What is Database

- Collection of data or information.
- Typically used to support OLTP
- DBMS store data in the database and enable users and applications to interact with the data.

Database Characteristics

- Security features to ensure that data can only be accessed by authorized users.
- ACID transactions to ensure data integrity.
- Query lang and API's to easily interact with data in the database
- Full text search.
- Private cloud, public cloud, hybrid-cloud and/or multicloud hosting options.

Why use a database?

- For storing data of diff types like :-
  - Medical Records
  - Financial Records
  - Articles & blog articles
  - IOT device readings.

Database Examples :-

- Relational Database :- Oracle, MySQL, Microsoft SQL Server
- Document Database :- MongoDB, and Couch DB
- Key-Value Database :- Redis and Dynamo DB
- Wide-Column Database :- Cassandra and HBase
- Graph Databases :- Neo4j and Amazon Neptune.

(2)

OLAP data warehouses and data lakes

- Both DW and DL are meant to support OLAP
- OLAP systems are typically used to collect data from a variety of sources
- Data is used to power a range of analytical use cases ranging from business intelligence and reporting to forecasting.

What is data warehouse?

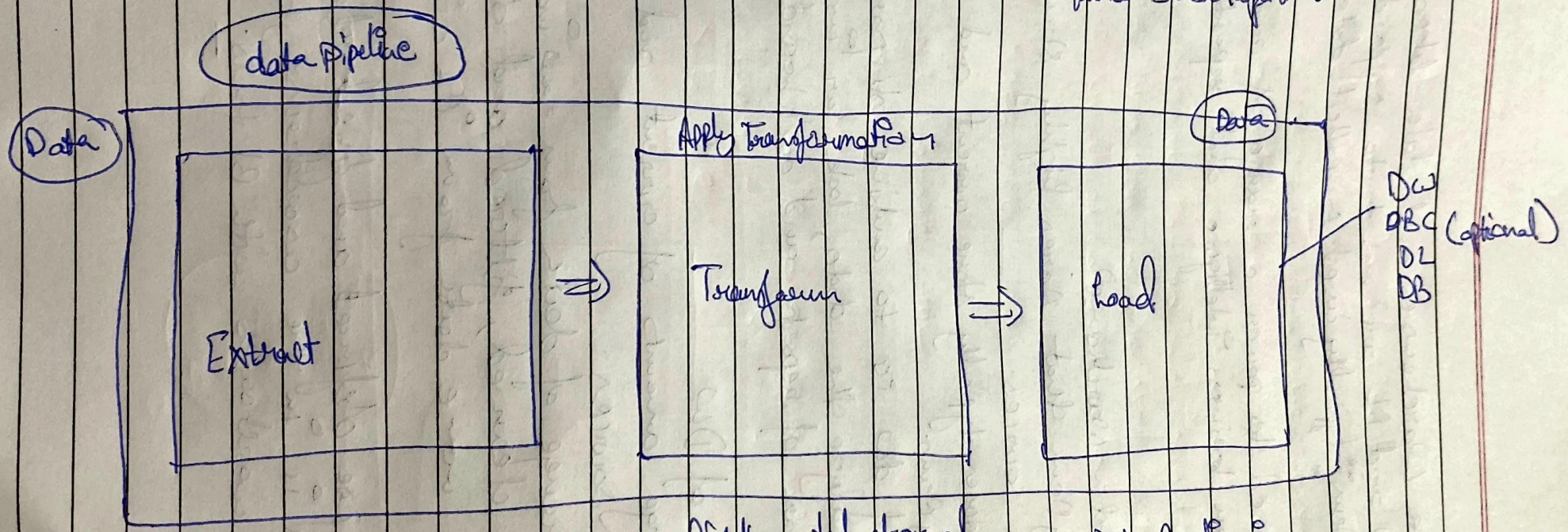
- It is a system that stores highly structured information from various sources.
- Data warehouses typically store current and historical data from one or more systems.
- The goal of dw is to combine disparate data sources in order to analyze the data, look for insights and create BI in the form of reports and dashboards.

Characteristics of DW.

- Can store large amounts of current and historical data from various sources.
- Contain a range of data from raw ingested data to highly curated, cleansed, filtered and aggregated data.
- ETL process move data from its original source to the DW.
- They have a pre defined and fixed relational schema.
- Once the data is in the warehouse, business analysts can connect data warehouses with BI tools.

Extract, Transform, Load (ETL) Process.

Sequence is also important.



Raw data from the sources

tools:

Highly curated, cleansed, filtered and aggregated data

Data loading in DW, DBC, Datalake, Database

(40)

Why use a DW?

⇒ They are a good option when we need to store large amounts of data and/or perform in-depth analysis of your data to generate business intelligence.

Examples of DWs:-

1. Amazon Redshift
2. Google BigQuery
3. IBM Db2 Warehouse
4. Microsoft Azure Synapse.
5. Snowflake.
6. Teradata Vantage.

What is data lake?

⇒ It is a repository of data from disparate sources that is stored in its original, raw format.

⇒ They also store large amounts of historical data or current data like DWs.

⇒ They can store data in variety of formats like JSON, BSON, CSV, TSV, Avro, ORC and Parquet.

⇒ Primary purpose is to analyze the data to gain insights.

⇒ With tools and technologies a data lake can also update the storage layer of a database.

⇒ Tools like Hadhurst, Presto, Dremio and Atlas Data Lake can give a database like views into the data stored in your data lake.

#

53

Data Lake Characteristics.

1) They store large amounts of structured, semi-structured and unstructured data.
2) We don't need to transform the data in order to add it to data lake.
3) They are flexible in nature, which helps data scientists to look for unexpected patterns and insights.
4) Data in data lakes can be processed with a variety of different systems and visualized with BI tools.

Examples of data lake

- AWS S3
- Azure data lake Storage Gen2
- Google Cloud Storage

Some technologies enable organising and querying data in data lakes, including:-

- MongoDB Atlas Data Lake
- AWS Athena
- Presto
- Starburst
- Databricks SQL Analytics.

6

Diff b/w Database, Data Warehouse and Data lake.

- A database stores the current data required to power an application
- DW stores current & historical data from one or more systems in a predefined and fixed schema.
- Data lake stores current & historical data in its raw form.

⇒ Basically, they all have their own purpose. Nearly every modern app. will require a database and the organizations can choose to complement their database with a data warehouse, data lake or both.

⇒ Differences table is in the document.