**Task 3**

**1. Design a data warehouse schema using a data modeling tool (e.g., dbdiagram.io or draw.io). Your schema should include tables for Orders, Products, Customers, and Sales.**

1. The dim\_customer table will hold customer business and profile-related information where only unique records will exist here. The primary key is customer\_code which is the customer ID.
2. The dim\_product table will hold material business attributes and only unique records will exist here.
3. The key columns of dim\_customer and dim\_product are interlinked with other entities in terms of relationships as shown in the figure below.
4. In fact\_sales, it will hold the sales-related information where it stores the sales transaction data and interlinked with dim\_customer and dim\_product using foreign keys.
5. In fact\_orders, it will hold the orders-related information where it stores the orders transaction data and interlinked with dim\_customer and dim\_product using foreign keys.

A diagram of a data flow

Description automatically generated with medium confidence

**2. Describe the batch processing pipeline for populating and updating the data warehouse with historical data. This should include steps for data extraction, transformation, and loading (ETL).**

Firstly, extract the data from various sources, an e-commerce company may have sources from CRM systems, Web Analytics, Sales systems, and others. The data transformation process will include processes like cleaning the data and transforming the data. It is important to check if there are any duplications, missing values, and outliers. After the data is cleansed, we can perform joining between tables, calculating key metrics, and creating derived KPI. The cleansed and transformed data then will be loaded into a data warehouse. This could be an incremental loading by updating the existing records and adding new rows if not exist.

For Data Streaming, I don’t have exposure to real-time data streaming hence I could not answer them..