

Agenda

a. Building an Analytical Data Warehouse for Faster and Efficient Insight Generation

b. OLTP - Introduction

c. Introduction to Data Warehousing

d. Benefits of Data Warehouse

e. Designing Data Warehouse

f. Business Objective & Granularity

g. Star Schema

h. Snowflake schema

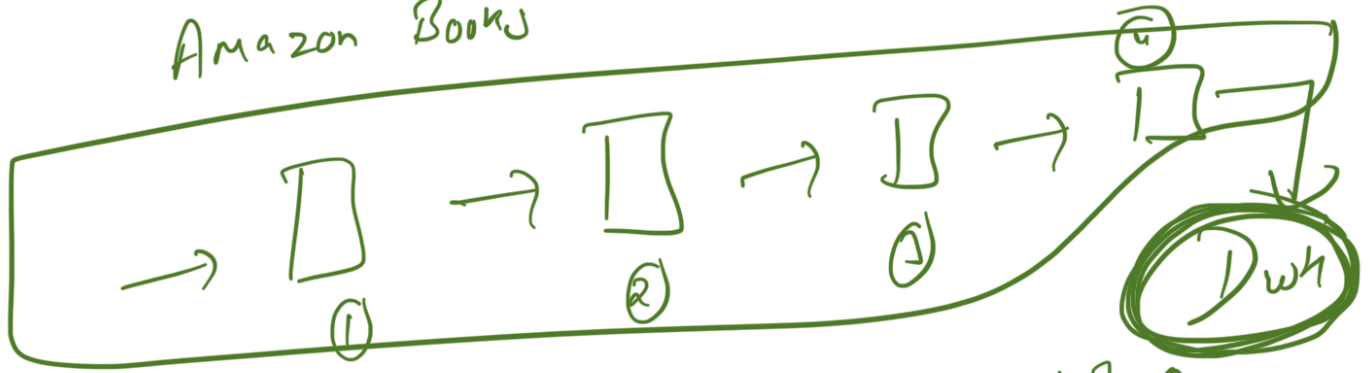
i. Galaxy/Fact Constellation Schema

j. Slowly Changing Dimensions

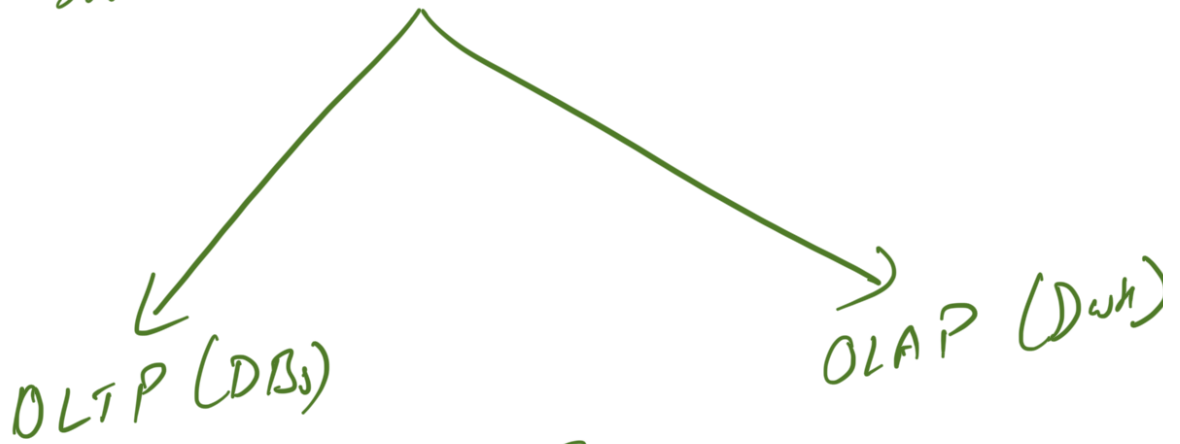
k. MAANG interview question



Amazon Books



- what are the top 10 books sold weekly? ↑
- ↳ what are the other products frequently bought with books?
- ↳ who are the top authors based on book sales?



→ OLTP → Online txn Processing

✓ C ✓ R ✓ U ✓ D ✓ (I)

→ DBs → RDBMS (ACID)*

↳ Oracle

↳ MySQL

↳ MSSQL

↳ DB2

↳ Sqlite
↳ postgres....

↳ others

and lot of things

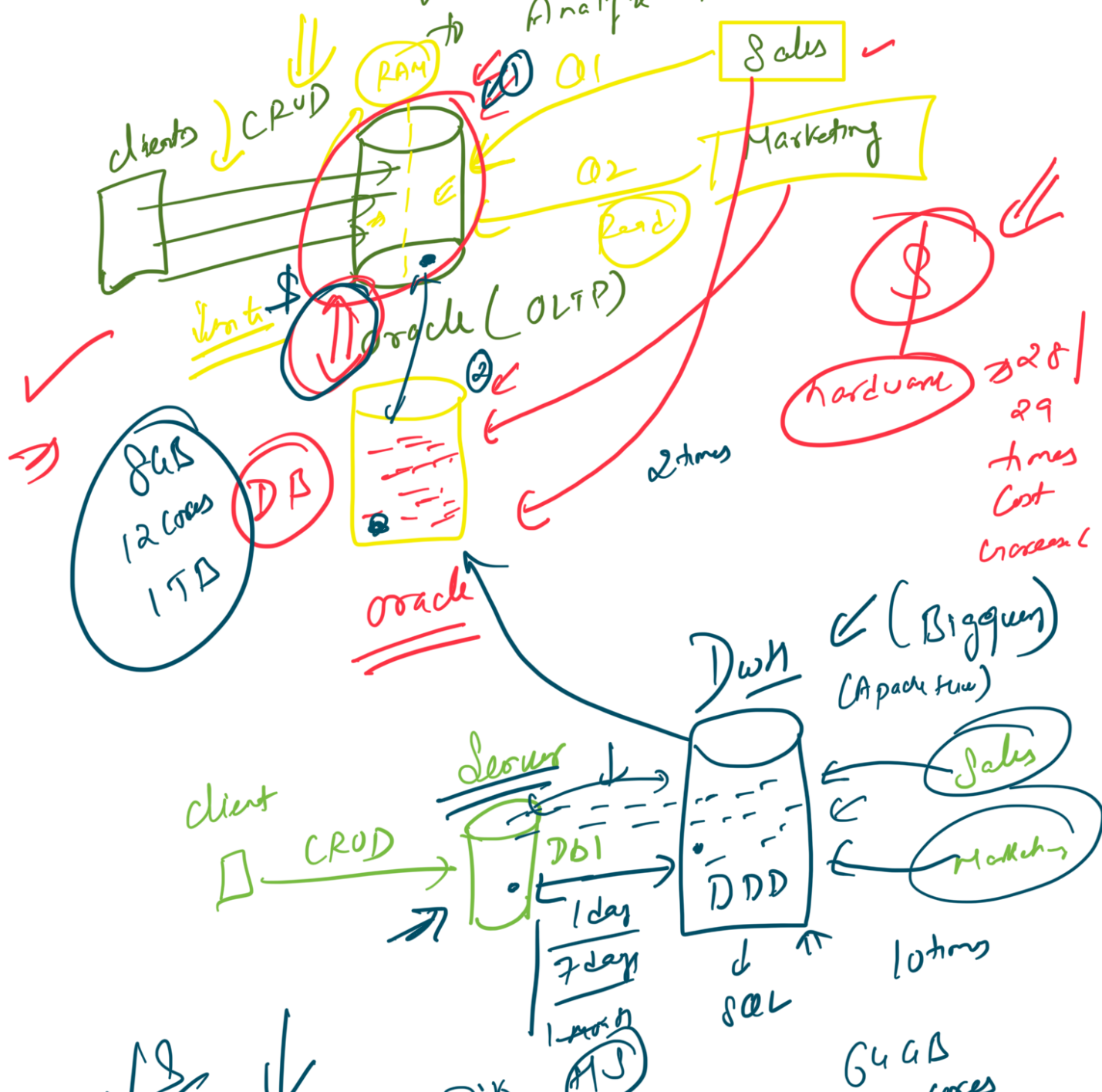
→ Limitation: DBs are not a J^u-analysis.

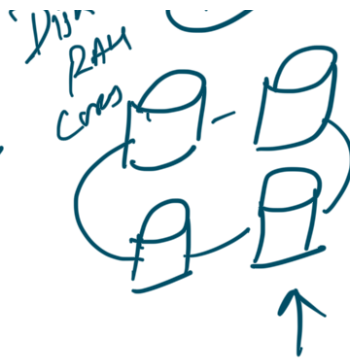
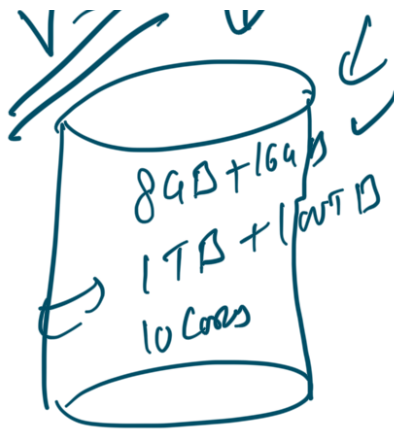
↳ DBs are not a J^u-analysis.

100

↳ DBS system can accessed by multiple users at same time.

↳ DBs are primarily designed for collecting large data fast and not necessarily to Analyse the data.

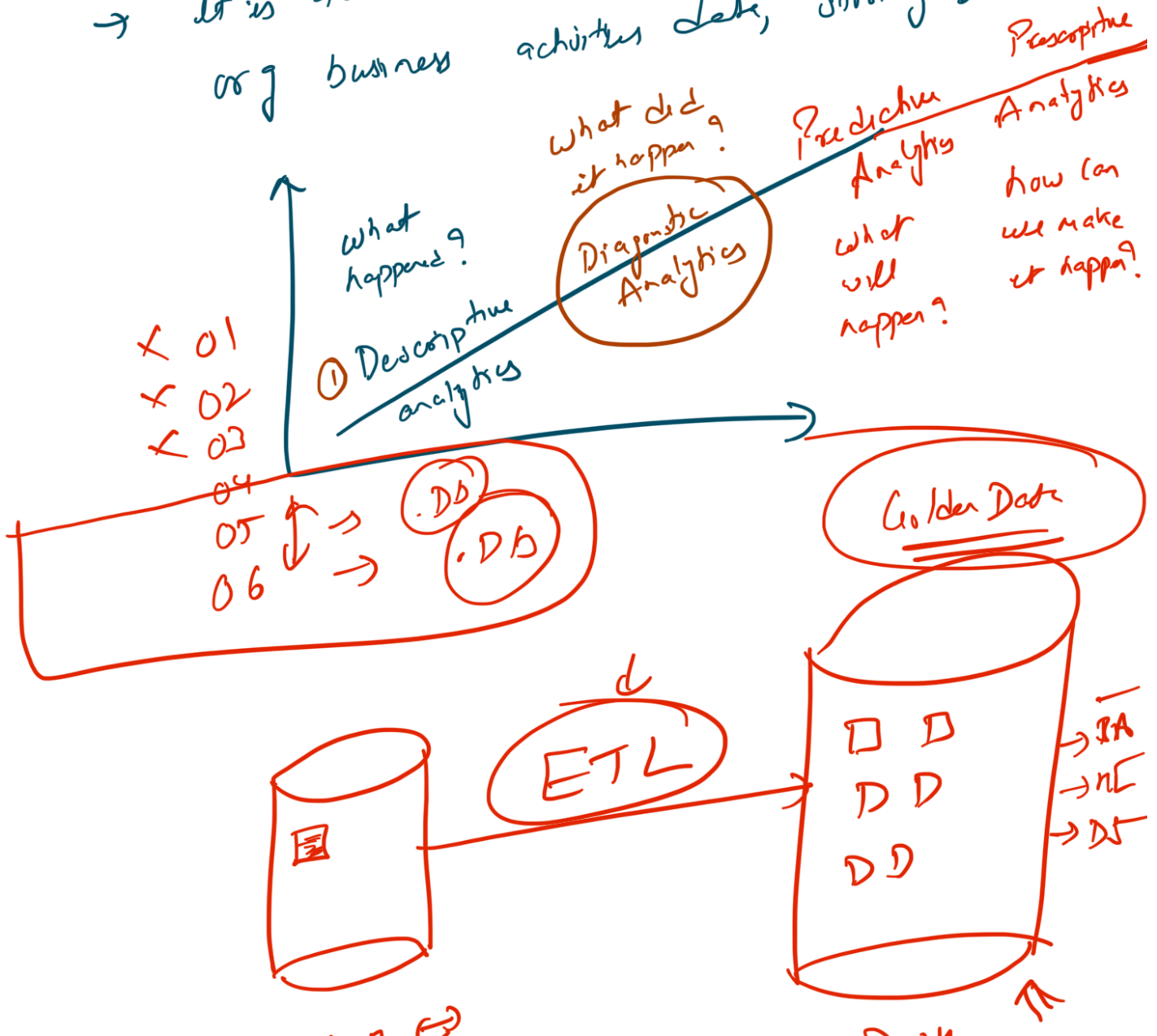




40+ GB
1 petabytes

Intro to Data Warehousing

→ It is the tool of collecting an org business activities data, strong.



DD

Don

Every Company will be a Data Company.

Why is dwh a better option?

① Business oriented
↳ intended to contain large amounts of business data

2.2 Billion
↓
4mins

30dec
35dec
22

102.8 Million → 3hrs → RTD
↳ 3.2 hrs

②

Fast Query Response time. (Architecture)

③

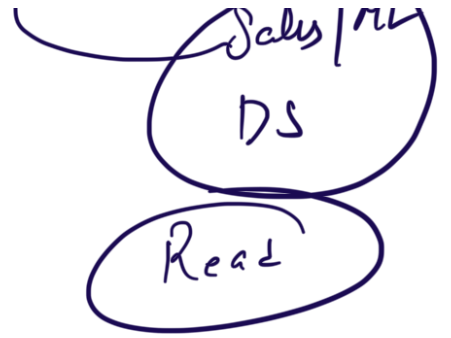
Data Quality is of high standard.

④

Integrated



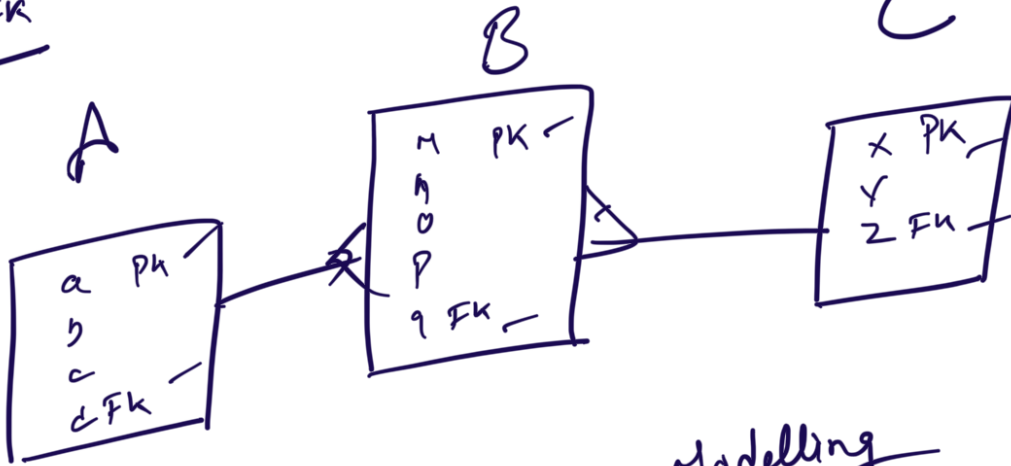
↓ ↓
WORM



→ normalization
→ PK / FK

Design a Data Warehouse

denormalization
PK
FK

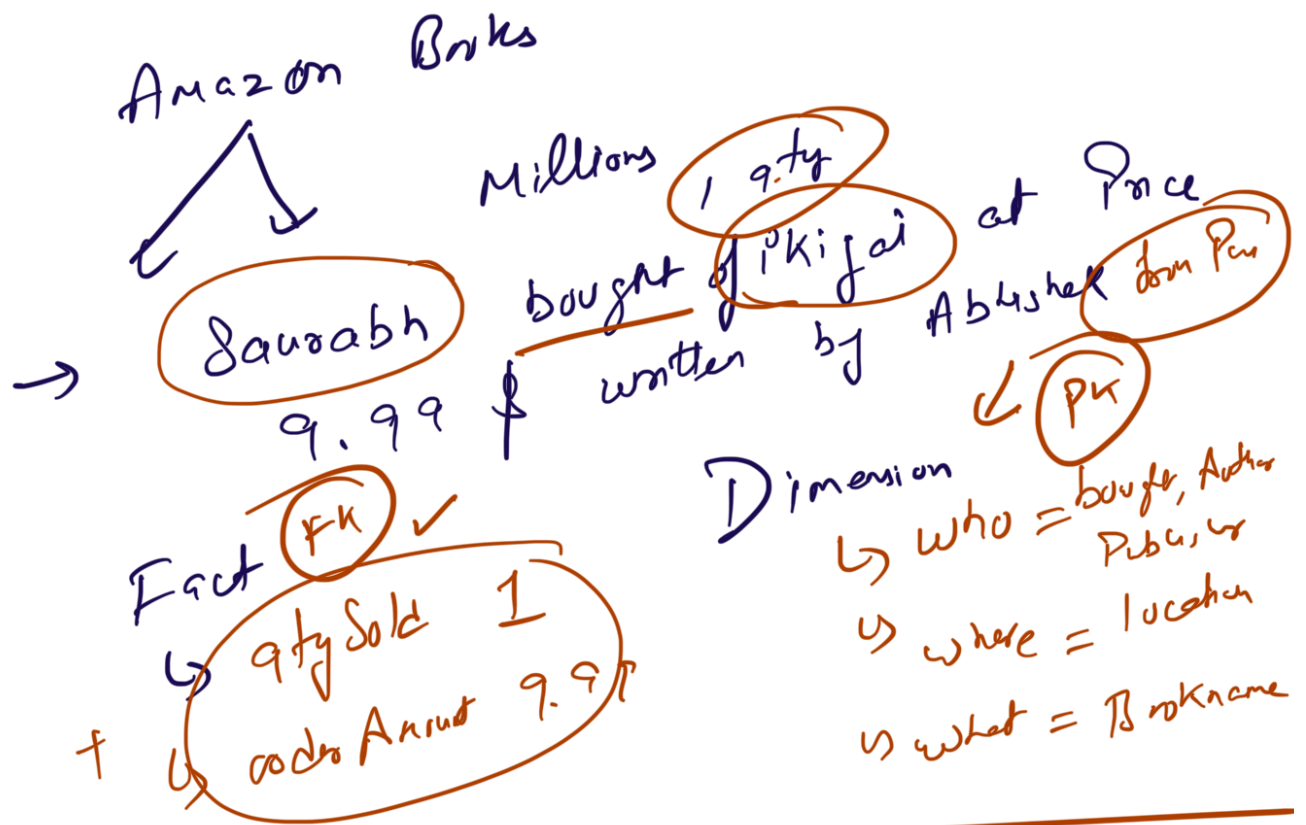


Dimensional Modelling

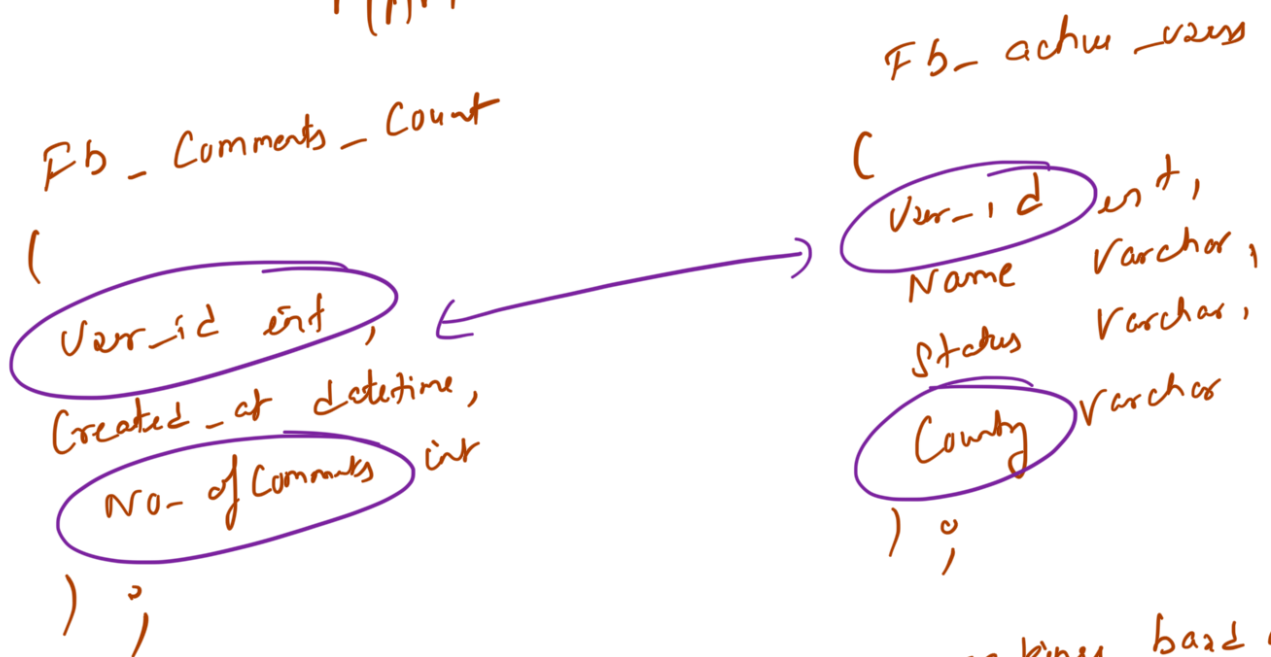
is a technique which include a set of methods
practices, & concepts to design and
store data in DWH.

- (1) Fact table = measurements / Measures or Facts for your business process.
- (2) Dimension table = Content surrounding a business process event.

→ Fact table ⇒ Columns
Numbers Foreign Keys



MAANG Question



Qn.1) Which Countries have order in rankings based on the no. of Comments between Dec 2021 vs Jan 2022? (Avoid gaps b/w ranks)

CN	No. of - Com
IN	110

→

CN	cnts
IN	110

1	USA	100
2	ZN	90
3	UK	80
4	FR	60

Dec 2021

↓
Cler

2	USA	100
3	FR	90
4	UK	85

Jan 2022

↓
Cler

1	FR	100
2		
3		
4		

Join
↓
Group by → Jan (no of months)
↓
dense_rank
↓
Comparison