System Requirements Specification Index

For

Snowflake ETL Query surge testing use case

Version 1.0

IIHT Pvt. Ltd.
fullstack@iiht.com

Problem Statement : Snowflake ETL Testing use case

Description : Use relevant methods operations toperform specified activities which are given in the instructions.

A school district wants to centralize student information from multiple schools for reporting and analysis purposes. Each school provides daily updates in flat files (CSV format) containing student information such as attendance, grades, and personal details. The goal is to load this data into Snowflake for easy access and reporting.

Objectives

1. **Extract** student information data from flat files.
2. **Transform** the data to ensure it is clean, standardized, and enriched.
3. **Load** the transformed data into Snowflake for further analysis and reporting.

. Extract Data from Flat Files

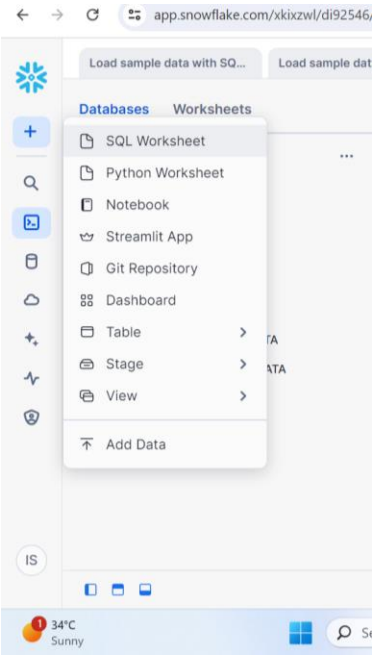
- **Source**: Daily CSV files from multiple schools.
- **Fields**:
- StudentID (NUMBER)
- Age (NUMBER)
- GenderEthnicity (NUMBER)
- ParentalEducation(VARCHAR)
- StudyTimeWeekly (VARCHAR)
- Absences(VARCHAR)
- Tutoring(VARCHAR)
- ParentalSupport(VARCHAR)
- Extracurricular(VARCHAR)
- Sports(VARCHAR)
- Music(VARCHAR)
- Volunteering(VARCHAR)
- GPA(NUMBER)

- Steps to login in the snowflake account

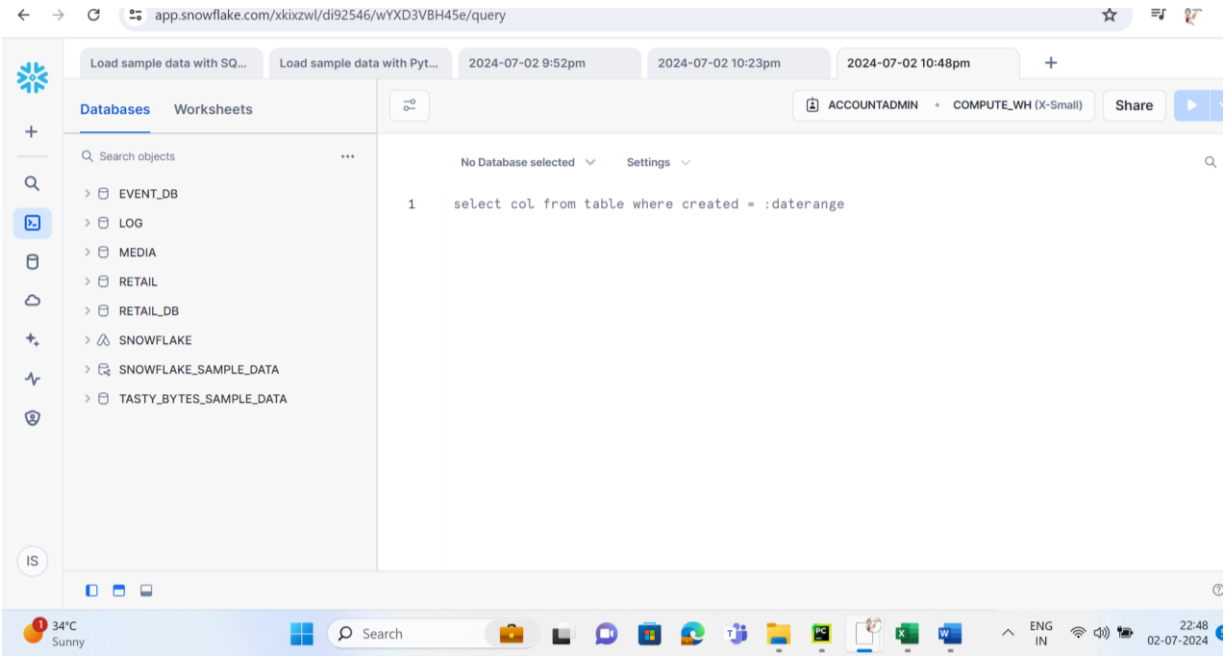
- Use the credential given to you though your host

Note :

PLEASE CREATE THE DATATYPES ACCORDING TO THE SPECIFICATIONS

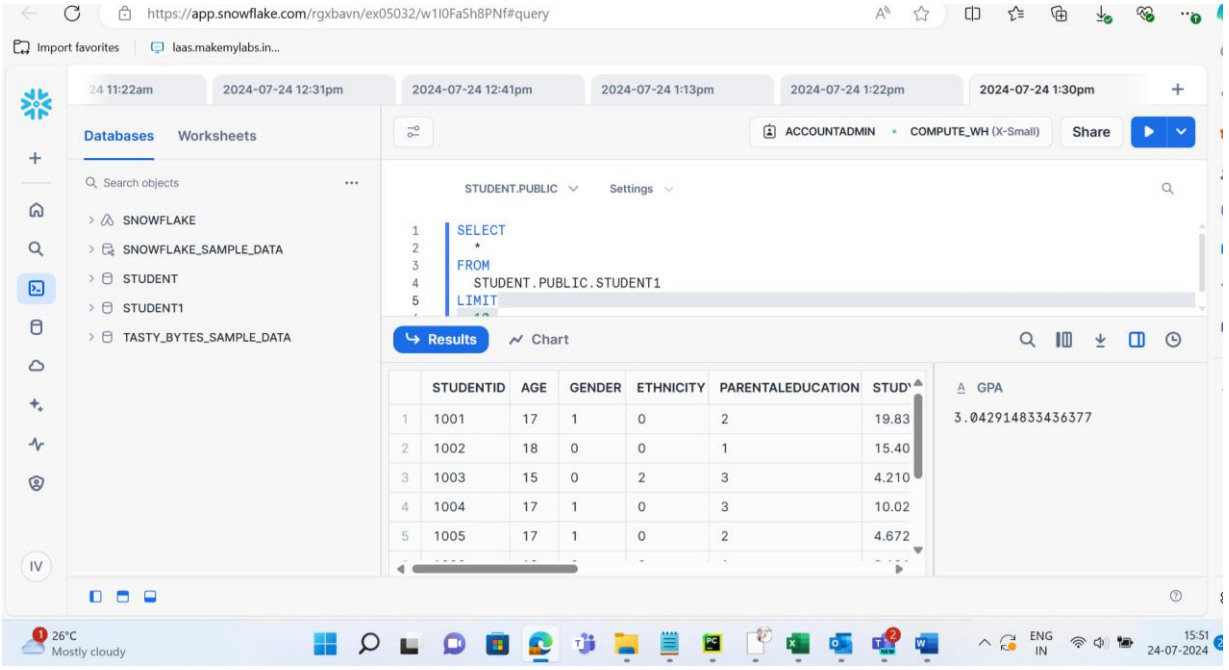


- Click on the sql worksheet to open the query editor

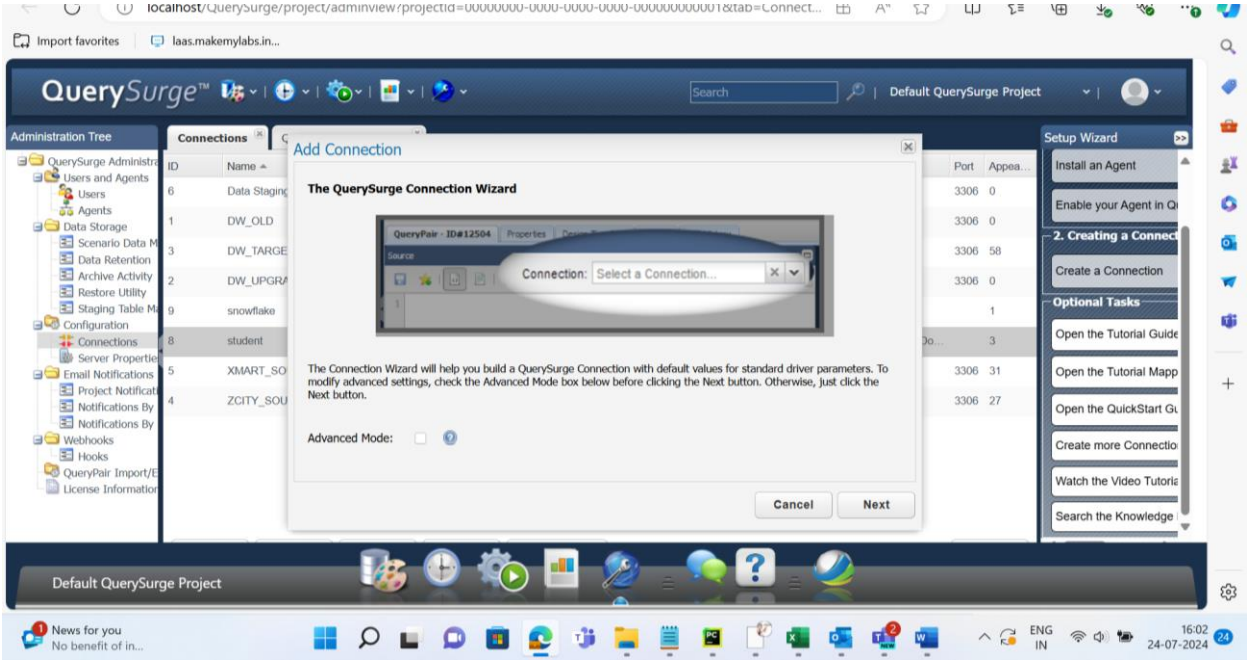


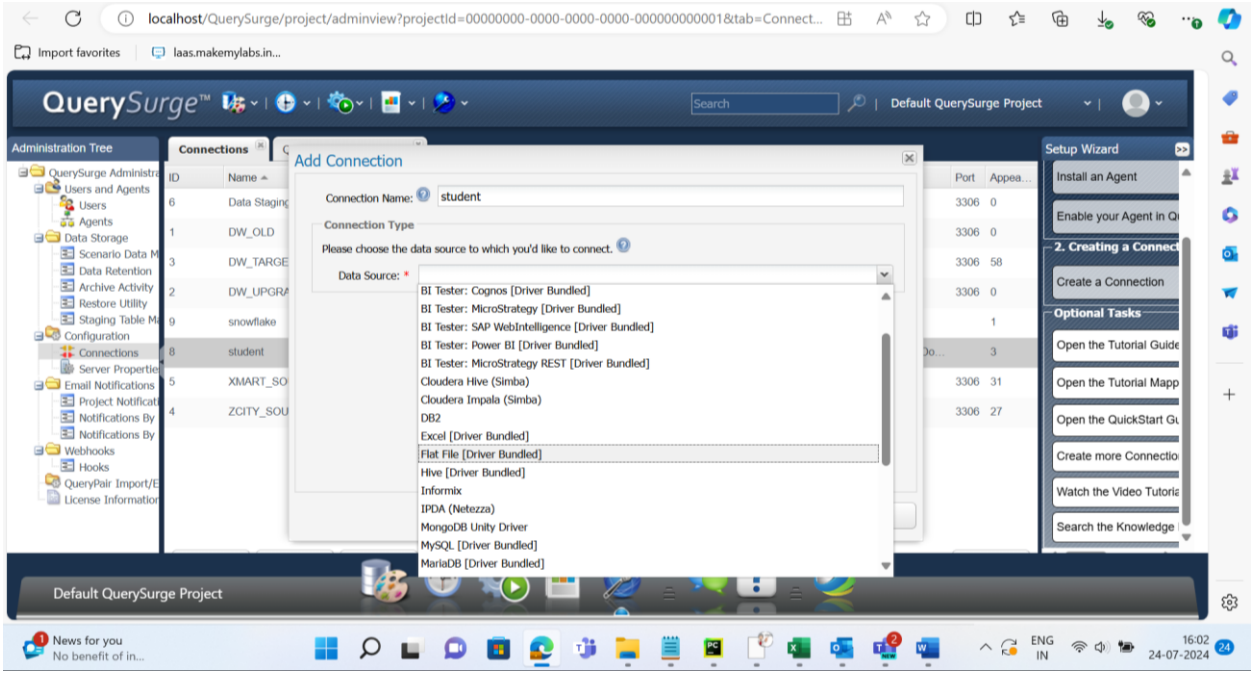
You will need to create the database and move inside the database

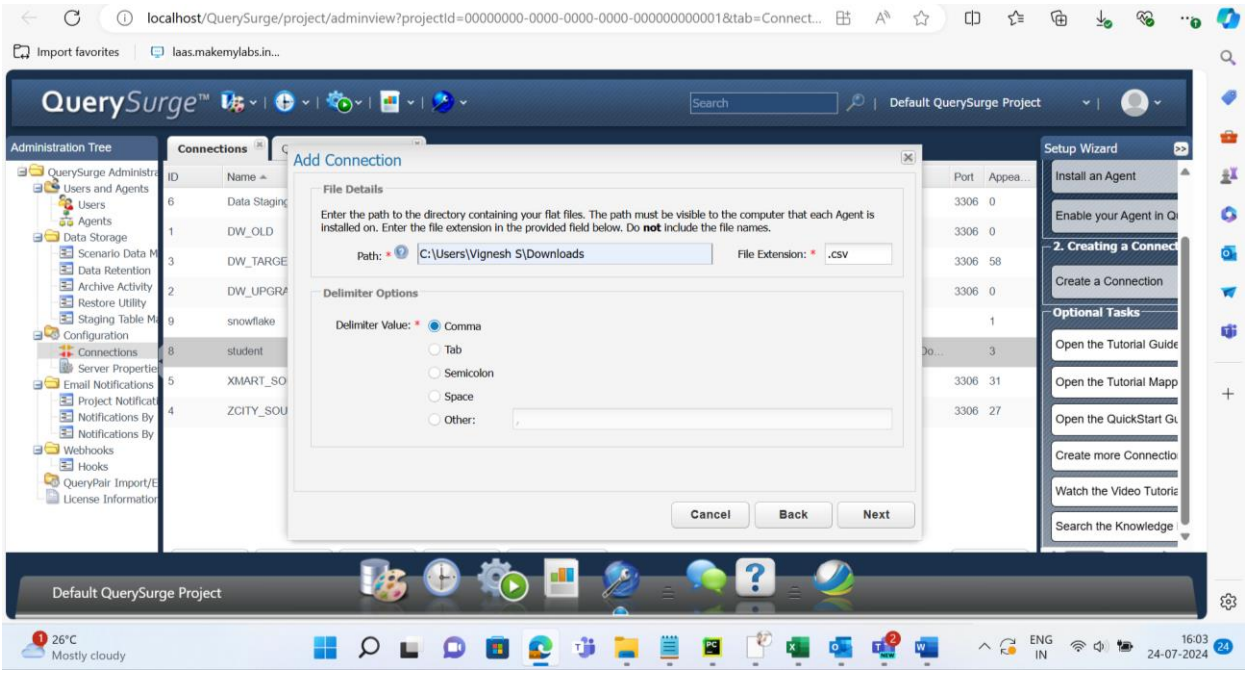
Once you have imported the table check for respective rows and columns are imported correctly.



In query surge we need to import flat file

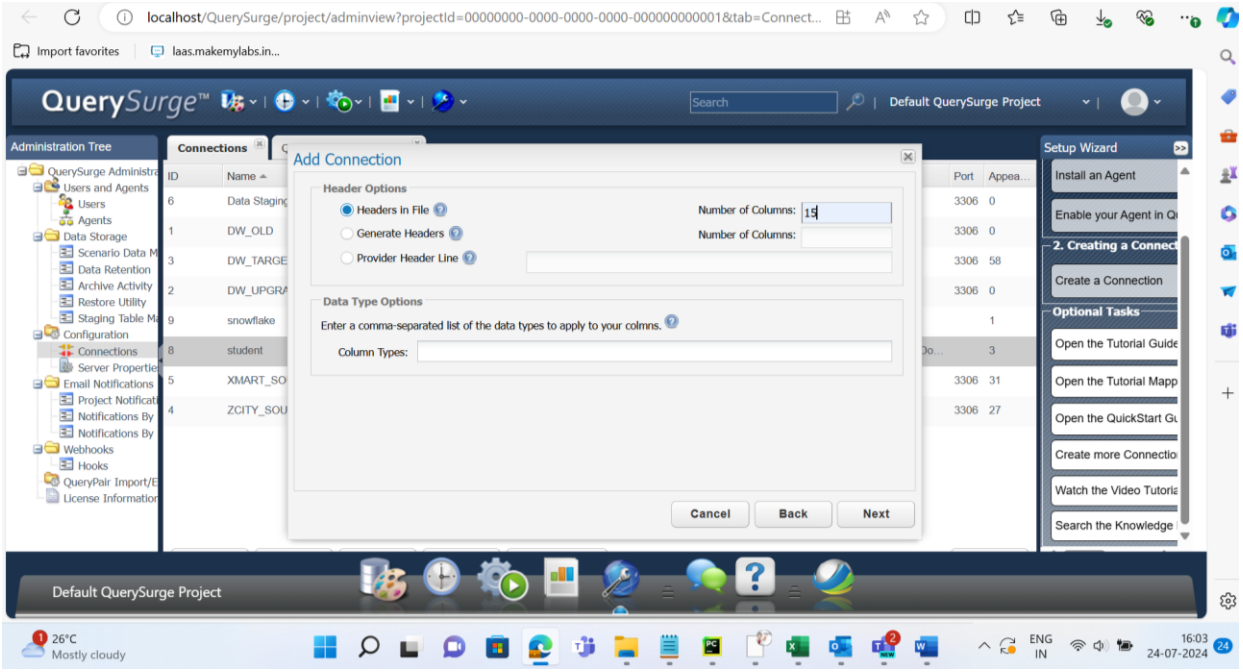




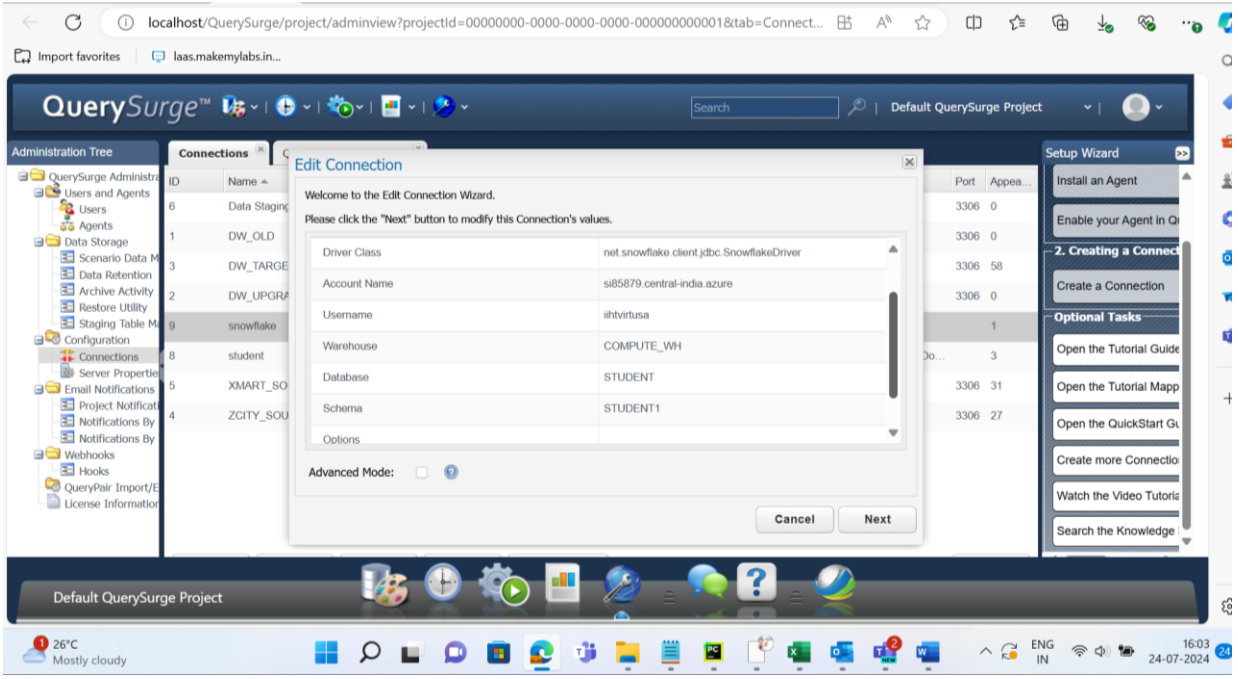


Give the path for the csv file

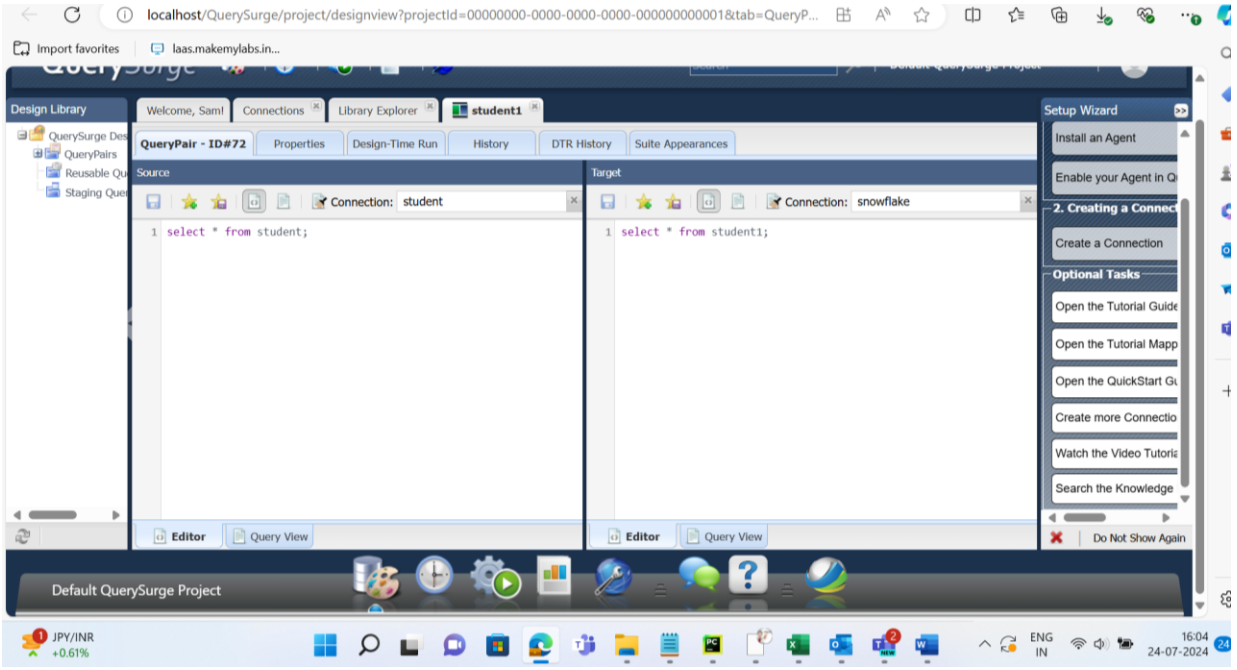
Provide the input columns in the corresponding input



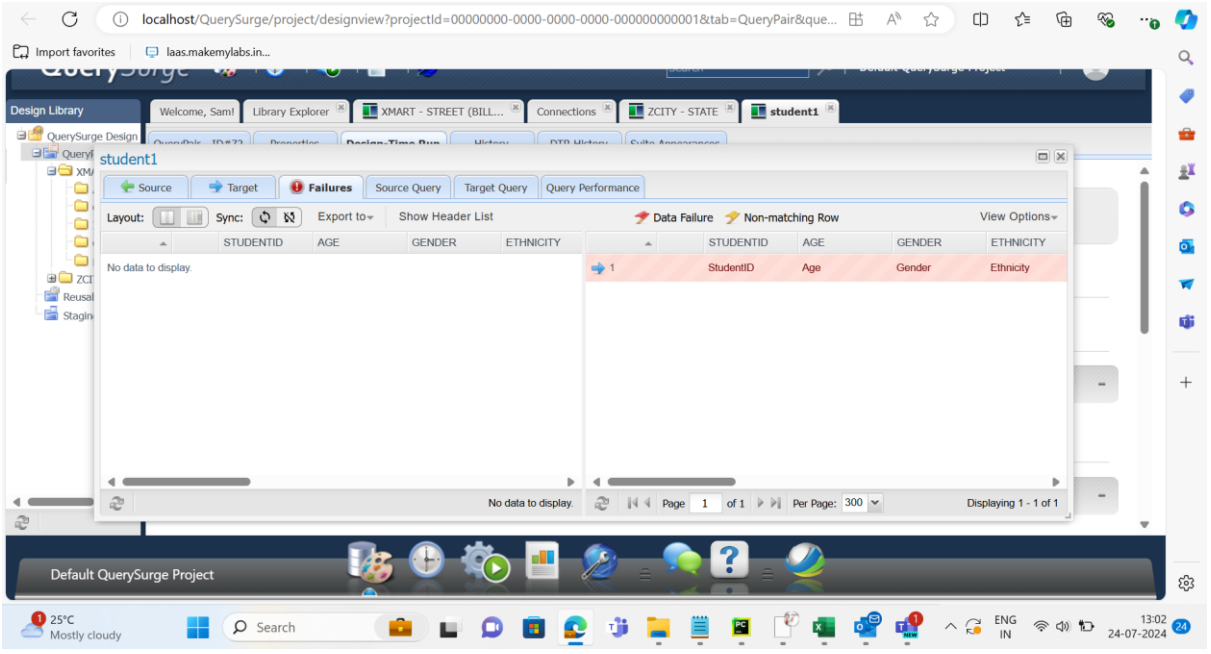
Snowflake connection

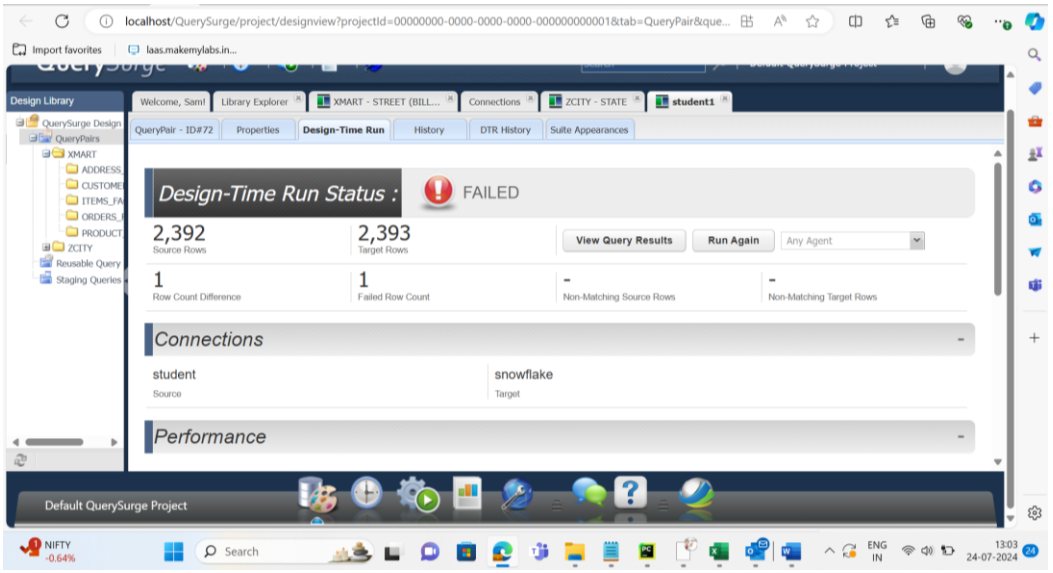


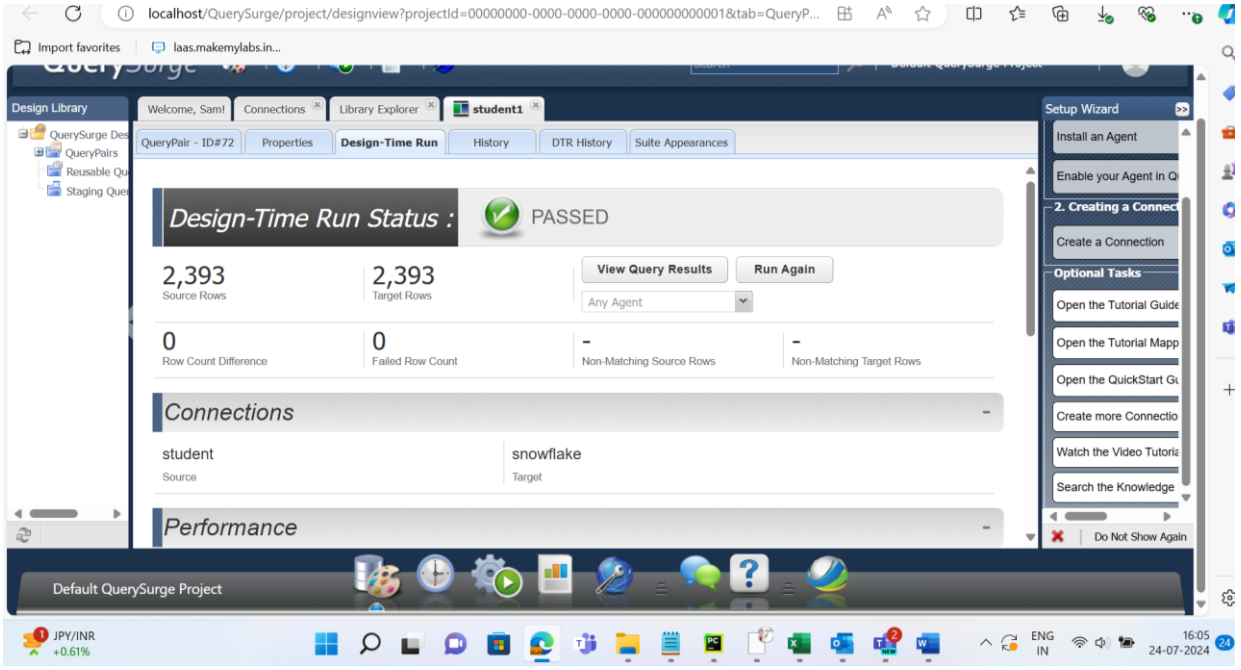
Perform two queries in source and target and run the query to validate the values in the tables

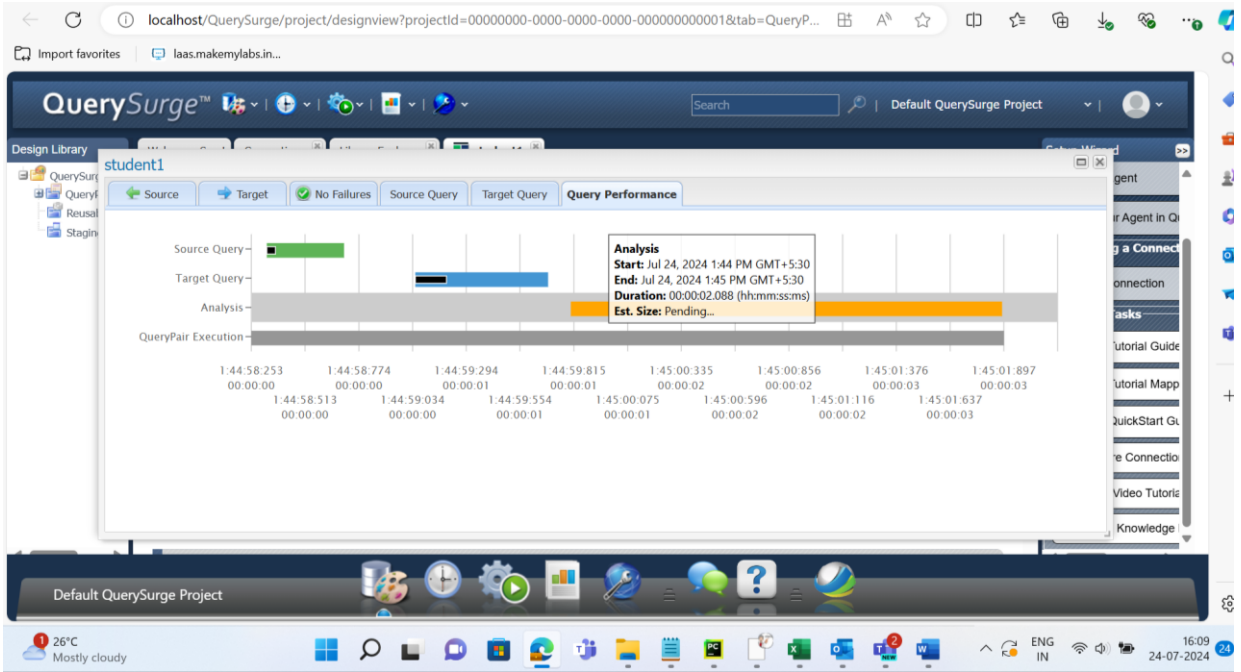


Example of failed data comparison




Design Library
Welcome, Sam!
Library Explorer
XMART - STREET (BILL...
Connections
ZCITY - STATE
student1
QueryPair - ID#72
Properties
Design-Time Run
History
DTR History
Suite Appearances
Design-Time Run Status :
FAILED
2,392
Source Rows
2,393
Target Rows
View Query Results
Run Again
Any Agent
1
Row Count Difference
1
Failed Row Count
-
Non-Matching Source Rows
-
Non-Matching Target Rows
Connections
student
Source
snowflake
Target
Performance
Default QuerySurge Project


Design Library
Welcome, Sam!
Connections
Library Explorer
student1
QueryPair - ID#72
Properties
Design-Time Run
History
DTR History
Suite Appearances
Setup Wizard
Install an Agent
Create a Connection
Optional Tasks
Design-Time Run Status :
PASSED
2,393
Source Rows
2,393
Target Rows
View Query Results
Run Again
Any Agent
0
Row Count Difference
0
Failed Row Count
-
Non-Matching Source Rows
-
Non-Matching Target Rows
Connections
student
Source
snowflake
Target
Performance
Do Not Show Again
Default QuerySurge Project


QuerySurge
Search
Default QuerySurge Project
Design Library
student1
Source
Target
No Failures
Source Query
Target Query
Query Performance
Source Query
Target Query
Analysis
QueryPair Execution
Analysis
Start: Jul 24, 2024 1:44 PM GMT+5:30
End: Jul 24, 2024 1:45 PM GMT+5:30
Duration: 00:00:02.088 (hh:mm:ss:ms)
Est. Size: Pending...
Default QuerySurge Project

Task A to be performed

1 Upload the data to query surge
2 Upload the data to snowflake
3 Configure query surge for snowflake and flatfile connection
4 Creating query to load the data from source to target
5 Compare two tables with source table and target table for changes
6 Check and modify using queries for query testing

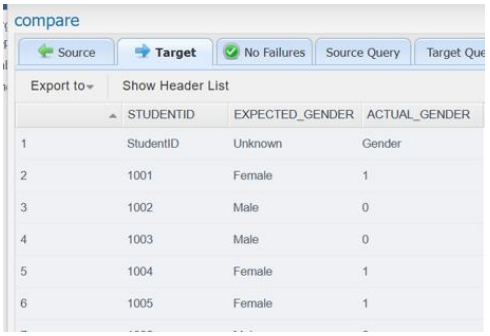
Task B to be performed

Data Completeness Testing

1. Query to count records in both source and target student tables:

Conditional Value Mapping

2. Check if the Gender values are mapped correctly between the source and target tables Where I is female and 0 male others unknown .



Data integrity testing

3. To check whether a specific column, such as GPA, matches between Snowflake and CSV
4. Check if there null values present in the dataset

Duplicate data testing

Using the existing tables check for duplicates based on a combination of multiple columns with StudentID, Age, and Gender

Data Transformation Testing

5. If GPA is greater than or equal to 3.5, classify as "High"; otherwise, "Low".

Filtering Transformation Validation

6. Only include students with GPA above 3.0

Execution Steps to Follow:

1. Open the snowflake console
2. Import the dataset from the document
3. Perform all the query respective to the question provided
4. Take screenshots of the query execution
5. Upload the code to the Github
6. Download the report from query surge

-----x-----