GridReplicate - Qlik Replicate

Architecture, Standards, and Guidelines

Version: 1.0

Last Updated: June 27, 2022

Author: Nishit Ajwaliya

Contents

[1. Qlik Replicate Logical Architecture 3](#_Toc107210783)

[1.1 Qlik Replicate Server (QRS) 3](#_Toc107210784)

[1.2 Qlik Enterprise Manager (QEM) 3](#_Toc107210785)

[2. Qlik Replicate Physical Architecture 5](#_Toc107210786)

[2.1 Non-Prod (DEV and TEST) 5](#_Toc107210787)

[2.2 Prod 6](#_Toc107210788)

[2.3 Qlik URLs 6](#_Toc107210789)

[2.3.1 Qlik Enterprise Manager 6](#_Toc107210790)

[2.3.2 Qlik Replicate Server 6](#_Toc107210791)

[3. Access Request 7](#_Toc107210792)

[4. Source Data Captured 7](#_Toc107210793)

[4.1 DB2z 7](#_Toc107210794)

[5. Endpoint Connections 8](#_Toc107210795)

[5.1 Naming Standard 8](#_Toc107210796)

[5.2 Source Endpoints 8](#_Toc107210797)

[5.2.1 DB2z 8](#_Toc107210798)

[5.3 Target Endpoints 9](#_Toc107210799)

[5.3.1 Snowflake 9](#_Toc107210800)

[6. Data Load/Sync Task 11](#_Toc107210801)

[6.1 Naming Standard 11](#_Toc107210802)

[6.2 Task Creation 11](#_Toc107210803)

[6.3 Audit Columns 12](#_Toc107210804)

[6.4 Target Schema 14](#_Toc107210805)

[6.5 Control Tables 15](#_Toc107210806)

[6.6 HEX Conversion (requires for CRIS only) 16](#_Toc107210807)

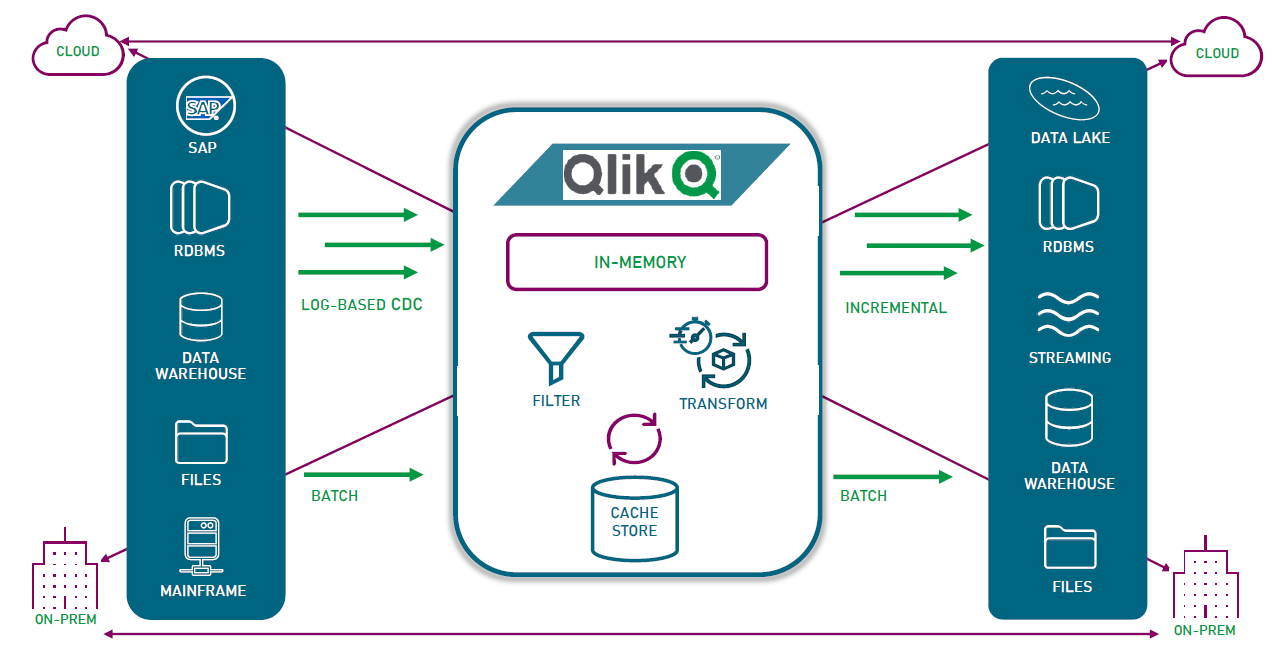
[Appendix 17](#_Toc107210808)

# Qlik Replicate Logical Architecture

## Qlik Replicate Server (QRS)

Qlik Replicate is a modern data replication tool and platform that provides data replication, ingestion and streaming across a wide variety of homogenous and heterogeneous databases, data warehouses, and big data platforms.

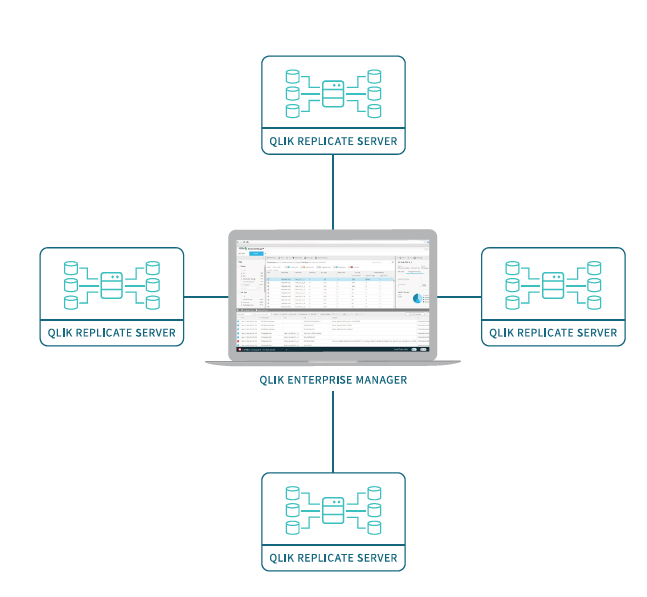
Qlik Replicate (formerly Attunity Replicate) empowers a high-performance data replication and ingestion platform that replicates, synchronizes, distributes, consolidates and ingests data across all major databases, data warehouses and Hadoop, on-premises and in the cloud. With a unique zero-footprint architecture that facilitates zero-downtime data migrations and database upgrades, Qlik Replicate moves your data easily, securely and efficiently with minimal operational impact.



Note: Currently Qlik Replicate is licensed to use for Source - CRIS and CSS DB2z and target -Snowflake. Please reach out to Mcauley, Sean (Thomas) [Thomas.Mcauley@nationalgrid.com](mailto:Thomas.Mcauley@nationalgrid.com) if there is a need/requirement/use case to support additional source and/or target databases.

## Qlik Enterprise Manager (QEM)

Qlik Enterprise Manager (formerly Attunity Enterprise Manager) simplifies the entire process. It’s a centralized platform with an intuitive graphical interface for efficiently managing data replication processes. It uses to design, execute, monitor, and analyze Qlik Replicate tasks across large and growing environments. It helps to improve the management of high-scale data consolidation efforts across dozens of Qlik Replicate servers and hundreds of end points. It helps to meet, or even exceed, business service-level agreements (SLAs). The solution also simplifies enterprise-wide management tasks by seamlessly integrating with larger IT systems and practices — a boost to your security and compliance.



# Qlik Replicate Physical Architecture

## Non-Prod (DEV and TEST)



## Prod



## Qlik URLs

## Qlik Enterprise Manager

DEV/TEST –

<https://qlikem-nonprod.nationalgrid.com/attunityenterprisemanager/> or <https://azuse-qlikpp02.preprod.nationalgrid.com/attunityenterprisemanager>

PROD –

tbd

## Qlik Replicate Server

DEV -

<https://azuse-qlikpp01.preprod.nationalgrid.com/attunityreplicate/>

TEST –

<https://azuse-qlikpp03.preprod.nationalgrid.com/attunityreplicate>

PROD –

tbd

# Access Request

Please send email to Chen, Howard [Howard.Chen@nationalgrid.com](mailto:Howard.Chen@nationalgrid.com) to request an access with following information:

Environments – DEV/TEST/PROD

Name and Network ID – <Your Name> and <Your Network ID that you use to login into NG >

Role – Configurator or Viewer

Note: Only very limited people will have access to Qlik CDC tool so unless you are working on configuration, please don’t make access request.

# Source Data Captured

## DB2z

Qlik Replicates reads change data information from the database log to reduce the performance impact on the database and application. It expects that table that requires the data sync has data capture turned on. Make sure that DB2z table that is in scope for CDC data sync, has DATA CAPTURE turned on.

With the help of DB2 DBA, we should be able to validate if data capture is turned on by using following query:

*SELECT*

*SUBSTR(CREATOR,1,8) AS SCHEMA*

*,SUBSTR(NAME,1,20) AS TAB\_NAME*

*,DATACAPTURE*

*FROM SYSIBM.SYSTABLES*

*WHERE CREATOR in ('CRS','@CTBD','CRS2'*

*ORDER BY 3 DESC,1,2*

*WITH UR;*

Note: Make sure appropriate schemas are updated in highlighted section of the above query.

If the value for *DATACAPTURE is ‘Y’* means data capture is ON for the table. Blank value means data capture is OFF for the table.

# Endpoint Connections

## Naming Standard

<Environment>\_<Database>\_<Application>

Where environments can be DEV or TEST or PROD

Eg

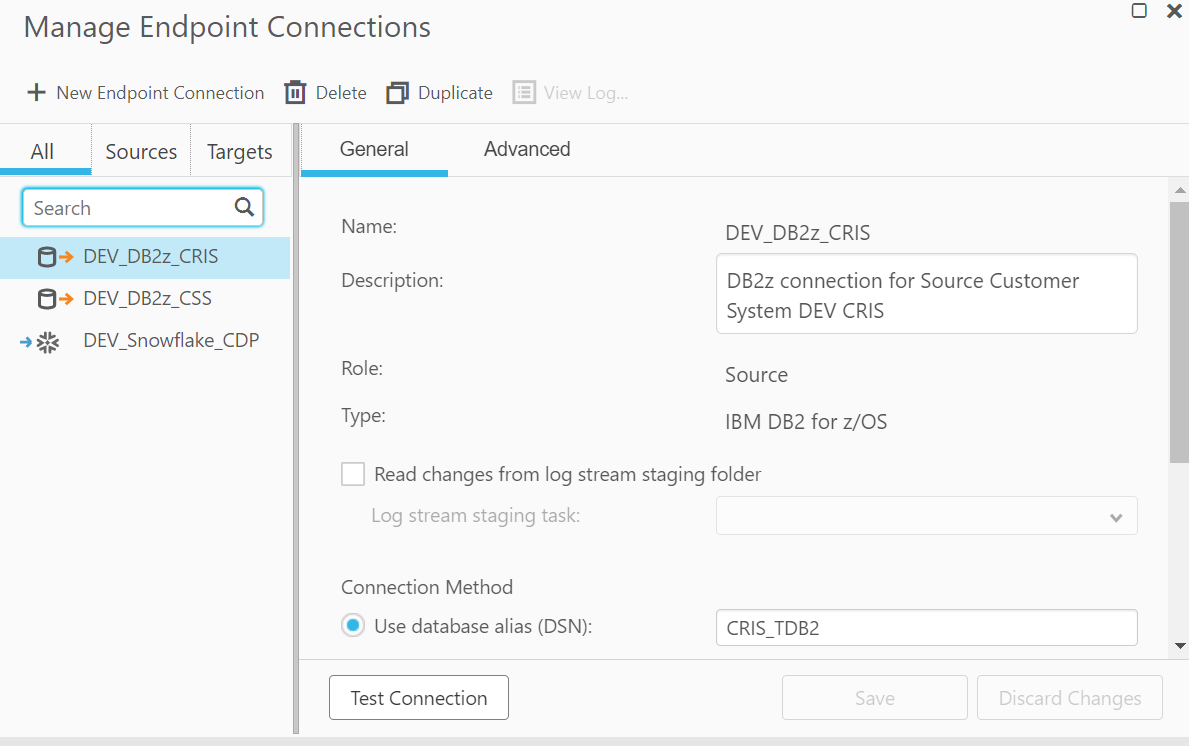
DEV CRIS Connection will be named - DEV\_DB2z\_CRIS

Snowflake TEST connection for CDP will be named – TEST\_Snowflake\_CDP

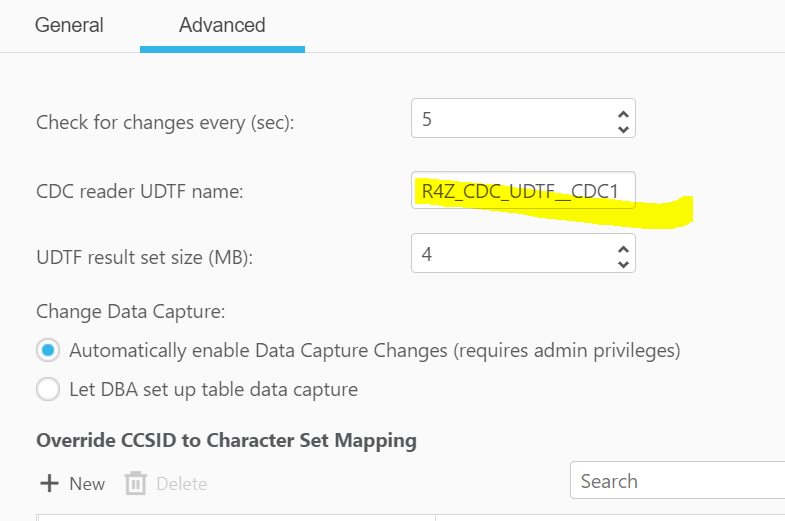
## Source Endpoints

### DB2z

There are 2 source endpoint connections configured one for each mainframe source systems ICRIS and CSS) as shown below:



* “CDC Reader UDTF” is configured on the Advanced tab with appropriate name. This UDTF name is configured in Qlik R4Z Mainframe installation that allows Qlik Replicate to read change data from DB2z log file and sync to Snowflake near real time.
* Eg: DEV/TEST CSS configuration is below:



CSS Dev/Test - R4Z\_CDC\_UDTF\_\_CDC1

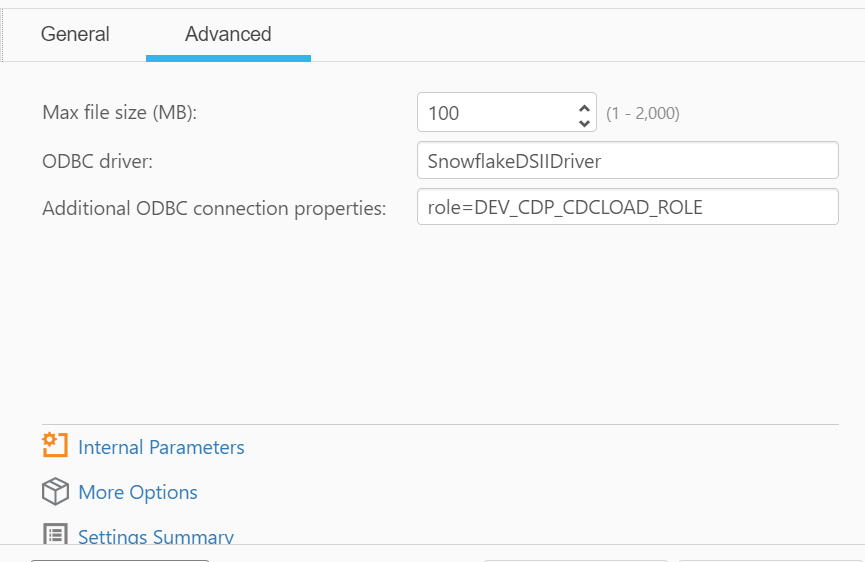
CRIS Dev (CDB2) - R4Z\_CDC\_UDTF\_\_CDC1

CRIS TEST (TDB2) - R4Z\_CDC\_UDTF\_\_CDB2

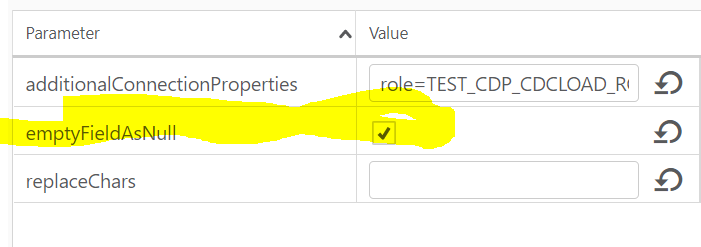
## Target Endpoints

### Snowflake

There is one target Snowflake endpoint connection configured with default Snowflake role as shown below:



For CREATE\_TIMESTAMP to work during incremental load, “emptyFieldAsNull” parameter is required to be configure as shown below in "Internal Parameters":



# Data Load/Sync Task

## Naming Standard

<Environment>\_<Application>\_<Source>\_To\_<Target1>\_<Optional Target2>

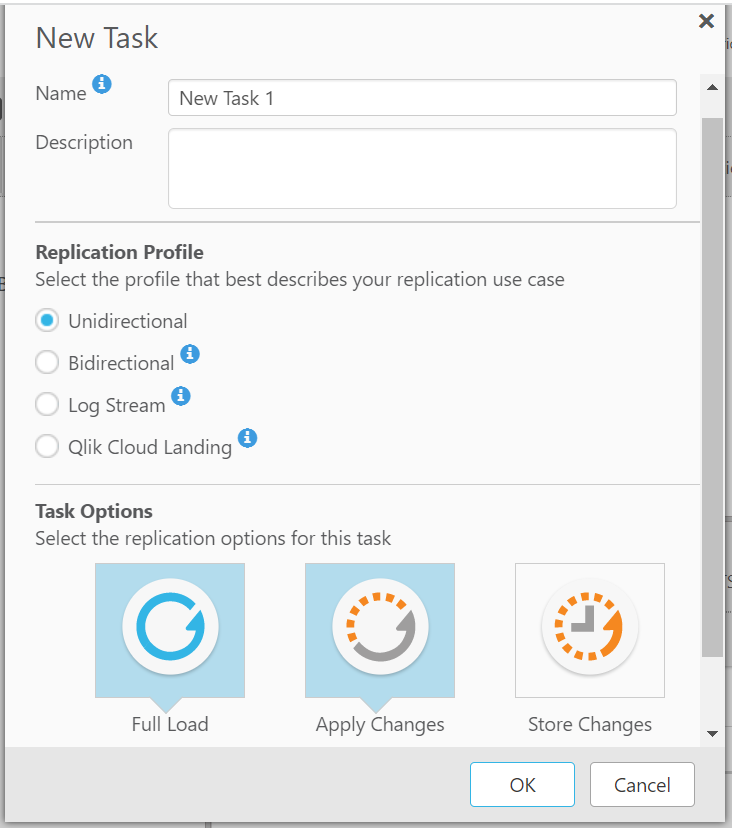
where environment can be DEV or TEST or PROD

Eg:

DEV\_CDP\_CSS\_To\_Snowflake

## Task Creation

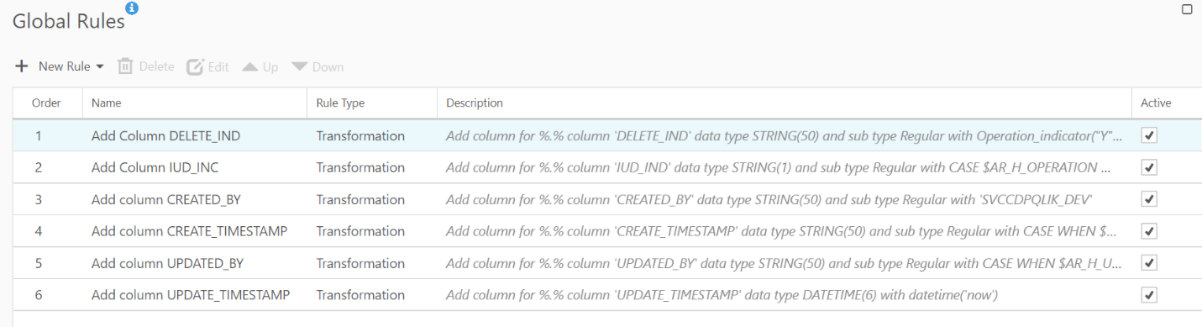
* Create new task with default configuration with both FULL Load and Apply Changes options.



* Once Task is created, Name cannot be changed so make sure to not make mistake. If mistake has made in the name, then there are 2 choices:
  1. Delete the task and create a new task with correct name
  2. Export the task, modify name, and import the task. Qlik admin help is required for importing task.
* Once task is created, drag and drop appropriate source endpoint connection and target endpoint connection. Then select all source tables that are in scope for the raw data sync to Snowflake.

## Audit Columns

* Audit columns will be handled using Global rules
* Create 6 global rules to add 6 Audit columns as shown below:



* JSON configuration for the Global Rules:

"global\_manipulation": {

"name": "global\_manipulation",

"rules": [

{

"action": "ADD\_COLUMN",

"name": "Add Column DELETE\_IND",

"expression": "'DELETE\_IND'",

"column": {

"where\_column\_name": "%",

"new\_data\_type": "kAR\_DATA\_TYPE\_STR",

"new\_length": 1,

"computation\_expression": "Operation\_indicator(\"Y\",\"N\",\"N\")",

"where\_sub\_data\_type": "KAR\_SUB\_DATA\_TYPE\_UNSPECIFIED",

"new\_sub\_data\_type": "KAR\_SUB\_DATA\_TYPE\_UNSPECIFIED"

},

"sub\_action": "RENAME\_EXPRESSION"

},

{

"action": "ADD\_COLUMN",

"name": "Add Column IUD\_INC",

"expression": "'IUD\_IND'",

"column": {

"where\_column\_name": "%",

"new\_data\_type": "kAR\_DATA\_TYPE\_STR",

"new\_length": 1,

"computation\_expression": "CASE $AR\_H\_OPERATION\nWHEN 'UPDATE' THEN 'U'\nWHEN 'INSERT' THEN 'I'\nWHEN 'DELETE' THEN 'D'\nEND",

"where\_sub\_data\_type": "KAR\_SUB\_DATA\_TYPE\_UNSPECIFIED",

"new\_sub\_data\_type": "KAR\_SUB\_DATA\_TYPE\_UNSPECIFIED"

},

"sub\_action": "RENAME\_EXPRESSION"

},

{

"action": "ADD\_COLUMN",

"name": "Add column CREATED\_BY",

"expression": "'CREATED\_BY'",

"column": {

"where\_column\_name": "%",

"new\_data\_type": "kAR\_DATA\_TYPE\_STR",

"new\_length": 50,

"computation\_expression": "'SVCCDPQLIK\_DEV'",

"where\_sub\_data\_type": "KAR\_SUB\_DATA\_TYPE\_UNSPECIFIED",

"new\_sub\_data\_type": "KAR\_SUB\_DATA\_TYPE\_UNSPECIFIED"

},

"sub\_action": "RENAME\_EXPRESSION"

},

{

"action": "ADD\_COLUMN",

"name": "Add column CREATE\_TIMESTAMP",

"expression": "'CREATE\_TIMESTAMP'",

"column": {

"where\_column\_name": "%",

"new\_data\_type": "kAR\_DATA\_TYPE\_STR",

"new\_length": 50,

"new\_precision": 0,

"new\_scale": 0,

"computation\_expression": "CASE\nWHEN $AR\_H\_OPERATION = 'INSERT' then\ndatetime('now')\nELSE\n''\nEND",

"where\_sub\_data\_type": "KAR\_SUB\_DATA\_TYPE\_UNSPECIFIED",

"new\_sub\_data\_type": "KAR\_SUB\_DATA\_TYPE\_UNSPECIFIED"

},

"sub\_action": "RENAME\_EXPRESSION"

},

{

"action": "ADD\_COLUMN",

"name": "Add column UPDATED\_BY",

"expression": "'UPDATED\_BY'",

"column": {

"where\_column\_name": "%",

"new\_data\_type": "kAR\_DATA\_TYPE\_STR",

"new\_length": 50,

"computation\_expression": "CASE \n WHEN $AR\_H\_USER != '' THEN $AR\_H\_USER \n ELSE 'SVCCDPQLIK\_DEV' \nEND",

"where\_sub\_data\_type": "KAR\_SUB\_DATA\_TYPE\_UNSPECIFIED",

"new\_sub\_data\_type": "KAR\_SUB\_DATA\_TYPE\_UNSPECIFIED"

},

"sub\_action": "RENAME\_EXPRESSION"

},

{

"action": "ADD\_COLUMN",

"name": "Add column UPDATE\_TIMESTAMP",

"expression": "'UPDATE\_TIMESTAMP'",

"column": {

"where\_column\_name": "%",

"new\_data\_type": "kAR\_DATA\_TYPE\_TIMESTAMP",

"new\_length": 0,

"new\_precision": 0,

"new\_scale": 6,

"computation\_expression": "datetime('now')",

"where\_sub\_data\_type": "KAR\_SUB\_DATA\_TYPE\_UNSPECIFIED",

"new\_sub\_data\_type": "KAR\_SUB\_DATA\_TYPE\_UNSPECIFIED"

},

"sub\_action": "RENAME\_EXPRESSION"

}

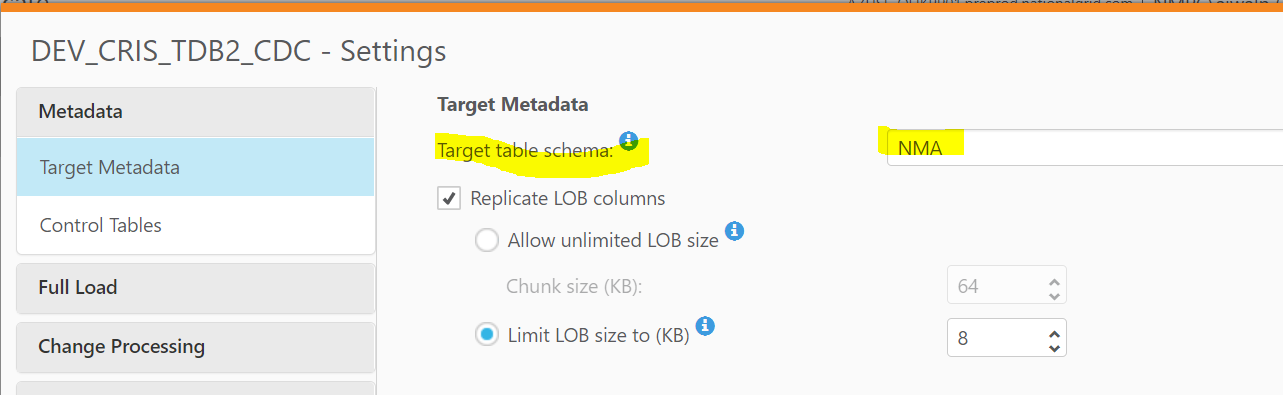
]

},

* + - DELETE\_IND is not a National Grid standard, but it is required for Qlik Replicate to track soft delete and not to physically delete Snowflake record when DB2z record is deleted. Without this column appropriate global rules, Snowflake record will be physically deleted if source deletes any record. DELETE\_IND will have 3 possible values - Y, N and NULL.
      1. Value Y means record is deleted from source.
      2. Value N means record is not deleted from Source.
      3. Value Null will be set for initial load, and it means record is not deleted from Source.
    - CREATE\_TIMESTAMP data type is VARCHAR and not TIMESTAMP. The data type must change for Qlik Replicate to appropriately keep actual CREATE DATE of the Snowflake record when source record is either deleted or updated once created in Snowflake.

## Target Schema

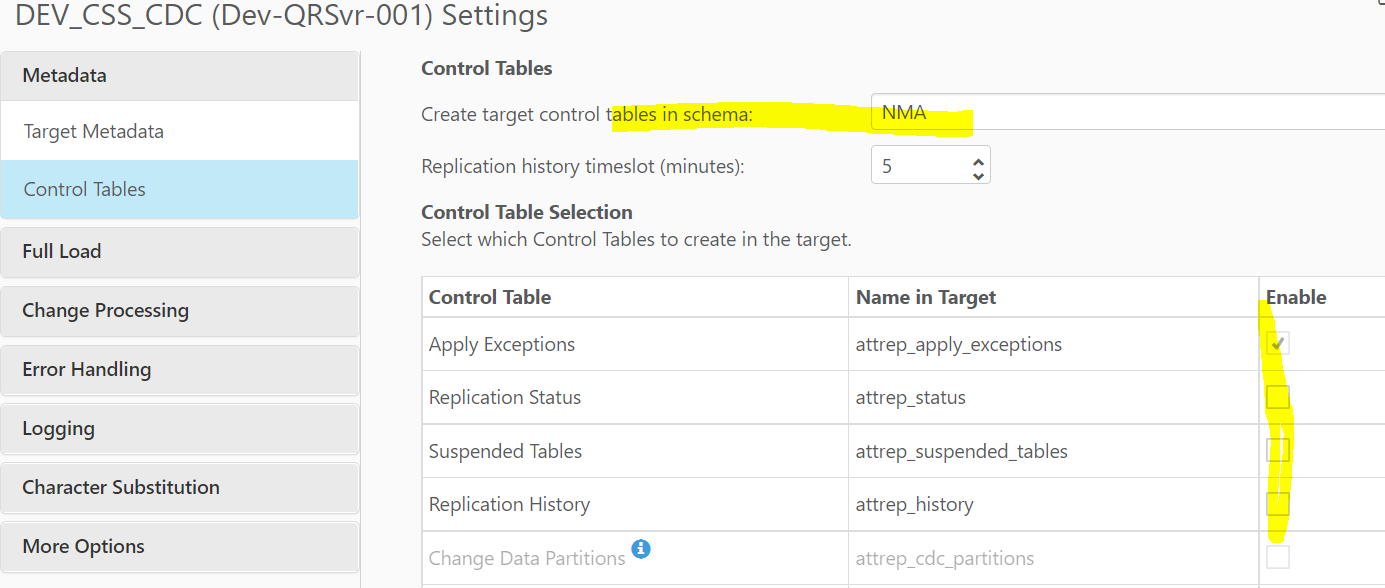
* + Qlik Replicate allows to configure separate schema name for target using the Task Metadata from Task settings as shown below:



* + If Target table Schema is not configured, then Qlik Replicate will assume Source schema as target schema
  + Please set following target database to:
    - 1. CRIS\_CDC for CRIS
      2. CSS\_CDC for CSS
  + If Snowflake does not have configured Schema, then Qlik Replicate will create it automatically when Task is run
  + If Snowflake does not have configured table, then Qlik Replicate will create it automatically when Task is run
  + If DB2z table is altered after the data sync to Snowflake, Qlik Replicate will alter the Snowflake table automatically

## Control Tables

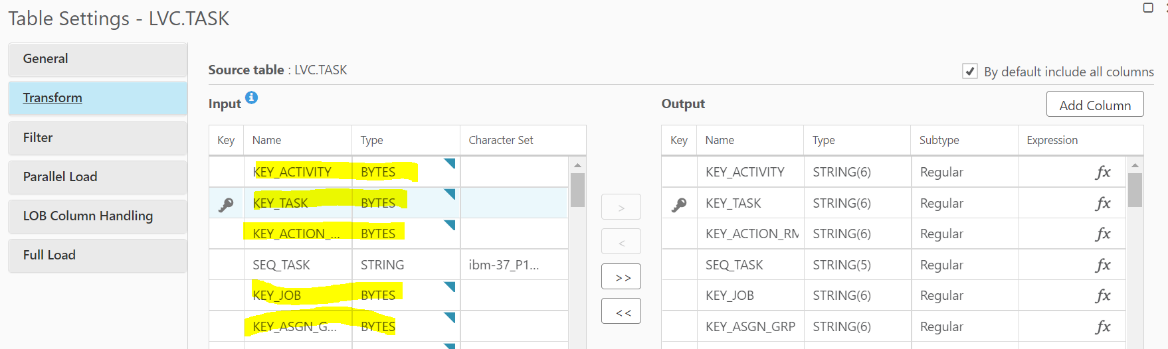
* + Snowflake schema for Qlik Replicate control tables to track replicate exceptions, history, status can be configured from Control Tables tab of the Task Settings as shown below: Snowflake Log tables schema update



* + If schema name is left blank, then Qlik Replicate will create control tables into PUBLIC schema
  + Please make sure that attrep\_history and attrep\_status control tables are NOT enabled (checked) unless needed. If we enabled any one of the tables, Qlik Replicate will perform INSERT/UPDATE every minute so Snowflake Warehouse will never suspend, and it will be very costly for the National Grid.

## HEX Conversion (requires for CRIS only)

* + CRIS has columns that stores HEX value, and it requires to convert them before it is sync into Snowflake
  + From the Task, select/double click the appropriate table where we need HEX conversion. Select transformation tab and select the column that needs Hex conversion. Modify the data types from STRING to BYTES as shown below. Save it.



# Appendix

Meeting recording

<https://nationalgridplc-my.sharepoint.com/personal/nishit_ajwaliya_us_nationalgrid_com/_layouts/15/onedrive.aspx?id=%2Fpersonal%2Fnishit%5Fajwaliya%5Fus%5Fnationalgrid%5Fcom%2FDocuments%2FRecordings%2FQlik%20%2D%20Development%20Env%2D20220624%5F093444%2DMeeting%20Recording%2Emp4&parent=%2Fpersonal%2Fnishit%5Fajwaliya%5Fus%5Fnationalgrid%5Fcom%2FDocuments%2FRecordings&ga=1>