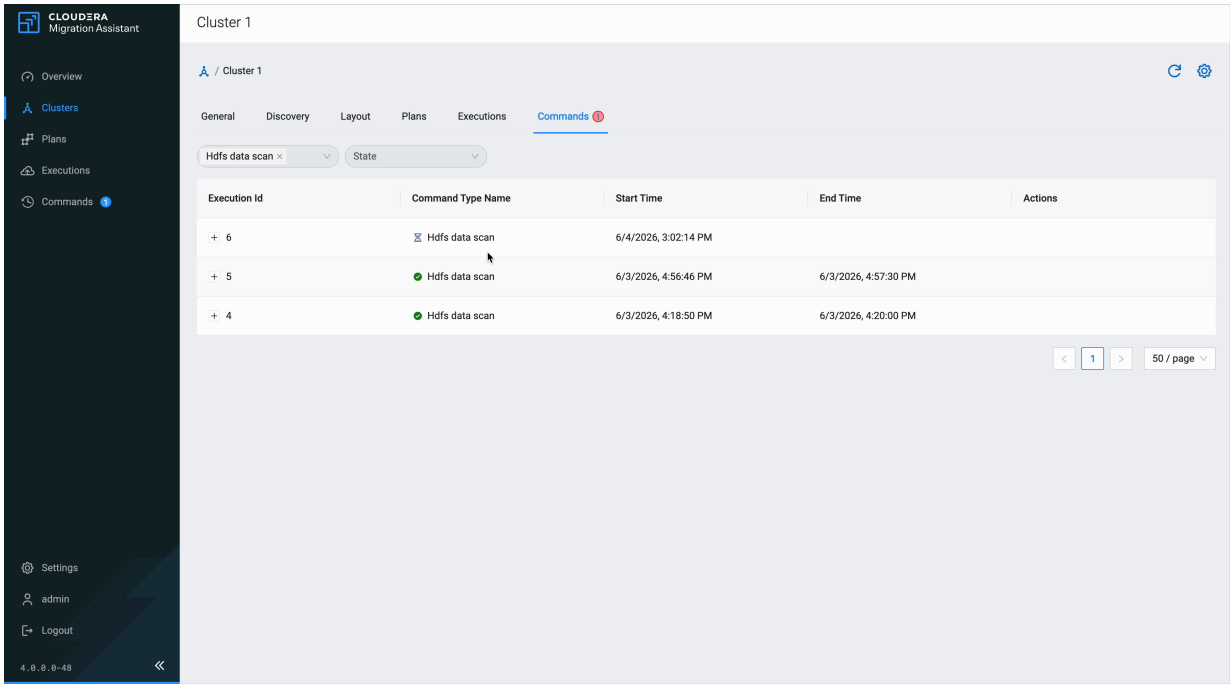
4. The scan starts and the status changes to Initializing. Track scan progress by clicking the Commands tab, which shows a badge with the number of active commands.

Figure 45: scan in progress



5. On the Commands tab, expand a command to view its Output Log, Error Log, and Configuration.

Figure 46: scan commands



Note: Scan results are also available in the data directory on the source cluster at <CMA_ROOT_DIR>/data/sources/ClusterName/Source_ID/cluster-scan/ on the host.

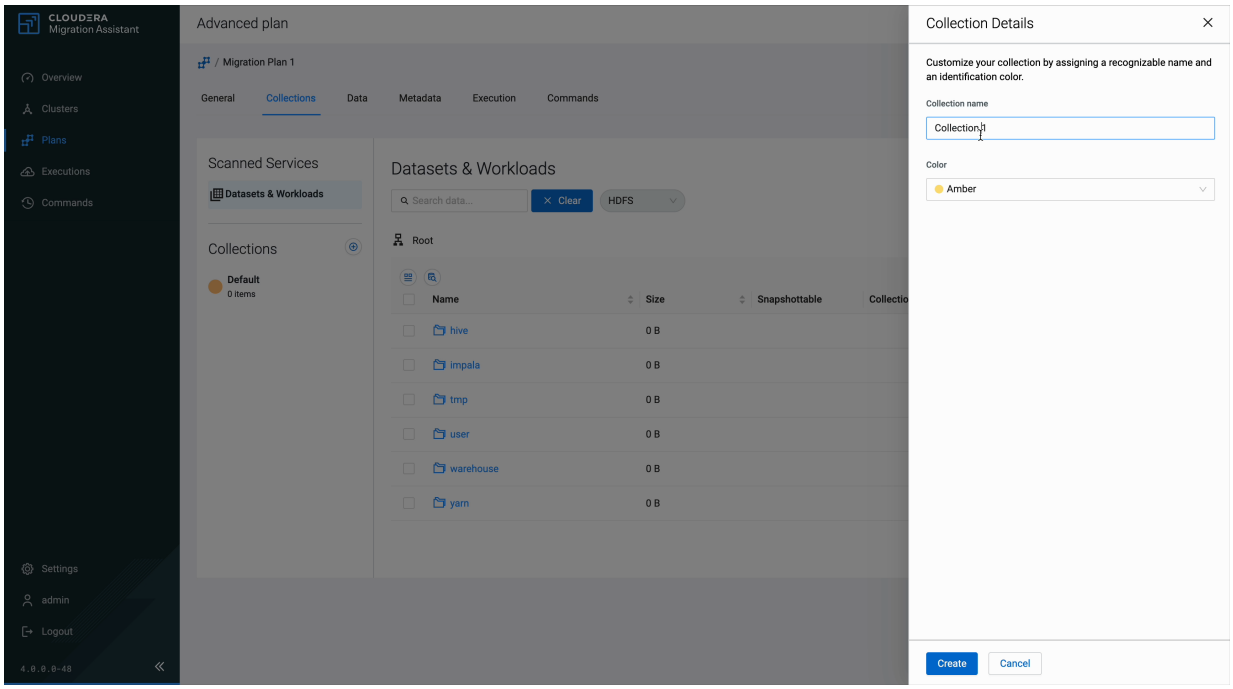
Procedure

Phase 5 — Create collections (optional)

Procedure

1. Click the + icon next to Collections in the sidebar.
2. In the **Collection Details** drawer, enter a Collection name and select a Color for identification.

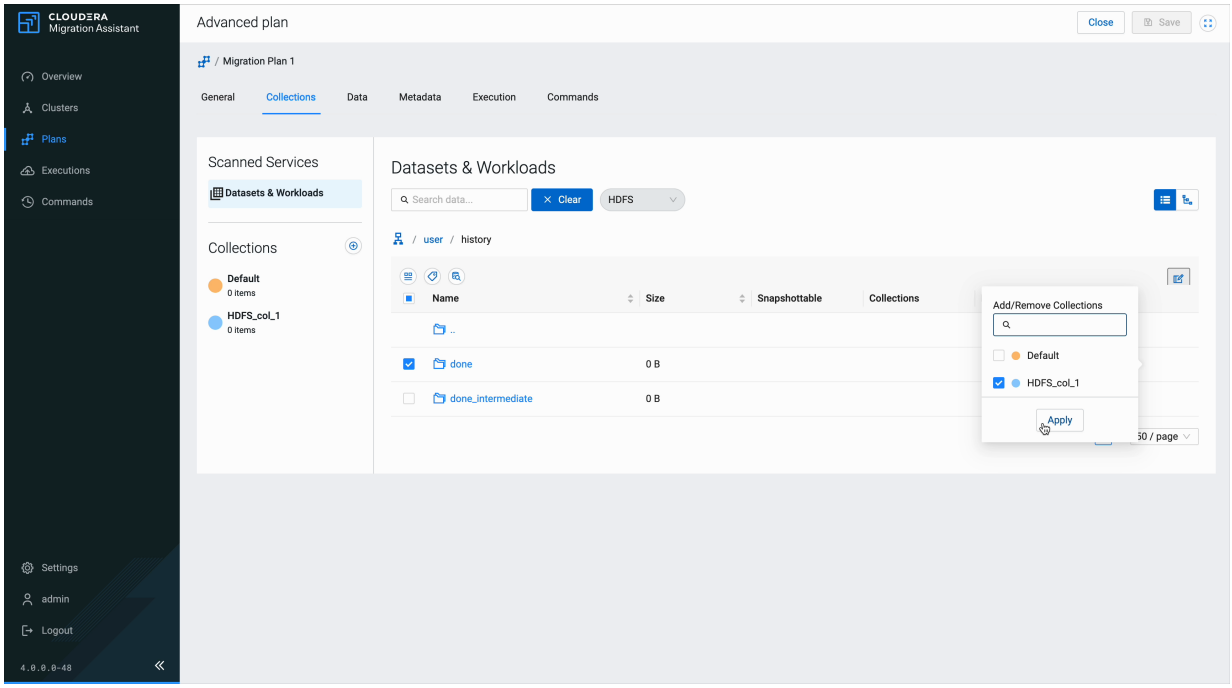
Figure 47: create collection



3. Click Create.

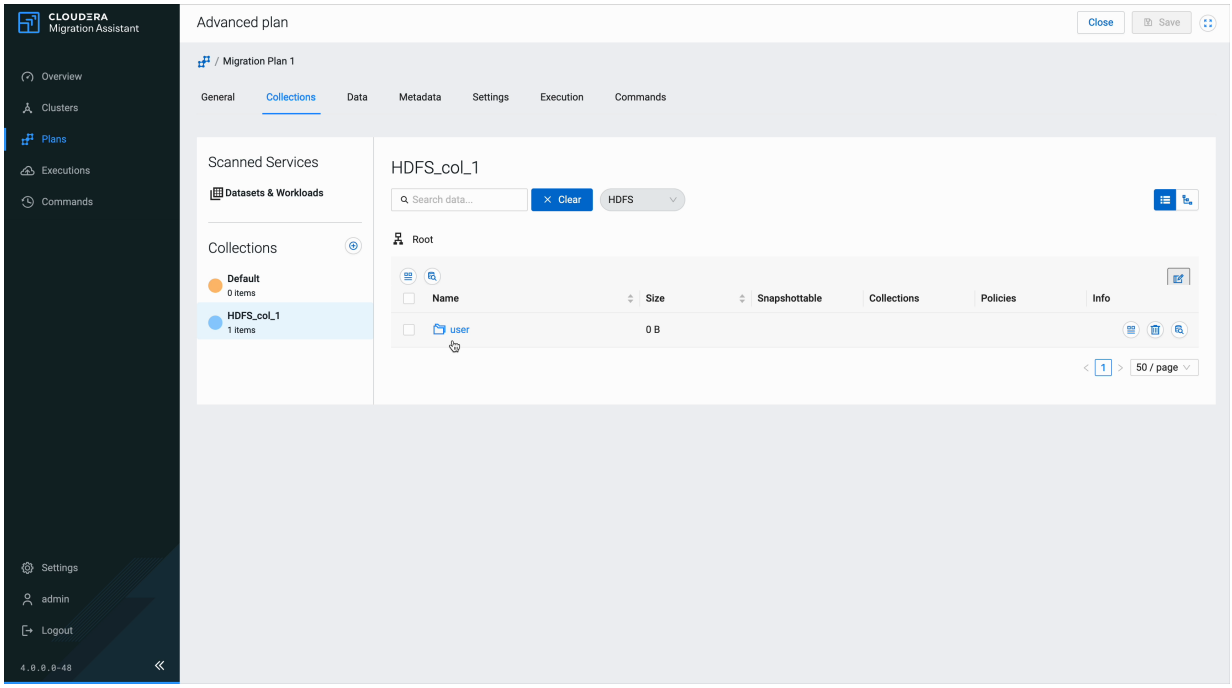
   The new collection appears in the sidebar with 0 items.
4. Browse the data tree to find the items you want to migrate. Click folder names to navigate deeper into the hierarchy.
5. Select items using the checkboxes next to each item.
6. Click the collection icon that appears on the selected item row.
7. In the **Add/Remove Collections** popup, check the collections you want to assign the item to.
8. Click Apply.

Figure 48: assign to collection



9. Click a collection name in the sidebar to view its contents. The data browser switches to show only the items assigned to that collection.

Figure 49: view collection



Procedure

Phase 6 — Run the migration workload

Procedure

Step 1 — Scan the cluster

1. Identify the cluster whose name suffix reads -cde and open it.
2. Click Start Scanning.
3. Select Scan for Airflow.
4. Click Scan selected and wait until the progress indicator displays success.

Reading scan results

Discovery displays CDE Service # Virtual Cluster # Airflow DAGs. Expand each branch to inspect individual items.

Table 12: Discovery page columns

Column	Description
DAG Name	job display name.
DAG ID	Unique ID (dag_id) inside the Python script.
Cron	Configured schedule expressions.
Active	Indicator of the DAG schedule status.
Operators	Inferred operators, such as BashOperator.
Triggers	Outbound trigger graph.
Triggered By	Inbound trigger graph.
Resource Pools	resource pools allocated for mounted artifacts.

You can use the following interface controls to change how the data is displayed:

- Data – Switches to a file-centric tree, which is ideal before creating mappings.
- Metadata — Switches to the logical view of DAG summaries.
- List/tree toggle – Switches between hierarchical tree and flattened browsing list.

Note: Blank virtual-cluster rows often mean Knox authentication stalled during scan. Check for SKIP PED hints inside the command logs.

Step 2 — Scan the Git Repository

5. Open clusters ending in -git.
6. Click Start Scanning.
7. Select Git Repository Scan, then click Scan selected.

The repository tree displays Name and Size metadata. Empty repositories are acceptable because executions bootstrap directories.

Step 3 — Create a Migration Plan

8. On the Plans tab, click Create Plan.
9. Select a descriptive name aligned with Git branch naming rules (cma/<planName> prefixes created branches).
10. Select the -cde source and -git target clusters.
11. Select DAGs to Git Repository as transition type, then finalize creation.

Step 4 — Label DAGs on the Collections tab

Collections act as a filter that controls which DAGs are visible on downstream configuration tabs.

12. In the Plan tab, go to Collections.
13. Expand the hierarchies.
14. Select DAG rows to migrate.
15. Click Add to collection, or click the + icon to create a new collection.

Tip: Revisit Collections whenever scope shifts. Your updates propagate instantly to the downstream mapping tabs.

Step 5 — Map DAG Metadata

16. On the Metadata tab, review split panes listing (left) and Git (right) DAG stubs.
17. Map individually (Map) or bulk-select unfinished rows (Copy to Git).
18. Defaults land at /dags/<dagId>/<dagId>.py plus /include/<service>/<vc>/<pool>/... artifact bundles.

    Unmap removes targets while preserving shared artifacts when safe.

Warning: rejects conflicting dag_id collisions across virtual clusters until you resolve overlaps in Step 7.

Step 6 — Verify File Paths on the Data tab

19. Go to Data to compare mirrored trees.
20. Verify that DAG payloads and supporting files have the expected prefixes.

    If you notice any incorrect file paths, go to the Metadata tab to unmap or remap those files.

Step 7 — Review DAG Source (Workloads tab)

21. Open Workloads whenever duplicate-ID badges block executions.
22. Select DAG entries to load Monaco editors with read-only scanned copy displayed in the left pane, editable version in the right pane.
23. Edit the right pane to rename dag_id, revise connections, or prune unsupported operators and save afterward.
24. Ensure that warnings clear across every DAG affected by overlaps.

Important: Renaming dag_id rewires filesystem targets automatically (/dags/<newDagId>/...).

Step 8 — Execute Git Publishing

25. In Execution, click Create Execution.
26. Monitor the sequential pipeline as it automatically processes the following stages:
    a. **Upload resources** – The retrieve scanned artifacts locally (skipped when unnecessary).
    b. **Prepare resources** – The merges edited DAG payloads with staging artifacts.
    c. **Create Git change** – The Git-role agents clone staging work, publish branch cma/<planName>, commit files, then open Pull Requests, or Merge Requests or fallback pushes depending on your Git configuration.

Step 9 — Inspect Provider Output

27. Expand the completed **Create Git Change** rows to view direct links to your repository.
28. Click the outbound links to open the generated Pull Requests or Merge Requests in your Git provider interface.

    Merged repositories carry DAG artifacts on tracked base branches and workspaces continue governing runtime workloads.

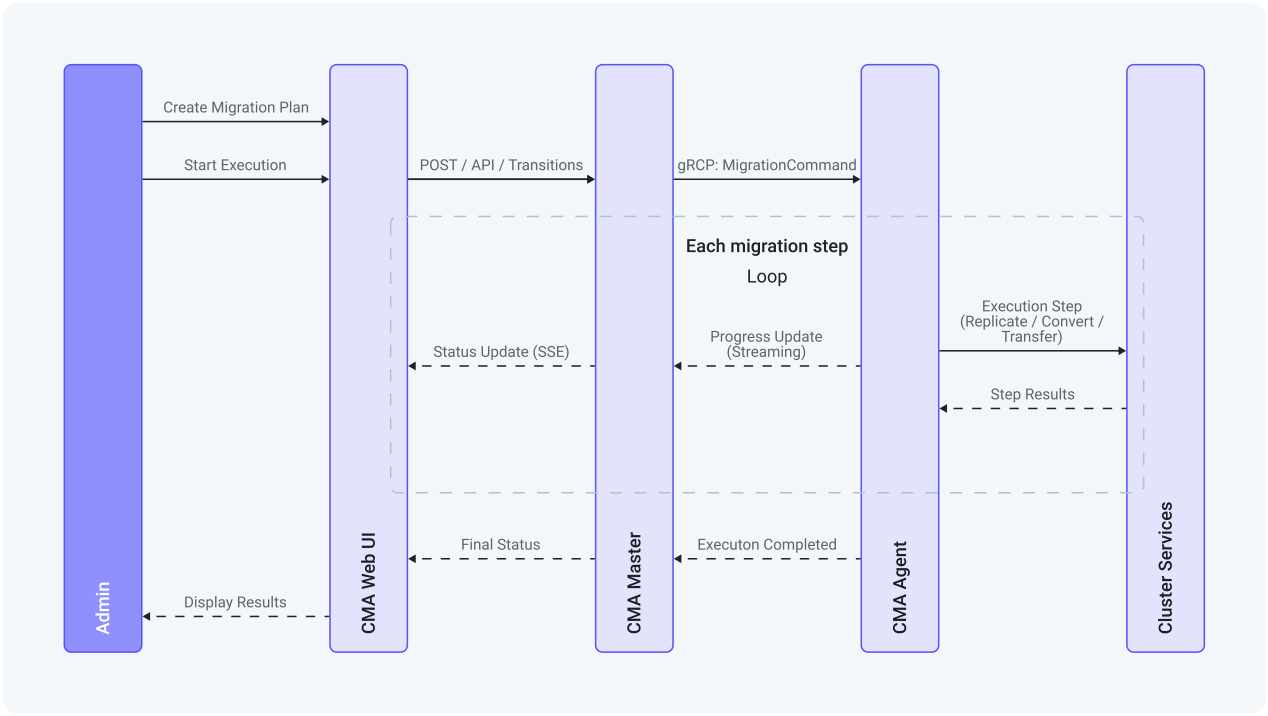
Discovering Ranger policies end to end [Technical Preview]

Deploy Cloudera Migration Assistant, register a Ranger-enabled cluster, and browse discovered policies.

About this task

This end-to-end workflow combines deployment, platform preparation, discovery, and workload steps. Detailed reference procedures are reused from the Administrator and Migration guides.

Figure 50: Migration flow



Before you begin

Verify supported platforms and migration paths in *Supported platforms and migration paths* before you start.

Procedure

Phase 1 — Deploy the CMA Master and CMA Agents

Procedure

1. Download both CSD files to the /opt/cloudera/csd/ directory on the Server host.

   ```
   wget -P /opt/cloudera/csd/ https://archive.cloudera.com/cma/VERSION/csd/
   CMA_MASTER-VERSION-BUILD.jar
   ```

   ```
   wget -P /opt/cloudera/csd/ https://archive.cloudera.com/cma/VERSION/csd/
   CMA_AGENT-VERSION-BUILD.jar
   ```

   automatically detects the CSD files.
2. Change the ownership of the CSD files.

   ```
   chown cloudera-scm:cloudera-scm /opt/cloudera/csd/CMA_MA
   STER-VERSION-BUILD.jar /opt/cloudera/csd/CMA_AGENT-VERSION-BUILD.jar
   ```
3. Restart for the changes to take effect.

   ```
   systemctl restart cloudera-scm-server
   ```
4. Log into .
5. Restart the Cloudera Management Service.
6. Go to Hosts Parcels .
7. Click Parcel Repositories & Network Settings.

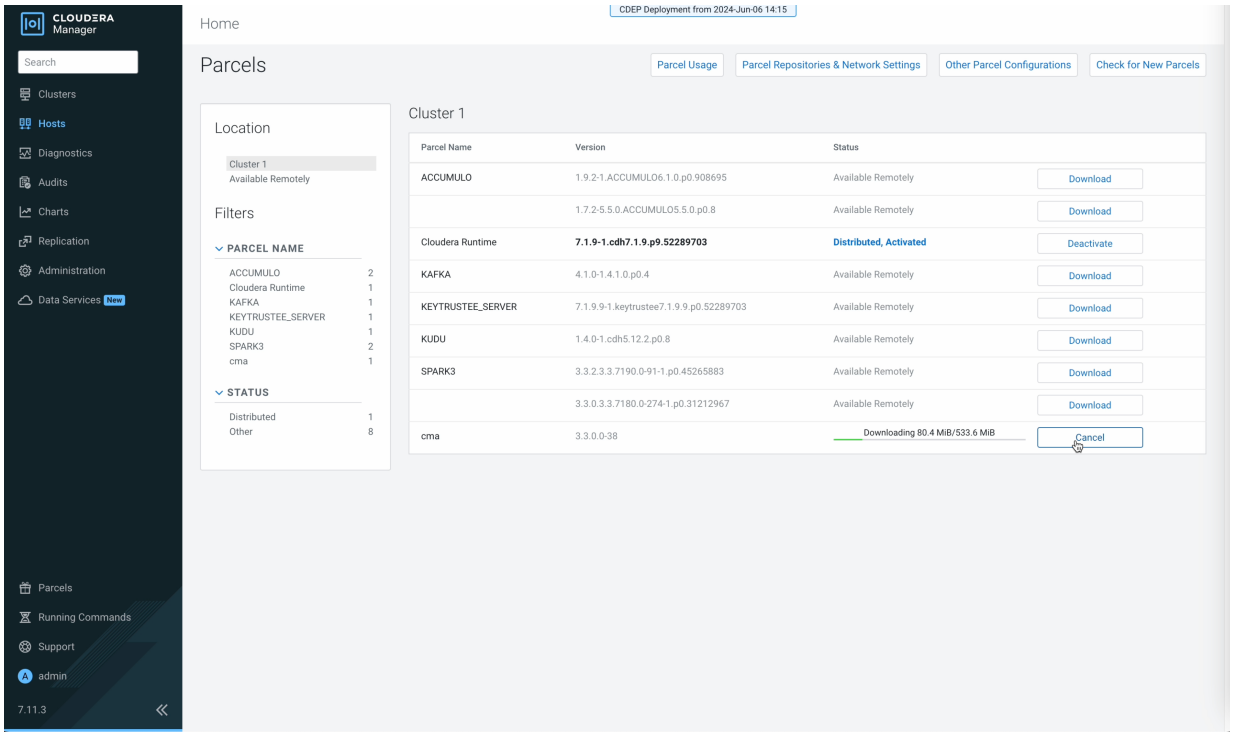
8. Add the Remote Parcel Repository URL for .

```
https://archive.cloudera.com/cma/VERSION/parcels/
```

Note: Make sure that the Remote Parcel Repository URL uses an HTTPS link. To install a different version of the parcel, change the URL as needed.

9. Click Save & Verify Configuration to commit the change.
10. Click Close.
    You are redirected to the Parcels page.
11. Search for , and click Download to download the parcel to the local repository.
12. After download completes, click Distribute to distribute the parcel to all clusters.

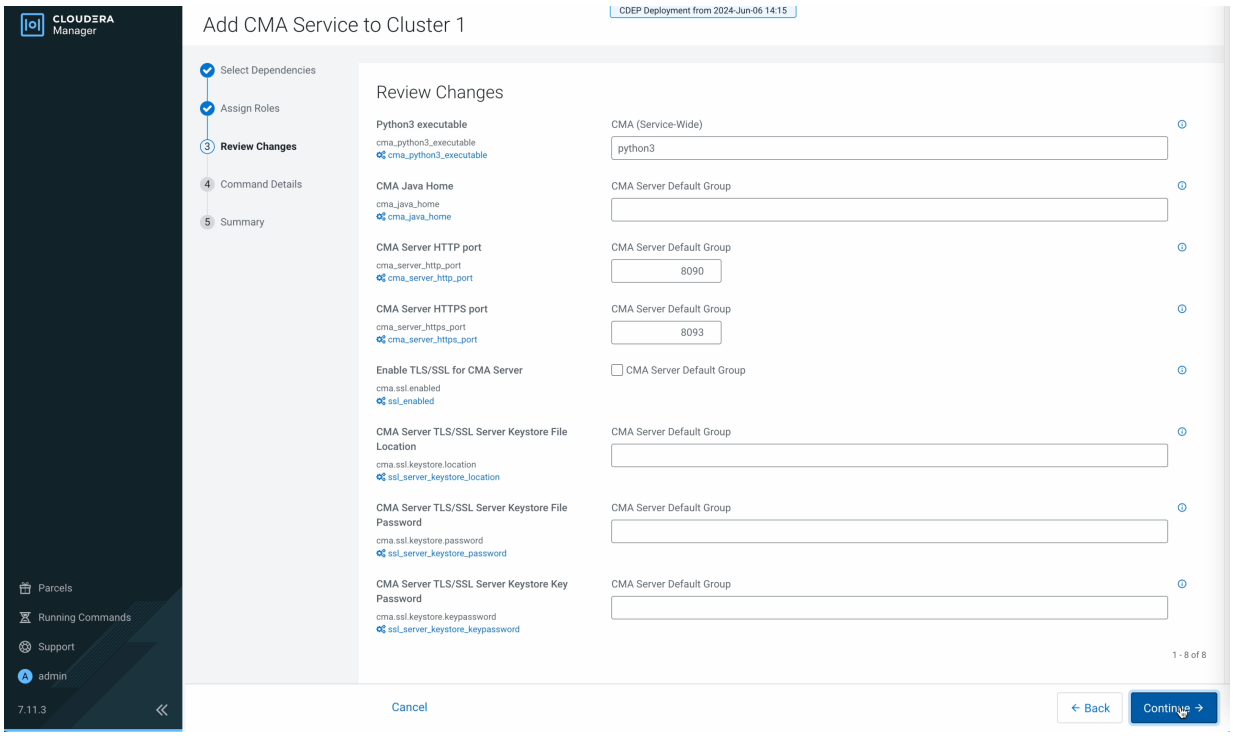
Figure 51: Distribute the parcel



13. Click Activate to activate the parcel.
14. Click OK when confirmation is required.
15. Click Clusters in the left navigation pane.
16. Select Add Service from the drop-down menu to the right of your cluster.
17. From the list, select the service type to add, then click Continue. The single parcel provides two service types: CMA_MASTER (central control node) and CMA_AGENT (cluster-side executor).

    The **Add Service wizard** opens.

Figure 52: Add service

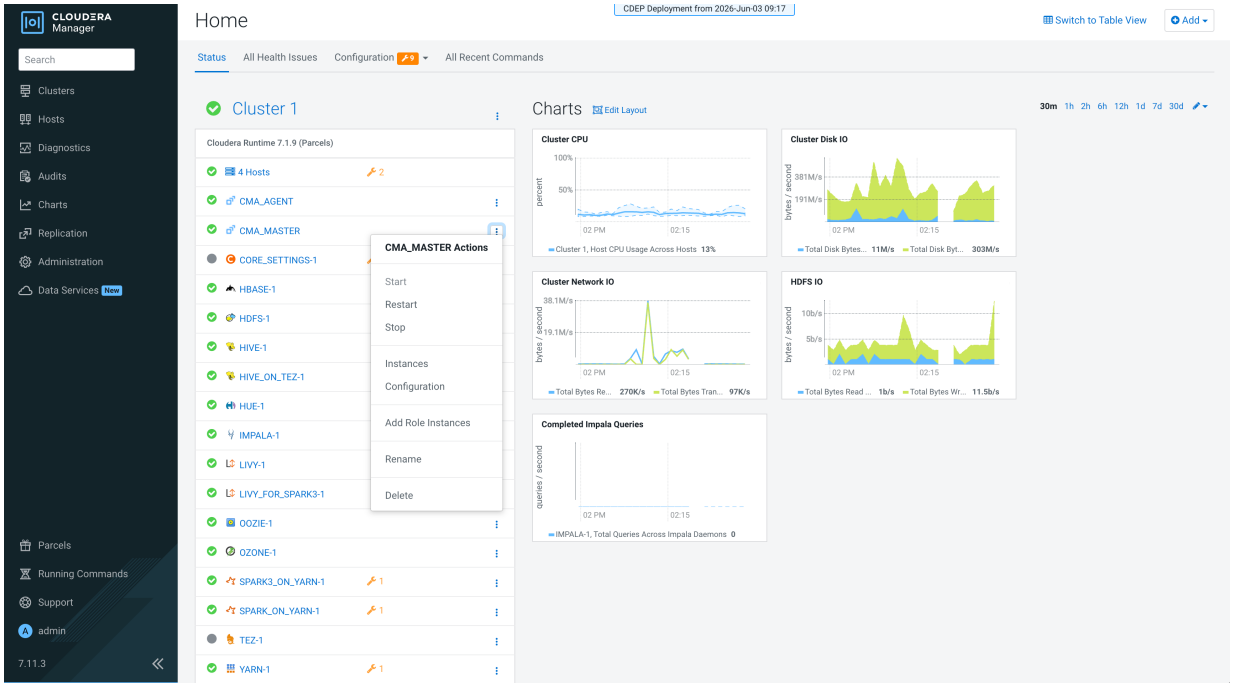


18. Assign the service roles to the hosts where Python 3.11 and JDK 17+ are installed, and click Continue.

19. Review service configurations and click Continue. At minimum, set OAuth2 client credentials (cma_client_secret, cma_agent_client_secret). If the is on a different cluster from the , also set Gateway URL (cma_gateway_url).

The first run of the service starts. When the command finishes, the service is added to the cluster.

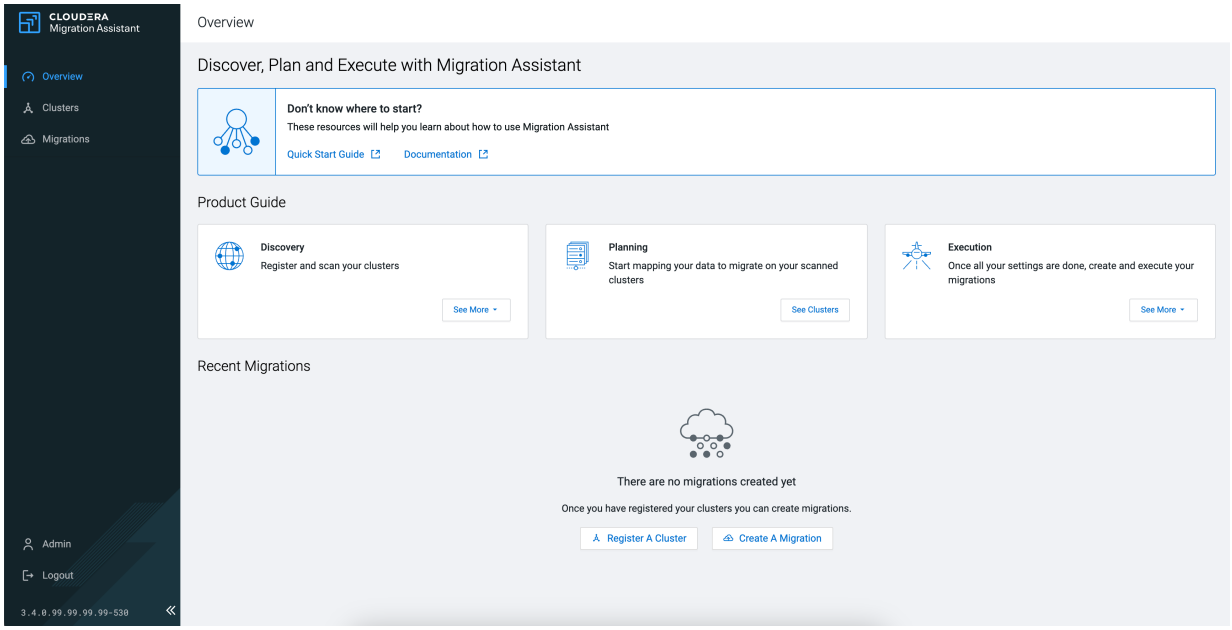
Figure 53: service started



20. Repeat steps 16–19 to add the other service type if needed (for example, add CMA_AGENT after CMA_MAST ER, or vice versa).

21. Go back to the cluster homepage, open the service page, and click the Server UI tab to open the UI.

Figure 54: landing page



22. Set the service dependency to none in the service configuration in Cloudera Manager because Cloudera Manager cannot discover the Master automatically across clusters.
23. Set the **Gateway URL** to cma_gateway_url in the service configuration.

```
curl -sk https://<master-host>:8093/api/agent/config
```

Procedure

Phase 2 — Configure Cloudera Replication Manager on the target cluster

Procedure

1. Enable the HDFS to Ozone flag on the Server host.

   For HDFS to Ozone migrations, requires the CMF_FF_API_H2O_REPLICATION feature flag to be enabled. Without this flag, the API endpoints for HDFS to Ozone replication are not available.

   Note: This flag is only required for HDFS to Ozone migrations. It is not required for HDFS-to-HDFS, Hive, or HBase replication.

   a) Append the following line to /etc/default/cloudera-scm-server on the Server host:

   ```
   export CMF_FF_API_H2O_REPLICATION=true
   ```

   b) Restart the Server.

   ```
   sudo systemctl restart cloudera-scm-server
   ```

   c) Wait for to become available again on port 7180 before proceeding with the setup.
2. Register the instance in the UI to allow to locate and communicate with the replication service.
   a) In the Web UI, go to Settings.
   b) Go to the App Access section.
   c) Configure your connection fields using one of the following validation methods:

- If has already detected a service on one of the registered clusters, select it from the Suggestions drop-down list to auto-fill the form.
- Manually enter your target replication parameters in the following fields:
  - App URL — The full URL of the REST API, including protocol, host, and port, for example, https://rm-host.example.com:9011.
  - App username — The username for authenticating with the API.
  - App password — The password for authenticating with the API.

d) Click Save and Push to Agents.

Procedure

Ranger policy discovery is technical preview and supports scan and browse only; it does not export policies to a target cluster.

Procedure

Phase 3 — Register source and target clusters

Procedure

1. In on the source cluster, select Add Service and select CMA_AGENT.
2. Configure the required properties.

Table 13: configuration properties

Property	Description
cma_gateway_url	URL of the Gateway, for example, https://master-host:8093. **Required when the is on a different cluster than the .** The agent uses this URL to discover the Master host, gRPC port, and Auth Server URI automatically through the discovery endpoint (GET /api/agent/config). When connecting from a different cluster, set the service dependency to none in .
cm_username	administrator user name.
cm_password	administrator password.
cma_user_home	Home directory for the cma-agent user, for example, /var/lib/cma-agent.

3. Set cma_gateway_url when the is on a different cluster than the .
4. Start the service.
5. Wait for the agent to connect to the and finish the registration automatically.

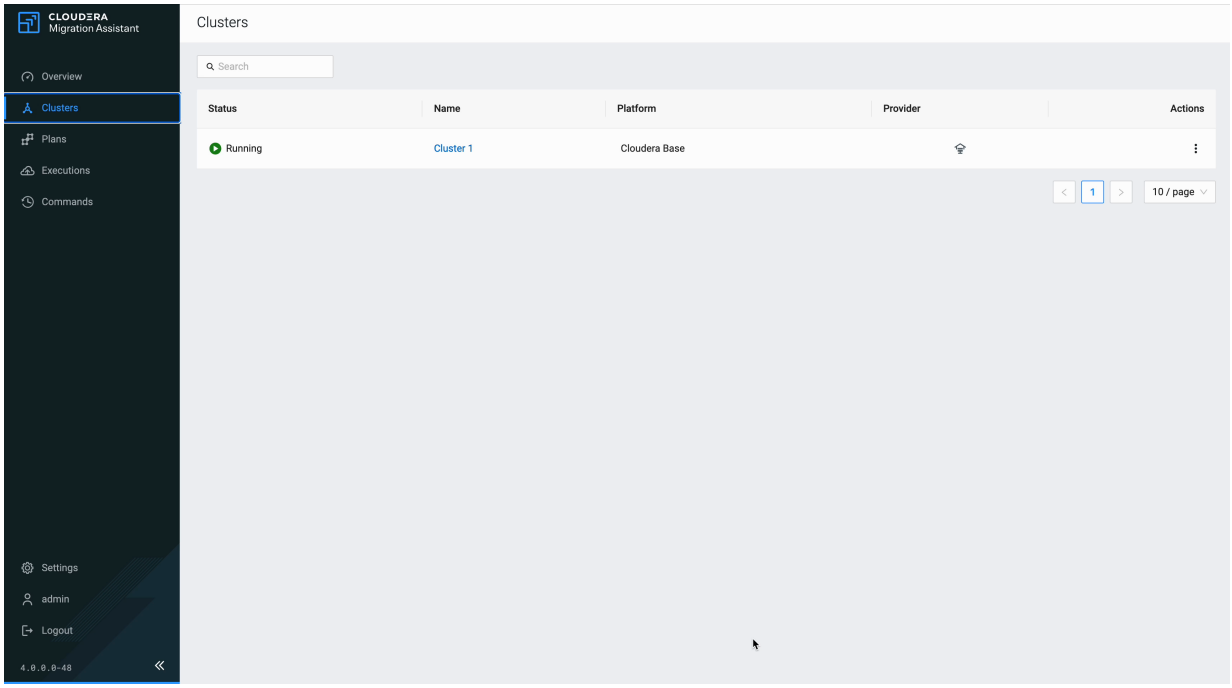
Procedure

Phase 4 — Scan clusters

Procedure

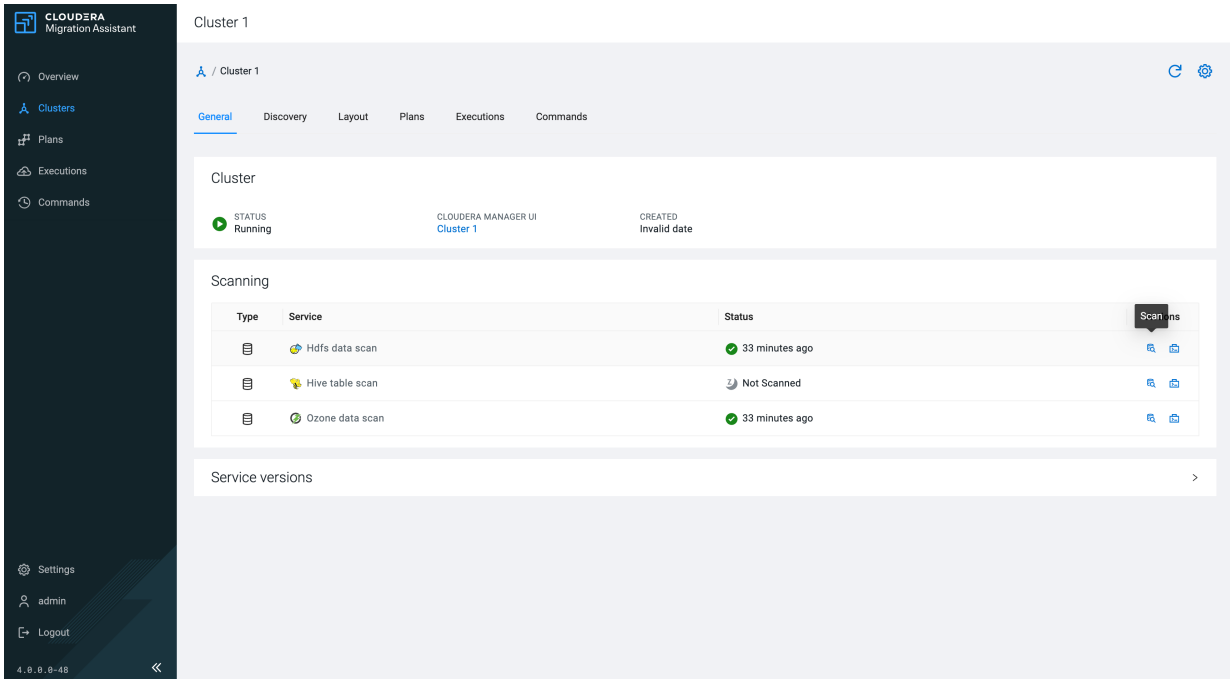
1. Navigate to the Clusters page and click the cluster you want to scan.

Figure 55: Clusters page



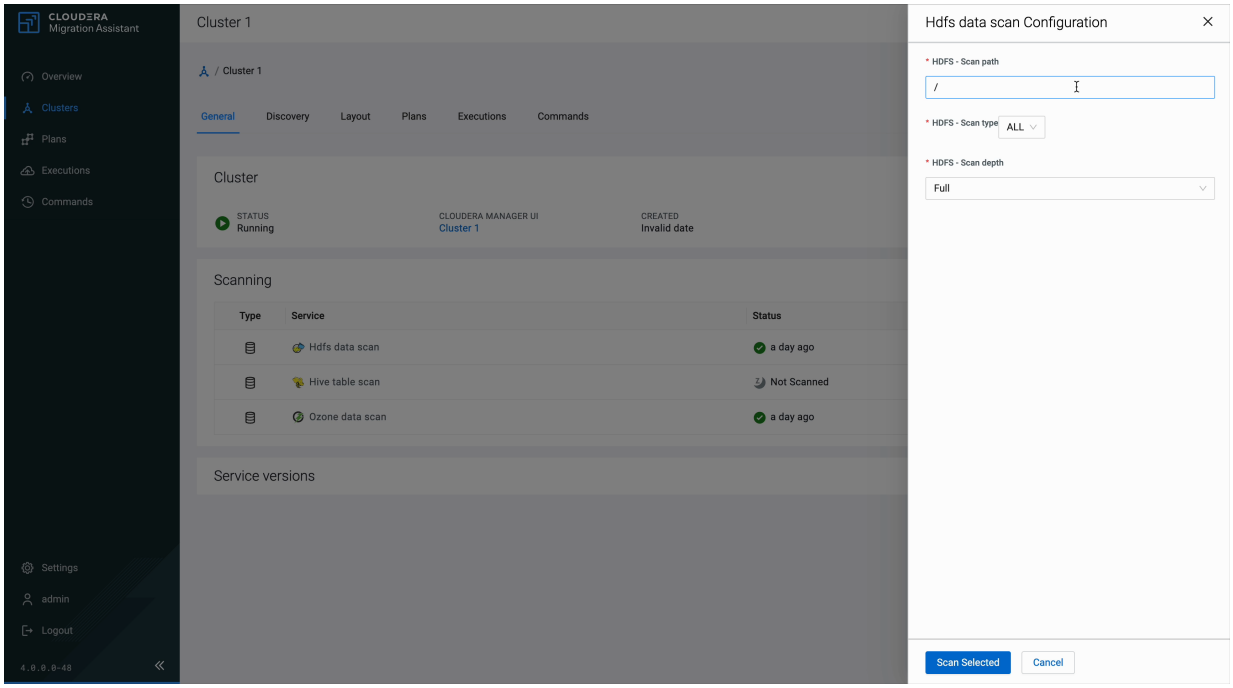
2. On the cluster detail page, find the Scanning section, which lists the available scan types for your cluster. Click the configuration icon on the right side of the scan type you want to trigger.

Figure 56: cluster scanning



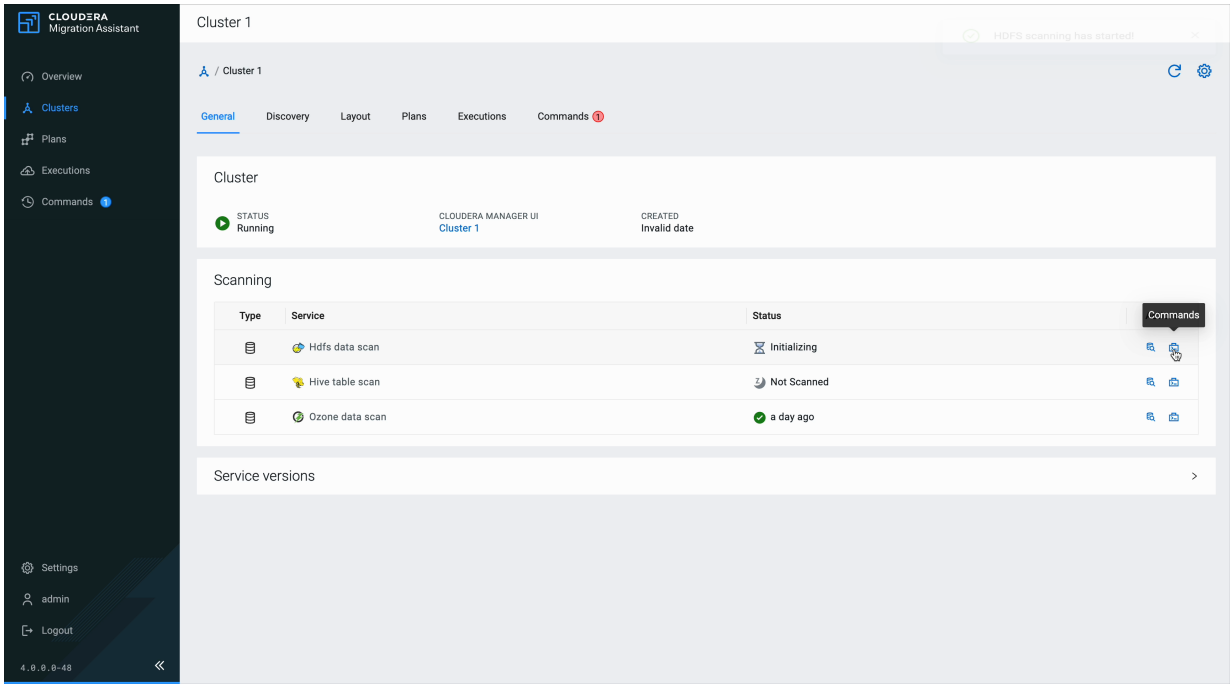
3. In the scan configuration drawer, configure the scan parameters (for example, scan path, scan type, scan depth) and click Scan Selected.

Figure 57: scan configuration



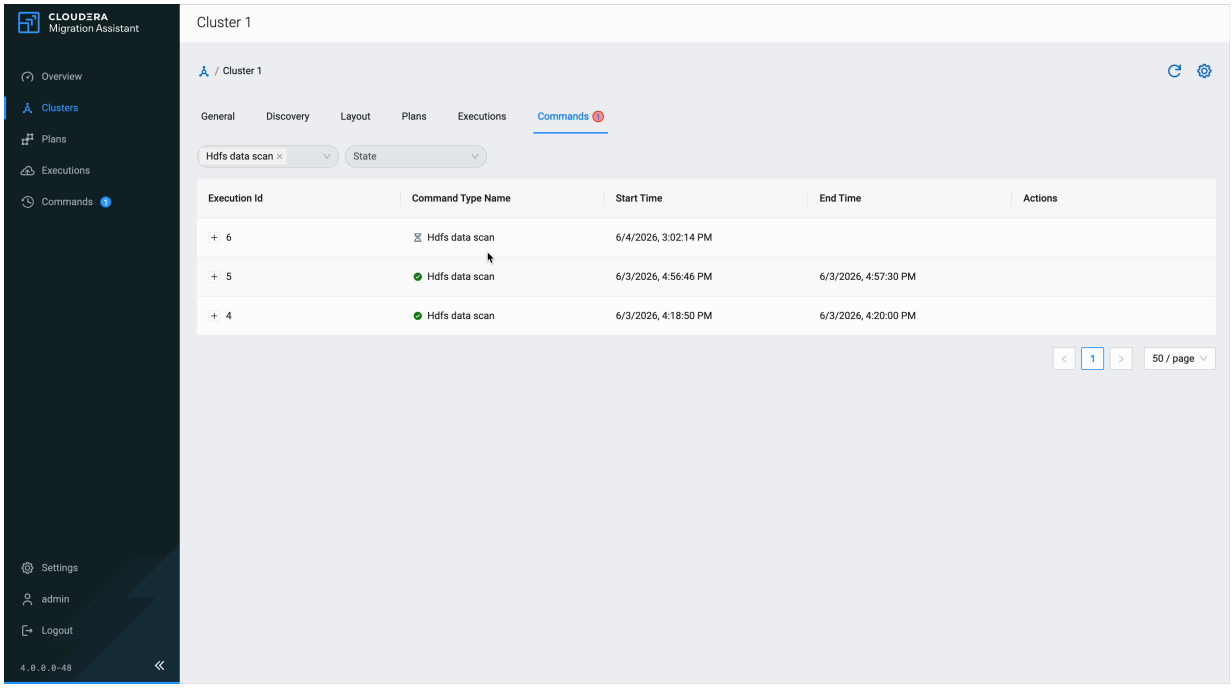
4. The scan starts and the status changes to Initializing. Track scan progress by clicking the Commands tab, which shows a badge with the number of active commands.

Figure 58: scan in progress



5. On the Commands tab, expand a command to view its Output Log, Error Log, and Configuration.

Figure 59: scan commands



Note: Scan results are also available in the data directory on the source cluster at <CMA_ROOT_DIR>/data/sources/ClusterName/Source_ID/cluster-scan/ on the host.

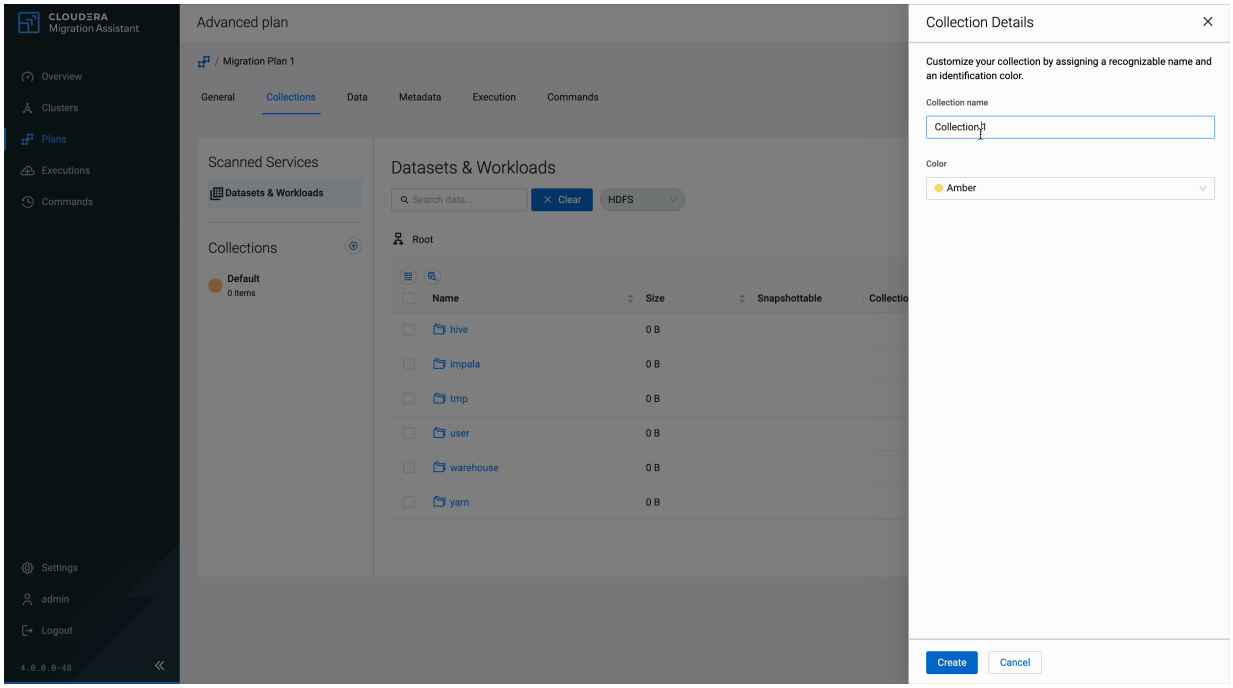
Procedure

Phase 5 — Create collections (optional)

Procedure

1. Click the + icon next to Collections in the sidebar.
2. In the **Collection Details** drawer, enter a Collection name and select a Color for identification.

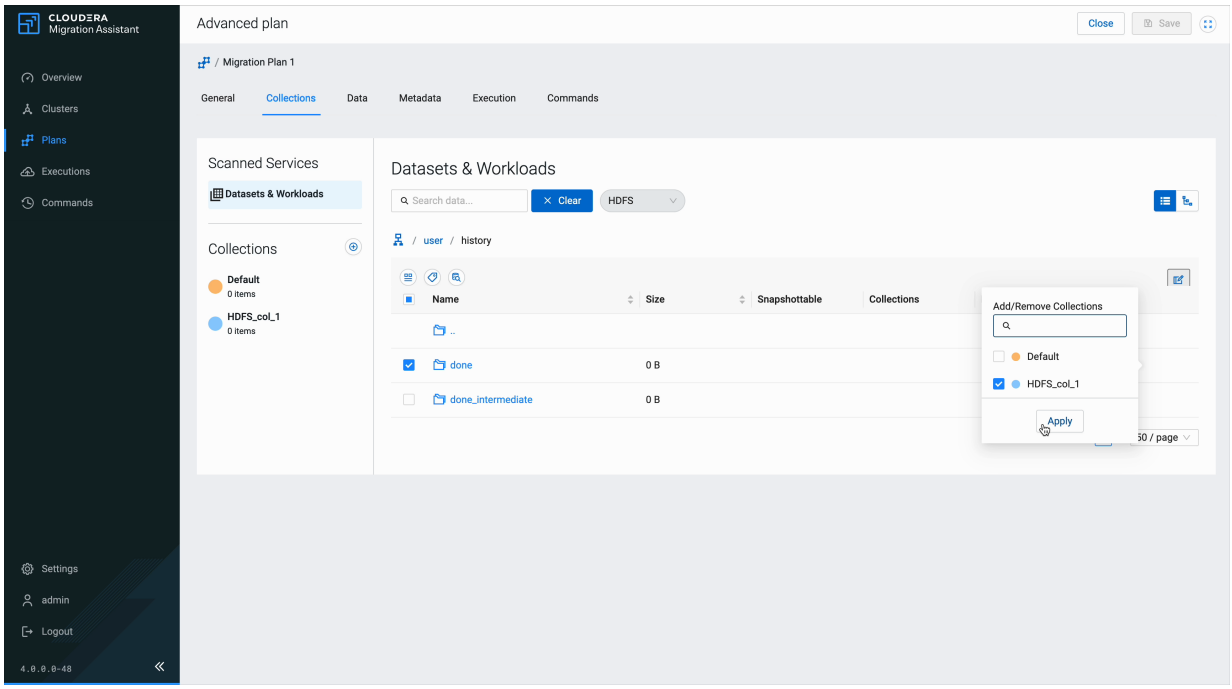
Figure 60: create collection



3. Click Create.

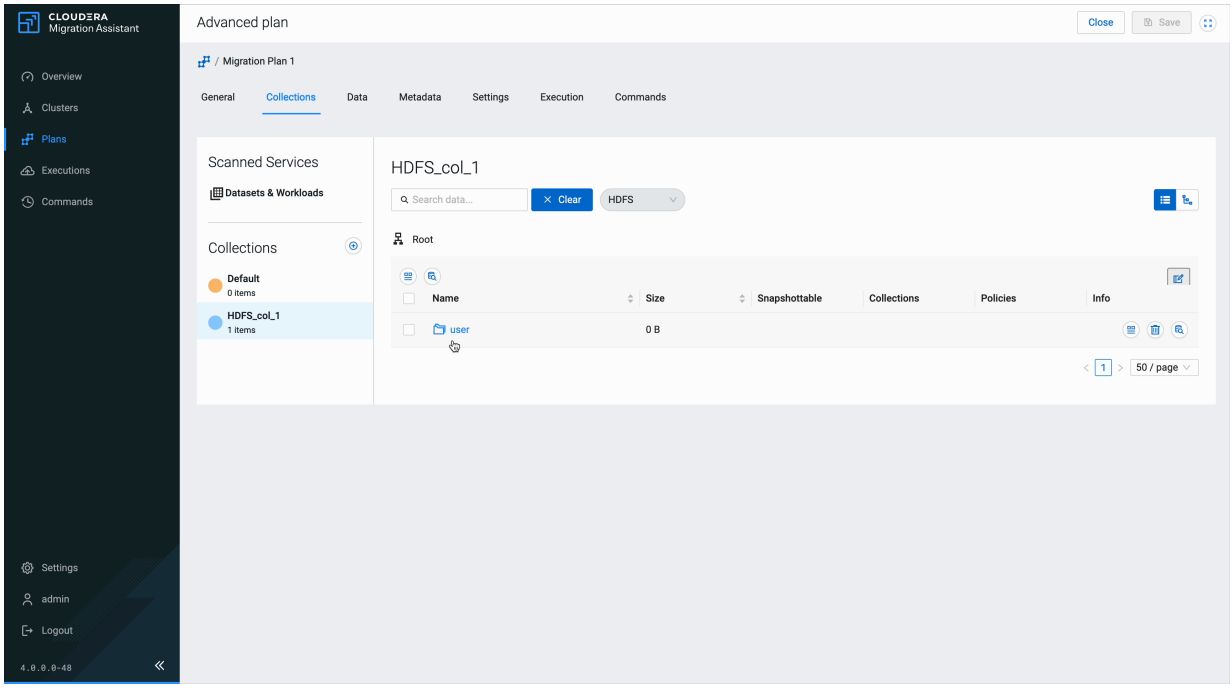
   The new collection appears in the sidebar with 0 items.
4. Browse the data tree to find the items you want to migrate. Click folder names to navigate deeper into the hierarchy.
5. Select items using the checkboxes next to each item.
6. Click the collection icon that appears on the selected item row.
7. In the **Add/Remove Collections** popup, check the collections you want to assign the item to.
8. Click Apply.

Figure 61: assign to collection



9. Click a collection name in the sidebar to view its contents. The data browser switches to show only the items assigned to that collection.

Figure 62: view collection



Procedure

Phase 6 — Run the migration workload

Procedure

Step 1 — Scan Ranger policies.

1. On the **Clusters** page, select the source cluster.
2. Click Start Scanning to display **Scan Settings**.
3. Select Ranger policy scan.
4. Select Scan selected and wait for Ranger services to finish synchronizing.

Note: The scan enumerates Ranger services, streams policies back to , and captures Ranger users, groups, and roles for reference.

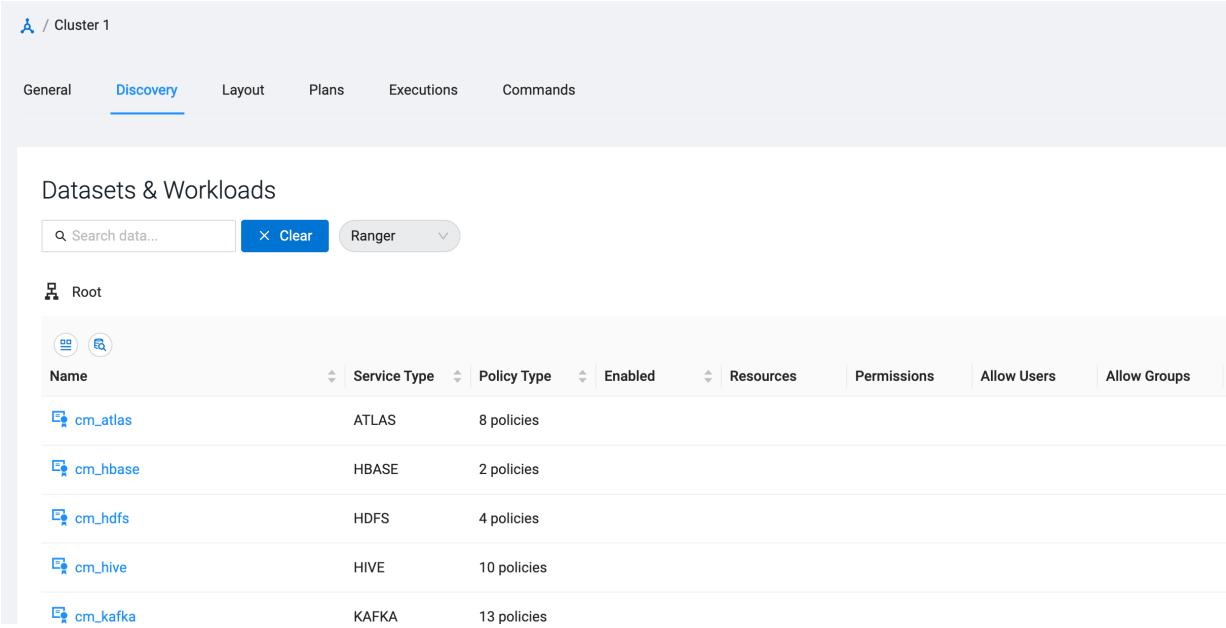
Step 2 — Browse Ranger policies in Discovery.

Finished scans displays policies inside Discovery.

5. Go to the cluster's Discovery tab.
6. Select Ranger Policies in the navigator.

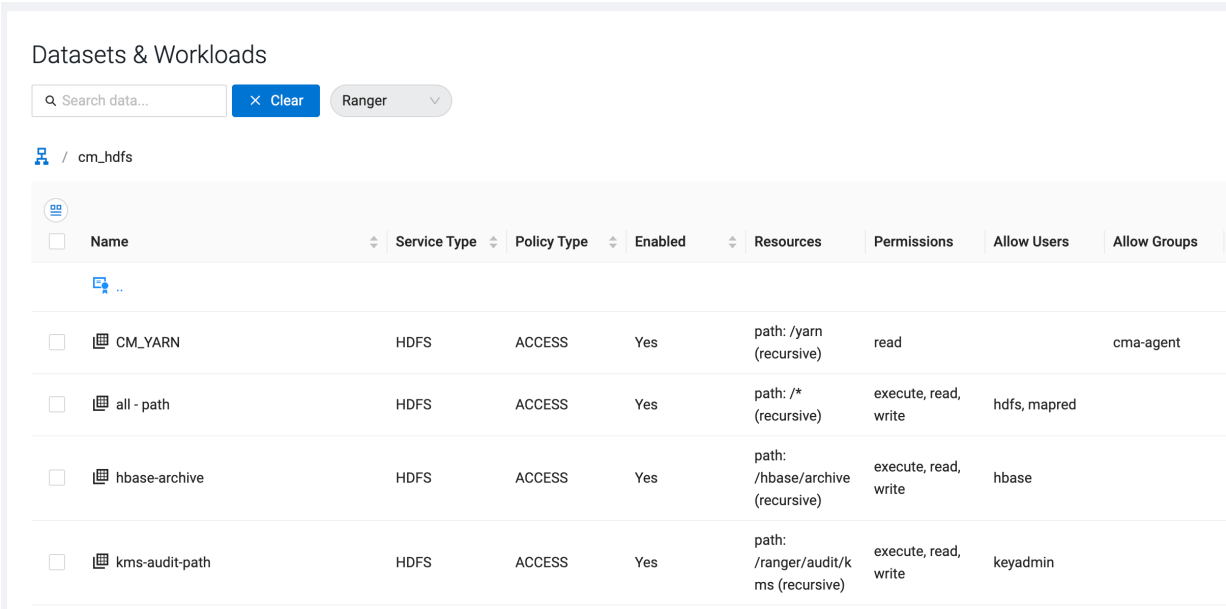
- **Root** — Lists discovered Ranger services.
- **Service** — Displays service metadata and policy totals.
- **Policy** — Renders individual rows in the policy table.

Figure 63: Ranger Discovery tree



7. Expand tree nodes for your Ranger service to view their associated policies.

Figure 64: Policies for one Ranger service



8. Review the default columns in the policy grid to evaluate your policy configurations.

Table 14: Default policy grid columns

Column	Description
Name	The Ranger-defined policy title.
Service Type	The Ranger service type, for example, HDFS, Hive, or HBase.
Policy Type	The operational category of the policy, which is access, masking, or row filtering.
Enabled	The activation state of the policy that is Yes or No.

Column	Description
Resources	The matched paths or tables, including recursive and exclusion directives.
Permissions	The granted privileges.
Allow Users	The users who are granted access by the policy rules.
Allow Groups	The user groups who are granted access by the policy rules.
Collections	The applied collection badges.

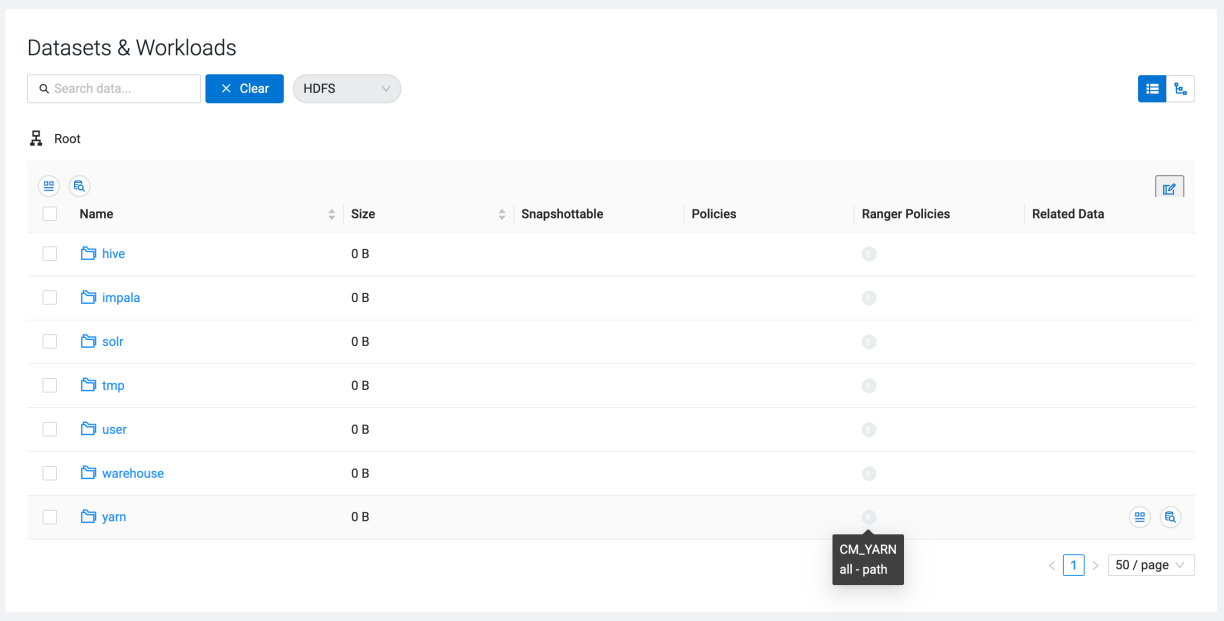
To view additional Ranger metadata, such as policy ID, deny rules, delegate admin, Ranger labels, add the corresponding columns to the grid using the column picker.

Step 3 — Review Ranger policy indicators.

Completed scans populate indicators on filesystem and relational assets.

9. Go to HDFS Locations or Hive Tables within Discovery.
10. Inspect the Ranger Policies column for the Ranger badge.
11. Open the badge to reveal matching policy titles.

Figure 65: Ranger policy indicator on HDFS



12. Understand matching logic when reconciling badges with policies.
    - **Exact match:** The policy resource matches the HDFS path or Hive table name verbatim.
    - **Recursive match:** The policy automatically includes all descendant files and subdirectories .
    - **Wildcard match:** The policy evaluates Ranger patterns, such as *, or ? against the resource path.

Step 4 — Organize policies into collections.

13. In Ranger Policies, select the policies to group.
14. Click Add to collection for the Default label.
15. Click +, enter a unique label and color, then click Create.
16. Apply collection filters whenever you narrow remediation workstreams.