Name : Parth Nandedkar
Date : 17 Jan 2024
Topic : Data Warehousing
Batch : Data Engineering Batch-1

Data Warehousing

Definition :

It is a collection of data designed to support management decision making by presenting a coherent picture of business conditions at a single point of time.

ETL (Extract, Transform, Load):

Data from source systems undergoes ETL processes to extract, transform, and load it into the data warehouse.

Transformation may include cleaning, aggregating, and conforming data to a common format.

Characteristics of Data Warehousing:

-**Subject-Oriented:**
Data warehouses are subject-oriented, meaning they focus on providing information about a specific subject, such as sales, finance, or customer data.

-Integrated:
Data integration is a key feature. Data from different sources is harmonised and transformed to ensure consistency across the warehouse.

-Time-Variant:
Data warehouses store historical data, allowing users to analyse trends and changes over time. This is essential for making strategic business decisions.

-Non-Volatile:
Once data is loaded into the warehouse, it is typically not updated or deleted. Instead, new data is added to maintain a historical record.

DSS:

DSS stands for Decision Support System, and its primary purpose is to assist decision-makers in making informed and timely decisions. The need for Decision Support Systems arises from the complexity and volume of data that organisations deal with, as well as the increasing pace of business operations.

DSS architectural styles:

-OLTP (Online Transaction Processing):
used by traditional operational systems (RDBMS).

-OLAP (Online Analytical Processing):
used by Data Warehouse.

OLTP (Online Transaction Processing):

-Purpose:
OLTP systems are designed for transaction-oriented processing. They handle day-to-day, routine operations such as data entry, updating, and retrieval of small amounts of data in real-time.

-Database Structure:
OLTP databases are normalised, meaning they are designed to minimise redundancy and maintain data consistency. This structure is suitable for frequent updates and inserts.

-Volume of Data:
OLTP systems deal with a large number of short, simple transactions. The volume of data processed is typically high, but the individual transactions are relatively small.

-Queries:
OLTP queries are simple and focused on retrieving specific records. The goal is to support efficient and rapid transaction processing.

-Concurrency Control:
OLTP systems require robust concurrency control mechanisms to manage simultaneous transactions from multiple users.
Response Time:
The response time for OLTP systems is critical. Users expect quick responses to their transactions, making low-latency processing a priority.

-Example:
Examples of OLTP applications include online banking systems, point-of-sale systems, order processing systems, and airline reservation systems.

OLAP (Online Analytical Processing):

-Purpose:
OLAP systems are designed for complex, read-intensive queries and data analysis. They support decision-making and business intelligence activities.

-Database Structure:
OLAP databases are typically denormalized or partially denormalized. This structure is optimised for querying and reporting, allowing for faster analytical processing.

-Volume of Data:
OLAP systems deal with a large volume of historical and aggregated data. These systems are optimised for complex analytical queries that involve aggregations and comparisons.

-Queries:
OLAP queries are complex and involve aggregations, grouping, and data slicing. The goal is to provide insights into trends, patterns, and relationships within the data.

-Concurrency Control:
OLAP systems prioritise read consistency over high concurrency. They are optimised for data analysis and reporting rather than simultaneous transaction processing.

-Response Time:
While OLAP systems aim for reasonable response times, the emphasis is more on providing in-depth analysis, even if it takes longer than OLTP systems.

-Example:
Examples of OLAP applications include data warehouses, executive information systems, and business intelligence tools used for strategic decision-making.

Summery :

Data warehousing plays a crucial role in enabling organisations to leverage their data for strategic decision-making. It provides a unified and historical view of data, empowering business users with the information they need to understand trends, patterns, and insights. While there are challenges, the benefits of having a well-designed data warehouse make it a fundamental component of modern data infrastructure.