**Bachelor of Science in Data Science**

**Course Code: DS-3003**

**Course Name: Data Warehousing**

**Course Project**

**Building and Analysing Data Warehouse Prototype for**

**Electronica Business Chain**

**Name:** Afaq Alam

**Roll No:** 21i-1700

**Section:** N

***Table of Contents:***

1. *Project Overview*
2. *Data Warehouse Schema*
3. *HYBRIDJOIN Algorithm*
4. *OLAP Query Outputs*
5. *Shortcomings of HYBRIDJOIN*
6. *What I learned from this project*

**Project Overview:**

Given a scenario of a Business chain named Electronica, we were to create a data warehouse prototype for it. Moreover, we had to perform Extraction Transformation and Loading (ETL) in java. The database and data warehouse were created in MySQL. After ETL, we had to perform OLAP queries on the data warehouse. In ETL, we had to use HYBRIDJOIN Algorithm, introduced by M. Asif Naeem in 2011.

**Data Warehouse Schema:**



**HYBRIDJOIN Algorithm:**

Initialize Variables:

Fetch master data tuples.

Set w = 999, proc = 0.

Process Transactions:

Do:

Iterate over transactions:

Extract transaction columns.

Insert customer tuple.

Break if w reaches 0.

Get oldest join attribute from the queue.

Retrieve master data tuple and transaction tuple.

Iterate over each tuple:

Join transaction and master data.

Add joined output to finalOutput.

Insert joined output into the database.

Display processing status.

Remove oldest join attribute from the queue.

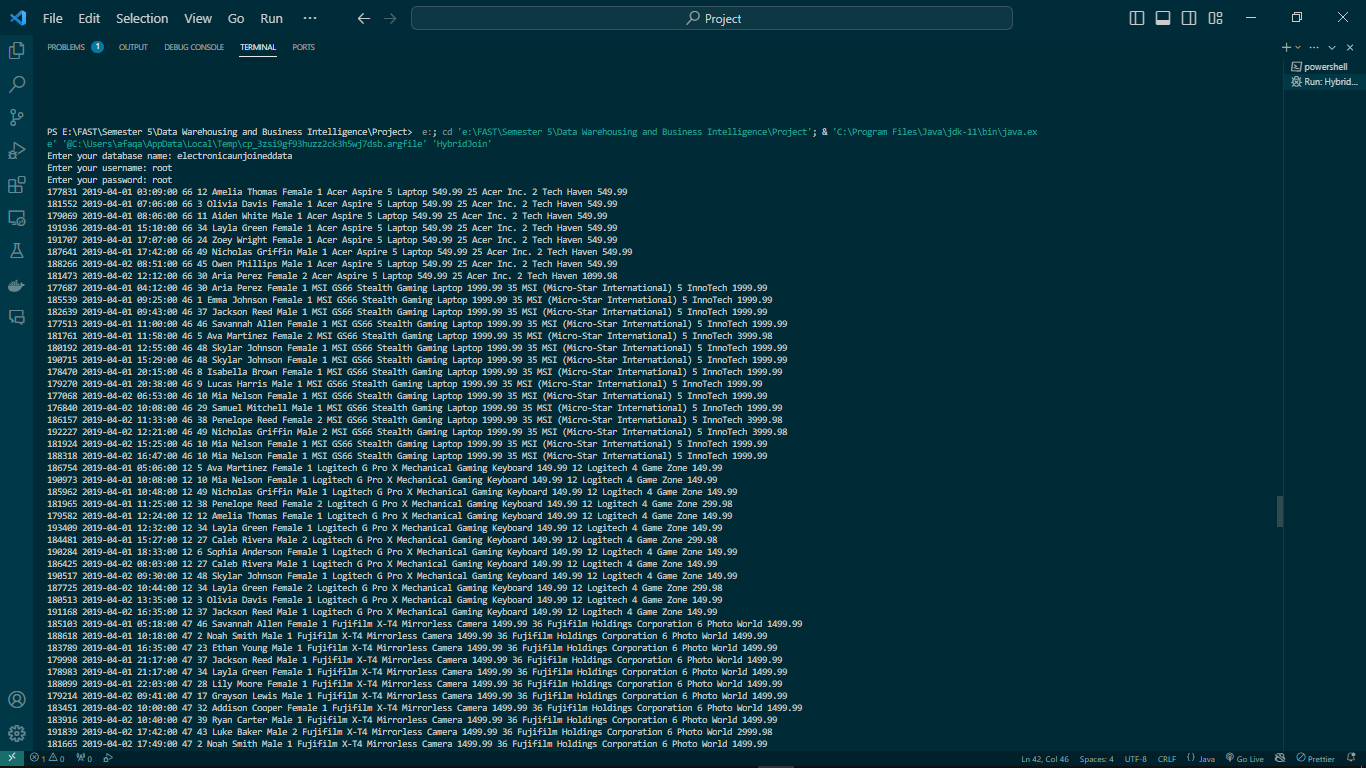
While the queue is not empty.

Display Process Completion:

Move console cursor up and clear line content.

Print "Process Completed!"

First 50 tuples of the join output:



**OLAP Query Outputs:**

**Shortcomings of HYBRIDJOIN:**

Hybrid join is very slow if multi-threading is not implemented.

**What I Learned from this Project:**

The project overall was a fun and challenging experience. It forced me to learn java, a new language and I found java quite easy and user friendly, as the community is quite active for help. Moreover, I understood the process of creating a data warehouse and performing ETL in real life.