**Project Report**

CAMP Feb BI V3 Batch 1 Group 4

**INDEX**

1. Project Description
2. Unit Test Report
   1. Task 1: Download and Load Data
      1. Download Dataset
      2. Creating S3 Bucket with folders
      3. Uploading data into folder1
      4. Uploading data into folder2
   2. Task 2: Creating Database
   3. Task 3: Creating Schema
   4. Task 4: Creating Datamart Schema
   5. Task 5: Creating Tables
      1. Creating flight table
      2. Creating flight\_scd table
      3. Creating target\_flight table
   6. Task 6: Creating External Stage
      1. Creating flight\_stage stage
      2. Display the files in the flight\_stage\_1
      3. Display the files in the flight\_stage\_2
   7. Task 7: Creating Snowpipe
      1. Creating the Snowpipe and copy data from pipe to flight table
      2. Refreshing the pipe
      3. Displaying the data from the flight table
   8. Task 8: Creating Stream
      1. Create a stream flight\_stream
      2. Display the (initial) content of the stream
   9. Task 9: Create SCD
   10. Task 10: Creating Task
       1. Create the task with merge operation
       2. Resume the task
       3. Display the Inserted, Updated, Deleted values of the consumer table

**PROJECT DESCRIPTION**

The project involves data ingestion and analysis from public data hub kaggle (make a hyperlink).

Steps involved in performing the data ingestion:

1. Loading data to external stage from S3 Bucket
2. Ingesting data into the landing schema
3. Ingesting data into the consumer table
4. Perform Analysis on the dataset

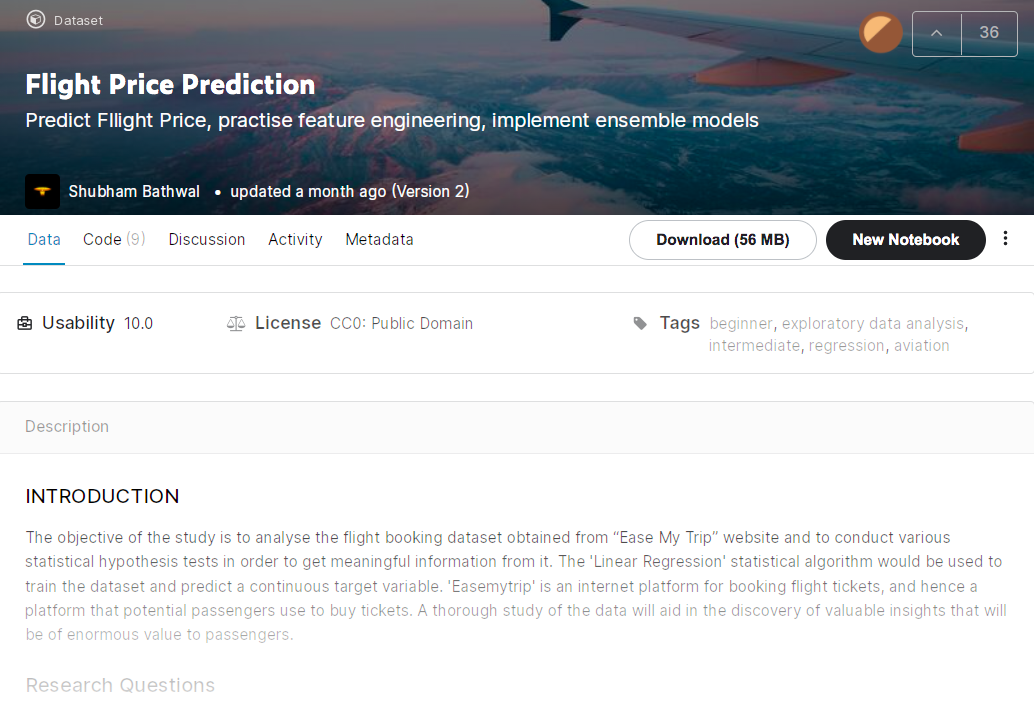
This project contains Data Definition Language (DDL) for creating,

* Database named SF\_Project.
* Three schemas named ITR\_RDS\_Landing for extracting data from the S3 Bucket, ITR\_RDS schema to storing the transformed data from ITR\_RDS\_Landing and ITR\_DIS schema for storing data after SCD (Slowly Changing Dimension).
* Three tables named as flight, which is the primary table, flight\_scd which is the table for storing the data for SCD part and target\_flight which the consumer table for flight\_scd.
* Storage integration named flight\_int, which will, stores a generated identity and access management (IAM) entity for your external cloud storage.
* External Stages named flight\_stage\_1 for loading data from external source to table in our case the data is loaded from S3 Bucket from Folder1 and another stages is created with the name flight\_stage\_2 for unloading the from same s3 Bucket but this time the data is in folder 2.
* Pipes in this project we created two pipes to copy data from stage to table and the pipes are named as flight\_pipe and flight\_pipe\_scd.
* Stream object records data manipulation language (DML) changes made to tables, including inserts, updates, and deletes, as well as metadata about each change, so that actions can be taken using the changed data.
* Tasks are used for continuous ELT workflows to process recently changed table rows here we created a task named Target\_Merger and Merge can be useful if the second table is a change log that contains new rows, modified rows, and marked rows in the target table.

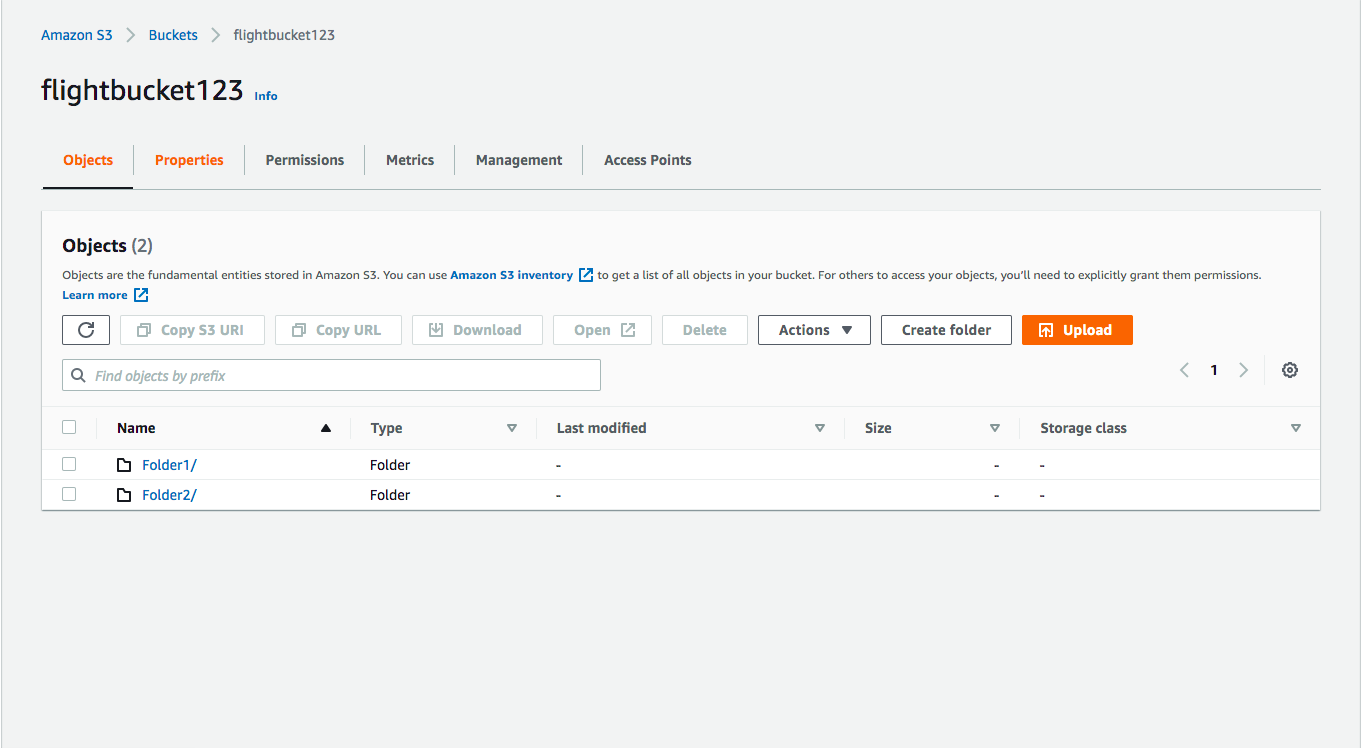
**UNIT TEST REPORT**

TASK 1: -

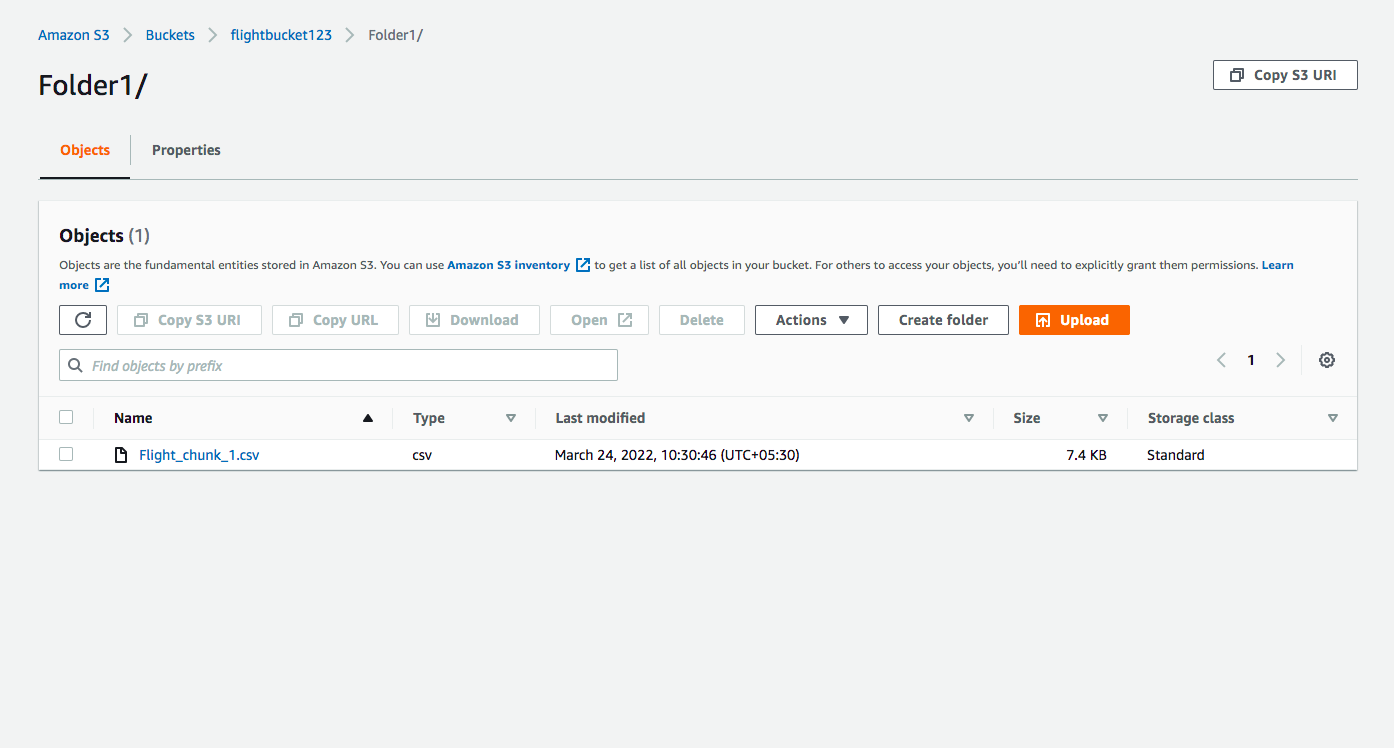
1. Downloading data from **[Kaggle](https://www.kaggle.com/)** hub.



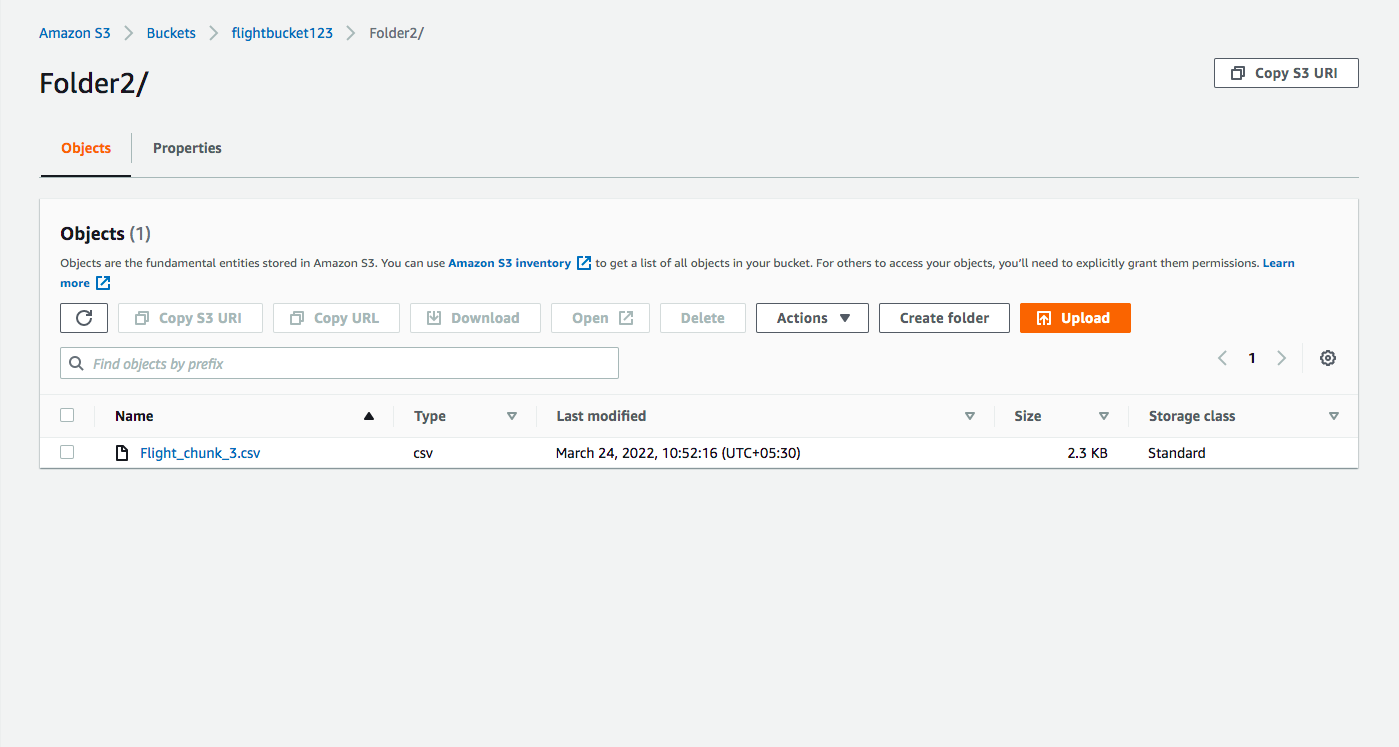
1. Creating two folders in S3 bucket (**flightbuckert123**).



1. Upload a chunk to pre-cleaned data named “**Flight\_chunk1.csv**” to the Folder1.

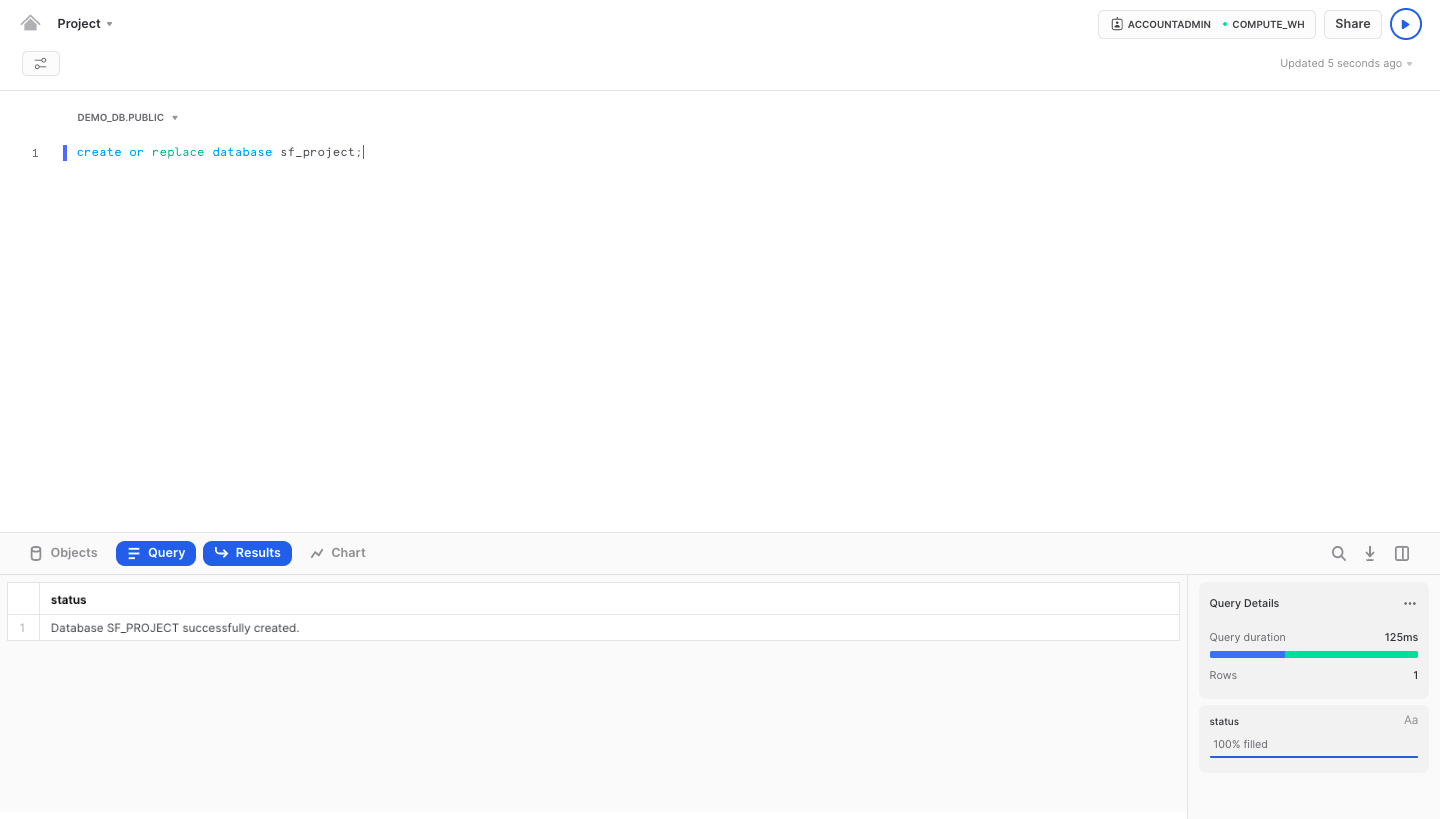


1. Upload a chunk to pre-cleaned data named “**Flight\_chunk3**.csv” to the Folder2.



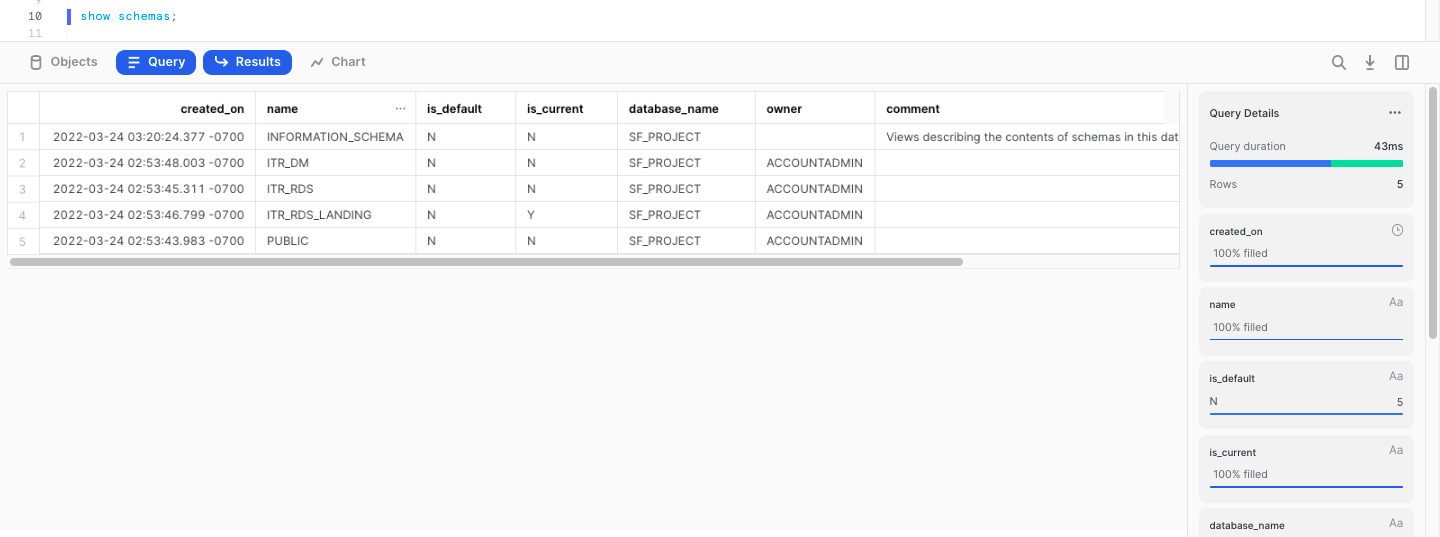
Task 2: -

1. Creating a database **“SF\_PROJECT”**.



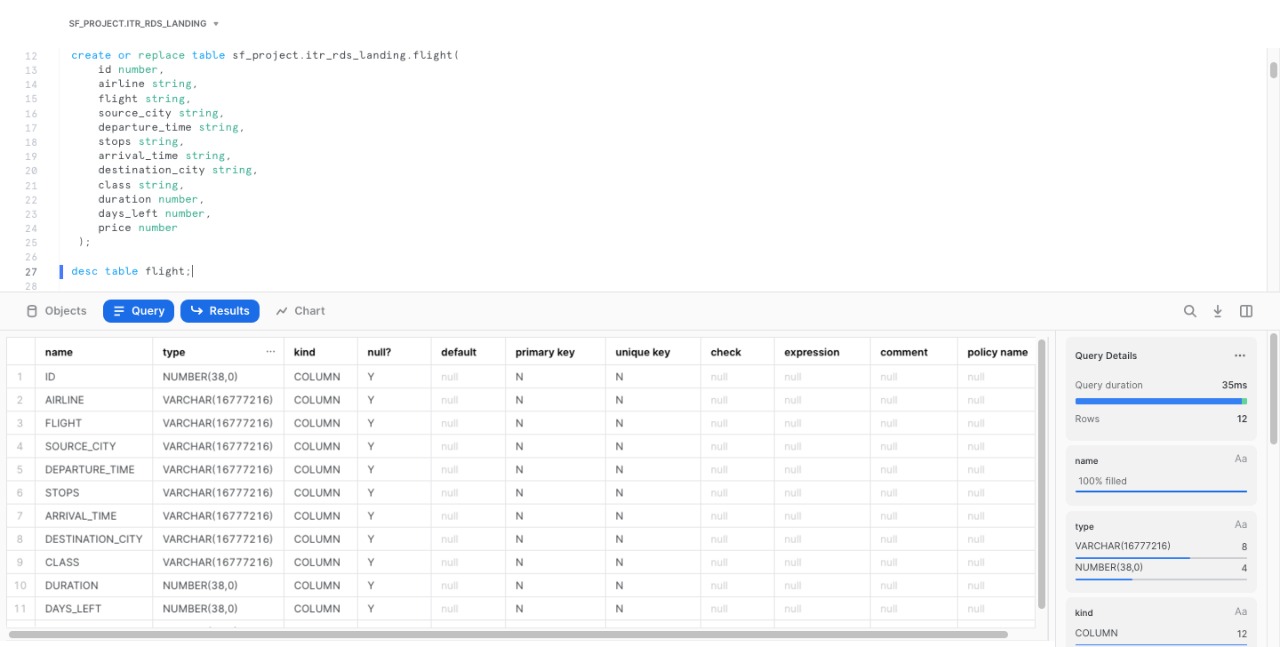
Task 3 & 4: -

1. Using **“SHOW SCHEMA”** to display all the schemas.

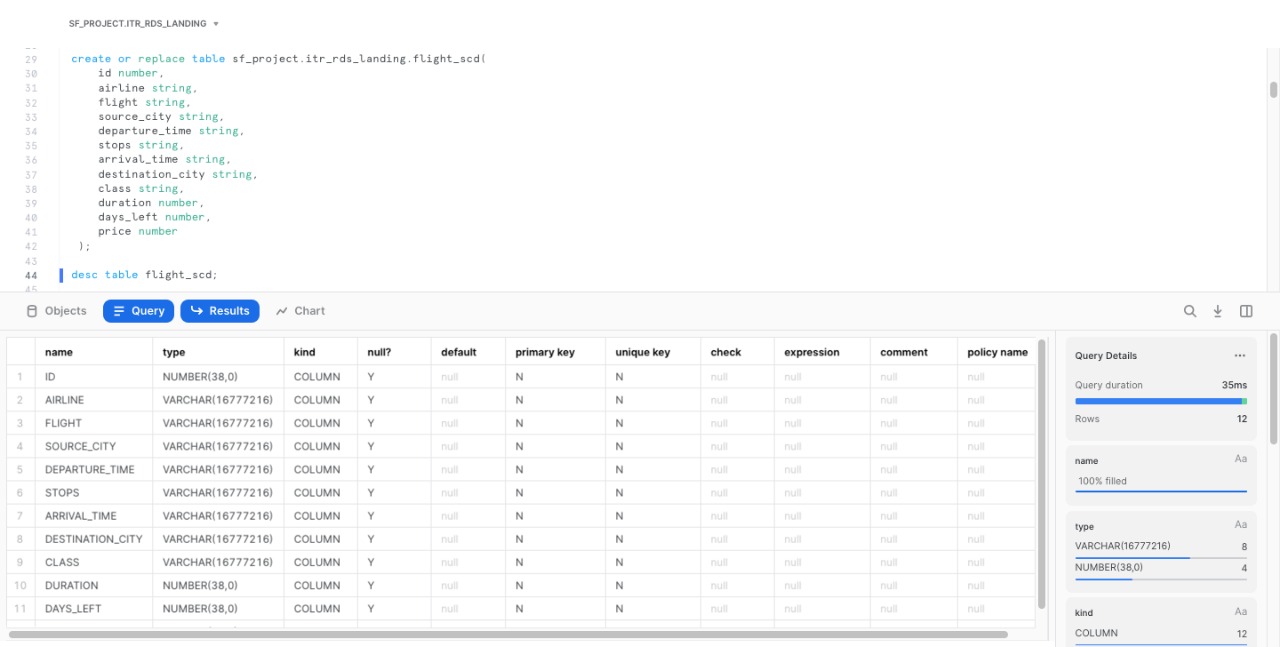


Task 5: -

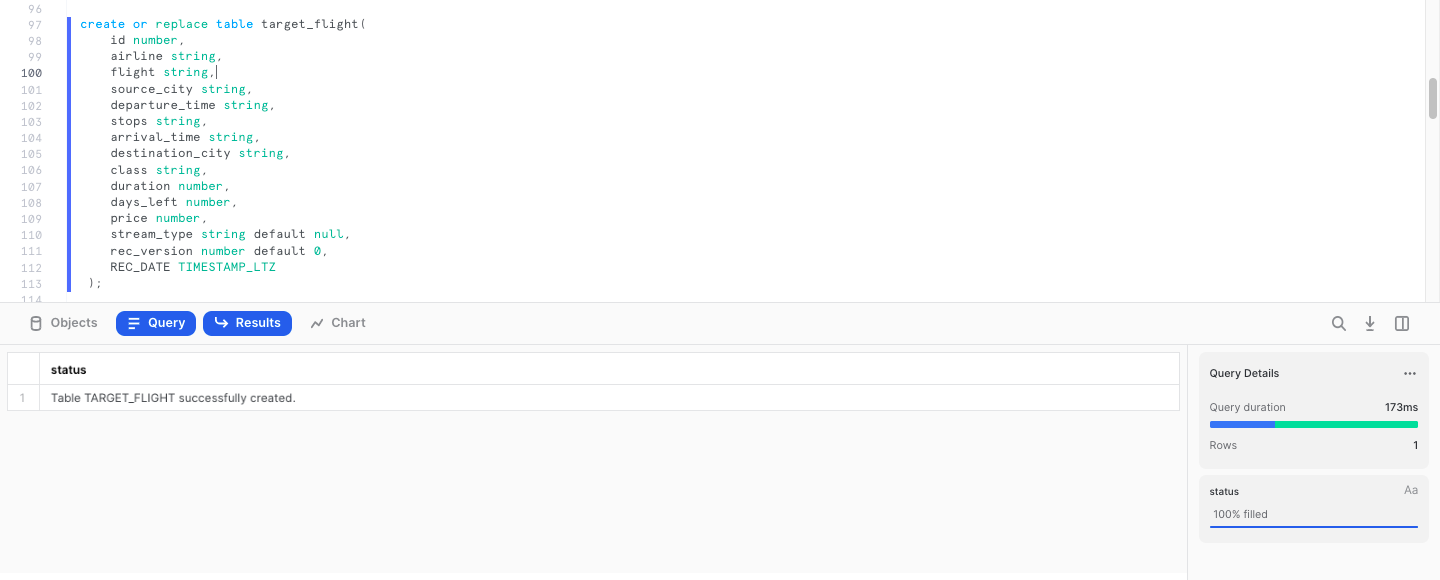
1.Creating and describing the table **“FLIGHT”** which will unload the data from the **S3 Bucket**.



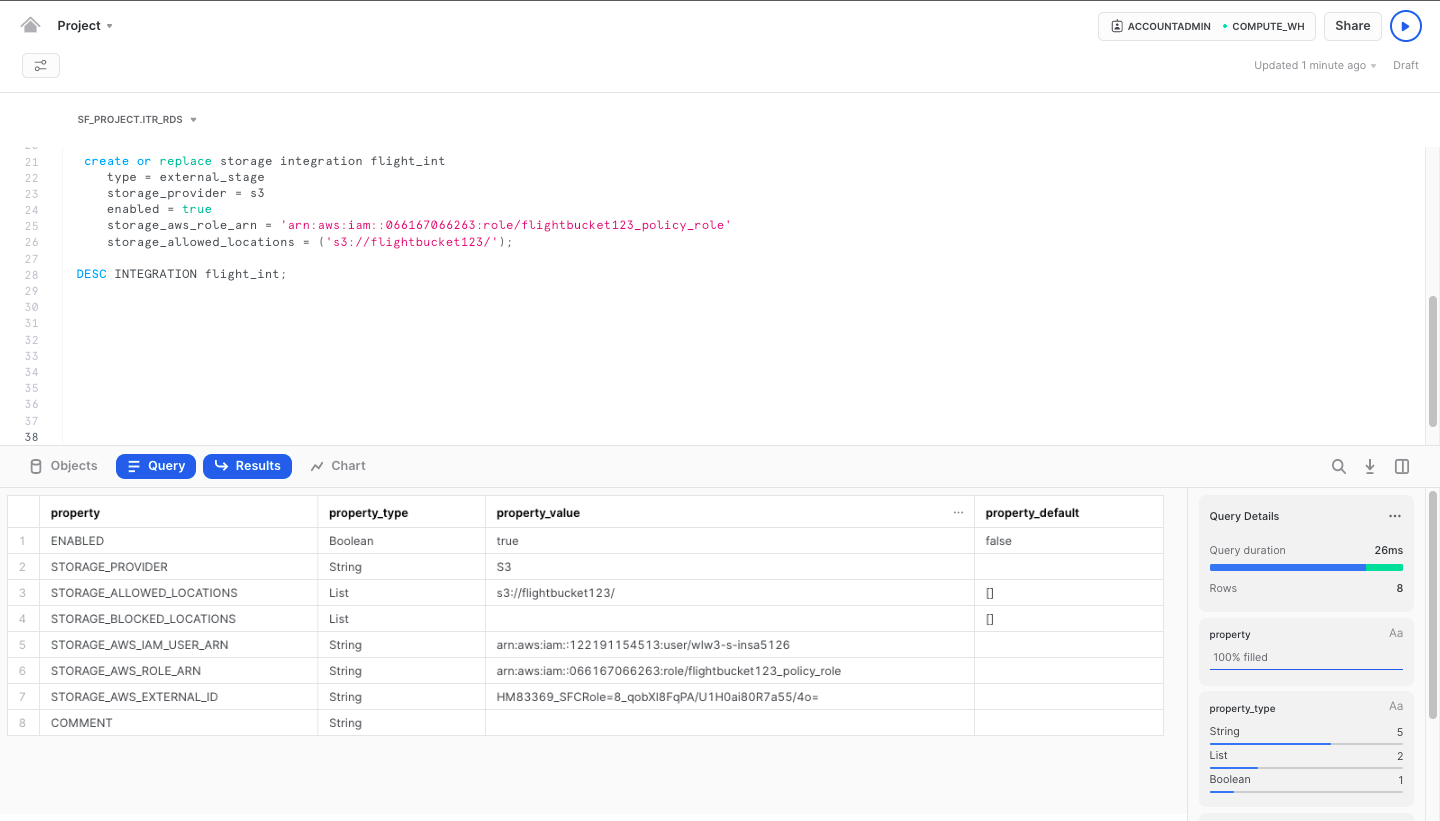
1. Creating and describing the table **“FLIGHT\_SCD”** which will be used to demonstrate SCD and Stream.



3.Creating consumer table **“TARGET\_FLIGHT”**.

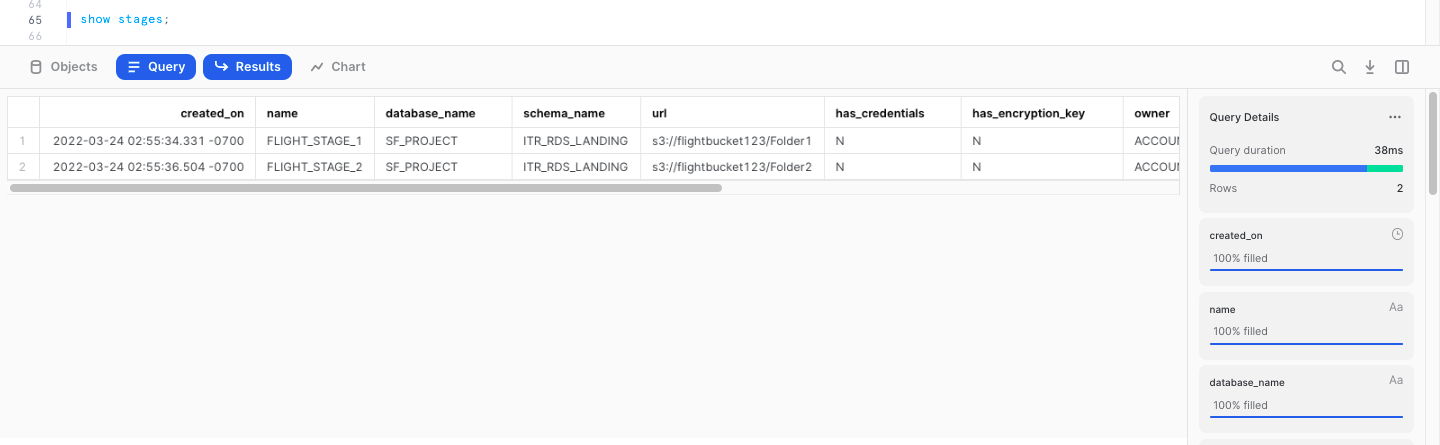


1. Creating and describing the storage integration object **“FLIGHT\_INT”** which will stores a generated identity and access management **(IAM) entity** for your external cloud storage.

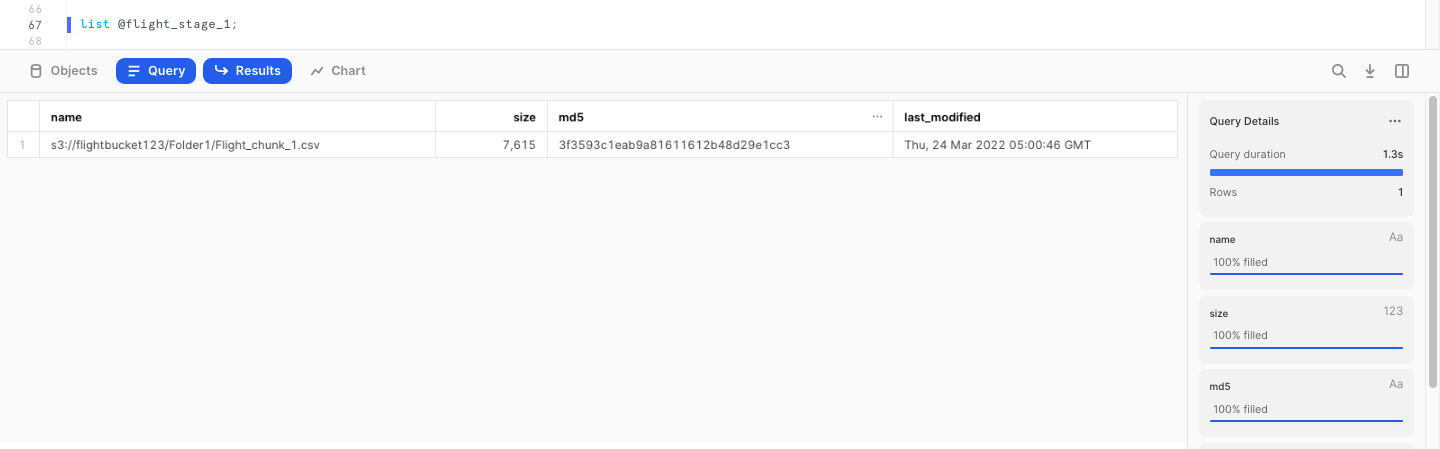


Task 6: -

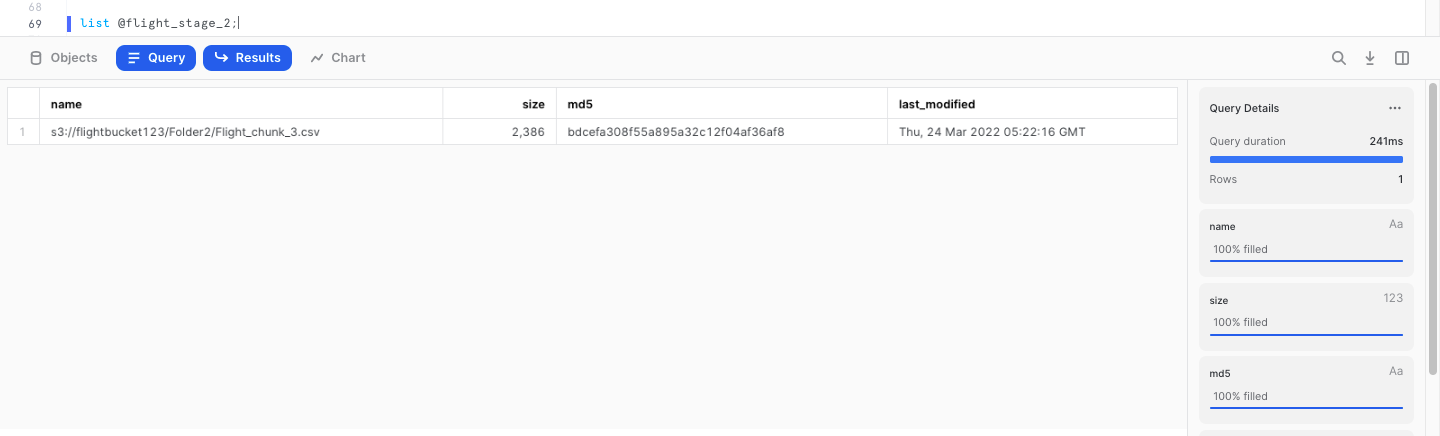
1. Displaying the external stages by using **“SHOW STAGES”** command. There are two stages one for loading data from external source to table in our case the data is loaded from S3 Bucket from **Folder1** and another stages is created for unloading the the from same s3 Bucket but this time the data is in folder 2.



1. Displaying list of all the files that is/are in **“FLIGHT\_STAGE\_1”** stage.

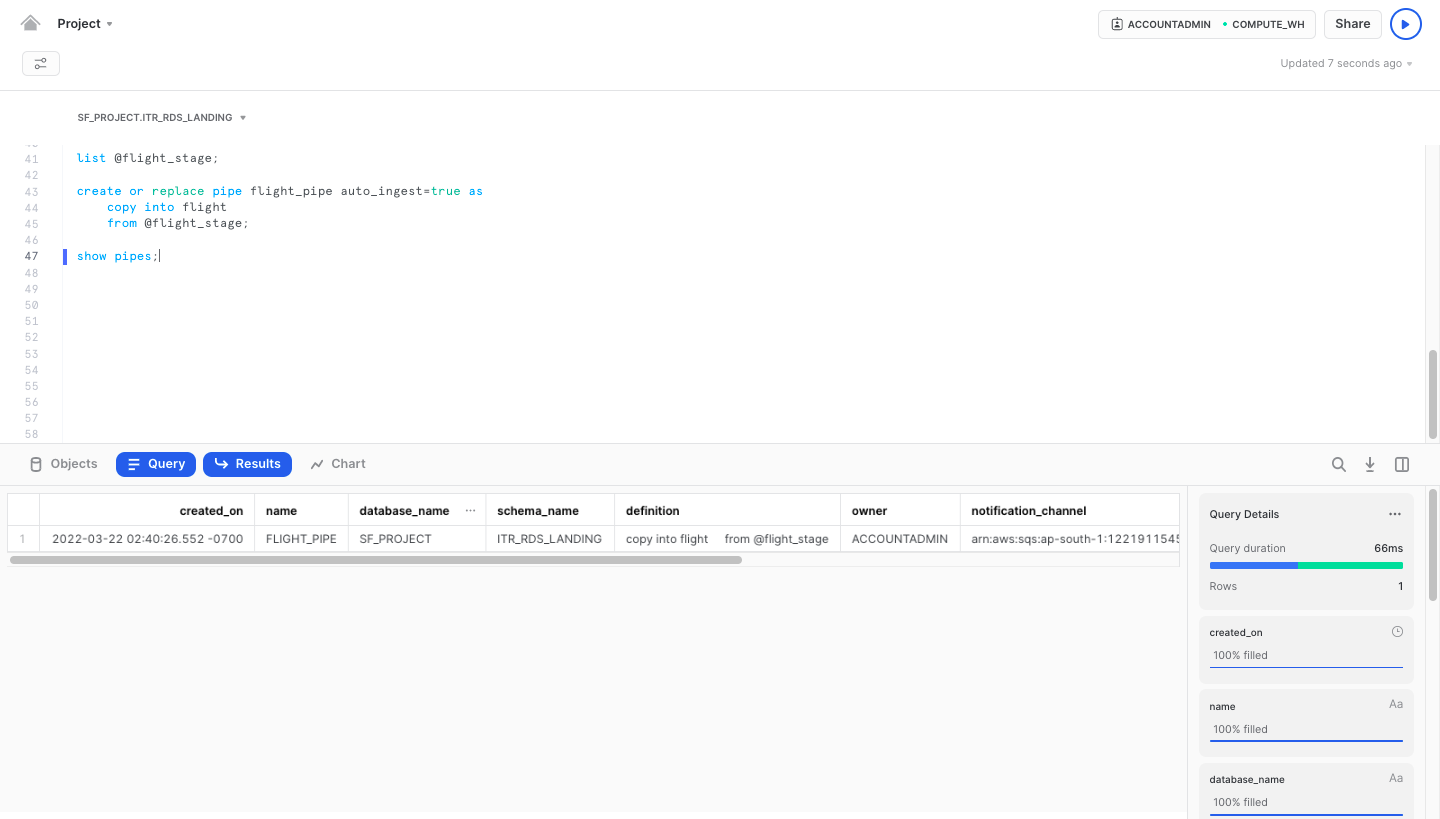


3.Displaying list of all the files that is/are in **“FLIGHT\_STAGE\_2”** stage.

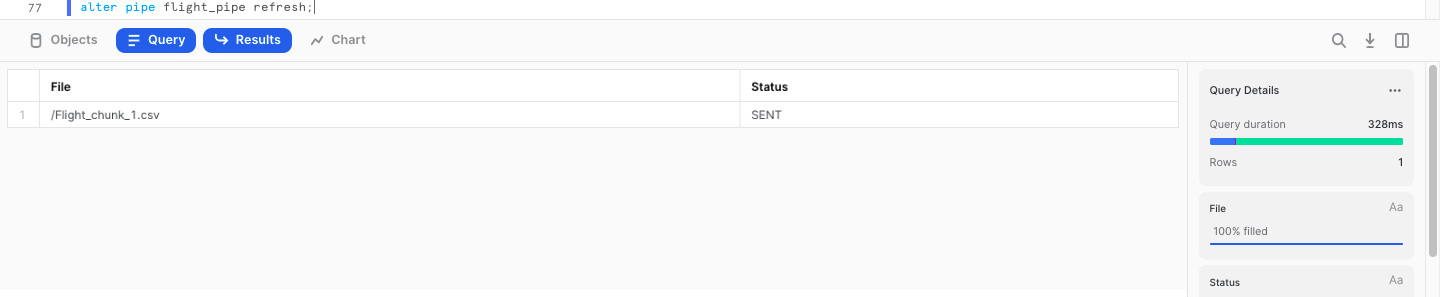


Task 7: -

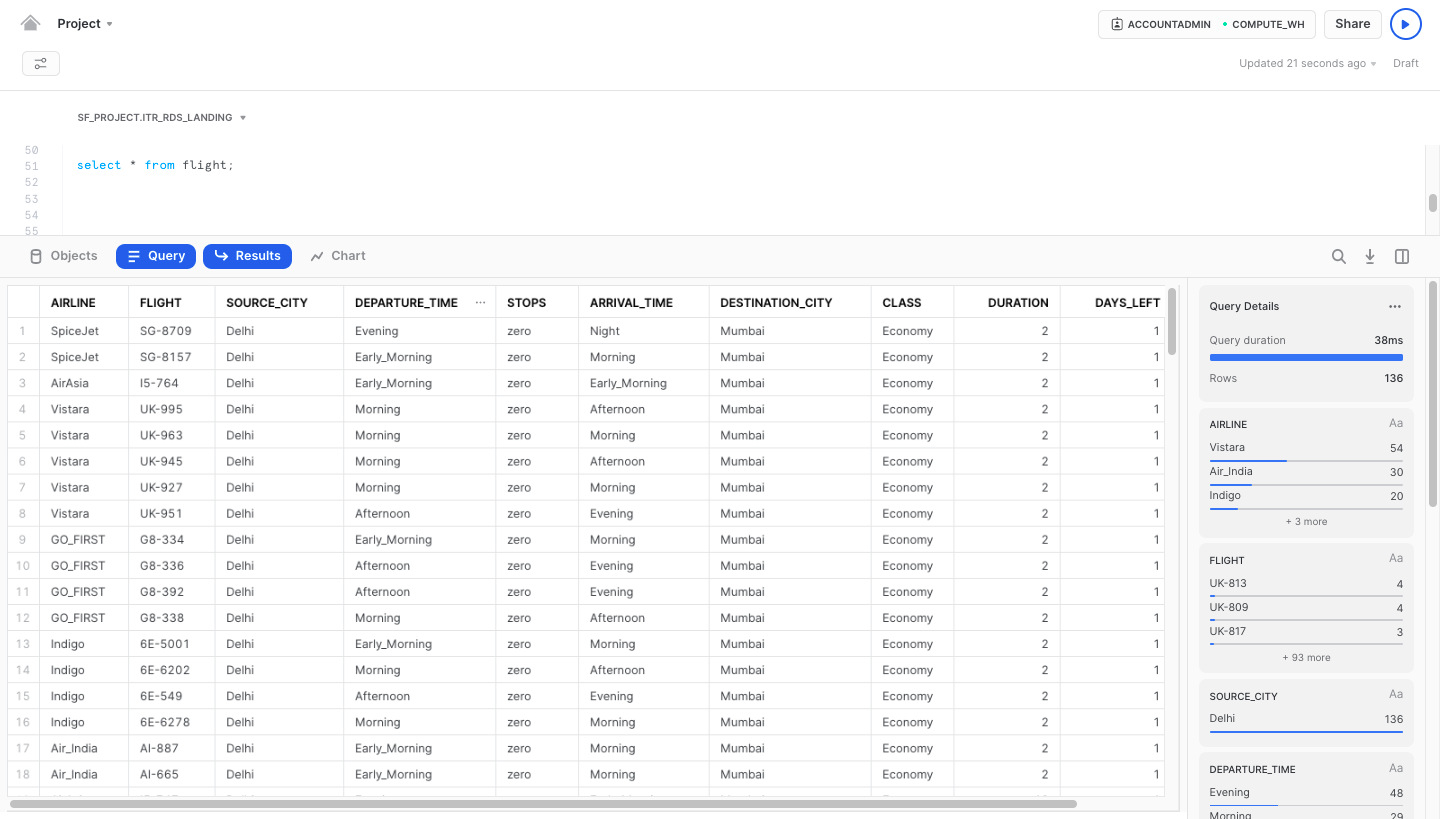
1.Creating **Snowpipe** to to copy data from stage to table.



2.**Refreshing** the Pipe so that data gets loaded into the table.



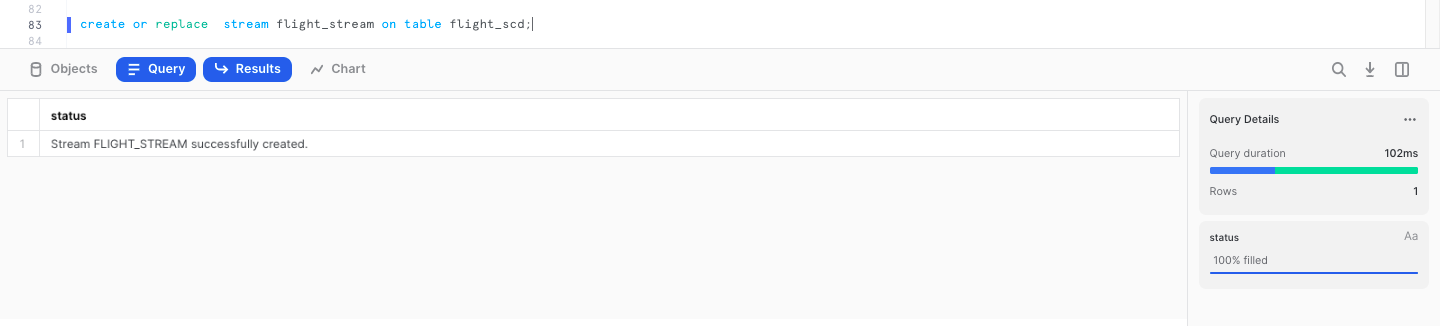
3.Displaying the contents loaded into the table named **“FLIGHT”**.



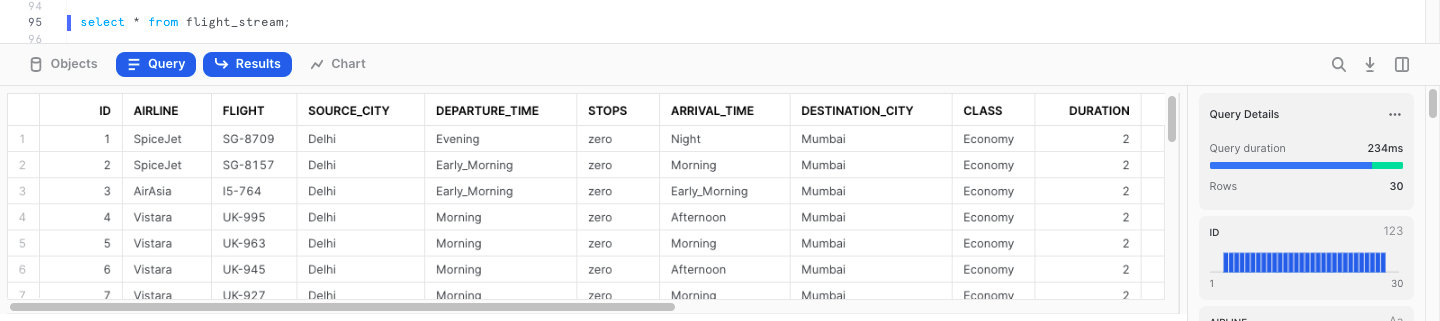
Task 8 & 9: -

1. Creating a stream **“FLIGHT\_STREAM”** in the current schema, which will records data

manipulation language **(DML)** changes made to tables.

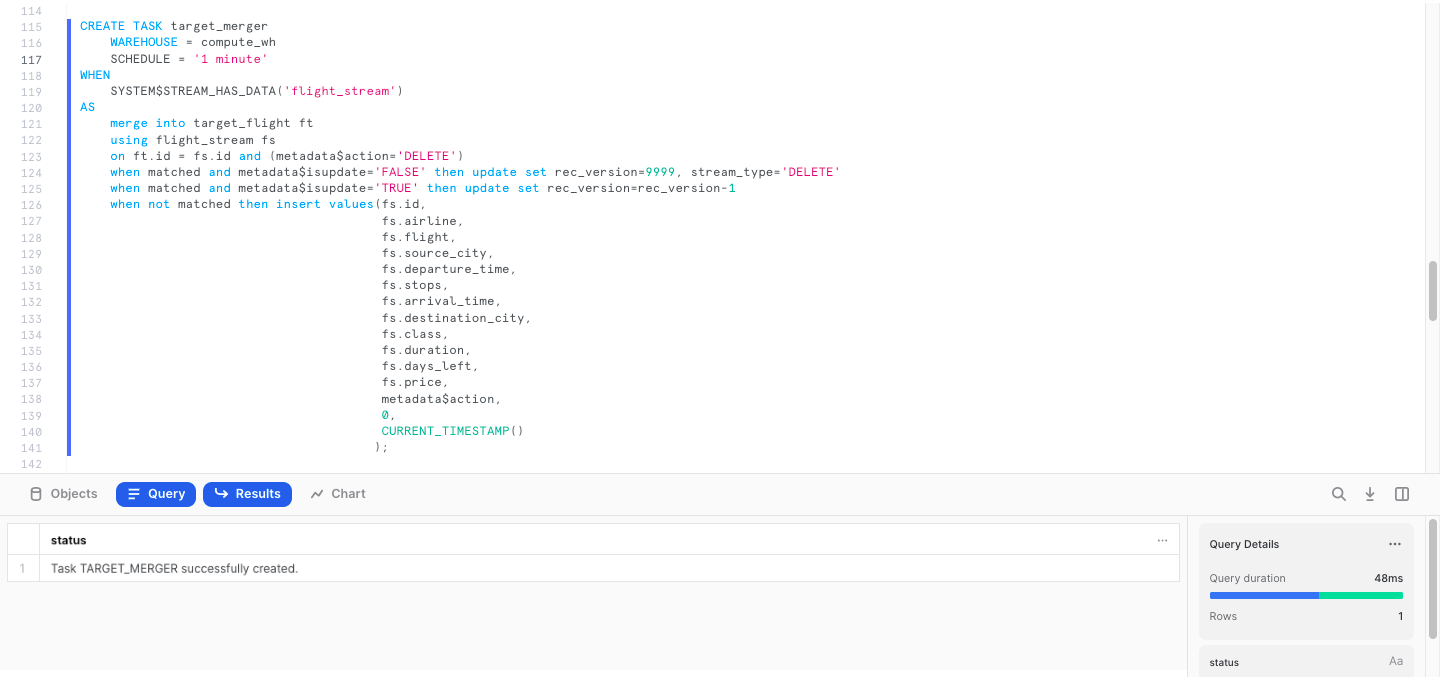


2.Displaying the contents loaded into the **Stream**.

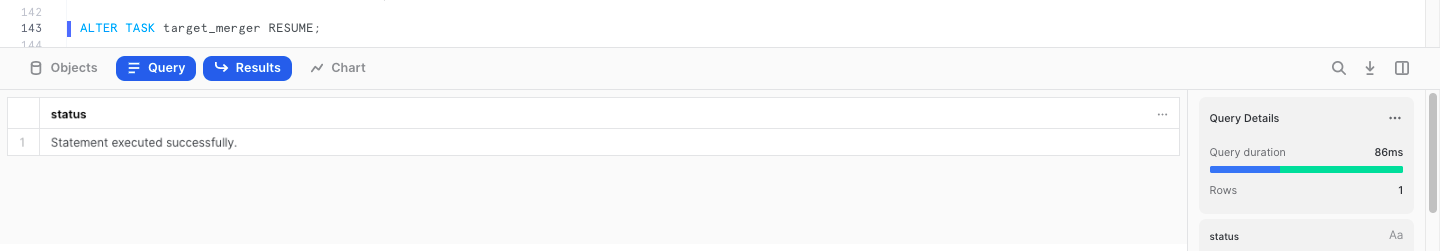


Task 10 & 9: -

1.Creating Task to merge the stream data with consumer table named **“TARGET\_FLIGHT”**.



2.Using **“ALTER TASK“** Command to resume the task **“TARGET\_MERGER”**.



3.Displaying contents of the **consumer table** including **recorded version, record date and stream type** after Updating, Deleting, and Inserting operations.

