Case Study

on

**Snowflake**

**Case study by,**

Mouli S,

Batch – 02,

moulisankar2002@outlook.com

**Contents**

1. Understandings of Case Study ……………………………………………3
2. Case Study Questions ………………………………………………………….4-14
   1. Question 1 …………………………………………………………4
   2. Question 2 …………………………………………………………7
   3. Question 3 …………………………………………………………11

**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**](https://s3.apsoutheast2.amazonaws.com/snowflake-essentials/ingesting%20data/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

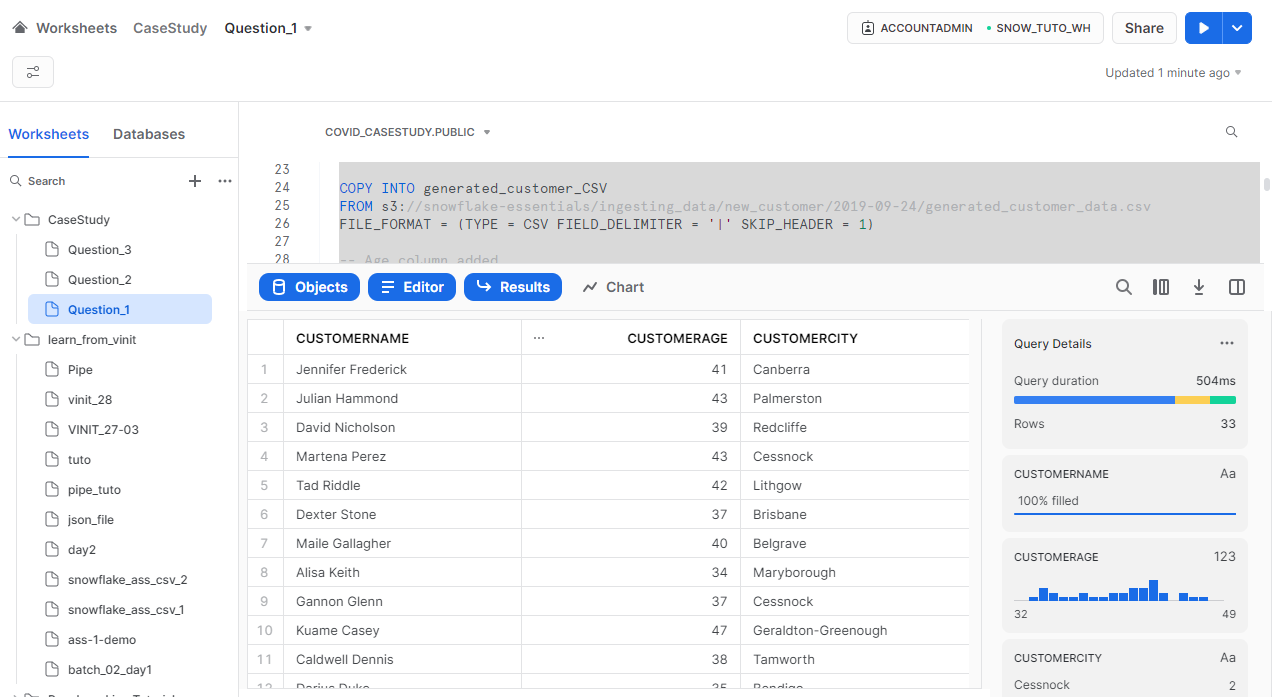
WHERE AGE > 30 AND AGE < 50;

SELECT \* FROM generated\_customer\_mv;

SHOW TABLES;

ALTER ACCOUNT SET AUTO\_CLUSTERING = ON;

**Output:**

****

**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%2F4Idsk%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 = ',');

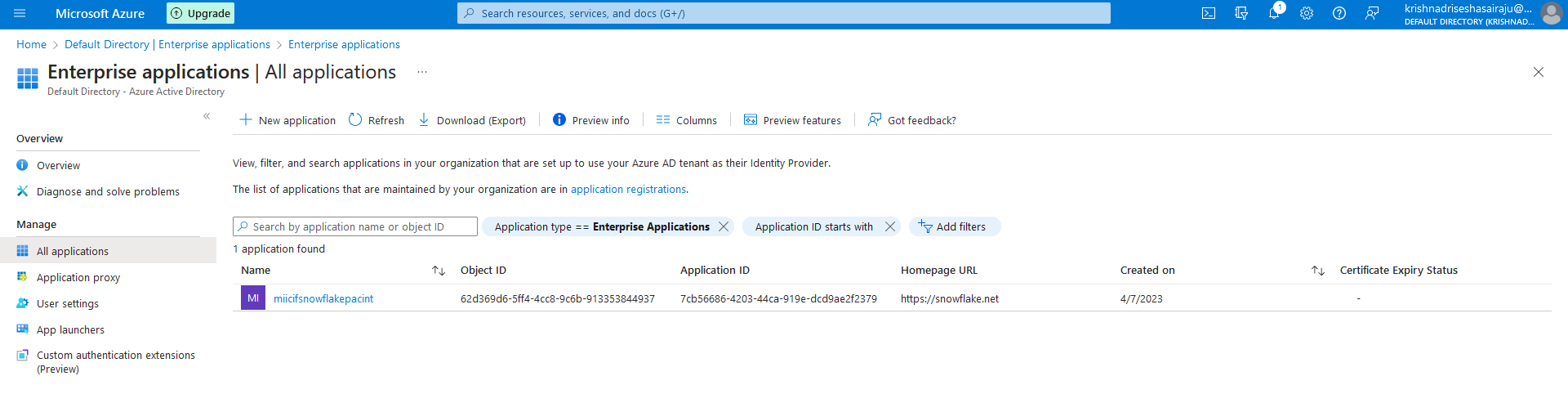
SHOW PIPES;

-- Load the files to the database

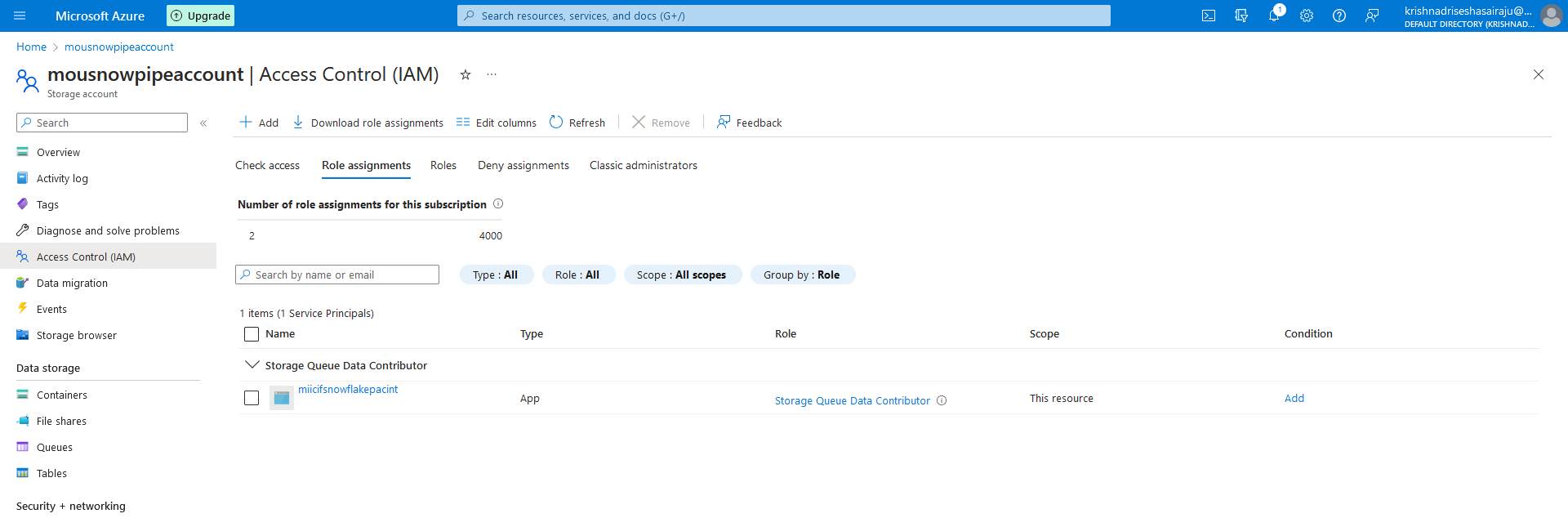
ALTER PIPE "COVID\_CASESTUDY"."PUBLIC"."AZURE\_SNOWFLAKE\_PIPE" REFRESH;

SELECT \* FROM "COVID\_CASESTUDY"."PUBLIC"."COVID19\_DATA";

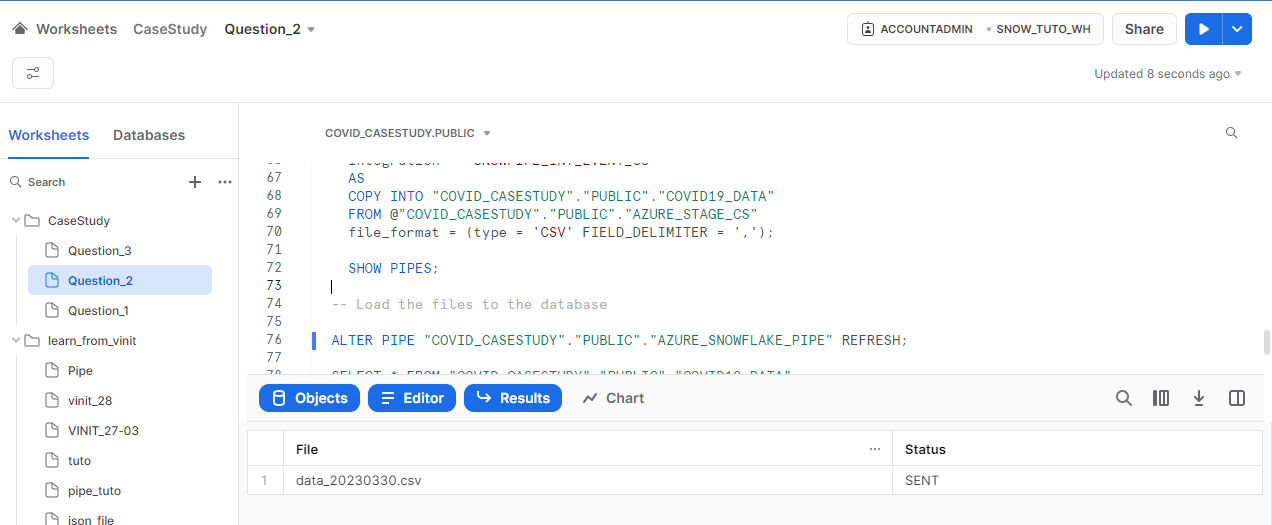
**Azure Active Directory:**

****

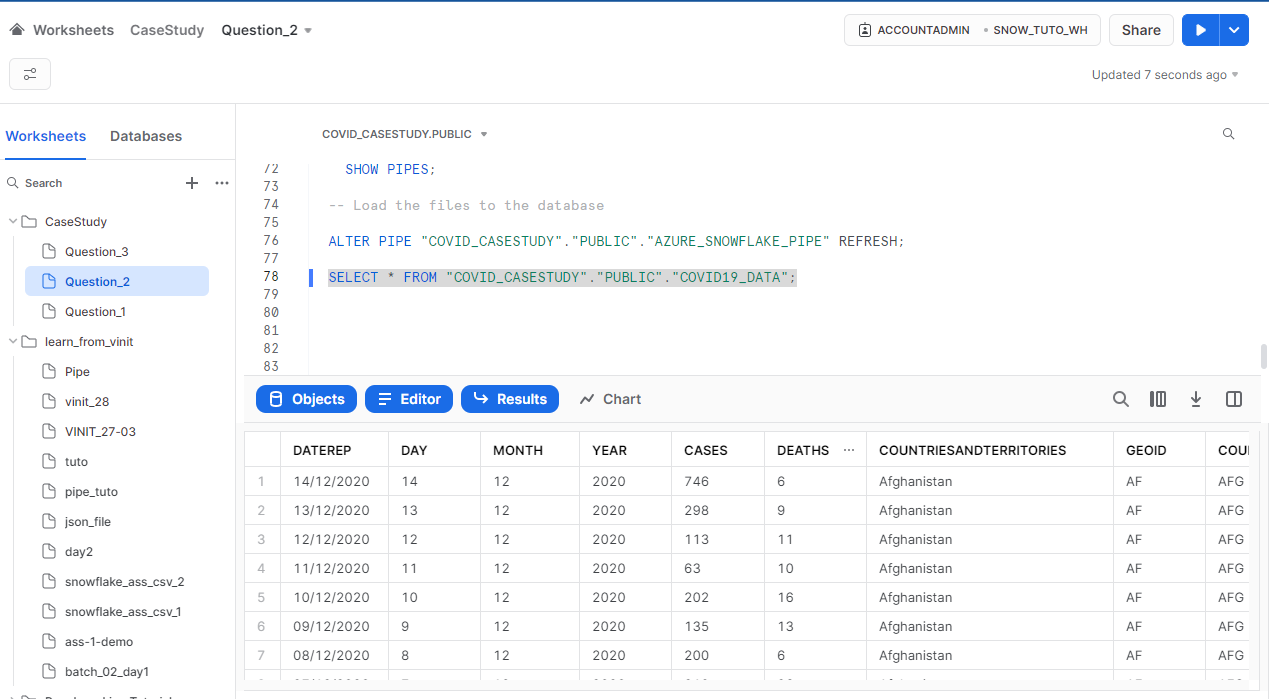
**Role Assignment for Snowflake:**

****

**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;

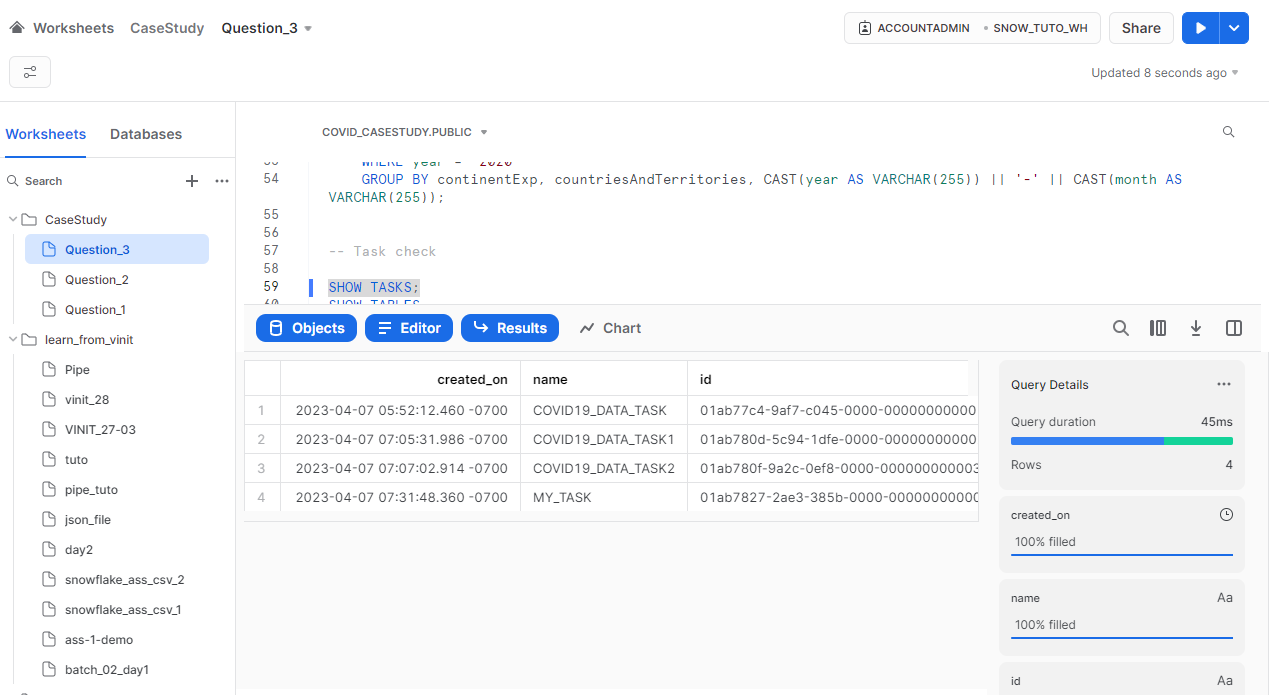
--CHECK TASK HISTORY

SELECT \* FROM TABLE(INFORMATION\_SCHEMA.TASK\_HISTORY()) WHERE NAME = 'MY\_TASK';

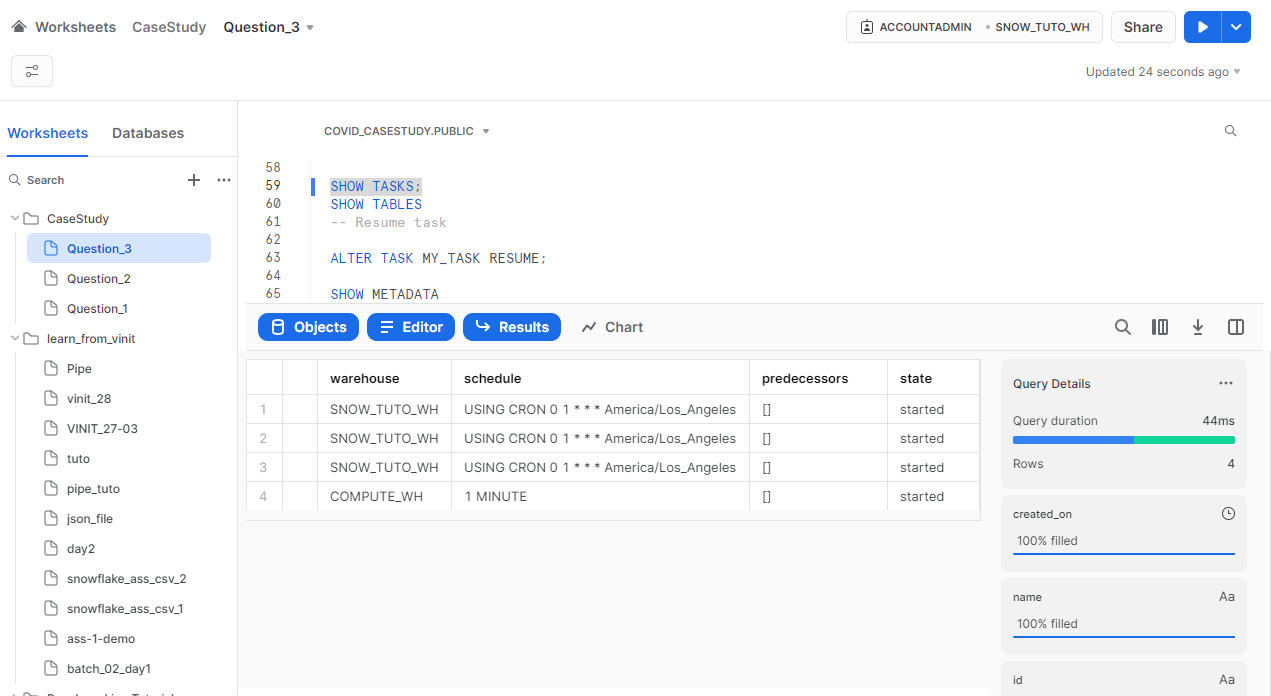
--DISPLAY

SELECT \* FROM OUTPUT\_TABLE;

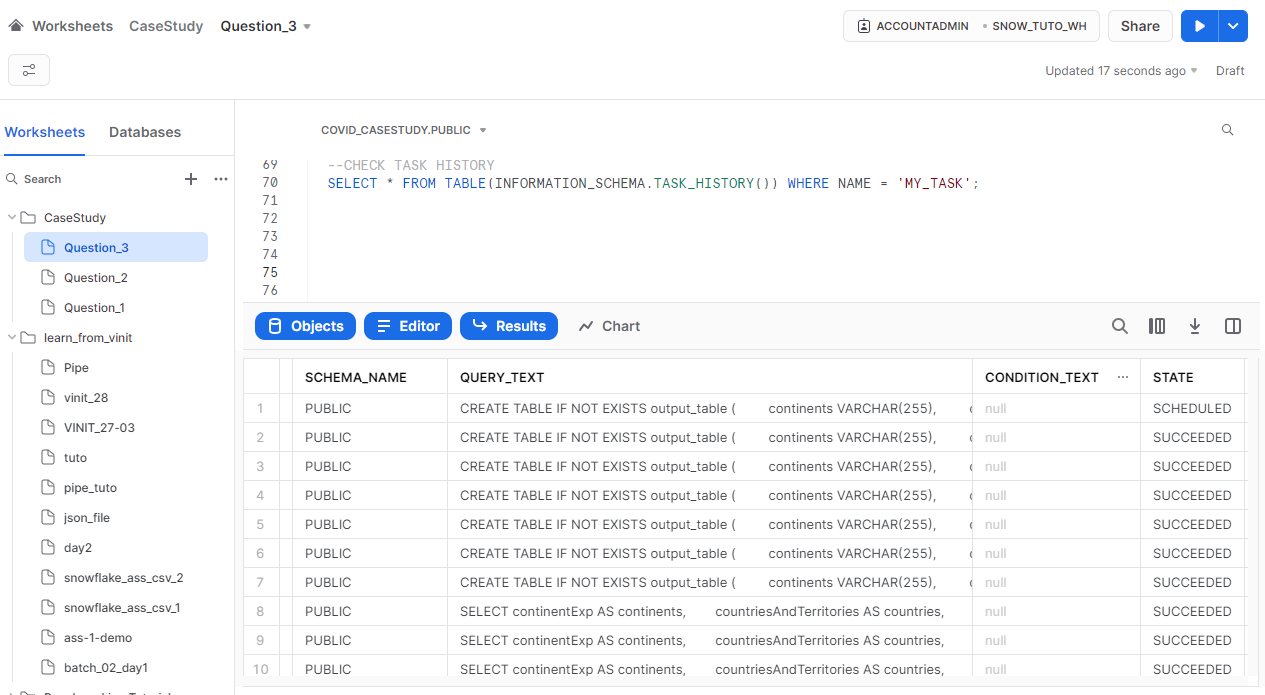
**Created Task:**

****

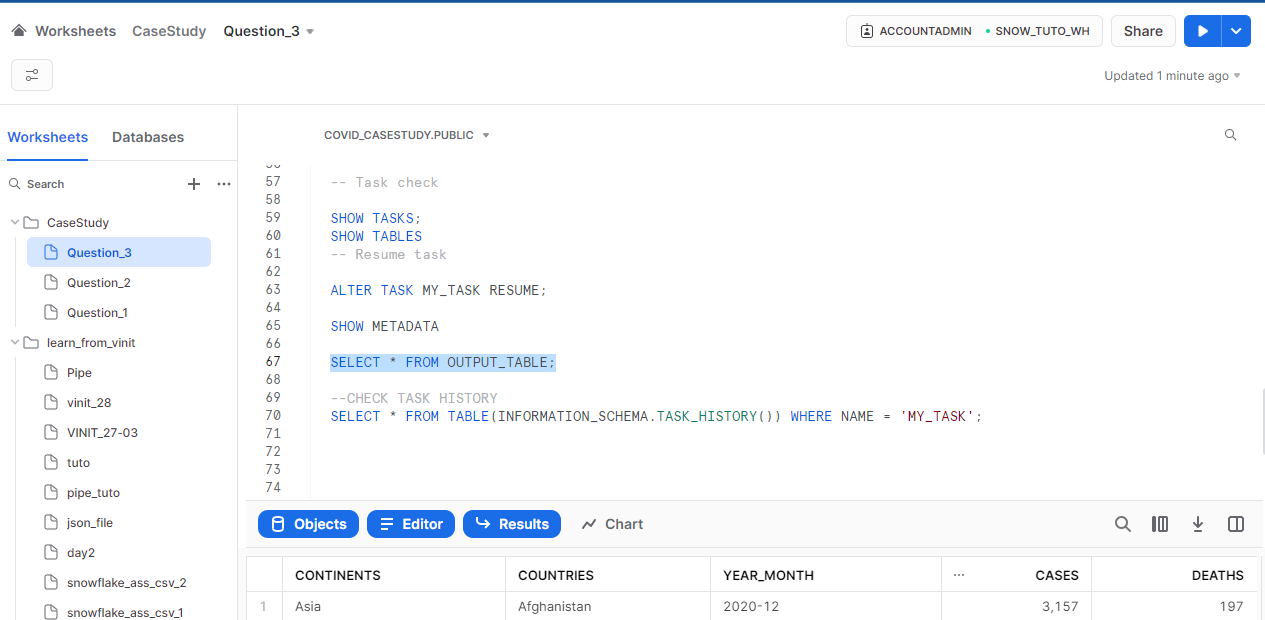
**Resumed Task:**

****

**Task History:**

****

**Output Table named as ‘output\_table:**

****