
System Requirements Specification Index

For

Snowflake ETL
Query surge testing
use case

Version 1.0

Problem Statement : Snowflake ETL Testing use case

Description : Use relevant methods operations to perform 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

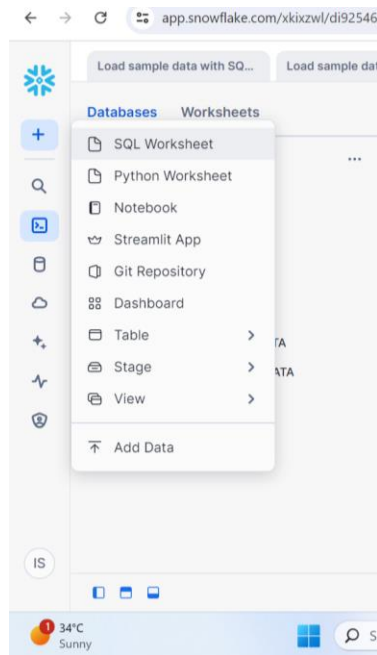
1. **Extract** student information data from flat files.
2. **Transform** the data to ensure it is clean, standardized, and enriched.
3. **Load** the transformed data into Snowflake for further analysis and reporting.

. Extract Data from Flat Files

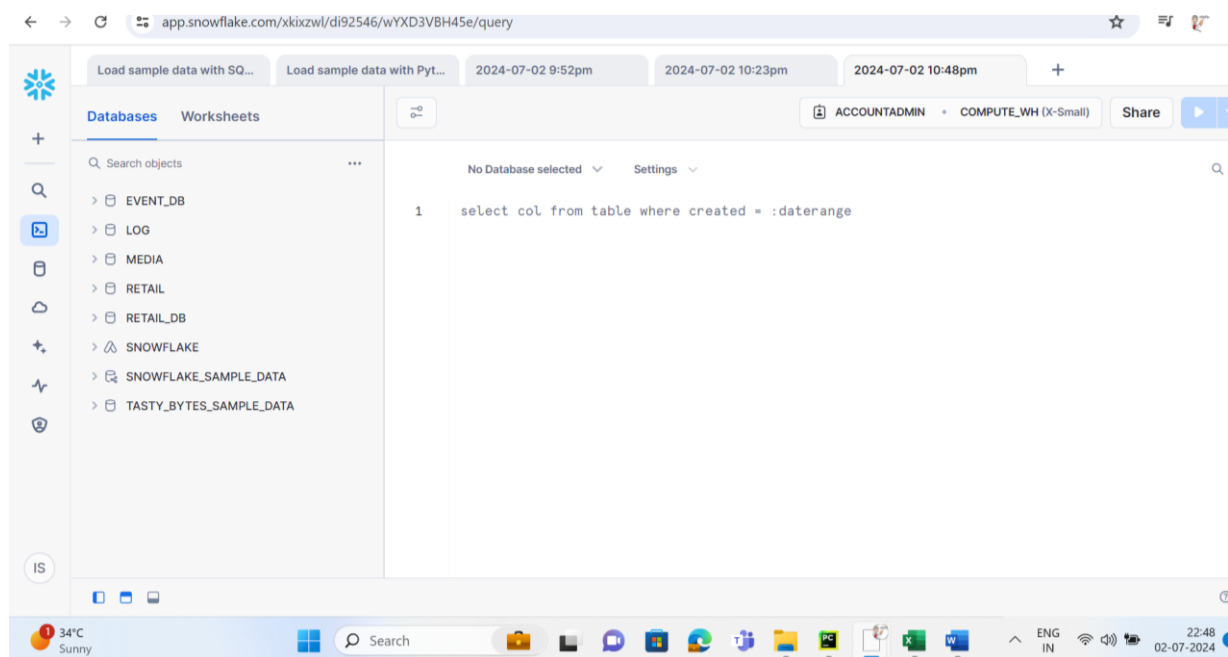
- **Source:** Daily CSV files from multiple schools.
- **Fields:**
 - StudentID (NUMBER)
 - Age (NUMBER)
 - GenderEthnicity (NUMBER)
 - ParentalEducation(VARCHAR)
 - StudyTimeWeekly (VARCHAR)
 - Absences(VARCHAR)
 - Tutoring(VARCHAR)
 - ParentalSupport(VARCHAR)
 - Extracurricular(VARCHAR)
 - Sports(VARCHAR)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

The screenshot shows the Snowflake web interface. The URL is <https://app.snowflake.com/rgxbavn/ex05032/w110FaSh8PNf#query>. The interface includes a sidebar with navigation options like Databases, Worksheets, and Search objects. The main area displays a SQL query:

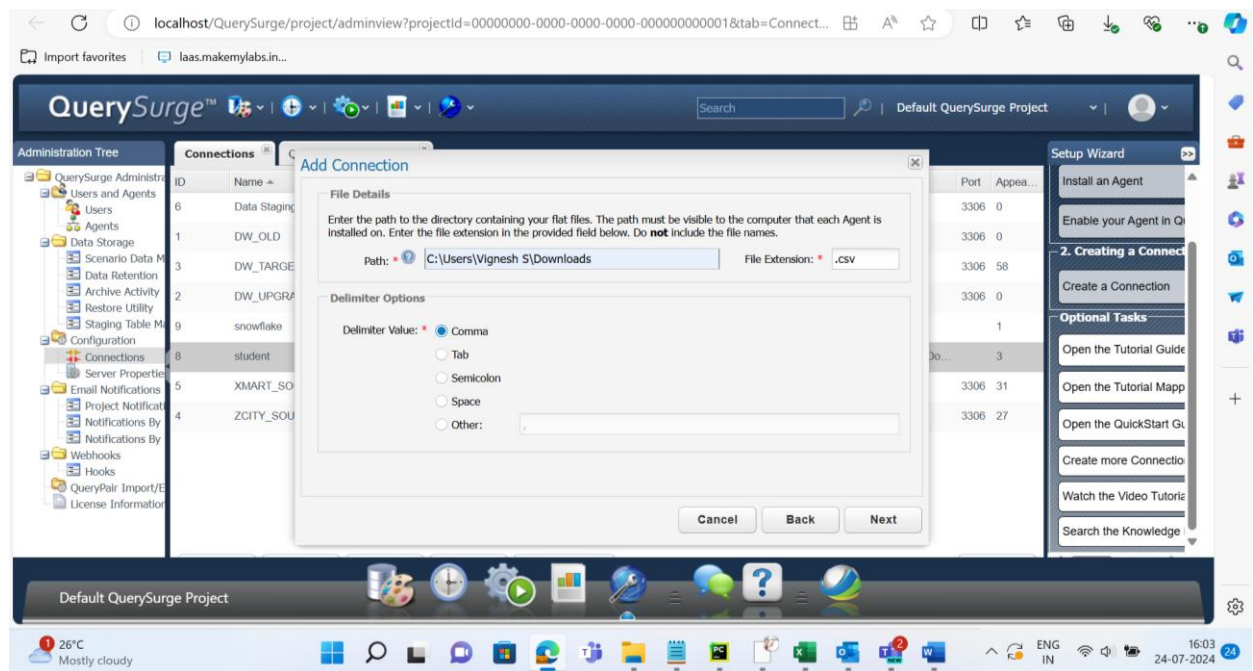
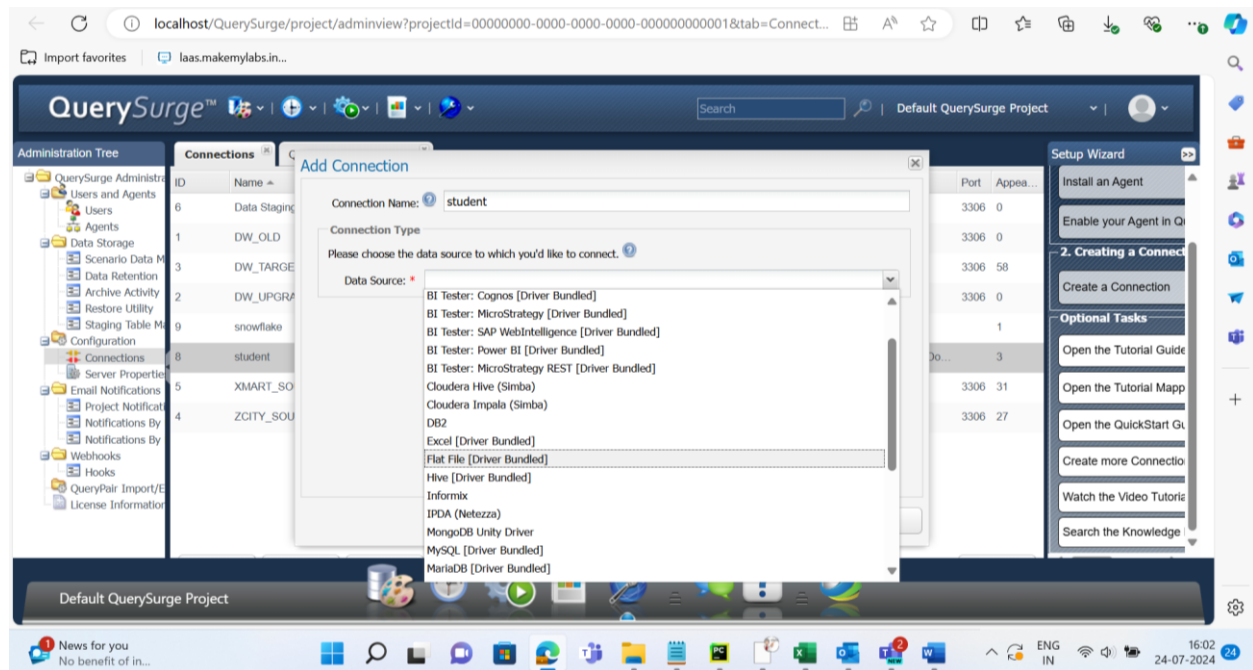
```
1 SELECT
2 *
3 FROM
4 STUDENT.PUBLIC.STUDENT1
5 LIMIT
```

. Below the query, the results are shown in a table format. The table has columns: STUDENTID, AGE, GENDER, ETHNICITY, PARENTALEDCATION, STUD, and GPA. The results are as follows:

	STUDENTID	AGE	GENDER	ETHNICITY	PARENTALEDCATION	STUD	GPA
1	1001	17	1	0	2	19.83	3.042914833436377
2	1002	18	0	0	1	15.40	
3	1003	15	0	2	3	4.210	
4	1004	17	1	0	3	10.02	
5	1005	17	1	0	2	4.672	

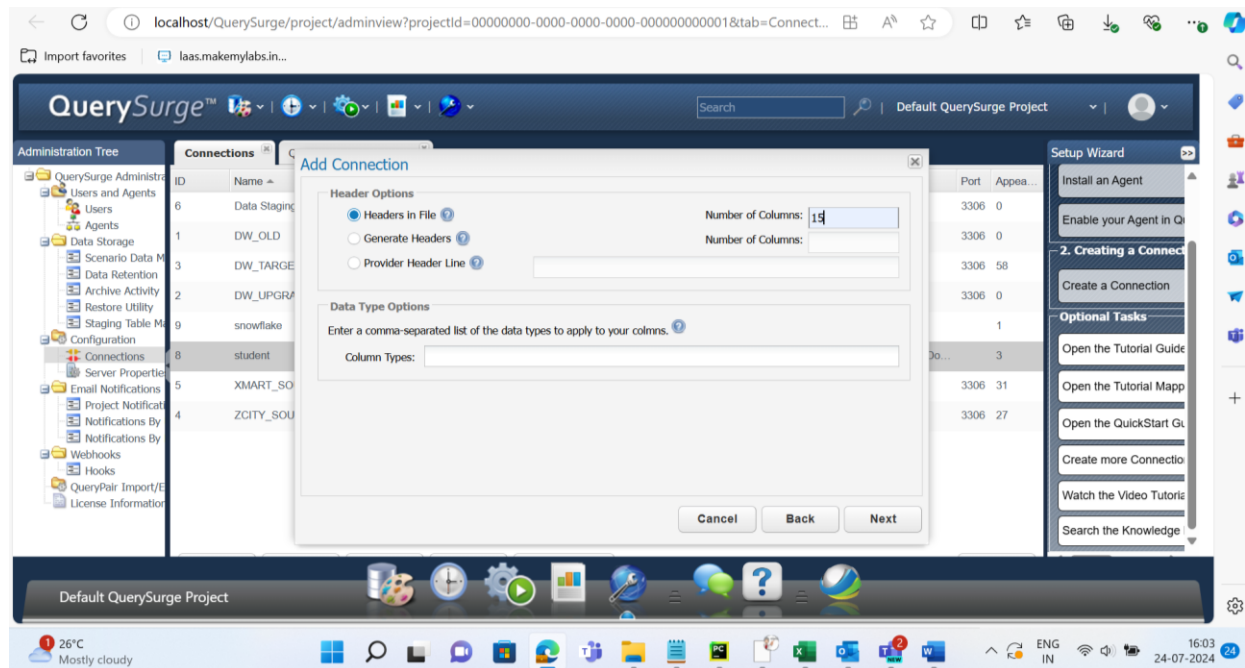
In query surge we need to import flat file

The screenshot shows the QuerySurge web interface. The URL is localhost:QuerySurge/project/adminview/projectid=00000000-0000-0000-0000-000000000001&tab=Connect.... The interface includes a sidebar with navigation options like Administration Tree, Connections, and Setup Wizard. The main area displays the 'Add Connection' wizard. The wizard has a title 'The QuerySurge Connection Wizard' and a description: 'The Connection Wizard will help you build a QuerySurge Connection with default values for standard driver parameters. To modify advanced settings, check the Advanced Mode box below before clicking the Next button. Otherwise, just click the Next button.' The wizard includes a 'Connection: Select a Connection...' dropdown menu and an 'Advanced Mode' checkbox. The 'Next' button is visible at the bottom right of the wizard.

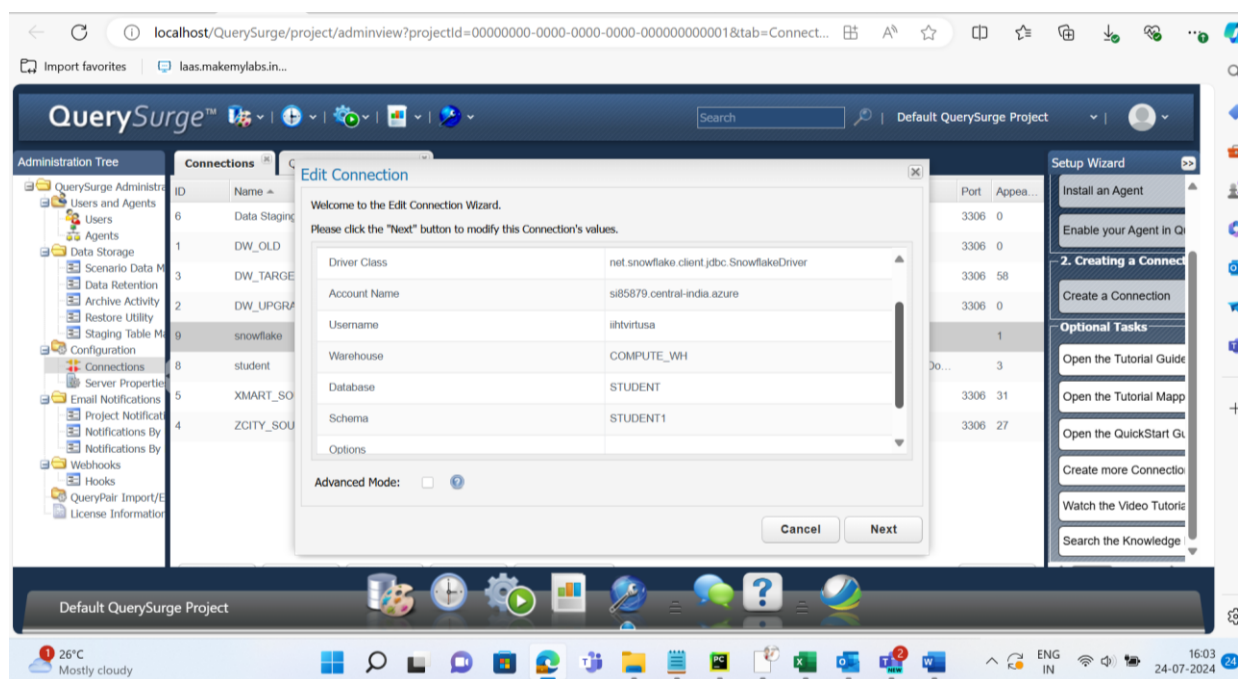


Give the path for the csv file

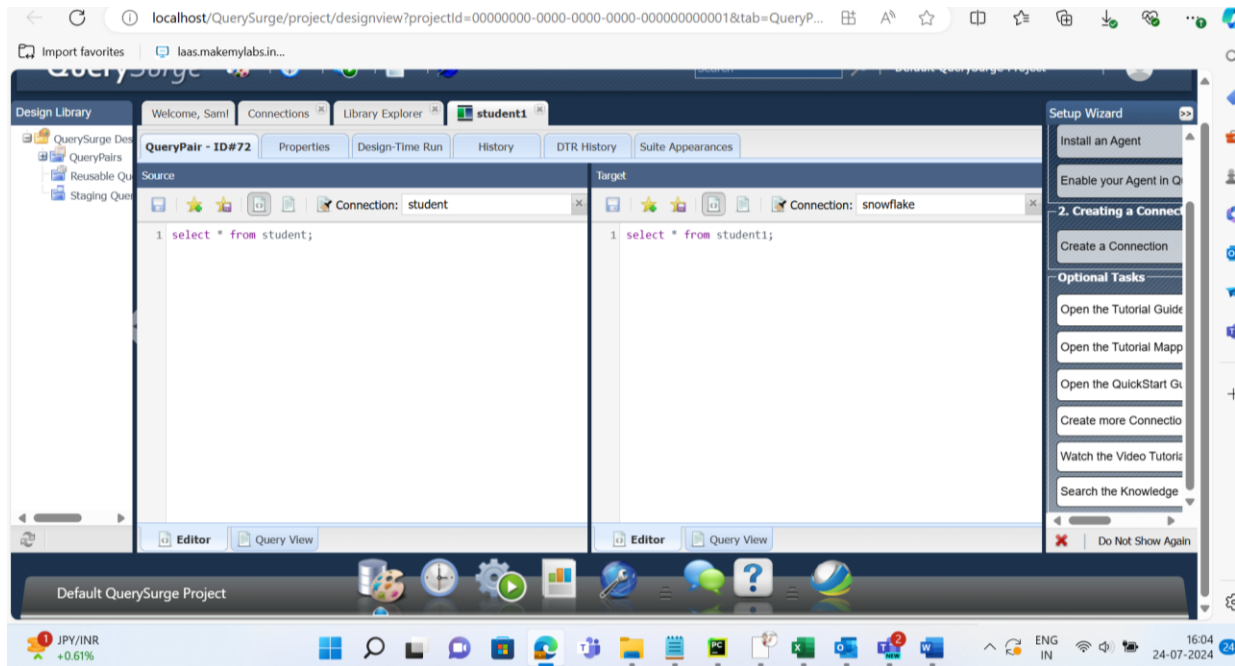
Provide the input columns in the corresponding input



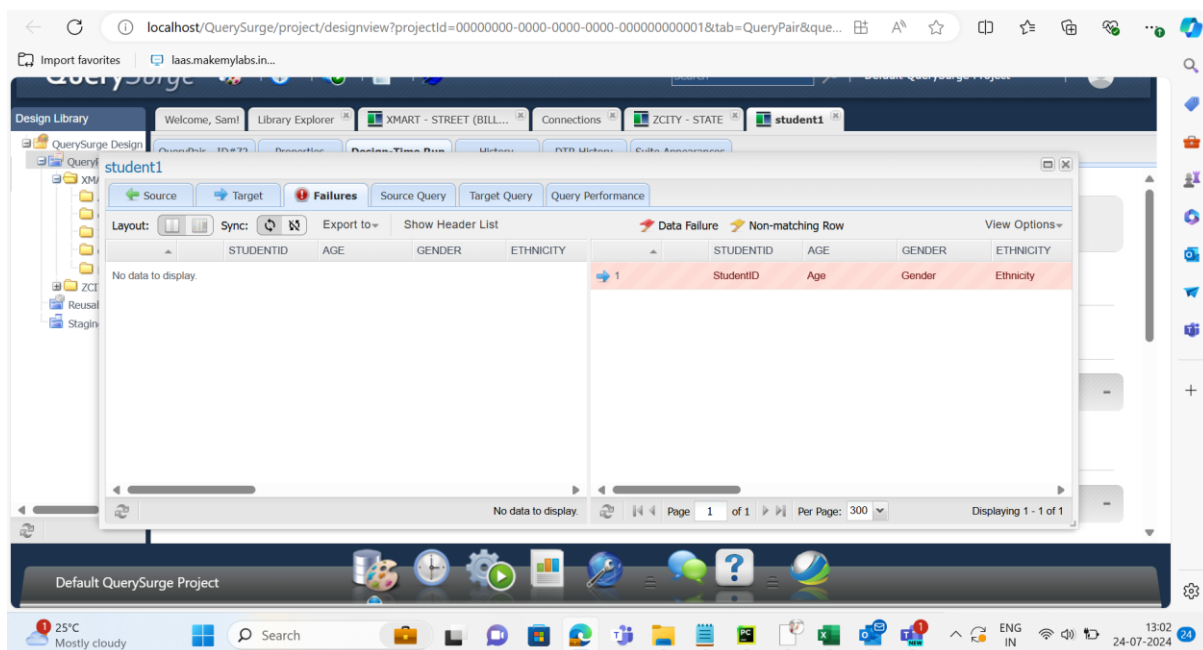
Snowflake connection



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



Example of failed data comparison



localhost/QuerySurge/project/designview?projectId=00000000-0000-0000-0000-000000000001&tab=QueryPair&que...

Design Library

Welcome, Sam! Library Explorer

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

13:03 24-07-2024

localhost/QuerySurge/project/designview?projectId=00000000-0000-0000-0000-000000000001&tab=QueryP...

Design Library

Welcome, Sam! Connections Library Explorer

QueryPair - ID#72 Properties Design-Time Run History DTR History Suite Appearances

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

Default QuerySurge Project

16:05 24-07-2024

Setup Wizard

- Install an Agent
- Enable your Agent in Q
- 2. Creating a Connect
- Create a Connection
- Optional Tasks
- Open the Tutorial Guide
- Open the Tutorial Mapp
- Open the QuickStart G
- Create more Connecti
- Watch the Video Tutori
- Search the Knowledge
- Do Not Show Again

localhost/QuerySurge/project/designview?projectId=00000000-0000-0000-0000-000000000001&tab=QueryP...

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...

1:44:58:253 1:44:58:774 1:44:59:294 1:44:59:815 1:45:00:335 1:45:00:856 1:45:01:376 1:45:01:897

00:00:00 00:00:00 00:00:01 00:00:01 00:00:02 00:00:02 00:00:03 00:00:03

1:44:58:513 1:44:59:034 1:44:59:554 1:45:00:075 1:45:00:596 1:45:01:116 1:45:01:637

00:00:00 00:00:00 00:00:01 00:00:01 00:00:02 00:00:02 00:00:03

Default QuerySurge Project

26°C Mostly cloudy 16:09 24-07-2024

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

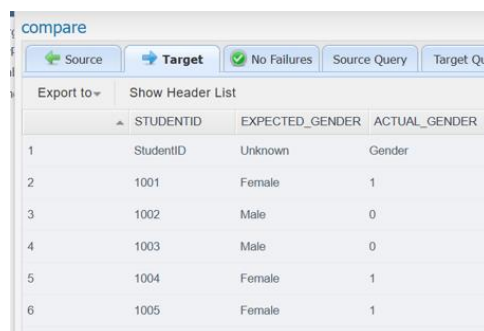
Task B to be performed

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 1 is female and 0 male others unknown .



	STUDENTID	EXPECTED_GENDER	ACTUAL_GENDER
1	StudentID	Unknown	Gender
2	1001	Female	1
3	1002	Male	0
4	1003	Male	0
5	1004	Female	1
6	1005	Female	1

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-----