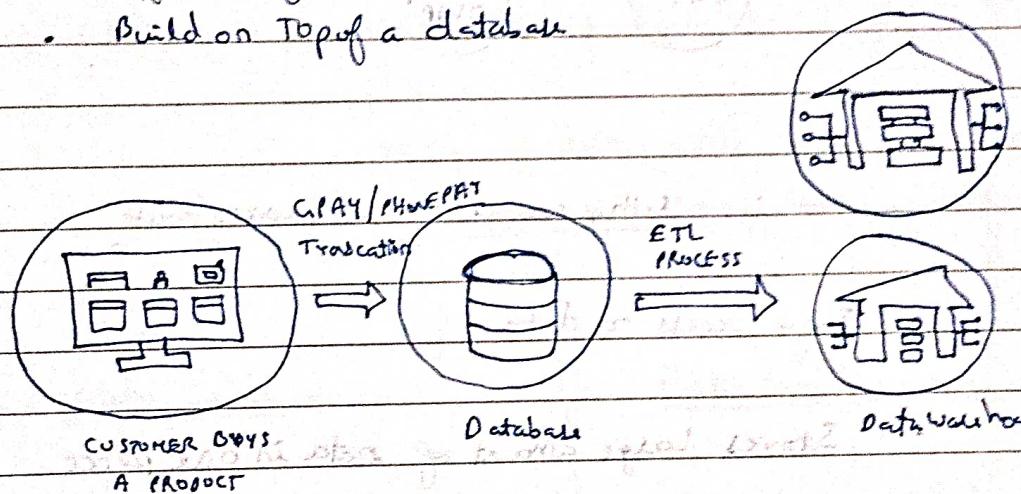


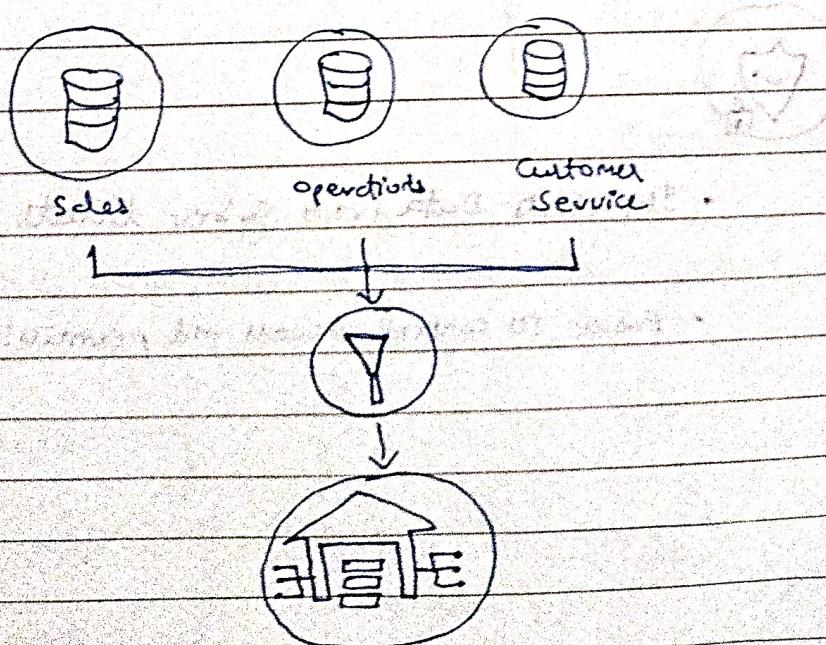
Data Engineering and Datawarehouse.

Data warehouse → A large warehouse that is filled with data rather than physical products.

- Large in Size
- Different from a database
- Build on Top of a database

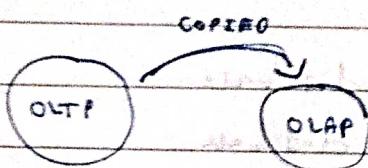


Online Transaction Processing (OLTP) Database



Online Analytical Processing (OLAP)

- Used for reading and analyzing large amount of data
- More data source → More complex of data warehouse



Why we use Data warehouse

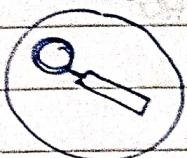
- Easy access to data
- Stores large amount of data in one place.
- Data base → Data Analysis (X)
Data warehouse → Data Analysis (M)



- It protects Data from cyber threats
- Easier to control Access and permission

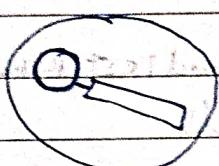
DATA WAREHOUSE

Data LAKE VS DATA WAREHOUSE



STRUCTURED DATA

Type of data with defined Schema
(Ex: Table in relational Database)



UNSTRUCTURED DATA

Type of data with no fixed Schema
(Ex: plainText, image, audio or video file)

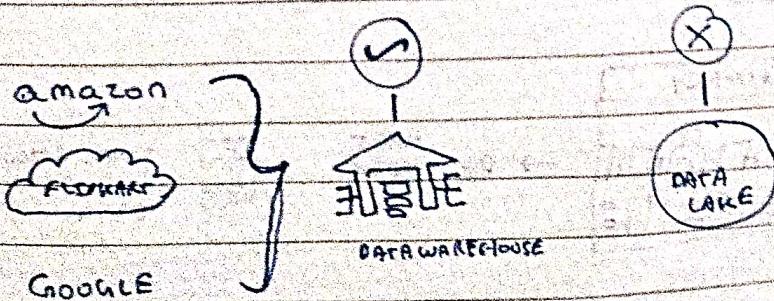
Data Lake

- Stores Structured and Unstructured data
- Better for storing raw data

Data Warehouse

- Stores Structured data
- Better for storing data that has been cleaned, transformed and organized (ETL) process

① Data warehouse is better than Data Lake.



DATAMART (VS) DATA WAREHOUSE

DATAMART → SMALL DATA MART

HOSPITAL → Data Warehouse : stores all patient information

PHARMACY → DATA MART : Blood Test results, Medicine details

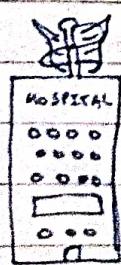
Data Mart

- Small, focused
- Target a specific group

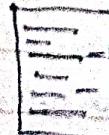
Data warehouse

- Large, Centralized repository of Data

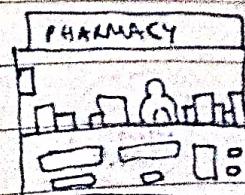
(v) Data Mart and Data warehouse both are Unique



→ DATA WAREHOUSE



LARGE AMOUNT OF DATA



→ DATA MART



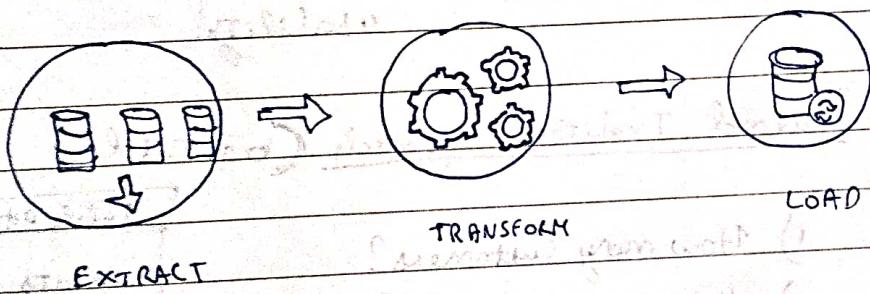
SMALL AMOUNT OF DATA

Extract, Transform, Load (ETL)

- 1) Gather data from a variety of sources
- 2) Cleaned Transform the data
- 3) Load the data into a destination

(Datawarehouse, DataMart, DataLake)

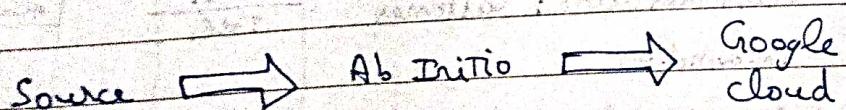
ETL Process



ETL → Securedata

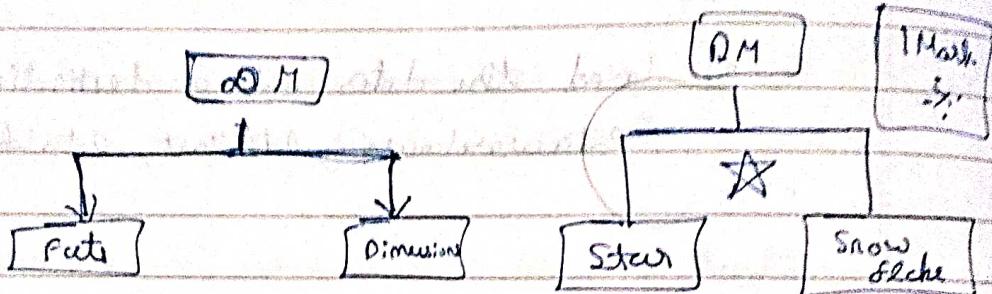
TOOLS

- SSIS
- Informatica
- Ab initio



Dimensional modelling.

- A technique used to organize and structure data in a data warehouse.



Dimensional modelling → High performance and better usability.

Normal Traditional Approach (OneTable)

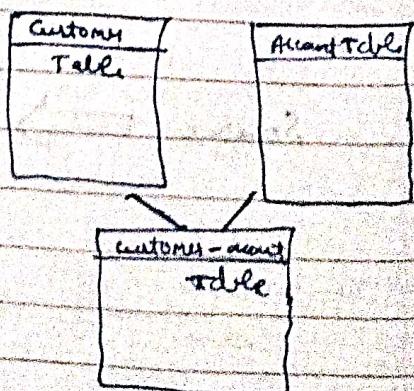
- 1) How many customers?
- 2) How many accounts?
- 3) How many Account type?

- Difficult answering these questions

Table Database
Customer
First Name
Last Name
Contact No.
Email
Address
Account Type

DM Box Approach (MultipleTables)

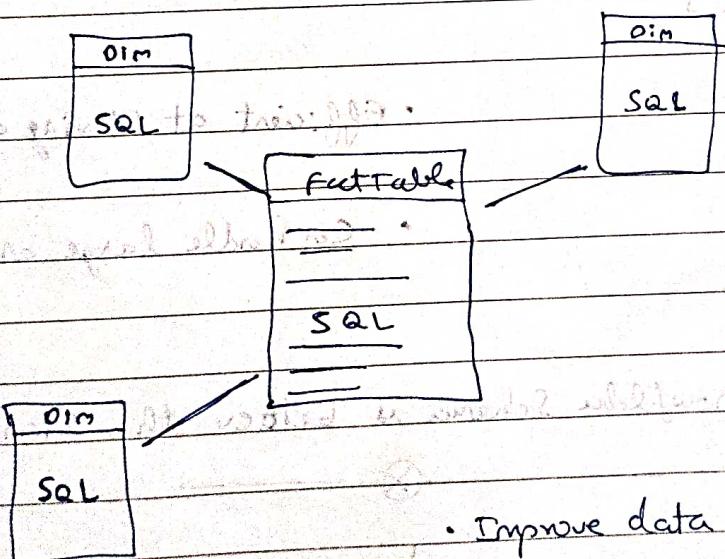
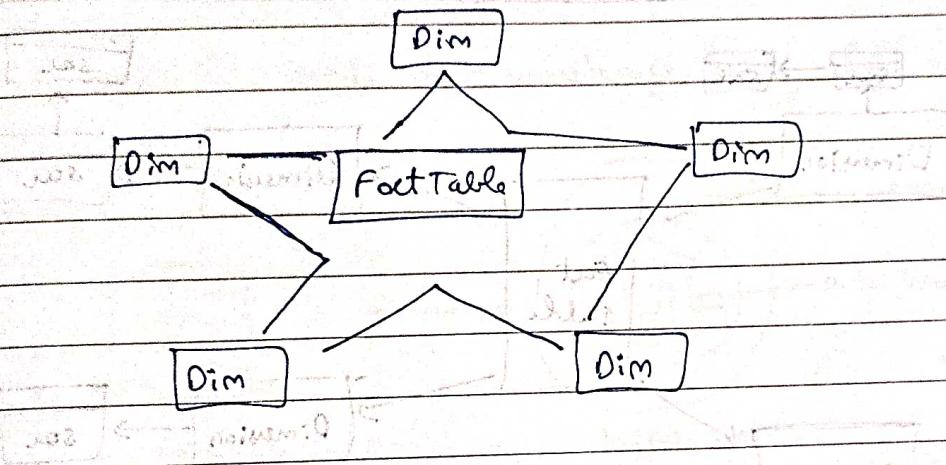
- 1) Customer table: Customer information
- 2) Account table: account information
- 3) Customer-account table: relationship between customer's and account's



Fact - ~~Q~~

Individual piece of data that can be stored in a data warehouse.

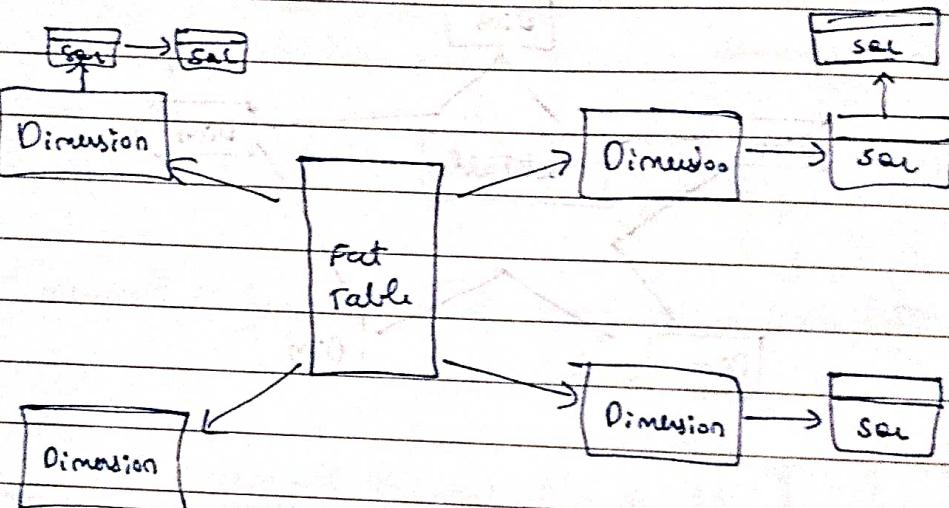
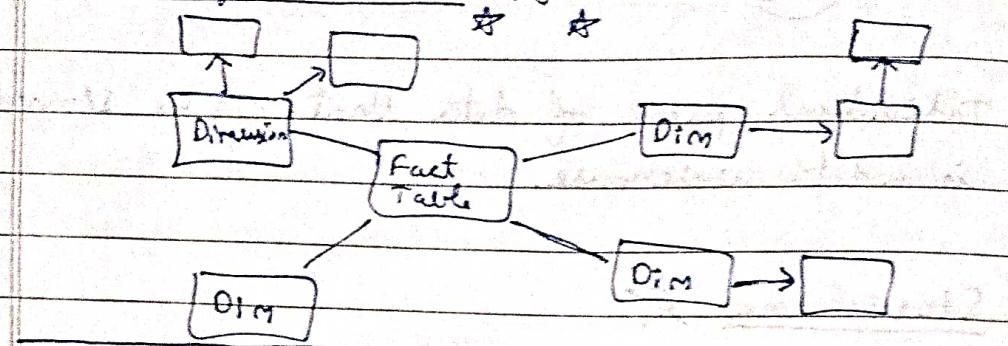
Star schema *



- Improve data Organization
- Improve query Performance

① Snowflake Schema is better than Star Schema.

Snowflake Schema



- Efficient at storing and retrieving data

- Can handle large amounts of data

- Snowflake Schema is better than Star Schema

Star Schema

- Suitable for small amounts of data

- Fast performance

Snowflake Schema

- Suitable for large amounts of data

- Slow performance

SQL Server

SSMS → SQL Server Management Studio

Install → SQL Server

Install → SSMS

Next → Create datawarehouse using SQL Server.

