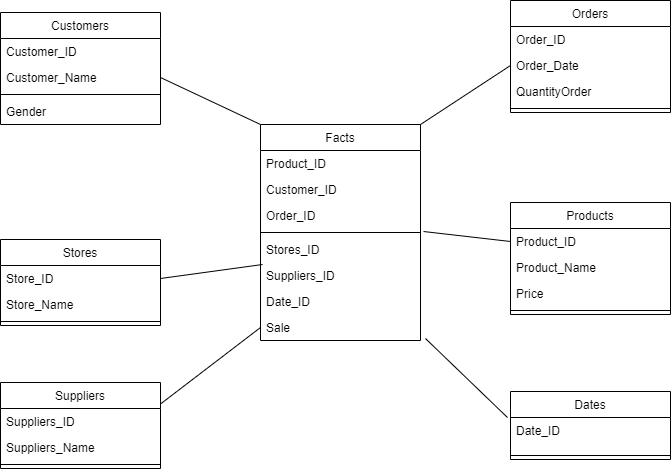
Schema:



PROJECT OVERVIEW:

In this project, I successfully implemented a Data Warehouse utilizing two main data components: Transactions and Masterdata. For the ETL (Extract, Transform, Load) process, I employed a Hybrid Algorithm, as per the project instructions, to efficiently join Transactions data with Masterdata before sending it to the Warehouse.

Hybrid Join OVERVIEW:

The Hybrid join was implemented in Java, leveraging a Queue and a MultiHashTable. The Queue was divided into 10 partitions, aligning with Masterdata partitions. Within the Queue, I stored pairs of Transaction\_id and Product\_id. Concurrently, the MultiHashTable used Product\_id as the key, with the entire tuple serving as its values.

During the Hybrid join process, Masterdata identified matching values in the MultiHashTable, facilitating the necessary join operation. Subsequently, the combined data was sent to the Data Warehouse for further processing.

What I Learned:

Through this project, I gained proficiency in Java programming language and acquired practical insights into the complete Data Warehouse lifecycle, particularly in the realms of ETL processes and aggregation techniques