DATA WAREHOUSE REPORT

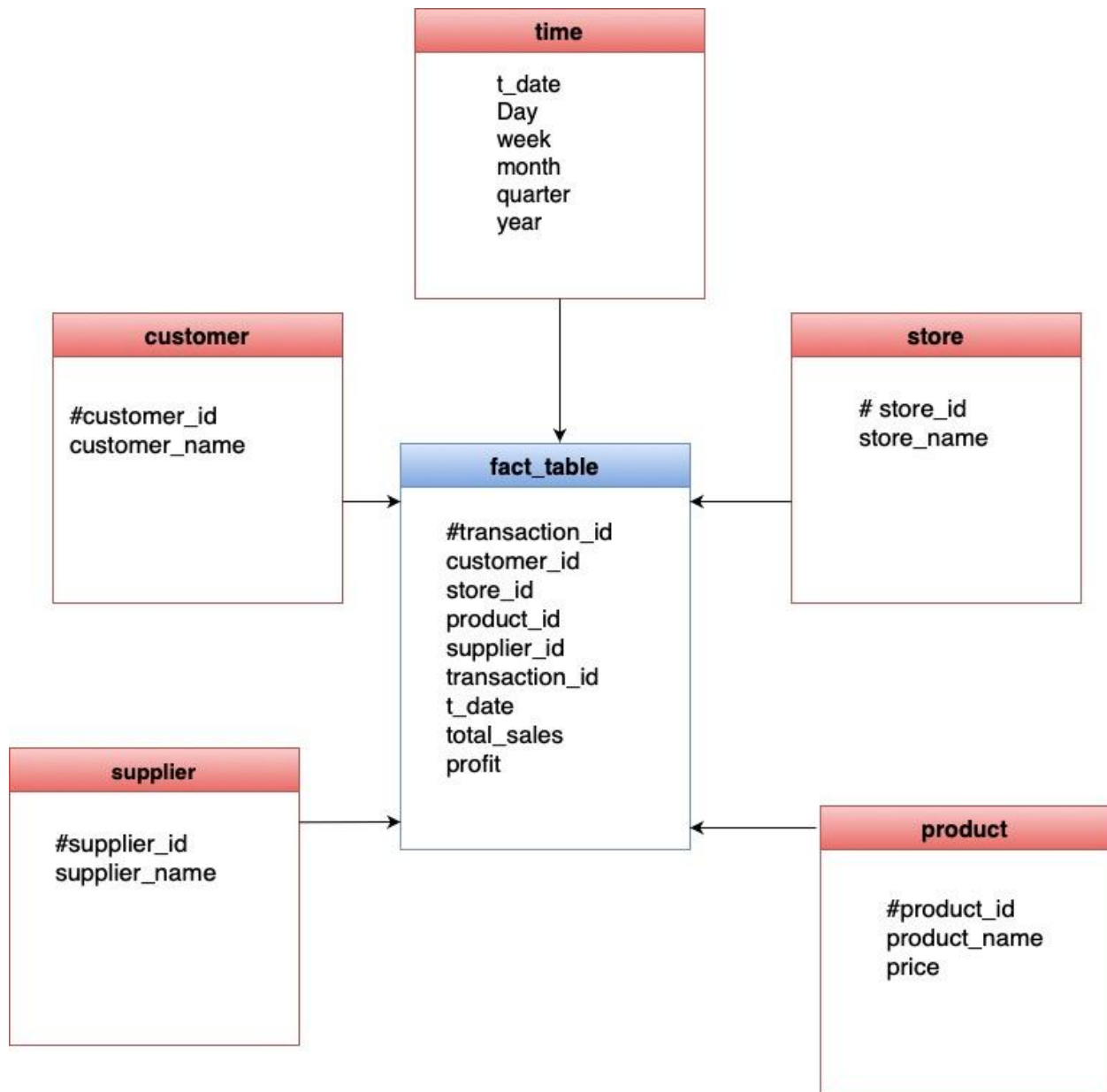**Aqsa Rahman**

Roll number: 19I-1908
SECTION : DS-N

Project Overview:

In this project I created a data warehouse for the metro. I started with loading the data of metro's(suppliers, customers,stores,and products) from the database. After loading the data i transformed it by applying mesh join algorithm and then loaded it into the warehouse which i created through the star schema by analyzing the database so that OLAP queries could be performed over it.

Star Schema

The star schema for the warehouse is as follows.

time
- t_date
- Day
- week
- month
- quarter
- year

customer
- #customer_id
- customer_name

store
- # store_id
- store_name

fact_table
- #transaction_id
- customer_id
- store_id
- product_id
- supplier_id
- transaction_id
- t_date
- total_sales
- profit

supplier
- #supplier_id
- supplier_name

product
- #product_id
- product_name
- price

Mesh Join Algorithm

Mesh Join Algorithm basically makes the join between data from to streams efficient and optimized.Here the data is chunked and these chunks of data from two streams are joined together and transferred into a hashtable from where the data is loaded into the warehouse finally.

Shortcomings of Algorithm

If the data is huge then input and output operations would slow down so it might be a disadvantage for a bigger company as they have very huge data such as Amazon and macys.

It is difficult to manage retrieval of data from different resources.

Learning from the project

This project helped us in learning the entire ETL process in the warehouse which included how to handle data coming from different sources,formatting the different kinds of data. Designing schema and creating warehouses for any company that sells different products.It also helped us to learn how to optimize the joining and transferring of data to the warehouse.