Mistake made in Data Interview

1) Impulsive Answer

→ Star Method

S : Situation

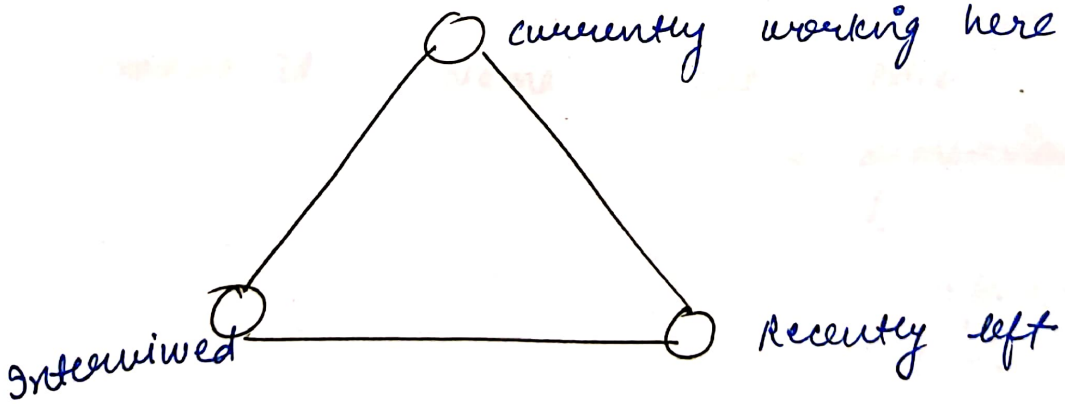
T : Task

A : Action

R : Result

2) No Research

→ Triangle Technique



3.) Using same example multiple times

4.) All talk & no show

5.) Being inflexible.

6.) Being too hard on ourselves

7.)

Various Terms :

(1) ~~Table~~

Date | Product Id | customer id | quantity

↳ fact table

(2)

Product Id | Name | cat | price

↳ Dimension table.

↓

tells information from product dimension

(3) customer id | Name | age | Gender

↳ customer id (Dimension table)

④ Date Day of week Month Year Quarter

↳ Calender Table

Note :-

① Fact table :→ Tall & Narrow

② Dimension table :→ short & wide

Relationship b/w

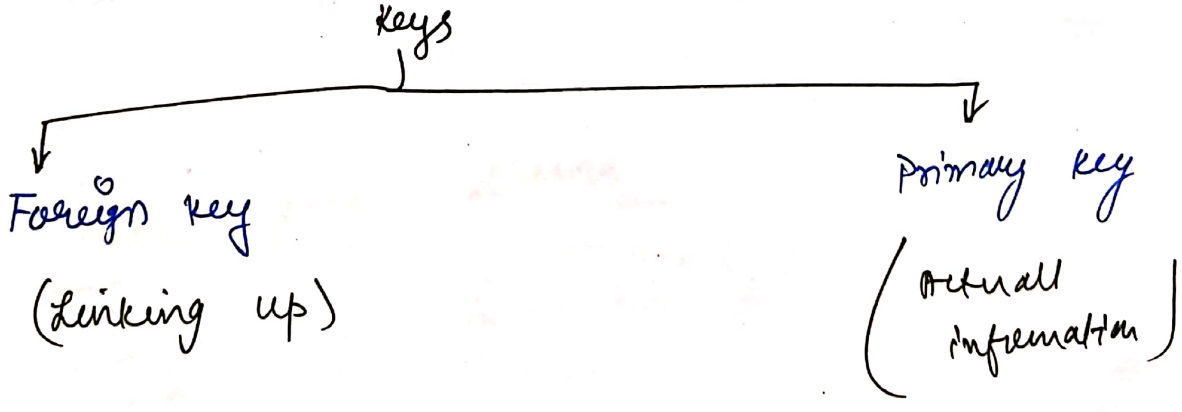
fact & Dimension

Many ~ One

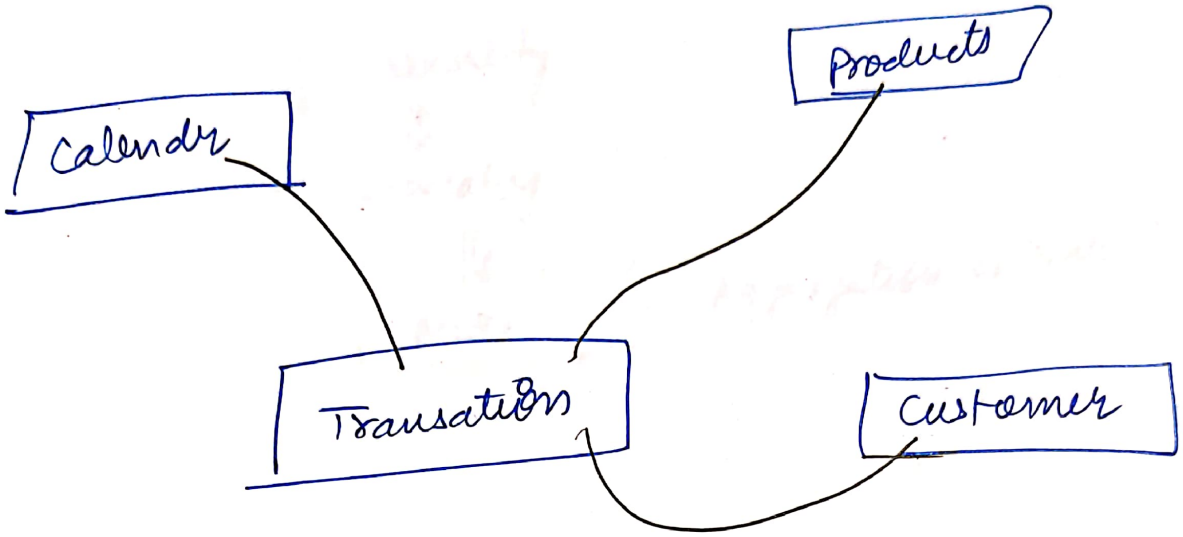
Many-one relation.

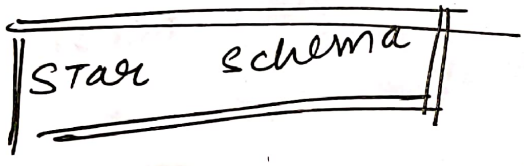
↓

These relationships are referred to as keys.



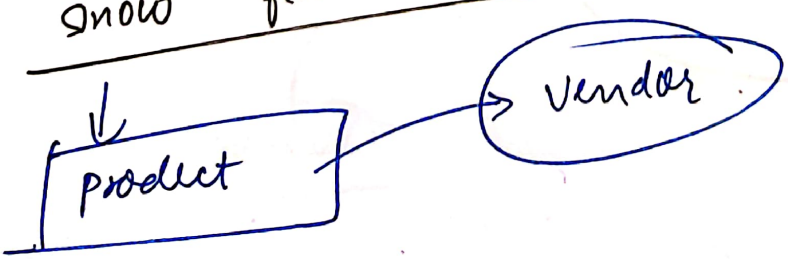
Modelling





* A central fact table & multiple dimension table on the outer adjust is referred as :⇒ star schema.

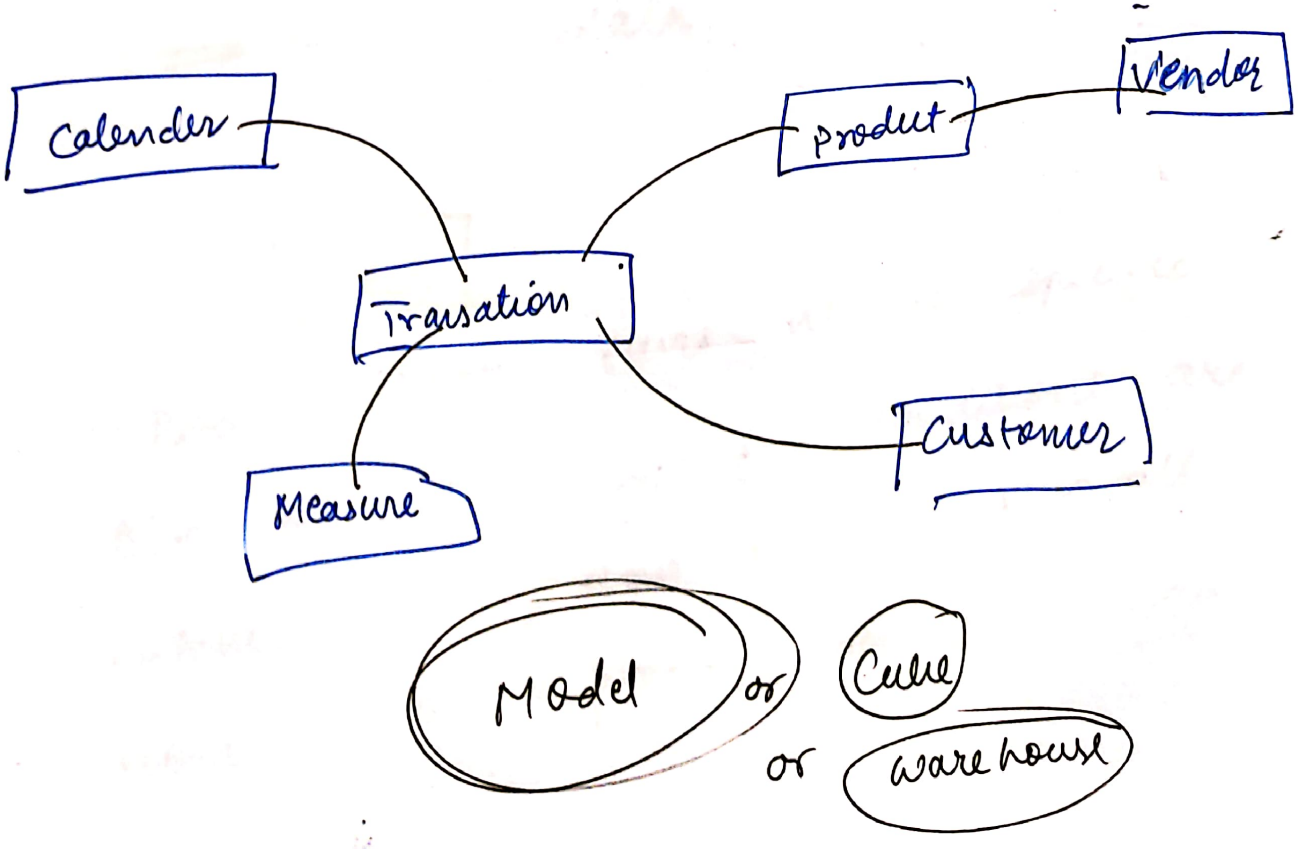
* Snow flake schema



Measures & Values

Sum of Quantity colum
⇓
calculating it
⇓
Measure [or Aggregation or Value]

Slicer
↓
break down the table & look it by individual category
or filter



Query or Power Query

⇒ for any given transaction table
↓
we need to connect to another data table for calculation (sales)

ETL

(Extract Transform Load)

Extract :- ⇒ go to different source system, collect the data.

Transform ⇒ them ⇒ size or nature of data

Load ⇒ loading them into final data

Batch

→ Process that runs at a specific time or when specific conditions are matched/meet, then it will run all your ETL process and give you

final output data so that you can update your report.

Data Pipeline

⇒ How you are getting your data before generating Reports.

Source - Refresh

Source :- Place where data is originally maintained

Refresh :- Process we ask our report to be updated based on the latest date.

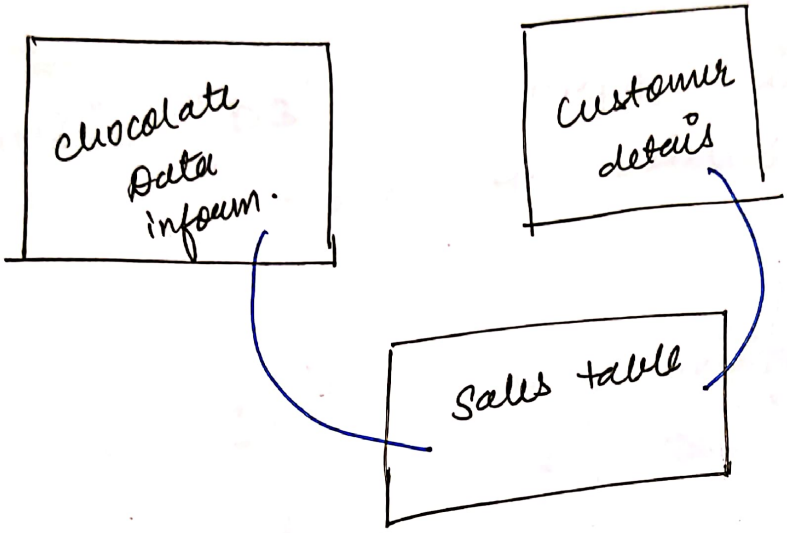
Technical Team

Data Base

Data Lake

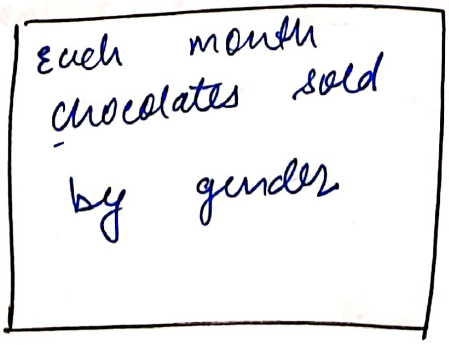
Data Warehouse

1.) Data Base (DB)



2.) Data Warehouse (DW)

→ special type of data-base → primarily designed so that u could do analysis and reporting.



⇓

figure out some trends to give discounts & analysis ↓ and what sales to promote

Note :-

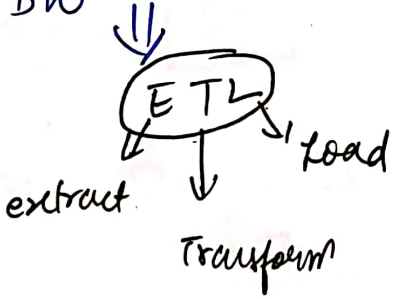
1) Within "DB" → purpose is to keep track of Business

2) Within "DW" → they periodically archive the data

↓

just like prev. year question paper

* Process of taking data from DB & loading it to DW



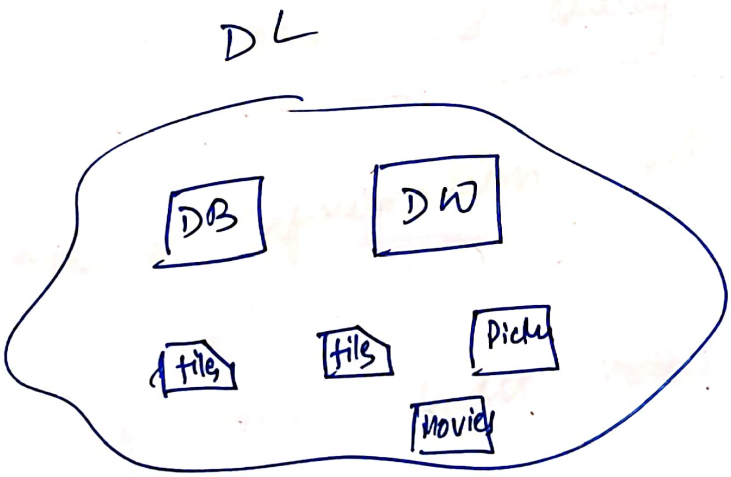
Data Lake (DL)

→ within DB & DW → data stored in form of Table

Table → bunch of rows & columns
↓
each column explains → bunch of data

→ "DL" → stored everything in semi-structured manner asking ques afterward can be used structured approach & unstructured approach.

unstructured approach
↓
asking computer to figure it out



Examples

(1) Data Base :- → SQL server
→ ORACLE
→ MS - ACCESS
→ Excel Spreadsheet

(2) Data warehouse :- → SQL Server
→ Power BI
→ Excel

(3) Data Lake :- → AWS
→ Azure
→ Big Query
(store it in cloud)
(maintain online → cloud)

SQL (Structured Query Language)

* Access Information within these three :->
  - SQL
  - forms & all (~~traditional reporting system~~)
  - Power BI
  - Python script
  - depends on situation & which type.