dremio

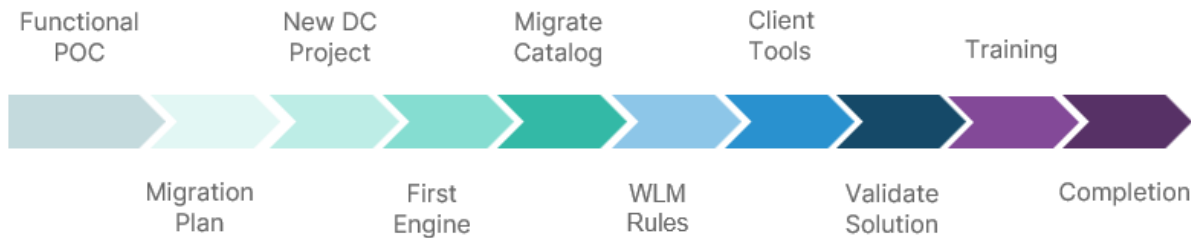# Migration Strategy from Dremio Software to Dremio Cloud

Introduction

The purpose of this document is to present Dremio's high-level strategy for moving from Dremio Software to Dremio Cloud as a timeline of events and activities.

Users can complete the required steps themselves or with assistance from Dremio Professional Services if they are active Dremio Enterprise Edition customers.

High-level Timeline of Activities

The following diagram depicts the activities needed to complete such a migration. Each of these activities is discussed in more depth in subsequent sections.

Functional POC — New DC Project — Migrate Catalog — Client Tools — Training

Migration Plan — First Engine — WLM Rules — Validate Solution — Completion

Activities

The following sections detail the activities that need to be performed to complete the migration.

Functional Proof of Concept

Before committing to moving from Dremio Software to Dremio Cloud the customer should perform a lightweight proof of concept to demonstrate that Dremio Cloud will be fit for purpose and that it will support their needs from a data source connectivity and client connectivity standpoint.

Migration Plan

Since every customer's requirements and use cases are unique, Dremio recommends hosting a working session with all relevant customer stakeholders and technical staff to understand and document the exact migration requirements.

The benefit of this session is that it helps to understand what the customer believes constitutes a successful migration and thus will prepare them with all the information they need to plan and successfully deliver the migration engagement, with the target of getting into production with Dremio Cloud as soon as possible according to their timelines. It also enables the customer to determine an estimate for the level of effort they will need to complete the migration exercise in full.

Based on the outcomes from the working session, a migration plan should be created that contains the full set of tasks that are required to complete the migration to Dremio Cloud. This migration plan should also list the success criteria that are being worked towards so that we know explicitly when the project is done. An example of the tasks one might put into the migration plan is referenced in Appendix A of this document. The migration plan may also be supplemented by a prerequisite tasks checklist, things that should be completed and checked

before beginning the migration. An example of a set of prerequisite tasks is referenced in Appendix B of this document.

Customers should review the migration plan and success criteria collectively as a team, make any comments or adjustments to the plan as necessary and then agree sign-off on the plan with the Economic Buyer in writing via email immediately after.

Create a New Dremio Cloud Project

It is recommended that a new project be created in the same Organization that was used in the earlier functional proof of concept, thus ensuring customers are starting with a clean and empty Dremio Cloud project. Please refer to this page for details on managing and adding projects.

Create First Engine

Create at least one new engine in the Dremio Cloud project, this engine will be responsible for executing queries submitted by users; later customers will create workload management rules that will route queries to specific engines based on their routing requirements. To create an engine, please refer to the guidelines on this page where there is an Adding Engines section.

Migrate Catalog Objects

The bulk of the work will be in leveraging Dremio REST APIs to export catalog objects out of the Dremio Software instance and importing them into the Dremio Cloud instance.

Catalog objects that will typically be migrated are sources (though it is recommended to recreate these manually in Dremio Cloud in case there are issues with connectivity upon initial creation), spaces, folders, tables, views, reflections, privileges, wikis and labels. Note that in Dremio Cloud the top-level artifact in your resource tree is no longer a space, instead it is an Arctic catalog, which will add an extra level of depth to the hierarchy of your catalog objects. It is anticipated in Dremio Cloud that a catalog will be an umbrella for what was a set of spaces in Dremio Software, so where in Dremio Software you may have had a resource whose fully qualified name was myspace.myfolder.myview, in Dremio Cloud the fully qualified name of that same resource will be mycatalog.myspace.myfolder.myview. This will need to be taken into consideration during the migration since every object being migrated will need to add a reference to the name of the catalog in its fully qualified name.

Dremio Professional Services has created tooling to assist in the migration of catalog resources and the full steps are documented in this companion document. Dremio recommends following the steps in the document to complete the catalog migration.

Set Workload Management Rules to Direct Queries to Engines

With the catalog items migrated customers need to set some Workload Management rules in Dremio Cloud to ensure when their queries start running that they are routed to the correct engines. Workload Management in Dremio Cloud is somewhat different from Workload Management in Dremio Software, it has been simplified to remove the notion of a queue and instead, rules are set to direct queries to specific engines. In addition, the routing by query cost has been removed so instead queries must be routed by constructs like user, role and job type. Please refer to this page to understand how to set Workload Management rules in Dremio Cloud.

Connect Client Tools

Whichever client tools are needed to consume the data in the migrated use cases should be connected to the new Dremio Cloud project, this will mean altering connection details in dashboards or scripts to point to the new Dremio Cloud project. Details of how to connect a broad range of client applications to Dremio Cloud can be found here.

If there are doubts about whether any of the client tools currently connecting to Dremio Software will continue to work with Dremio Cloud then the connectivity of these should be proven during the functional proof of concept phase.

Validate Functionality and Optimize

With an end-to-end solution in place, dashboards can run and scripts can be executed to test that the data requests and data retrieval processes are working and feeding data to the client tools.

Reflections will have been migrated automatically if that option was chosen during the Migrate Catalog Objects phase, so now is the time to confirm that reflections are refreshing and that the refresh schedules for the reflections are set accurately. If reflections were not migrated and as a result, queries are running sub-optimally, now is also the time to introduce reflections strategically using Dremio's reflection best practices.

The same goes for metadata refresh settings, these will have been set in Dremio Software so please confirm they are set as expected and equivalently in Dremio Cloud. Customers can leverage Dremio's metadata refresh best practices for guidance on setting these appropriately if they are unfamiliar with them.

Training

Although customers may be very familiar with Dremio Software and a lot of the "look and feel" customers are used to remains consistent in Dremio Cloud, there are certainly features and functionality that differ between the two, for example, Arctic catalogs in Dremio Cloud have replaced the use of Spaces that you are used to in Dremio Software, therefore it is recommended at a bare minimum that every user of Dremio Cloud gets familiar with these differences by taking the following dedicated Dremio Cloud training modules that Dremio offers:

- Quick Start

In addition, there are 6 further training modules for anyone who wants a deeper dive into the full capabilities of Dremio Cloud:

- Sources & Tables
- Clients & Views
- Semantic Layers
- Running Queries
- Reflections
- Tuning Queries

Please speak with your Dremio Account Executive to arrange Dremio Cloud training.

Complete Remaining Migration Plan Tasks

Outside of the activities already discussed in the sections above, there may still be tasks captured in the migration plan that need completing. Ensure all remaining tasks are completed and that all success criteria are met before the project is closed. These additional tasks may include, but are not limited to, the following items:

- Additional Client Tool integration
- Identity Provider integration
- Role-Based Access Control implementation
- Further query optimizations
- Query monitoring

Appendix A - Sample Migration Plan Tasks

The following tasks constitute a typical Dremio Cloud migration project.

- Capture baseline query performance in the Software environment
- Perform [prerequisite tasks](#) as detailed by Dremio
- Sign up for Dremio Cloud and update to Enterprise Edition
- Dremio Cloud Project setup
- Setup Internal Users/Roles to ensure we have a local admin
- Initial Engine and Workload Management configuration
- Integrate with an Identity Provider to enable secure logins and Single Sign-On
- Setup required data source connections
- Migrate catalog resources
- Design and implement Role-Based Access Control
- BI tool integration and performance optimizations
- TLS client integration between Dremio Cloud and client tools
- Set up Job Analyzer to enable monitoring of query performance
- Capture the performance of equivalent queries in Dremio Cloud as were captured in the baseline and compare. If the queries are not performing well enough, you may need to perform these additional tasks:
  - Use case query review and optimization
  - Semantic Layer design review and optimization
  - Reflection refresh process optimization
  - Metadata refresh process optimization
- Update Engine and Workload Management configuration if necessary

Appendix B - Prerequisites for Dremio Cloud Migrations

The following prerequisites should be considered and delivered before embarking on a migration from Dremio Software to Dremio Cloud.

- Deploy - Decide on a main contact and platform administrator for your Dremio Cloud
- Deploy - Create an Azure Subscription for the executor nodes
- Deploy - Firewall openings to access Dremio
- Deploy - Create a storage account to be used as Distributed storage
- Deploy - Decide on your Authentication configuration - AAD / OAuth2
- Use case - If you will have Development, Test and Production projects, decide on your CI / CD environment for the development and migration of resources
- Use case - Ensure you open up the necessary connectivity to data lakes and sources
- Use case - Provide samples or copies of actual dashboards that currently run in Dremio Software so that the credentials can be updated to point to the relevant Dremio Cloud project
- Use case - For testing post-migration, ensure the business and use case owners are available