



# **Data Migration Streamsets Service Standard Playbook**

*Prepared by:*

DMS

*July 2020*



## Document History

Version	Description	Date	Changed by
1.0	Initial	7/18/2020	Shaunak A
1.1	Various Updates	7/22/2020	Venky G, Shaunak A

# 1. Data Migration Process

A process oriented view is provided below that details out actions of involved parties

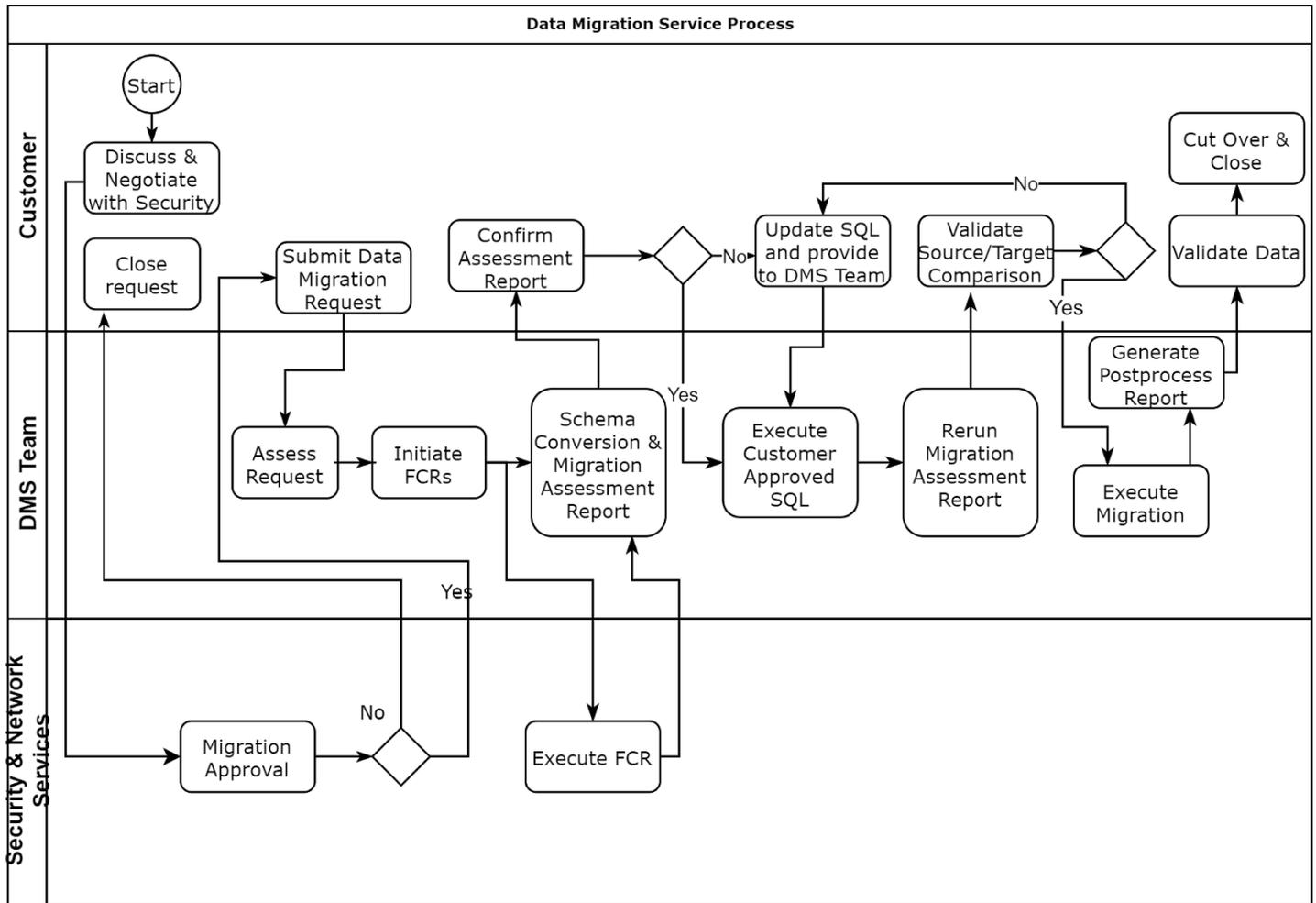


Image 3.2



## 2. Data Migration Features

This section will describe various features and operations of the Data Migration Service as implemented in the Database and Middleware Services subnets.

### 2.1 Service Configuration and Job Execution

The details of configuration are described in Section 4.

The overview is provided in the below steps:

1. Identify source and target databases' data center location
2. Choose appropriate service instance of the Streamsets service
3. Request firewall configuration to reach source and target databases
4. Request database accounts on source and target
5. Test connectivity to source and target endpoints and validate database account privileges
6. Optional - Convert schema(s) using the AWS Schema Conversion Tool (SCT) or execute the customer provided DDL on target database
7. Execute pre-processing jobs to validate source to target match
8. Execute the data migration job(s) to complete the data migration
9. Execute the post-processing pipeline to validate source to target data match
10. Consolidated all pre-, post- and migration reports to hand over to customer
11. Close the migration request

### 2.2 Supported Technologies

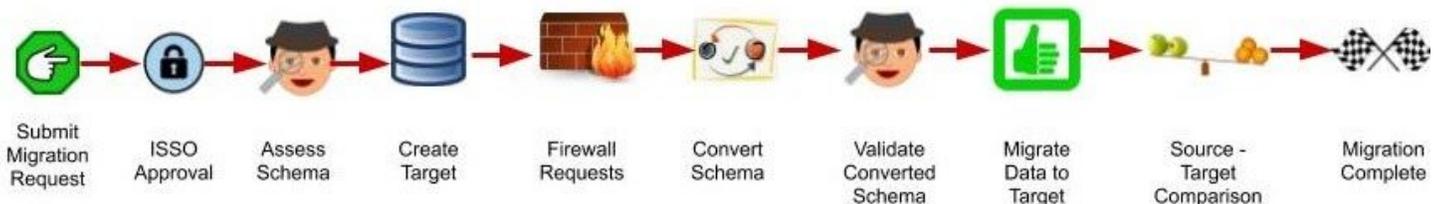
The following technologies are supported as source and target databases by this data migration service.

**Source:** Oracle, Sybase ASE, MySQL Postgres

**Target:** Oracle, MySQL, Postgres, SQL Server

## 3. Data Migration Process Life Cycle

The flowchart below will describe the lifecycle of a data migration request and fulfillment.





## 4. Migration High-Level Outline

Assuming the schema structure has already been migrated and to start migrating data from source database to target database using TMF solution, you execute the following steps. A summary of steps is provided below:

- Step 1: Configure Your Source Database
- Step 2: Configure Your Target Database
- Step 3: Configure and run Migration task

### Step 1: Prepare Source Database

Utilizing the Data Migration Service to migrate data from a source database requires some preparation and we also recommend a few additional steps as best practices.

- Source database – To migrate data, the Data Migration Service needs access to the source database, make sure to configure firewall rules to allow connections from the streamsets instance to the source.
- Privileges: Make sure the designated migration service user has following privileges

```
SELECT access on <source schema>  
SELECT access on DBA_TABLES  
SELECT access on all
```

### Step 2: Prepare Target Database

The following are key items to consider when launching your Target instance:

- The Data Migration Service account needs the following privileges on all databases to which data is being migrated.



ALTER, CREATE, DROP, INDEX, INSERT, UPDATE, DELETE, SELECT

### Step 3: Configure and run a Migration Task

Pipelines in the solution are classified and have to be executed in the same order:

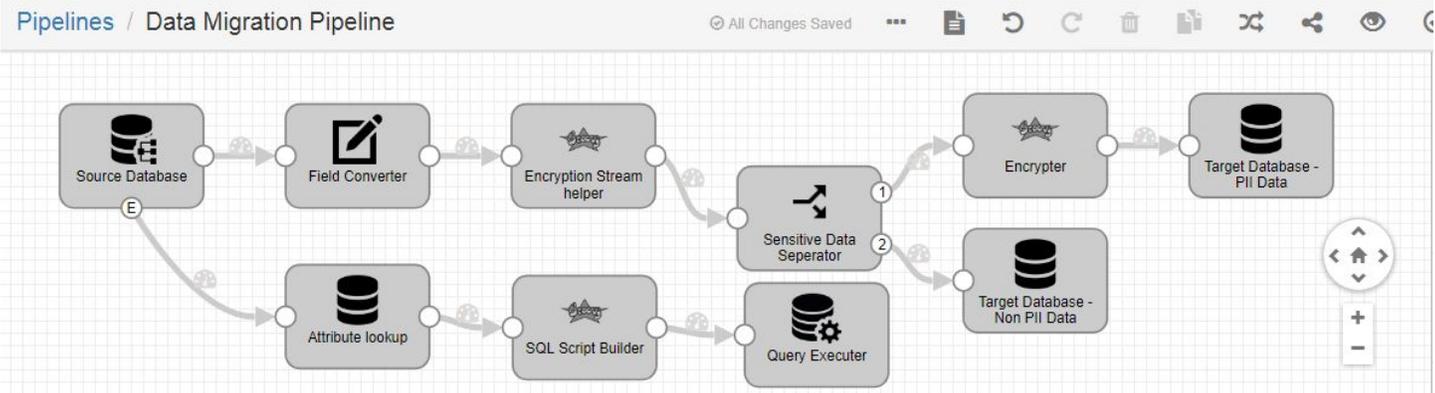
- **Pre-Migration Pipeline:** Generates report based on the metadata collected from source and destination
- **Data Migration Pipeline:** Migrates the data based on the user provided configuration for source and destination
- **Post Migration Pipeline:** Performs smoke testing once the migration is completed

Each of the pipelines expect a set of important parameters from the user to successfully initiate and complete the designated tasks at each stage of the migration. Copy these update parameters to the **Parameters** tab of the pipeline using **Bulk edit** mode:

```
[
  {
    "key": "ENCRYPTION_KEY",
    "value": ""
  },
  {
    "key": "ENCRYPTION_DETAILS",
    "value": "[]"
  },
  {
    "key": "SOURCE_SCHEMA",
    "value": "enter administrative schema (i.e., ADMIN)"
  },
  {
    "key": "SOURCE_IP",
    "value": "xxx.xxx.xxx.xxx"
  },
  {
    "key": "SOURCE_SID",
    "value": ""
  },
]
```



```
{
  "key": "SOURCE_USERNAME",
  "value": "enter source username (i.e., dms_user)"
},
{
  "key": "SOURCE_PASSWORD",
  "value": ""
},
{
  "key": "DESTINATION_SCHEMA",
  "value": ""
},
{
  "key": "DEST_IP",
  "value": "xxx.xxx.xxx.xxx"
},
{
  "key": "DEST_USERNAME",
  "value": "enter root level user (i.e., ROOT)"
},
{
  "key": "DEST_PASSWORD",
  "value": ""
},
{
  "key": "DEST_PORT",
  "value": ""
},
{
  "key": "SOURCE_PORT",
  "value": "xxxx"
},
{
  "key": "TARGET_TIME_ZONE",
  "value": "America/Chicago"
}
]
```



Once the parameters are updated, you can validate the pipeline, preview it for a sense of source and target data record loading and start the pipeline.