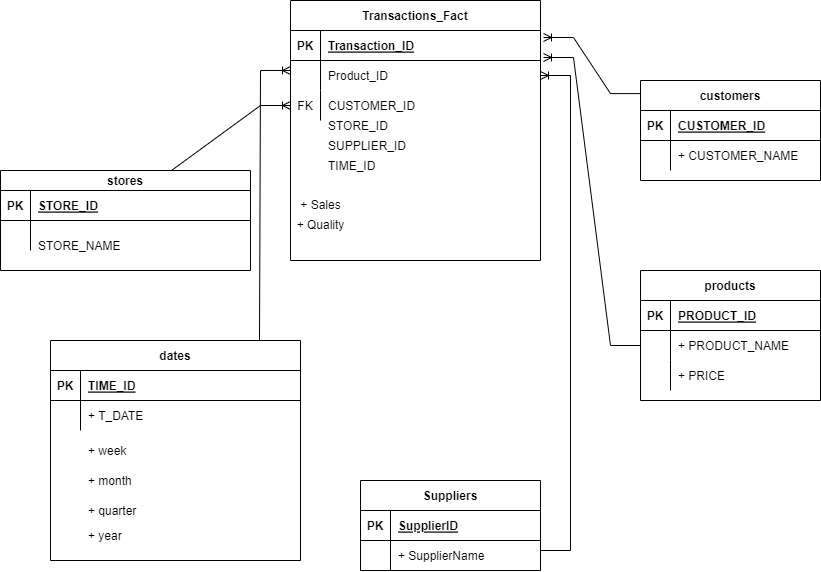
Name: Hassan Ali Ansari

Roll Number: 19I-1973

**Near Real Time Warehouse for Metro Cash & Carry**

**Star schema of the Warehouse:**



Project Description:

We have created a near real time Data ware house for Metro cash and Carry. We had 2 basically 2 sources of data. One is transactional data, and the other one is master data, The master data has further 2 tables, one is product data and the other one is one is Customer data. We are using MESH join algorithm to combine and transform these 2 data sources into a new table, we are calling that table a Hash table, and now form Hash table we are loading this transformed data into the warehouse.

Mesh Join:

In mesh join algorithm we are loading the data from the Transactional data to the Queue and then we will be we will be loading the data segment from the Master Data, then we are combining this data coming from both sources, we getting the product information, Customer information against the ID’s in transactional data attributes. Then this data is stored in a hash table and then the hash table is loaded into the DWH.

**Short Comings of Mesh Join:**

The master data is loaded into small chunks reading these chunks from the main memory, this can cause overlapping time of residence in the main memory and this can lead to staggered processing. Mesh join is a slow process when, for each transaction we have to traverse entire Master data tables.

**Learnings:**

1. Designing a warehouse schema.
2. Appling join on live data coming for OLTP, and the master data and transformation of that data.
3. Mesh Join
4. Data loading into a warehouse.