dremio

Dremio Software

Backup and Restore

Introduction

Dremio maintains a RocksDB database that includes information about the semantic layer developed in Dremio and usage information.

- Data Sources defined in Dremio
- Information about the Physical datasets in the defied sources
- Userspaces and folders, sub folders
- Virtual Datasets
- Users and Roles
- Access Privileges on datasources, physical and virtual datasets
- Reflections
- Materializations of reflections
- Job history
- Query Profiles etc

All this is crucial for the functioning Dremio deployment, as such we should back it up regularly as a fallback in case of metadata corruption or in case we need to bring up the same Dremio semantic layer in a new environment.

This whitepaper discusses the different options for backup and restore of the backup.

Backup Location Storage

Dremio recommends the backup location to be a mounted filesystem on the master coordinator node that will persist even if the physical server hosting the Dremio master coordinator crashes. This filesystem could additionally be copied into a different data center to protect against Disaster scenarios.

If the Dremio instance is hosted on VMs or pods in a cloud environment, we could backup the Dremio metadata to a data lake location, for example, on an S3, GCS, or AzureStorage.

Backup Frequency

File System Backup of Metadata DB should be done before and after a Dremio upgrade.

Dremio utility dremio-admin backup can be used to restore only until the point of time when the backup is taken. From that perspective, it is prudent to

- Schedule backup at least once a day.
- Depending on your recovery point objective, you could backup more frequently say (every 12 or every 6 hours or 4 hours); however, ensure the previous backup is completed before kicking off a new one
- Take a backup before and after every planned Dremio upgrade
- Perform a backup prior to any CI/CD deployment on any Dremio instance

File System Backup and Restore

File System Backup

The following requirements must be taken into account before using the file system backup.

- Dremio Service has to be down
- Locate the file system where Dremio metadata resides

The following steps are used to backup the file system

- Use Operating System utilities to backup the Dremio metadata directory above
- (example: cp -r /opt/dremio/data/db /backupdir/db_20230615_0203am)

File System Restore

The following requirements must be taken into account before restoring a backup.

- Dremio Instance has to be of the same version as when Backup was done.
- Dremio Service has to be down
- Locate the file system where Dremio metadata resides

The following steps are used to restore a backup from the file system

- Restore the contents of the db folder from the previous backup file
- You can reliably restore only if the Dremio service was down at the time of taking the backup.
- (example: cp -r /backupdir/db_20230615_0203am /opt/dremio/data/db)

Dremio Admin utility Backup and Restore

Dremio Admin Backup

The following should be considered when using the dremio-admin backup utility.

- The Dremio Service has to be up when dremio-admin backup is being performed.
- The target Dremio Instance should have access to the data sources that were configured in the source Dremio Instance

Dremio Admin Restore

The following should be considered when using the dremio-admin restore utility.

- Dremio service has to be down when dremio-admin restore is being performed.
- The Dremio instance must be the same version as the backed-up file.
- The metadata folder should be empty(.e.g /opt/dremio/data/db)
- The target Dremio instance should have access to the data sources that were configured in the source Dremio instance