Case Study

on

Snowflake

Case study by,

Mouli S,
Batch – 02,
moulisankar2002@outlook.com.

Contents

1. Understandings of Case Study

Coronavirus disease (COVID-19) is an infectious disease caused by the SARS-COV-2 virus.

Most people infected with the virus will experience mild to moderate respiratory illness and recover without requiring special treatment. However, some will become seriously ill and require medical attention. Older people and those with underlying medical conditions like cardiovascular disease, diabetes, chronic respiratory disease, or cancer are more likely to develop serious illness. Anyone can get sick with COVID-19 and become seriously ill or die at any age.

European Centre for Disease Prevention and Control (also known as ECDC) - an agency of the European Union tracks COVID-19 cases and vaccination status worldwide.

Reference Website - https://www.ecdc.europa.eu/en/

2. Case Study Questions

2.1 Question 1

Participants will create materialized view to copy data from $3 bucket for customer data (https://s3.apsoutheast2.amazonaws.com/snowflake-essentials/ingesting data/new_customer/2019-09 24/generated_customer_data.csv) and display out having customers whose age is greater than 30 years and less than 50 years with following columns:

a. CustomerName

b. CustomerAge (as on today, Integer)

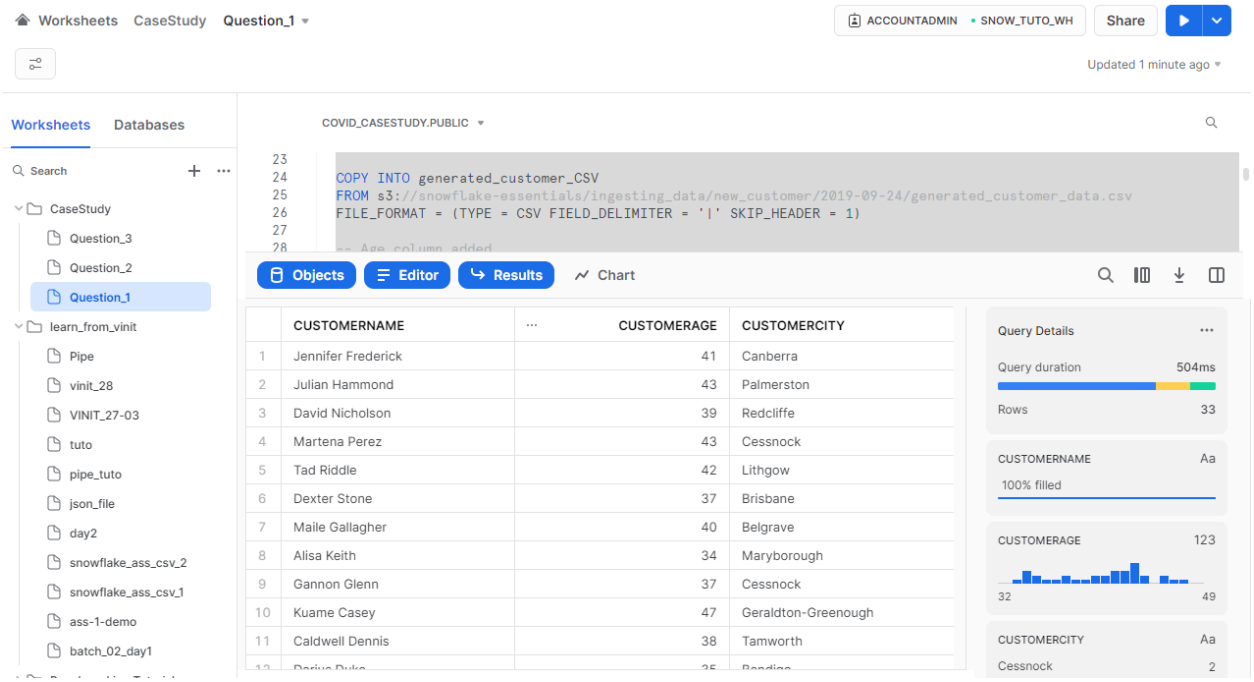
c. CustomerCity

SQL Query:

```
CREATE DATABASE Covid_CaseStudy;
-- Question 1
-- Table creation
CREATE TABLE generated_customer_CSV (
   Customer_ID INTEGER,
   Customer_Name VARCHAR(100),
   Customer_Email VARCHAR(100),
   Customer_City VARCHAR(100),
   Customer_State VARCHAR(100),
   Customer_DOB DATE
)
CLUSTER BY (Customer_ID);
```

```
-- Stage creation
CREATE STAGE my_s3_stage_CS url='s3://snowflake-essentials/'
COPY INTO generated_customer_CSV
FROM s3://snowflake-essentials/ingesting_data/new_customer/2019-09-
24/generated_customer_data.csv
FILE_FORMAT = (TYPE = CSV FIELD_DELIMITER = '|' SKIP_HEADER = 1)
-- Age column added
ALTER TABLE generated_customer_CSV
ADD COLUMN age INT AS (DATE_PART('year', CURRENT_DATE()) - DATE_PART('year',
CUSTOMER_DOB));
SELECT * FROM generated_customer_CSV;
-- Conditions
-- Materialized view creation with conditions
CREATE MATERIALIZED VIEW generated_customer_mv AS
SELECT CUSTOMER_NAME AS CustomerName,
AGE AS CustomerAge,
CUSTOMER_CITY AS CustomerCity
FROM generated_customer_CSV
WHERE AGE > 30 AND AGE < 50;
SELECT * FROM generated_customer_mv;
SHOW TABLES;
ALTER ACCOUNT SET AUTO_CLUSTERING = ON;
```

Output:


Worksheets
CaseStudy
Question_1
ACCOUNTADMIN
SNOW_TUTO_WH
Share
Updated 1 minute ago
Worksheets
Databases
Search
CaseStudy
Question_3
Question_2
Question_1
learn_from_vinit
Pipe
vinit_28
VINIT_27-03
tuto
pipe_tuto
json_file
day2
snowflake_ass_csv_2
snowflake_ass_csv_1
ass-1-demo
batch_02_day1
COVID_CASESTUDY.PUBLIC
23 COPY INTO generated_customer_CSV
24 FROM s3://snowflake-essentials/ingesting_data/new_customer/2019-09-24/generated_customer_data.csv
25 FILE_FORMAT = (TYPE = CSV FIELD_DELIMITER = '|' SKIP_HEADER = 1)
26
27
28 -- Age column added
Objects
Editor
Results
Chart
CUSTOMERNAME
CUSTOMERAGE
CUSTOMERCITY
1 Jennifer Frederick 41 Canberra
2 Julian Hammond 43 Palmerston
3 David Nicholson 39 Redcliffe
4 Martena Perez 43 Cessnock
5 Tad Riddle 42 Lithgow
6 Dexter Stone 37 Brisbane
7 Maile Gallagher 40 Belgrave
8 Alisa Keith 34 Maryborough
9 Gannon Glenn 37 Cessnock
10 Kuame Casey 47 Geraldton-Greenough
11 Caldwell Dennis 38 Tamworth
Query Details
Query duration 504ms
Rows 33
CUSTOMERNAME Aa
100% filled
CUSTOMERAGE 123
32 49
CUSTOMERCITY Aa
Cessnock 2

2.2 Question 2

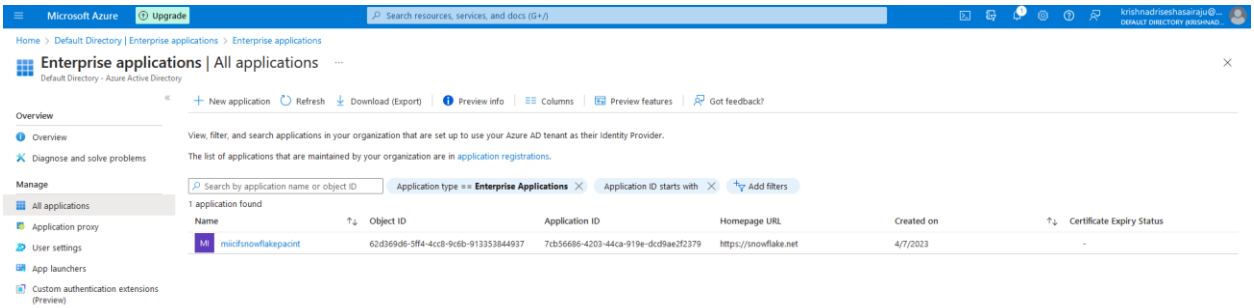
Design Job (using ADF) to read the file (data_20230330.csv) from blob storage and load into SQL Server.

SQL Query:

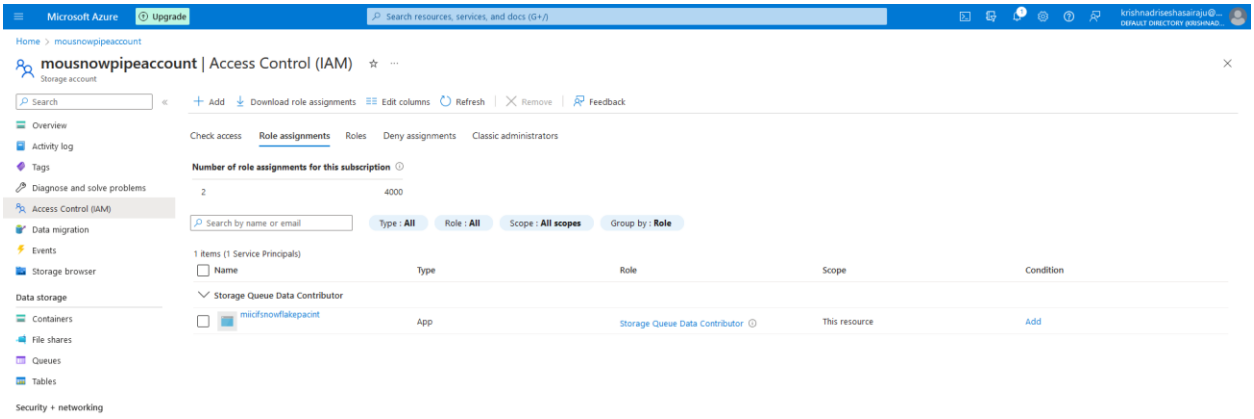
```
SELECT continentExp AS continents,
      countriesAndTerritories AS countries,
      CAST(year AS VARCHAR(255)) || '-' || CAST(month AS VARCHAR(255)) AS Year_Month,
      SUM(CAST(cases AS INT)) AS cases,
      SUM(CAST(deaths AS INT)) AS deaths
   FROM COVID19_DATA
   WHERE year = '2020'
   GROUP BY continentExp, countriesAndTerritories, CAST(year AS VARCHAR(255)) || '-' ||
CAST(month AS VARCHAR(255));
-- Performing the integration process
CREATE NOTIFICATION INTEGRATION SNOWPIPE_INT_EVENT_CS
ENABLED = TRUE
TYPE = QUEUE
NOTIFICATION_PROVIDER = azure_storage_queue
azure_storage_queue_primary_uri =
'https://mousnowpipeaccount.queue.core.windows.net/snowpipe-noti-queue'
AZURE_TENANT_ID = '94124bcb-dea0-47d3-bfba-57b969f441dd'
SHOW INTEGRATIONS;
DESC notification integration SNOWPIPE_INT_EVENT_CS;
CREATE OR REPLACE STAGE "COVID_CASESTUDY"."PUBLIC"."AZURE_STAGE_CS"
 url = 'azure://mousnowpipeaccount.blob.core.windows.net/snowflake-blob-src/'
 credentials = (azure_sas_token=
  '?sv=2021-12-02&ss=bfqt&srt=sco&sp=rwdlacupiytfx&se=2023-05-
09T19:43:45Z&st=2023-04-
07T11:43:45Z&spr=https&sig=Cxk0WRu%2BRYAec3CzDXEFtSAbkqYaAw4TEITCL%2F4Ids
```

```
k%3D'
);
ls @"COVID_CASESTUDY"."PUBLIC"."AZURE_STAGE_CS";
-- Creating the pipe
CREATE OR REPLACE TABLE "COVID_CASESTUDY"."PUBLIC"."COVID19_DATA" (
  dateRep STRING,
          day STRING,
          month STRING,
          year STRING,
          cases STRING,
          deaths STRING,
          countriesAndTerritories STRING     ,
          geoId  STRING,
          countryterritoryCode STRING,
          popData2019 STRING,
          continentExp STRING,
          "Cumulative_number_for_14_days_of_COVID-19_cases_per_100000" STRING
);
-- Creating the PIPE Between the Blob and Sowflake
CREATE OR REPLACE PIPE "COVID_CASESTUDY"."PUBLIC"."AZURE_SNOWFLAKE_PIPE"
  auto_ingest = true
  integration = 'SNOWPIPE_INT_EVENT_CS'
  AS
  COPY INTO "COVID_CASESTUDY"."PUBLIC"."COVID19_DATA"
  FROM @"COVID_CASESTUDY"."PUBLIC"."AZURE_STAGE_CS"
  file_format = (type = 'CSV' FIELD_DELIMITER = ',');
  SHOW PIPES;
-- Load the files to the database
ALTER PIPE "COVID_CASESTUDY"."PUBLIC"."AZURE_SNOWFLAKE_PIPE" REFRESH;
SELECT * FROM "COVID_CASESTUDY"."PUBLIC"."COVID19_DATA";
```

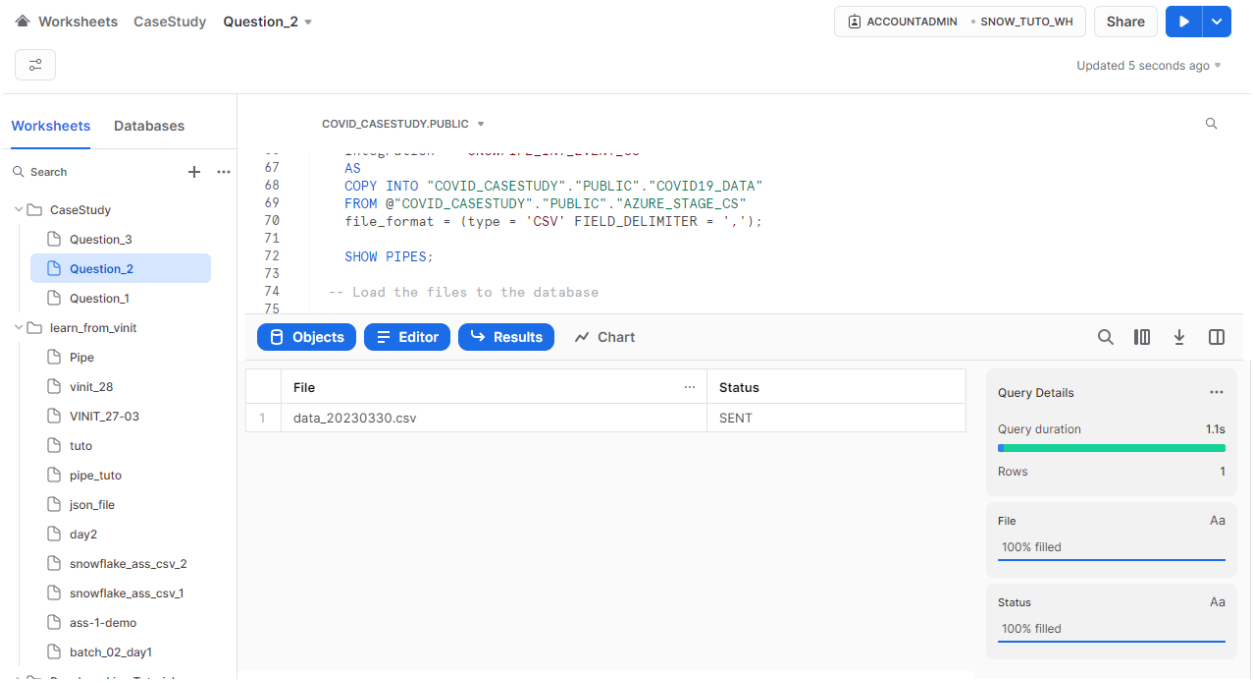
Azure Active Directory:



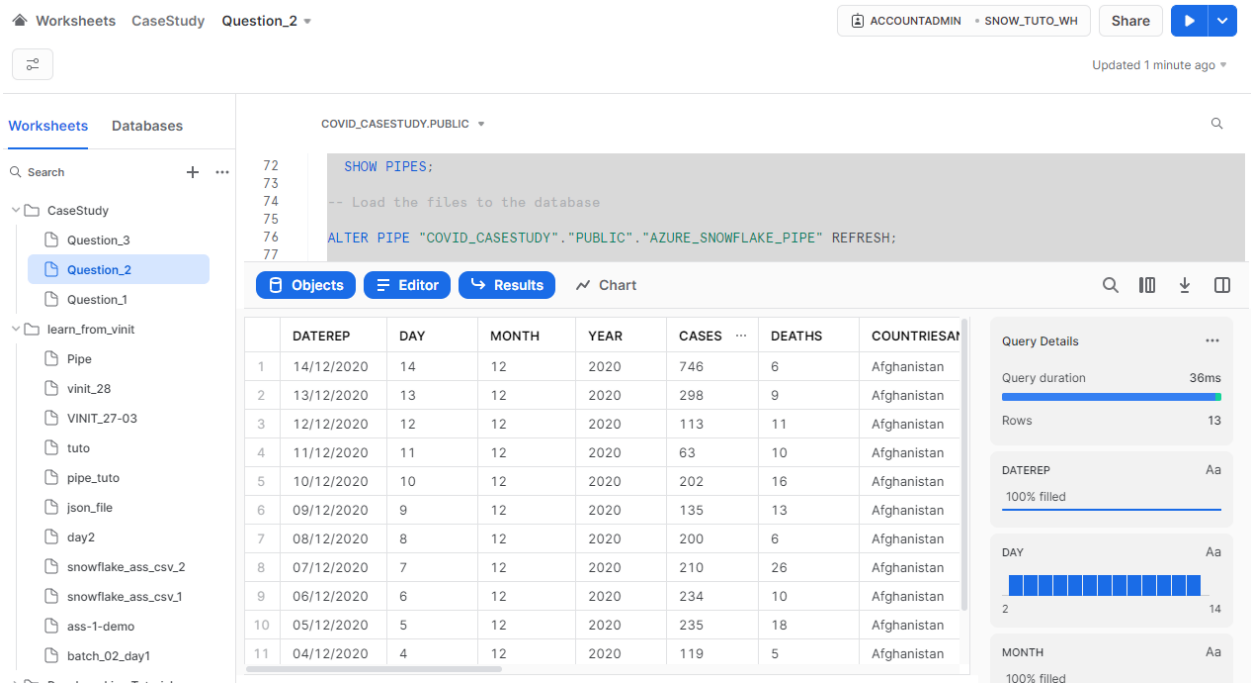
Role Assignment for Snowflake:



Pipe Status:



Output:



2.3 Question 3

Participants will create tasks for the snow pipe they have created in Question No. 2. Output table and file will consist of following columns for Year 2020:

a. Continent (e.g... Asia)

b. Countries (e.g... Afghanistan)

c. Year Month (e.g., 2020-01)

d. Cases (e.g., 0)

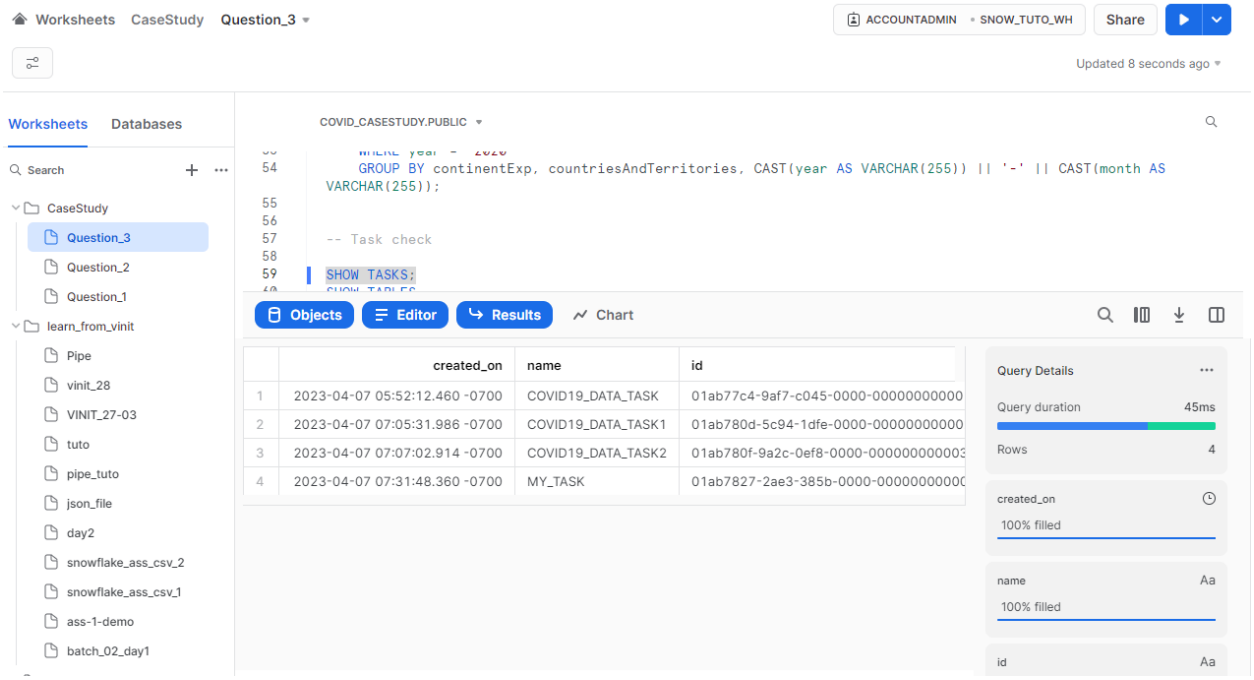
e. Deaths (e.g., 0)

SQL Query:

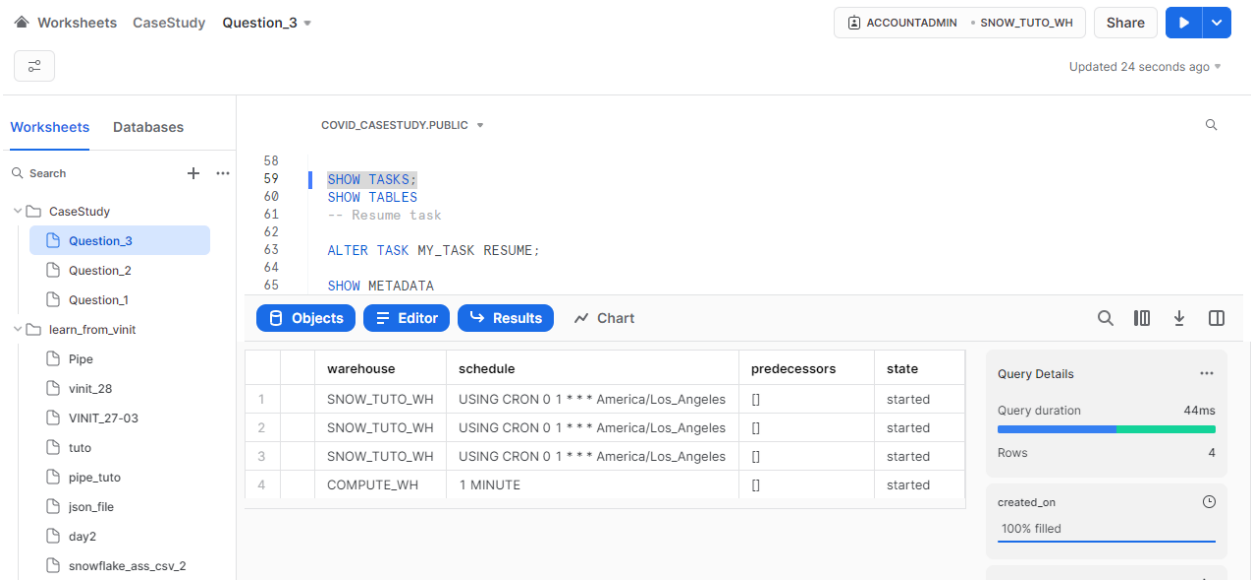
```
-- Create a task that references the stored procedure and specifies the schedule for the task to run
CREATE TASK COVID19_DATA_task2
WAREHOUSE = SNOW_TUTO_WH
SCHEDULE = 'USING CRON 0 1 * * * America/Los_Angeles'
AS
CALL COVID19_DATA_task1_pro();
CREATE OR REPLACE TASK my_task
  WAREHOUSE = COMPUTE_WH
  SCHEDULE = '1 MINUTE'
AS
  -- Create a target table to store the result
  CREATE TABLE IF NOT EXISTS output_table (
    continents VARCHAR(255),
    countries VARCHAR(255),
    Year_Month VARCHAR(255),
    cases INT,
```

```
    deaths INT
  );
  -- Insert the result of the query into the target table
  INSERT INTO output_table (continents, countries, Year_Month, cases, deaths)
  SELECT continentExp AS continents,
    countriesAndTerritories AS countries,
    CAST(year AS VARCHAR(255)) || '-' || CAST(month AS VARCHAR(255)) AS Year_Month,
    SUM(CAST(cases AS INT)) AS cases,
    SUM(CAST(deaths AS INT)) AS deaths
  FROM COVID19_DATA
  WHERE year = '2020'
  GROUP BY continentExp, countriesAndTerritories, CAST(year AS VARCHAR(255)) || '-' ||
CAST(month AS VARCHAR(255));
-- Task check
SHOW TASKS;
SHOW TABLES
-- Resume task
ALTER TASK MY_TASK RESUME;
--CHECK TASK HISTORY
SELECT * FROM TABLE(INFORMATION_SCHEMA.TASK_HISTORY()) WHERE NAME =
'MY_TASK';
--DISPLAY
SELECT * FROM OUTPUT_TABLE;
```

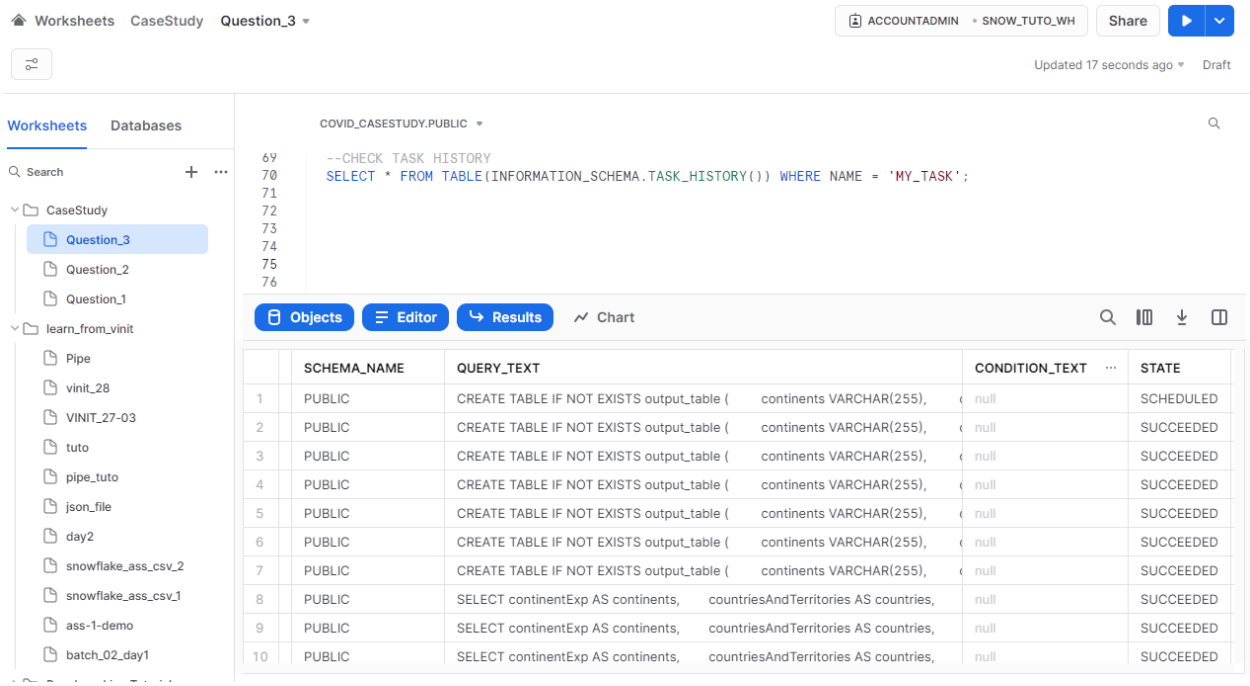
Created Task:



Resumed Task:



Task History:



Output Table named as ‘output_table: