**EXCHANGE DATA INTERNATIONAL (EDI)**

Deployment Document – EODP\_EQ

**Table of Contents**

|  |  |  |
| --- | --- | --- |
| S. No. | Title | Page No. |
| 1 | Initial Setup | 3 |
| 2 | Transformation Setup | 7 |
| 3 | Execution Setup | 9 |

**Initial Setup**

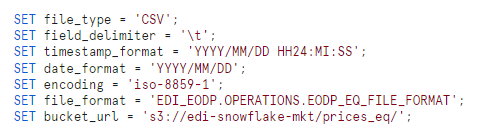
Note: Please do the one-time setup for the integration of AWS S3 and Snowflake before this if it's not done.

1. **Defining variables:**

Set the variables for specifying file type, field delimiter, date format, encoding, file format, S3 bucket URL, snowflake warehouse, and scheduled time.

*Refer to the* ***EDI\_EODP\_EQ\_SETUP.sql*** *file* ***Pre-requisite 1***

**Example query:**

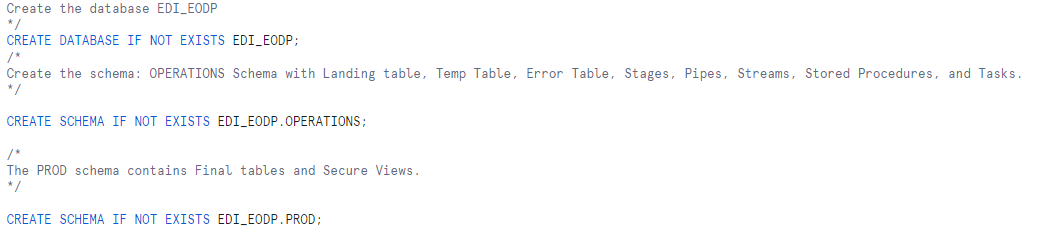
****

1. **Create Database and Schema:**

A new database is created for EDI\_EODP\_EQ with schemas as OPERATIONS and PROD to build tables, stages, pipes, streams, tasks, stored procedures and secure views.

*Refer to the* ***EDI\_EODP\_EQ\_SETUP.sql*** *file* ***Pre-requisite 2***

**Example query:**

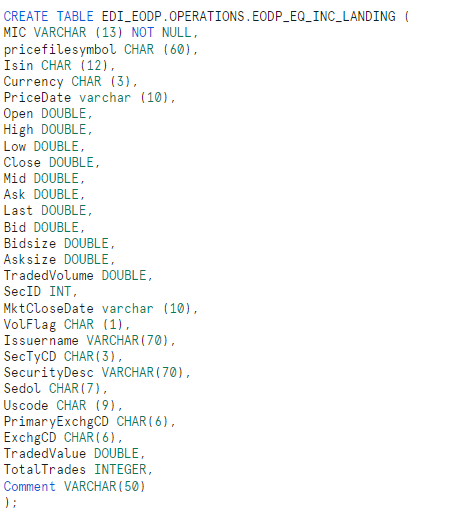


1. **[Create the tables](https://docs.snowflake.com/en/sql-reference/sql/create-table):**

Tables are the structure or skeleton for the data present. So that the end user would be easily understand the contents.

*Refer to the* ***EDI\_EODP\_EQ\_SETUP.sql*** *file* ***Pre-requisite 3*** *to create all the tables.*

**Example query:**

**

**LIST OF TABLES**

|  |  |
| --- | --- |
| **SCHEMA** | **TABLE NAME** |
| OPERATIONS | EODP\_EQ\_INC\_LANDING |
| OPERATIONS | EODP\_EQ\_INC\_TEMP |
| OPERATIONS | EODP\_EQ\_INC\_ERROR |
| PROD | EODP\_EQ\_INC |
| PROD | EODP\_LOOKUP |

We have **Landing tables** for data ingestion from AWS S3 to the snowflake environment via snowpipe.

**Temp tables** are temporarily storing the data from stream to process and dump it to the final table.

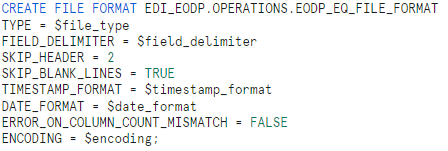
**Error tables** contain all the rows from file that failed with CREATED\_AT column to signify the timestamp when the file got failed.

## [Creating a File Format](https://docs.snowflake.com/en/sql-reference/sql/create-file-format):

## File formats are prefixed rules for snowflake to recognize the file's data and understand it. Create a file format to accept Tab separated file and to remove first two rows in the file (File name and column header in the file):

## *Refer to the* ***EDI\_EODP\_EQ\_SETUP.sql*** *file* ***Pre-requisite 4****.*

## Example query:

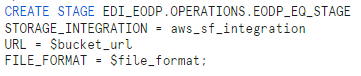
**

## **[Create a Stage](https://docs.snowflake.com/en/sql-reference/sql/create-stage):**

## Stages are responsible to get the data that has been uploaded in S3 bucket, in real time. Multiple Stages should be created for connecting folders of s3 bucket to Snowflake.

## *Refer to the* ***EDI\_EODP\_EQ\_SETUP.sql*** *file* ***Pre-requisite 5.***

## Example query:



To list all the list of files that's inside the stage from the S3 bucket: list @<stage name>. *Refer to the* ***EDI\_EODP\_EQ\_OPERATIONS.sql*** *file* under ***Queries related to External Stage.***

## **[Create a Pipe:](https://docs.snowflake.com/en/sql-reference/sql/create-pipe)**

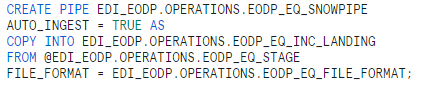
## Snowpipes, are responsible for copying the data into the table from External stage. Create a pipe with Auto-Ingest Enabled:

## *Refer to the* ***EDI\_EODP\_EQ\_SETUP.sql*** *file* ***Pre-requisite 6****.*

## 1: Switch to ACCOUNT ADMIN role

## **2**: Create Pipe

## Example query:

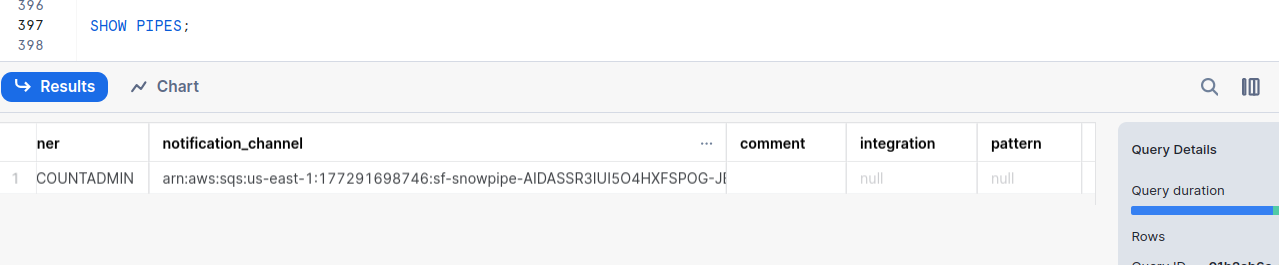


1. **Configure Event Notification**

### Event Notifications are responsible for notify the pipe about new data in stage. Execute the *SHOW PIPES* command.

### *Refer to the* ***EDI\_EODP\_EQ\_OPERATIONS.sql*** *file* under ***Queries related to*** ***Snowpipe.***

## Example query:



### Note the ARN of the SQS queue for the stage in the **notification\_channel** column. Copy the ARN to a convenient location.

### Follow the steps from option 1 step 4 in the docs from the link below to set up the event notification for the S3 bucket.

[Snowflake-AWS S3 automation](https://docs.snowflake.com/en/user-guide/data-load-snowpipe-auto-s3)

1. **[Create a Stream:](https://docs.snowflake.com/en/sql-reference/sql/create-stream)**

Stream gets the new data that has been added in to landing table.

*Refer to* ***EDI\_EODP\_EQ\_SETUP.sql*** *file* ***Pre-requisite 7***

**Example query:**

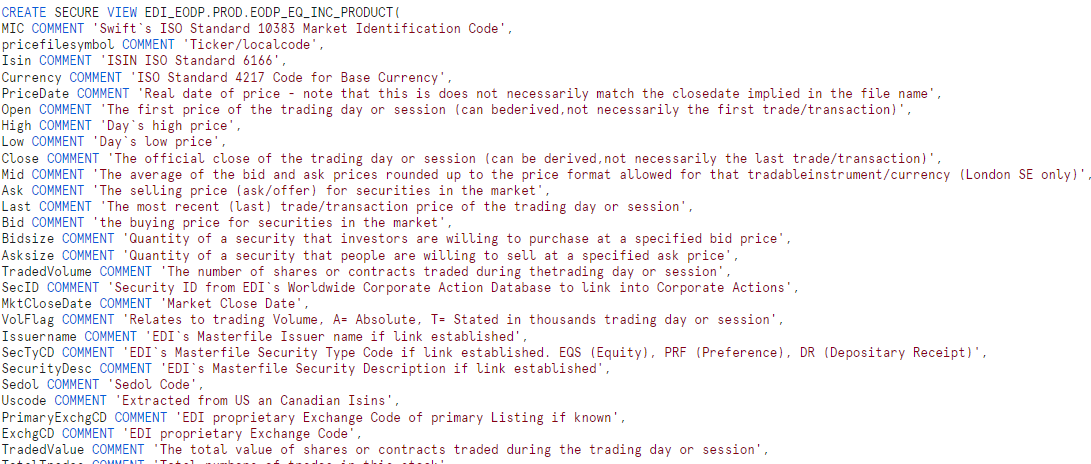
***eq_stream***

1. **Create Secure Views**

These are the queries to make secure views on top of the product data tables.

*Refer to* ***EDI\_EODP\_EQ\_SETUP.sql***  *file* ***Pre-requisite 8***

**Example query:**



**Transformations Setup**

**[Create a Stored Procedure:](https://docs.snowflake.com/en/sql-reference/sql/create-procedure)**

Stored procedure is the injection of python code which will standardize the data received from Temp table and send this data to Final table. In our case we have used the procedure to fix timestamp into a specific format.

Steps Done in Procedure

* Getting data from the TEMP Table
* Removing the Extra Columns for metadata created by the stream
* Adding the CREATED\_BY and CREATED\_AT columns.
* Conversion of all values in sedol to NULL
* Conversion of PricefileSymbol into an empty string from NULL
* Remove the EDI\_ENDOFFILE
* Standardize the date columns timestamp into required format for FINAL table
* Write the transformed data into FINAL table
* Clear the TEMP table upon successful Data Ingestion

*Refer* ***EDI\_EODP\_EQ\_TRANSFORMATION.sql*** *file*

**Example query:**

**

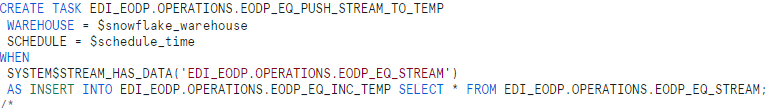
**Execution Setup**

**[Creating Tasks:](https://docs.snowflake.com/en/user-guide/tasks-intro)**

We create EODP\_EQ\_PUSH\_STREAM\_TO\_TEMP task which is responsible for regularly poll based on the scheduled interval if stream on top Landing table has received any new data and then insert that data into Temp table then trigger the EODP\_EQ\_PUSH\_TO\_EODP\_EQ\_INC task to call the stored procedure EODP\_EQ\_SP()

**1:** Creation of Task 1

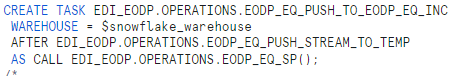
**Example query:**

******

*Refer* ***EDI\_EODP\_EQ\_EXECUTION.sql*** *file* ***Task 1***

**2:** Creation of Task 2

**Example query:**

******

*Refer* ***EDI\_EODP\_EQ\_EXECUTION.sql*** *file* ***Task 2***

**3*:*** *Execution of Tasks*

**Example query:**

***eq_task3***

*Refer* ***EDI\_EODP\_EQ\_EXECUTION.sql*** *file* ***Queries to initiate the tasks***

**WORKING ARCHITECTURE OF THE PROCESS**

