

Data Warehouse :-

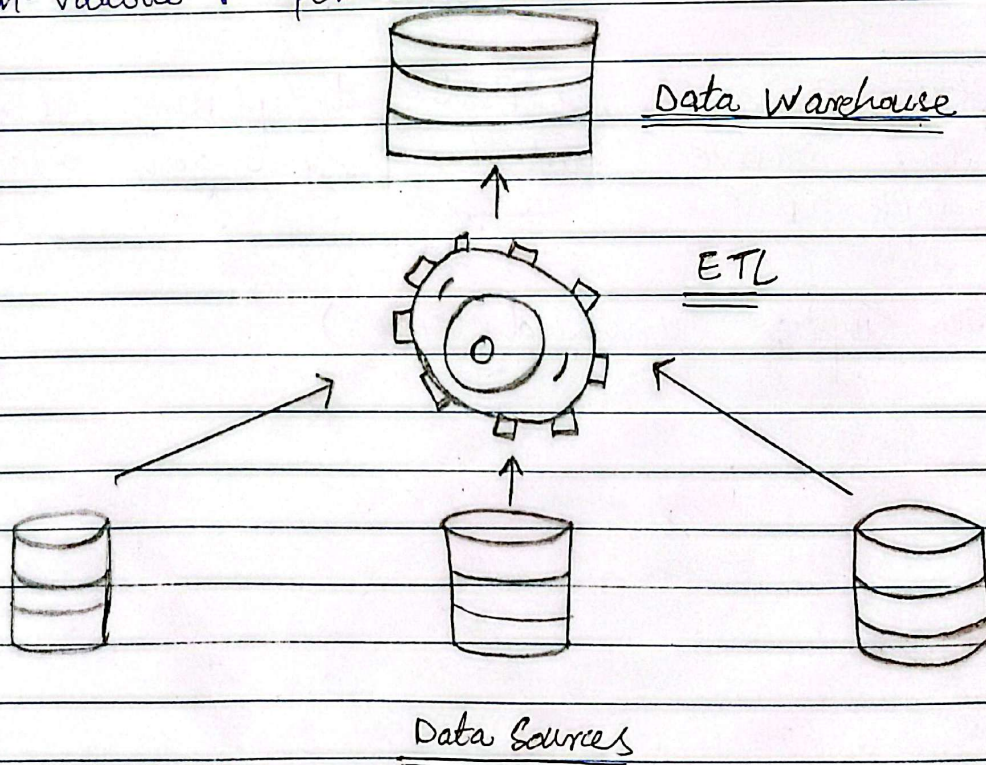
A subject-oriented, integrated, time-variant, and non-volatile collection of data in support of management's decision-making process.

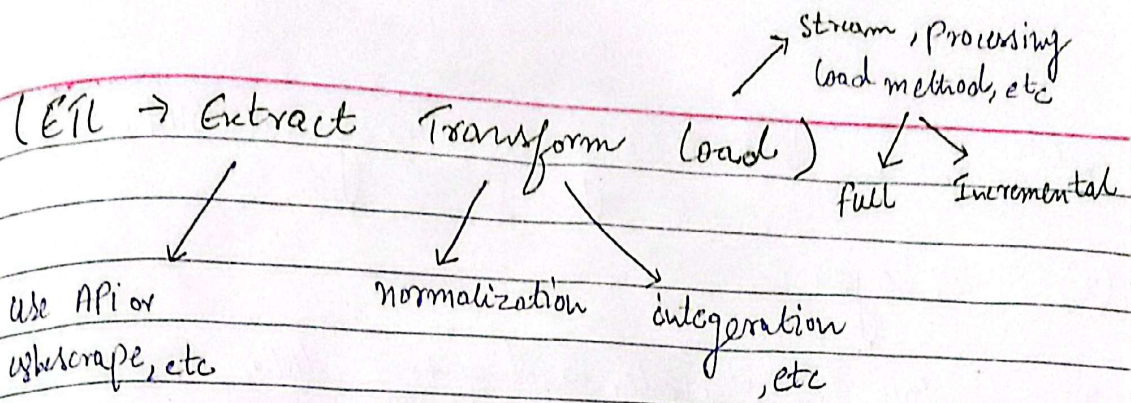
Used for OLAP → Online Analytical Process

reporting

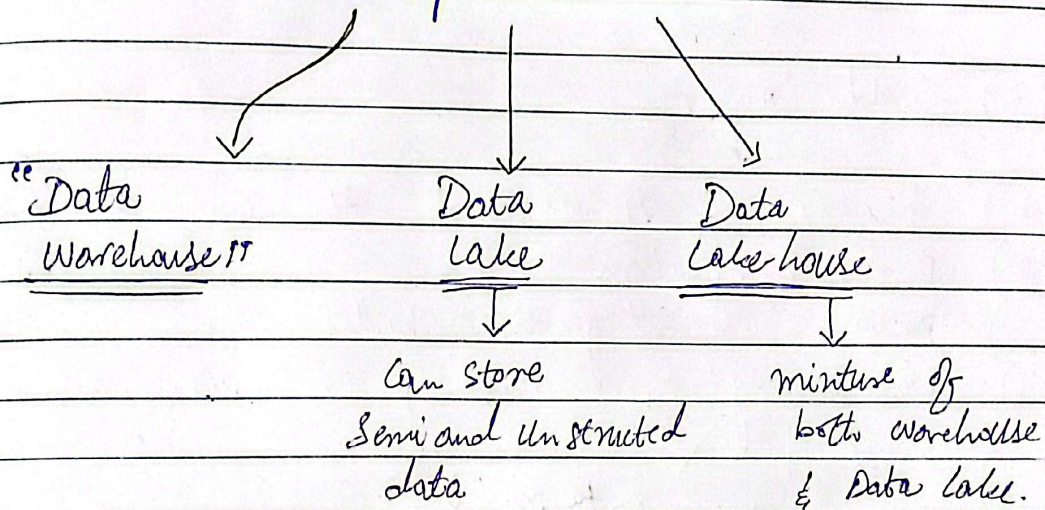
analytics

- time-variant : historical info can be found.
- integrated : organized (multiple sources into one)
- subject oriented : based on specific subject.
- non-volatile : permanent.





"Data Architectures"



- * Method 1 : Inmon → stage - EDW (3NF) - Data Marts
- " 2 : Kimball → stage → Data Marts
- " 3 : Data vault → stage - Row vault & Business vault - Data Mart
- " 4 : Medallion Architecture → Bronze - Silver - Gold

* Data Mart → Small subset of Data warehouse

1) Bronze Layer :-

- 1) Interview source system Export
- 2) Data Ingestion: A way to load the data into data warehouse from source.
- 3) Data completeness & Schema checker

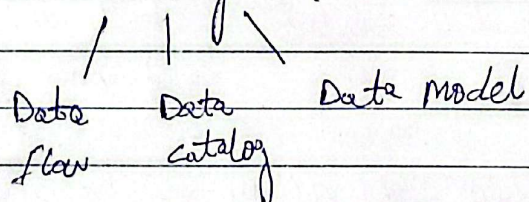
• 4) Data documenting & versioning in GIT.

(2) Silver Layer :-

- 1) Explore & understand the data.
- 2) Data cleaning
- 3) Data correctness check
- 4) Documenting & versioning in GIT.

(3) Gold Layer :-

- 1) Explore & Understand the Business objects.
- 2) Data Integration
- 3) Data Integration checks
- 4) Documenting & versioning in GIT.



Q) Data Modeling :-

is the process of creating a visual representation or blueprint of the data, its structure, and the relationships between different data elements within a database.

Types :-

conceptual → high level

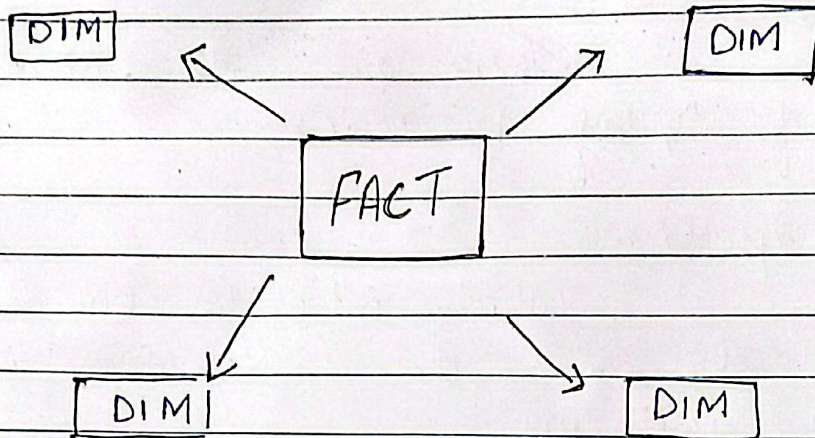
logical → This contains more details

physical → most detailed one

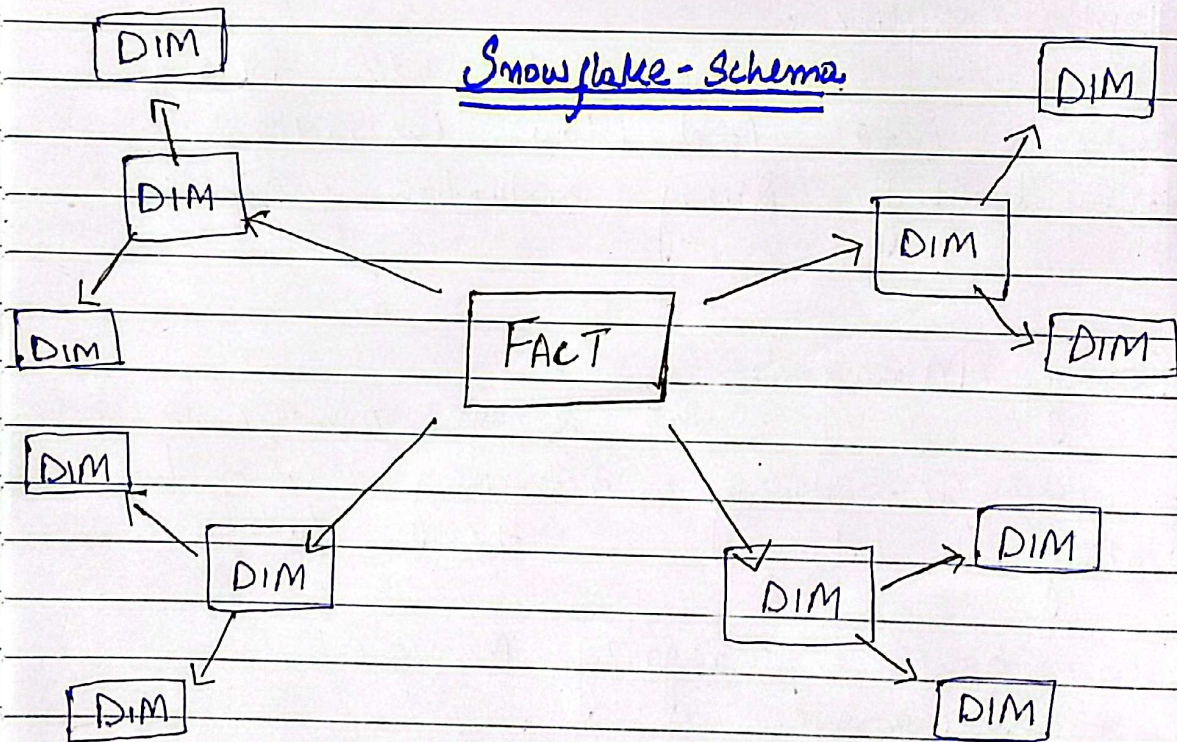
→ Serves as a blueprint or logical structure of database.

* Schema:-

Star-Schema



Snowflake-Schema



- Star Schema → Simple & Easy (big dimensions issues)
- Snowflake Schema → (More complex) large datasets storage efficient.

- Dimension :-

Descriptive information that give context to your data.

- like product info
who? what? where?

- Fact :-

Quantitative information that represents events.

- like Transactions
How much? How many?

- Surrogate key :-

System-generated unique identifier assigned to each record in a table.