Lecture-3

Difference between Database System and Data Warehouse–

Database

1. Used for Online Transactional Processing (OLTP) but can be used for other purposes such as Data Warehousing. This records the data from the user for history.
2. The tables and joins are complex since they are normalized (for RDMS). This is done to reduce redundant data and to save storage space.
3. Entity – Relational modelling techniques are used for RDMS database design.
4. Optimized for write operation.
5. Performance is low for analysis queries.

Data Warehouse

1. Used for Online Analytical Processing (OLAP). This reads the historical data for the Users for business decisions.
2. The Tables and joins are simple since they are de-normalized. This is done to reduce the response time for analytical queries.
3. Data – Modelling techniques are used for the Data Warehouse design.
4. Optimized for read operations.
5. High performance for analytical queries.
6. Is *usually* a Database.

It's important to note as well that Data Warehouses could be sourced from zero to many databases.

Database:

Used for Online Transactional Processing (OLTP).

- **Transaction-oriented.**
- **Application oriented.**
- **Current data.**
- **Detailed data.**
- **Scalable data.**
- **Many Users, Administrators / Operational.**
- **Execution time: short.**

Data Warehouse:

Used for Online Analytical Processing (OLAP).

- **Oriented analysis.**
- **Subject oriented.**
- **Historical data.**
- **Aggregated data.**
- **Static data.**

- Not many users, manager.
- Execution time: long.

Database Use Cases

Database use cases are related to the day-to-day transactional requirements in an organization. Some examples of database applications include:

- An airline using an online booking system
- A hospital registering a patient
- A bank adding an ATM withdrawal transaction to an account
- A website creating an order for a product it has sold

Data Warehouse Use Cases

Data warehouse use cases focus on providing high-level reporting and analysis that lead to more informed business decisions. Use cases include:

- Carrying out data mining to gain new insights from the information held in many large databases
- Conducting market research by analyzing large volumes of data in-depth
- An online business analyzing user behavior to make business decisions

Lecture - 4

The major differences between the **Databases** and **Data Warehouses** are as follows:-

FEATURES	DATABASE	DATA WAREHOUSE
Characteristic	It is based on Operational Processing.	It is based on Informational Processing.
Data	It mainly stores the Current data which always guaranteed to be up-to-date.	It usually stores the Historical data whose accuracy is maintained over time.
Function	It is used for day-to-day operations.	It is used for long-term informational requirements and decision support.
User	The common users are clerk, DBA, database professional.	The common users are knowledge worker (e.g., manager, executive, analyst)
Unit of work	Its work consists of short and simple transaction.	The operations on it consists of complex queries..
Focus	The focus is on "Data IN"	The focus is on "Information OUT"
Orientation	The orientation is on Transaction.	The orientation is on Analysis.
DB design	The designing of database is ER based and application-oriented.	The designing is done using star/snowflake schema and its subject-oriented.
Summarization	The data is primitive and highly detailed.	The data is summarized and in consolidated form.
View	The view of the data is flat relational.	The view of the data is multidimensional.
Access	The most frequent type of access type is read/write.	It mostly use the read access for the stored data.
Operations	The main operation is index/hash on primary key.	For any operation it needs a lot of scans.
Number of records accessed	A few tens of records.	A bunch of millions of records.
Number of users	In order of thousands.	In the order of hundreds only.

DB size	100 MB to GB.	100 GB to TB.
Priority	High performance, high availability	High flexibility, end-user autonomy
Metric	To measure the efficiency, transaction throughput is measured.	To measure the efficiency, query throughput