

How to migrate metadata to Kylin 4

Background

Apache Kylin have provided migration tool to support migrating metadata across different clusters since version 2.0. Recently, we have refined and added new ability to CubeMigration tool, The list of enhanced functions is showed as below:

1. Support migrating all cubes in source cluster
2. Support migrating a whole project in source cluster
3. Support migrating and upgrading metadata from older version to Kylin 4

This feature will be released in the Apache Kylin-4.0.0 official version.

How to use

'CubeMigrationCLI.java' CLI used positional arguments before, and is now changed to POSIX Style arguments, look at the following example:

```
./bin/kylin.sh org.apache.kylin.tool.CubeMigrationCLI -srcConfig ADMIN:KYLIN@host1:7070 -dstConfig ADMIN:KYLIN@host2:7070 -overwriteIfExists -srcProject default
```

All supported parameters are listed below:

Option	Has Parameter	Description
-srcConfig	true	[Required] The KylinConfig of the source cluster with 3 kinds of format: 1. uri e.g. user:pwd@host:port. 2. properties file e.g. /home/kylin/conf/kylin.properties 3. local metadata dir e.g. /home/kylin/meta_backups/meta_2021_04_21_18_29_04
-dstConfig	true	[Required] The KylinConfig of the target cluster with 3 kinds of format: 1. uri e.g. user:pwd@host:port. 2. properties file e.g. /home/kylin/conf/kylin.properties 3. local metadata dir e.g. /home/kylin/meta_backups/meta_2021_04_21_18_29_04
-allCubes	false	[Optional] migrate all cubes in source cluster.
-cube	true	[Optional] migrate single cube in source cluster, argument value is the cube name to migrate.
-srcProject	true	[Optional] migrate all cubes under specific project in source cluster, argument value is the project name to migrate.
-dstProject	true	[Optional] cube's new project home(will be created if not exists). If not set, keep the same as cube's original project. With this option, we can migrate a "single cube", "a project" or "all cubes" to a target project.
-migrateSegment	false	[Optional] migrate segment data
-overwriteIfExists	false	[Optional] overwrite target cluster's metadata if exists
-purgeAndDisable	false	[Optional] purge source cluster data
-realMigrate	true	[Optional] execute migration, parameter value: 1. true: execute migration 2. false: not execute migration
-copyAcl	false	[Optional] copy ACL

Please note:

1. The URI format of KylinConfig is not backward compatible, for it fetches metadata storage information by http request, older version may not be supported.
2. only one of "-cube", "-srcProject", "-allCubes" could be specified.
3. "-migrateSegment" will not be able to migrate segment data from older version to Kylin 4, because storage type is switched from HBase to Parquet.
4. CubeMigrationCli will reload target cluster's metadata, we don't have to restart target cluster.

Migrate metadata from HBase to Kylin 4

HBase will be removed after KYLIN 4.0. If you are migrating metadata stored in HBase to 4.x cluster, you are required to do an additional step:

1. Export metadata from older cluster to local dir using scripts in old software package:

```
./bin/metastore.sh backup
```

2. Use 'CubeMigrationCLI.java' CLI provided in 4.0 package to set source KylinConfig to metadata directory exported in step 1

```
./bin/kylin.sh org.apache.kylin.tool.CubeMigrationCLI -srcConfig ${exported_metadata_dir} -dstConfig USERNAME:PASSWORD@HostName:Port -allCubes ...
```

'CubeMigrationCLI.java' CLI will read exported metadata and change cube's information, including 'storage_type', 'query_engine', 'version' and 'cube signature', and then write into target metadata storage.