

Datawarehouse: Problem Statement

Let's consider a hypothetical case-study. Imagine U-Infra Pvt. Ltd. to be a construction company which has several malls pan Mumbai. Altogether, the company has got six malls operational in the city. With the rising number of outlets and the corresponding increase in sales, the higher management is facing difficulty in making business strategies. You are hired as a data engineer by the company to build a data warehouse solution for them.

Let's start designing the solution.

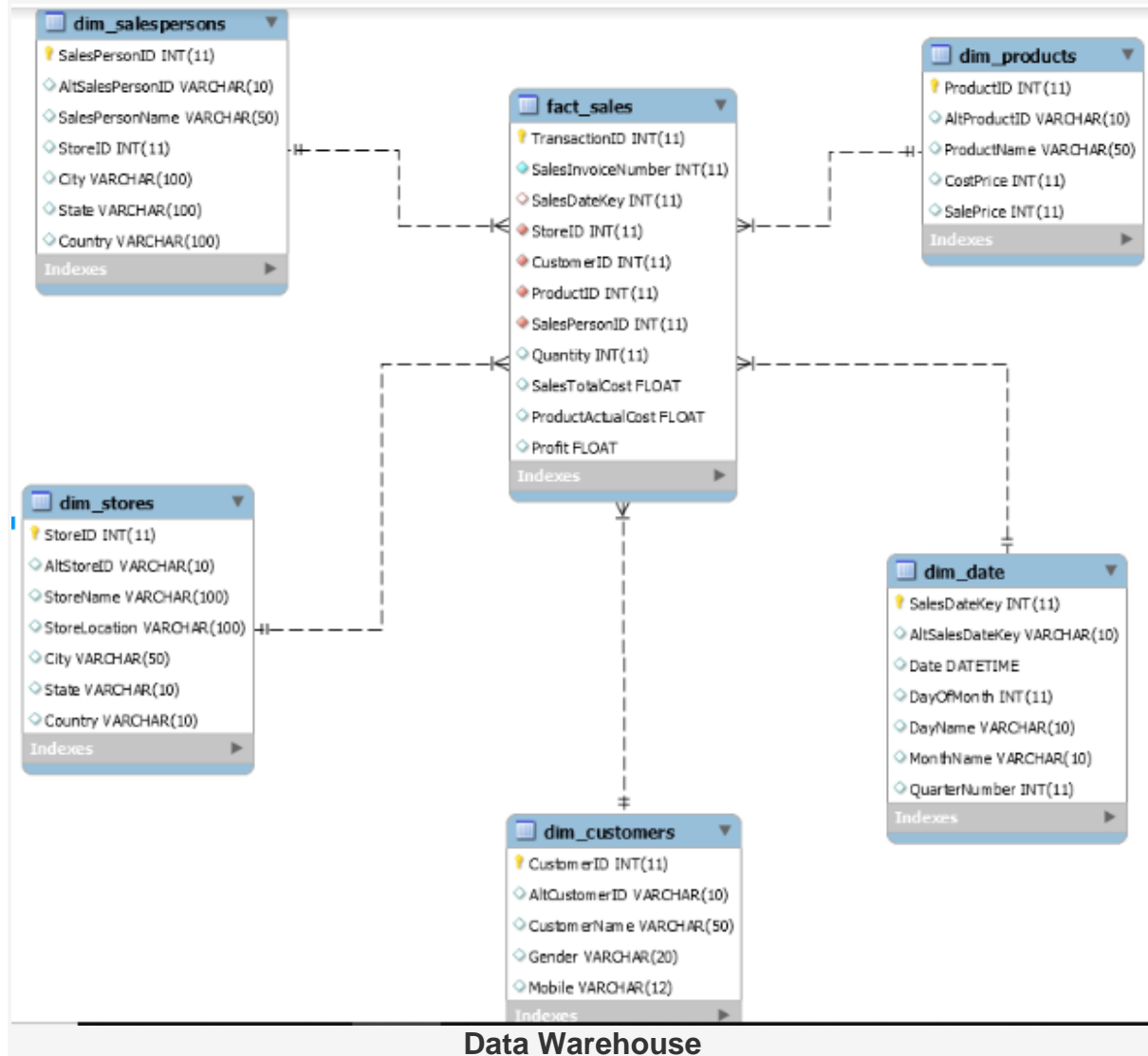
Data Warehouse Design

The first step should be to understand from the business development team what all queries they want to perform over the legacy data. Here is the list of queries that were collected from the BI team.

- What are the daily, weekly, monthly and a quarterly profit of each store?
- Which product has more demand on which location?
- What is the trend of sales by the period of the day over the week, month, and year?
- On what day are the sales higher?
- On every Sunday of this month, what are the sales and what is the corresponding profit?
- What is the trend of sales on the weekdays and weekends?
- Compare weekly, monthly and yearly sales to know growth and KPI.

The next step would be to identify the facts and dimensions. By looking at the business queries, the data warehouse should comprise of the following dimensions- Product, Customer, Store, Date, Salesperson. The Fact table should comprise of Foreign Keys

corresponding to each dimension along with measures such as- Quantity, Cost price, Sale Price and Profit. Here's how our data warehouse will look.



Sqoop Problem should assess the learner on the following concepts:

- Handling NULL via Sqoop
- Handling Mappers via Sqoop
- Importing Data in Binary File Format Via Sqoop
- Compression via Sqoop