****

**Name:** Fatima Azfar

**Roll Number:** 20i-2658

**Section:** DS-M

**Project Report**

**Submitted to: Dr. Asif Naeem**

**Project Overview:**

In this project we were instructed to work on the transactions and master sales database, load it into SQL and then apply the Hybrid Join algorithm and then write the respective OLAP queries.

Starting off, i created the tables in the SQL workbench and then I used java eclipse to do the rest of the code. I loaded the csv files, applied constraints and then inserted the data into the SQL tables. Converted the year format to mm/dd /yyyy to extract the data based on year.

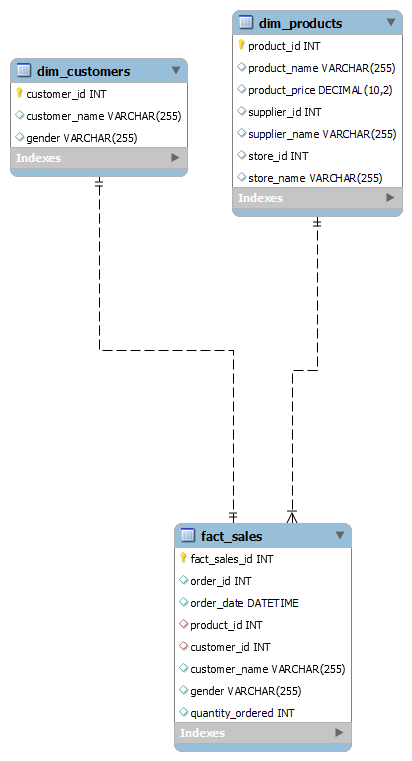
Errors I encountered on this step were:

1. Loading the data.
2. Incorrect timestamp format.
3. Filled values in the table.

Then implemented Hybrid Join by using the ETL method and used the concept of Multi-threading to estimate the Stream Arrival rate, to calculate the total sales and to estimate the service rate.

I implemented these in one file named as Hybrid Join and made separate static classes for them.

**Schema:**

****

**The fact table is the fact\_sales.**

**Hybrid Join Algorithm:**

Stream or load portions of the datasets into the in-memory hash tables. These portions should be small enough to fit into memory.

Build a hash table using the in-memory data.

Continue loading, sorting, and processing portions.

Moreover worked with multi threading and implemented ETL sideways.

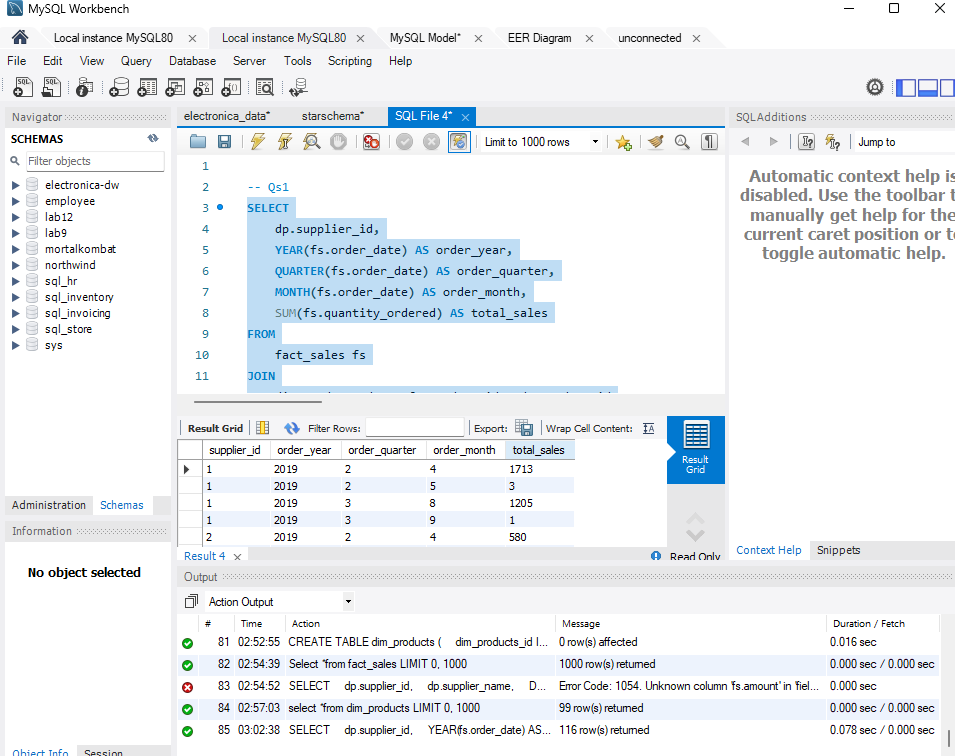
**StreamDataGenerator Class:**

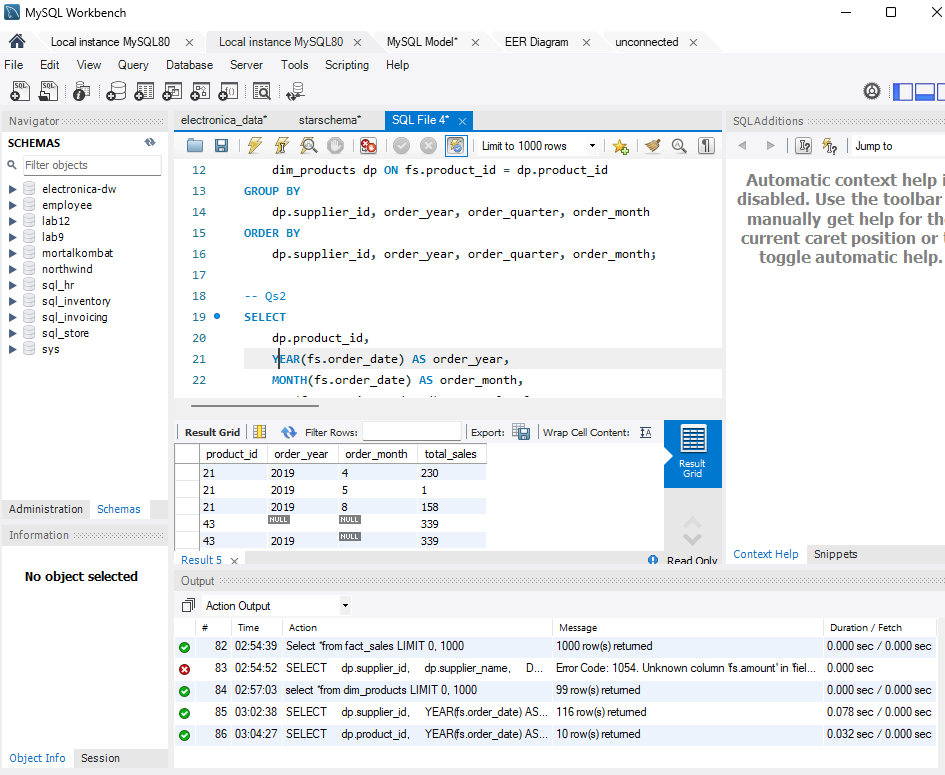
This class generates stream data and puts it into the transactionStreamBuffer.

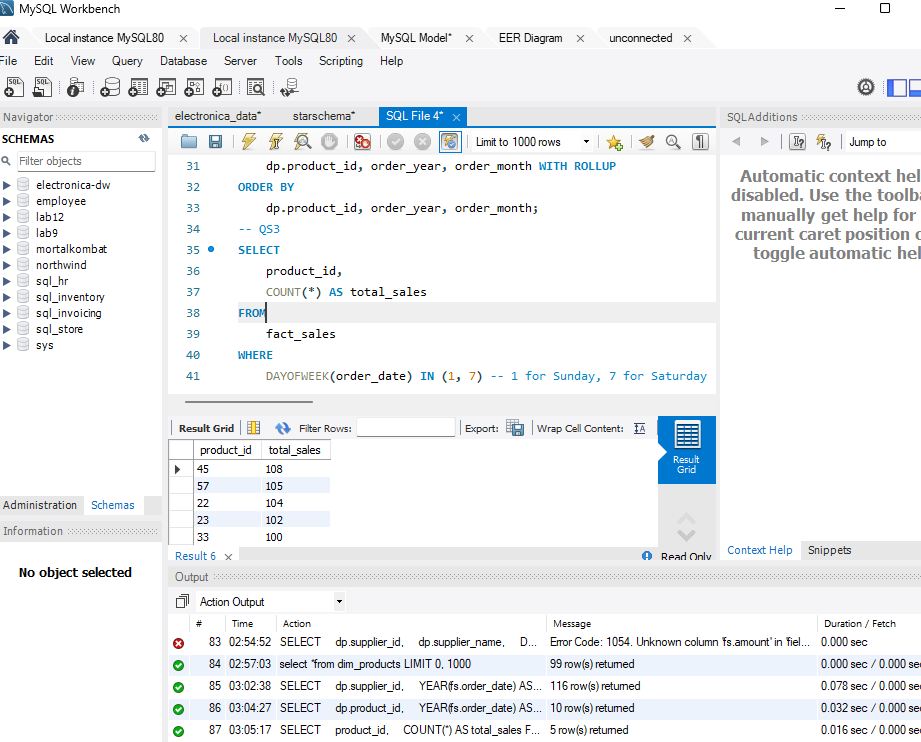
**SystemController Class:**

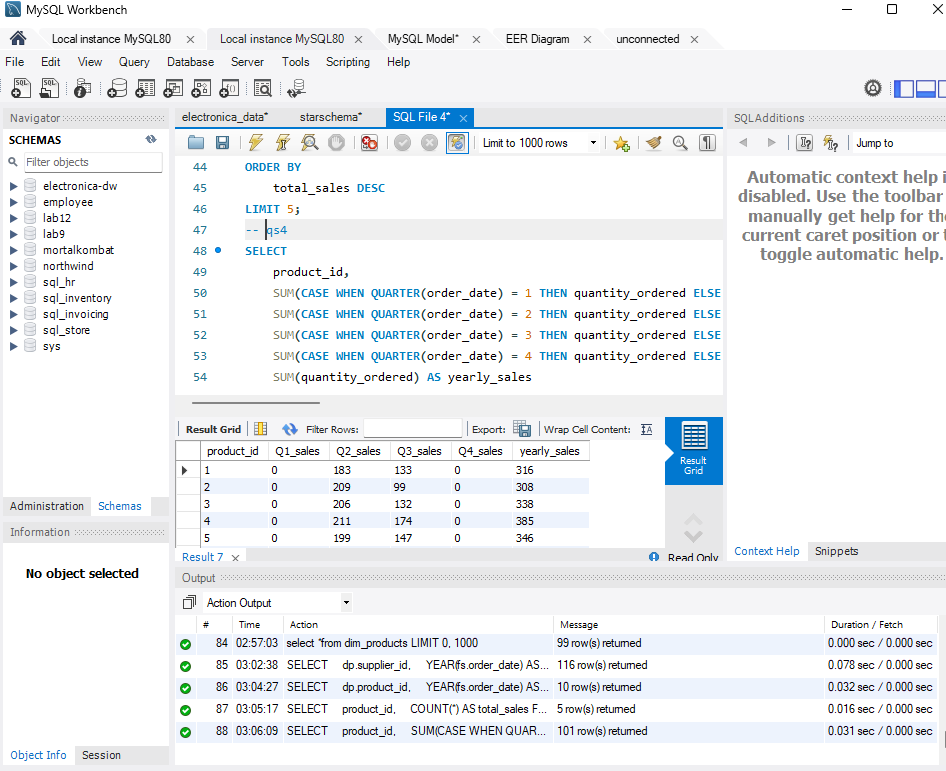
This class controls the system, estimating stream arrival rates and service rates.

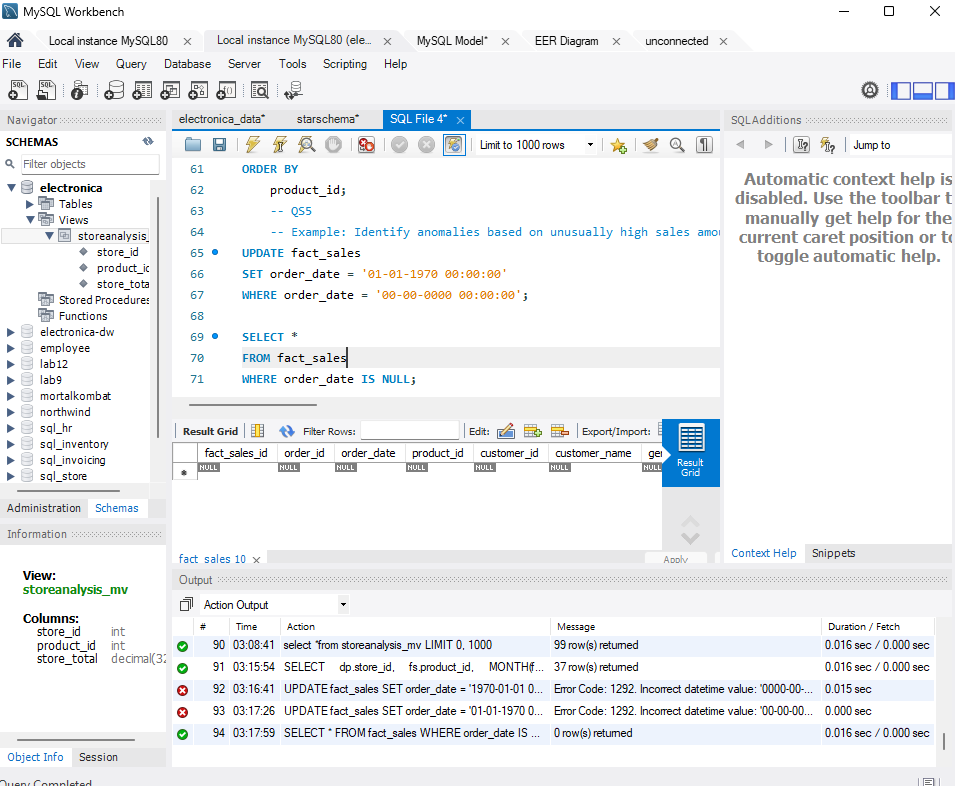
**OLAP Queries:**

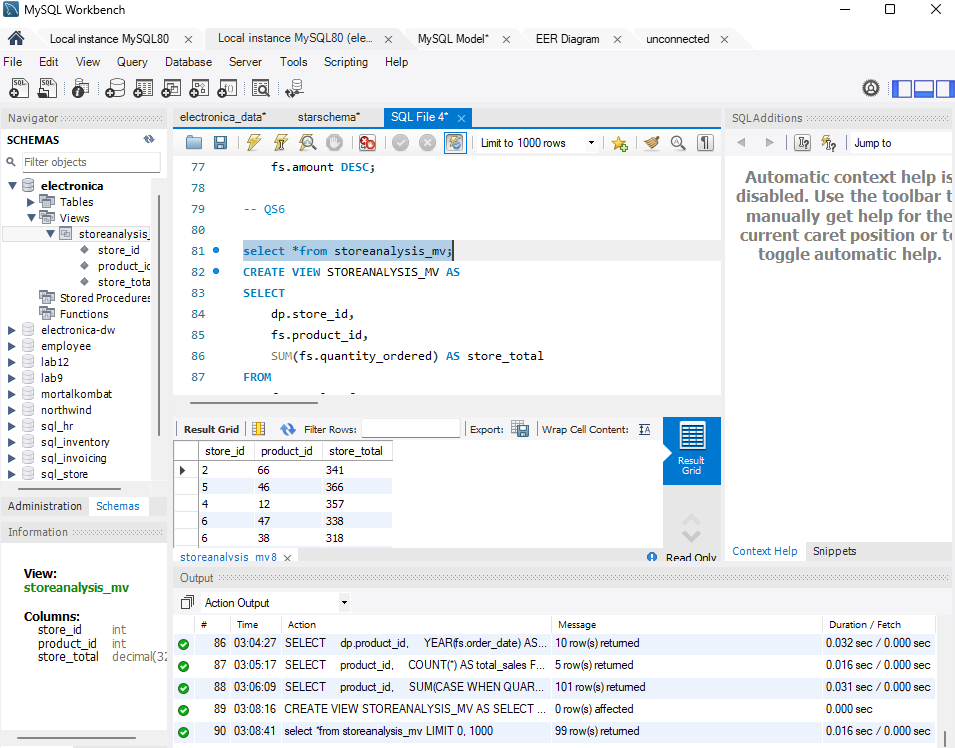
Q1:

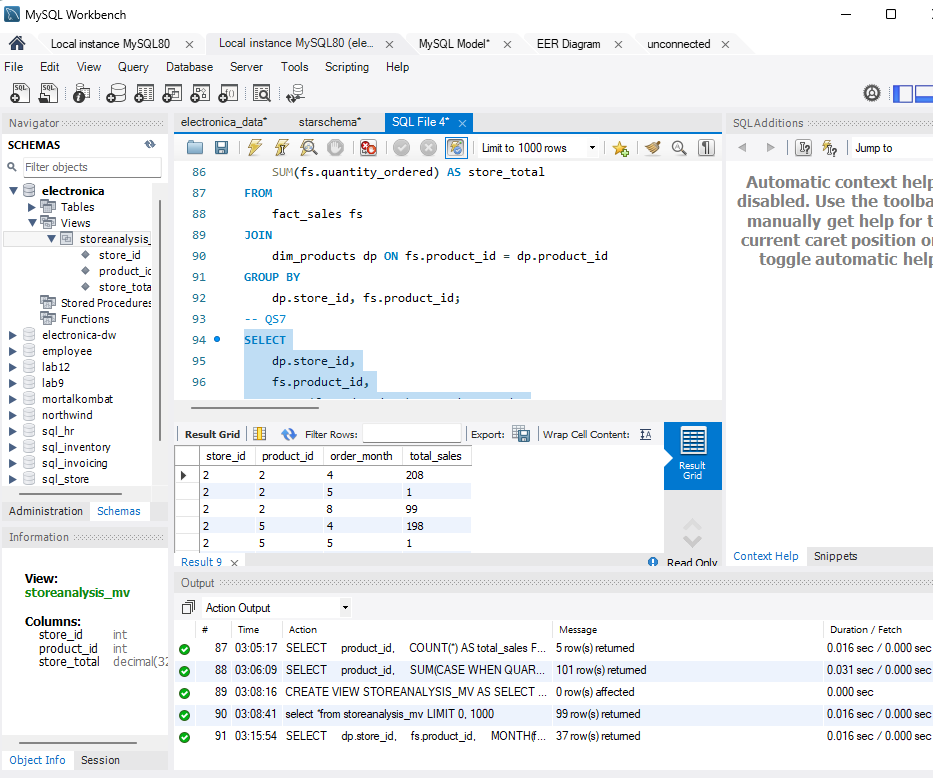
Q2: 

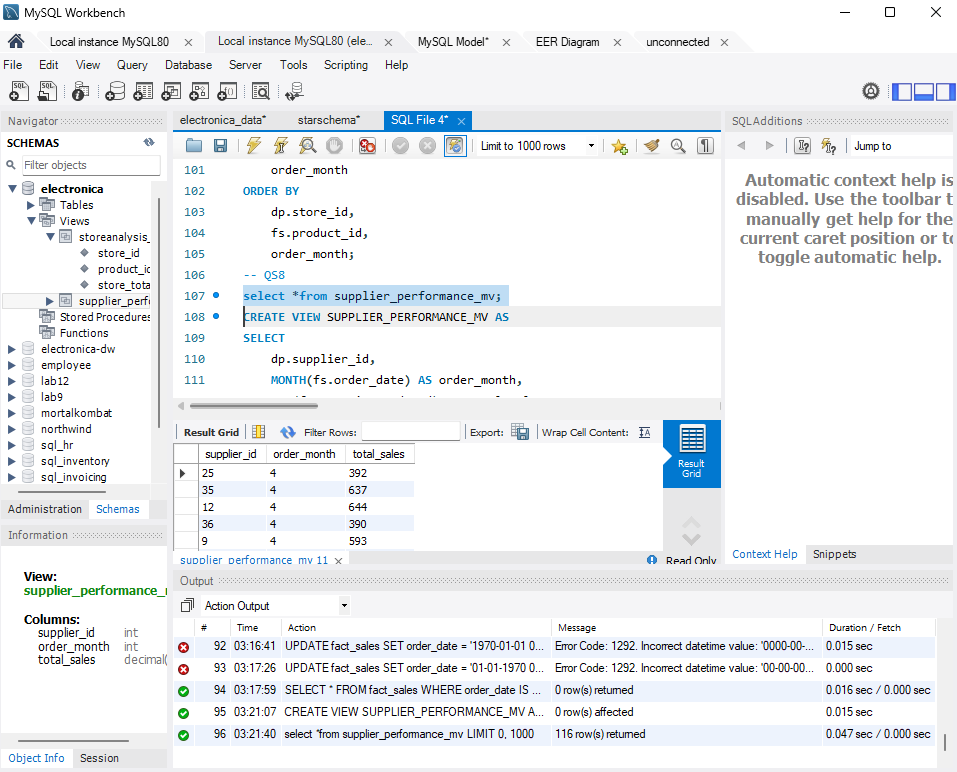
Q3:

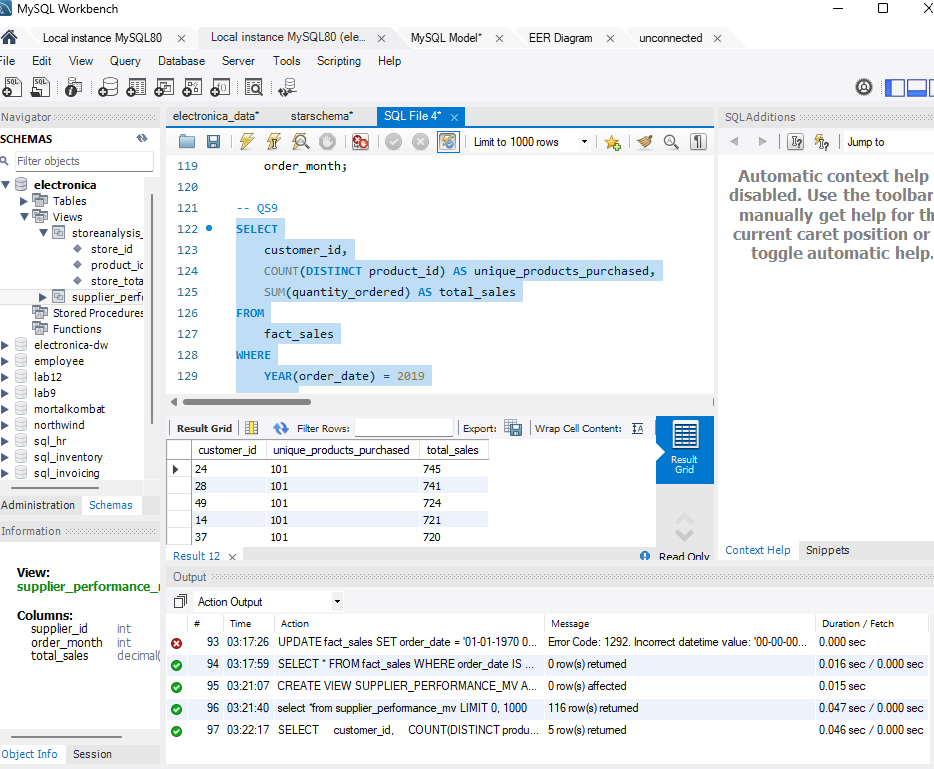
Q4:

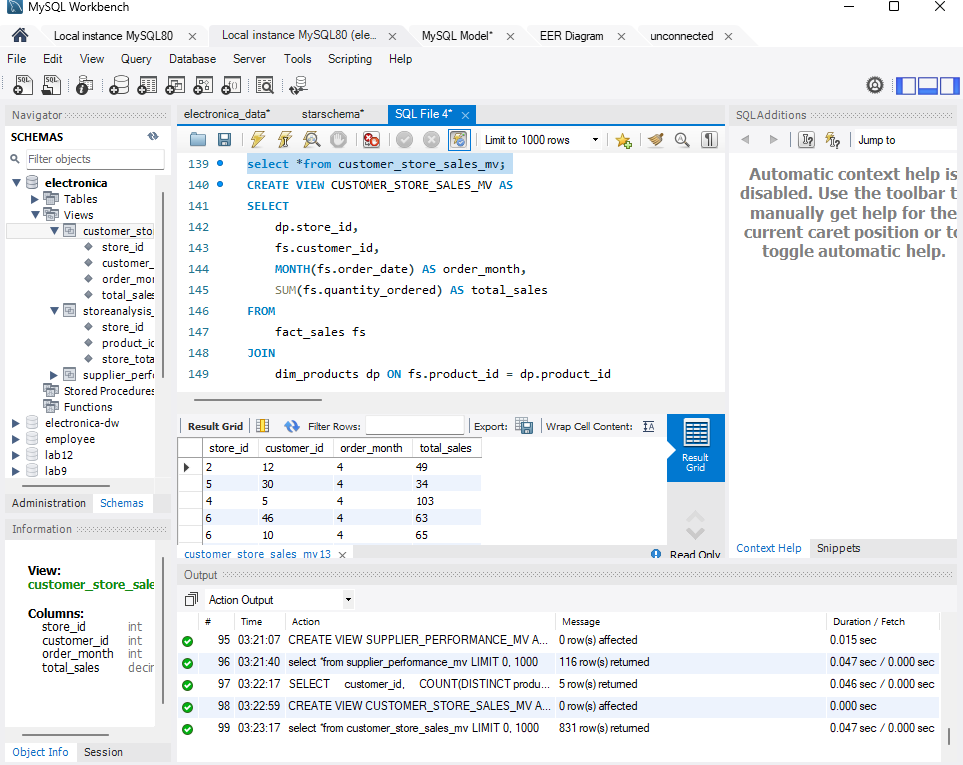
Q5:

Q6:

Q7:

Q8:

Q9:

Q10:

**Shortcomings of Hybrid Join:**

1. One of the main challenges with hybrid joins is their dependency on available memory.
2. The need to manage both in-memory and disk-based structures, handle partitioning of data, and optimize for various scenarios increases the implementation and maintenance overhead.

**Learning from the project:**

I learned how to load csv files into database and how to implement OLAP queries on them. Moreover, I learned to somehow built a data warehouse and learned the steps to implement the great Hybrid Join Algorithm. Learned to implement joins using multi threads and implement buffer w.r.t to the threads. It was an interesting project,

However I would like to acknowledge that I took help from the internet to understand some of the concepts of hybrid join and multi threads. I admit that I have full understanding of the code and the project is done by myself.