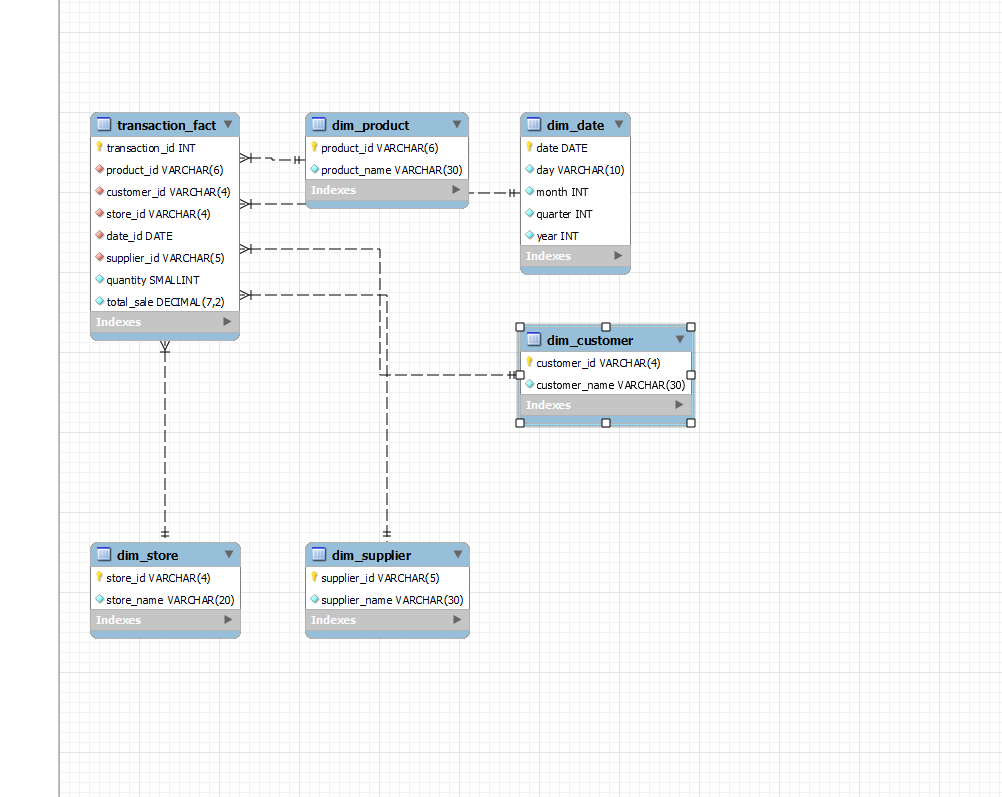
# **Data warehousing**

**Project Overview:** The project required to implement a near realtime datawarehouse. The schema for this project is star schema



**Drawback of meshjoin:**

Meshjoin needs to be modified if we are dealing with two tables in master data while maintaining only one hashmap and queue for both master data

**Procedure** MeshJoin()

**for** numTransaction **do**

Queue <- Transaction

**For** Queue **do**

HashTable<- <product\_id,transaction>

Diskbuffer1<-product

Diskbuffer2<-customer

**if** HashTable(diskbuffer1.product\_id) = diskbuffer1.product\_id **then**

**Enrich(**Transaction.tuple**)**

**loadTuple Transaction.tuple**

**end if**

**if** Queue is Full **then**

**removeFirstElement**

**end if**

**end for**

**end for**

We learned about the process of ETL on a basic level. How data is transformed and summarized to be stored in data warehouse.