Name - Aparna Bharti

Batch - Data Engineering batch 1

**Data Warehouse**

Data warehouse is a type ofdata management system that is designed to support business intelligence (BI) activities.

Data warehouses are used to perform queries and analysis and often contain large amounts of historical data. The data within a data warehouse is usually derived from a wide range of sources such as application log files and transaction applications,spreadsheets.

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.

**Data Warehouse features**

**1) Subject-oriented**

This means that the data is organised around specific subjects, such as customers, products, or sales. This allows for easy access to the data relevant to a specific subject.

It provides topic-wise information rather than the overall processes of a business.

**2) Integrated**

This means that the data is collected from various sources, such as transactional systems, and then cleaned, transformed, and into a single view. This allows for easy access and analysis of the data

The data warehouse can be accessed by any department within an organisation, and the data can be easily structured into spreadsheets or tables for analysis purposes.

**3) Time-variant**

which means that the data is stored with a time dimension. This allows for easy access to data for specific time periods, such as last quarter or last year

So inside a data warehouse, a time variant table can be structured almost exactly the same as the source table, but with the addition of a timestamp column.

**4) Nonvolatile**

This means that the data in the warehouse is never updated or deleted, only added

data that is stored in a persistent manner, meaning it remains intact and accessible even in the event of system restarts or power failures.

**DATA Warehousing applications**

**1) Information processing**

supports querying, basic statistical analysis, and reporting using crosstabs, tables, charts and graphs.

**2)Analytical processing**

multidimensional analysis of data warehouse data.supports basic OLAP operations, slice-dice, drilling, pivoting

**3) Data mining**

the process of extracting useful data from the databases. The data mining process depends on the data compiled in the data warehousing phase to recognize meaningful patterns

**OLTP**

On-Line Transaction Processing (OLTP) System refers to the system that manages transaction oriented applications. These systems are designed to support on-line transactions and process queries quickly on the Internet.

**For example:** the POS (point of sale) system of any supermarket is an OLTP System.

**Benefits of OLTP**

1. It has ability to handle many transaction requests simultaneously (called concurrency) and the ability to reliably backup and continue if part of the system fails (called atomicity)
2. It allows its users to perform operations like read, write and delete data quickly.

**Pitfalls of OLTP**

1. The data we get from OLTP is not suitable for data analysis.
2. To perform one simple transaction even with the normalised structure, we need to query multiple tables by using joins.

**OLAP**

OLAP (online analytical processing) is a computing method that enables users to easily and selectively [extract](https://www.techtarget.com/searchbusinessanalytics/answer/Examining-different-data-access-methods-OLAP-and-data-mining) and [query](https://www.techtarget.com/searchdatamanagement/definition/query) data in order to analyse it from different points of view

**Benefits of OLAP**

1. OLAP is the technology that many business intelligence programs use, and it automatically calculates its figures and trends
2. Unlike flat or two-dimensional data, such as sales over time, the multiple dimensions in an OLAP cube show how more aspects of a business affect each other

**Pitfalls of OLTP**

1. Pre-modeling is a must. Regarding the business data, the traditional OLAP tools do not allow for the immediate analysis without pre-modeling.

**Types of OLAP**

**1) ROLAP(Relational OLAP)**

In ROLAP data is stored in a relational database. In essence, each action of slicing and dicing is equivalent to adding a “WHERE” clause in the SQL statement. ROLAP can handle large amounts of data. ROLAP can leverage functionalities inherent in the relational database.

**2) MOLAP(Multidimensional OLAP)**

A MOLAP system is based on a native logical model that directly supports multidimensional data and operations. Data is stored physically into multidimensional arrays, and positional techniques are used to access them.

**RDBMS**

A relational database is a collection of information that organises data in predefined relationships where data is stored in one or more tables (or "relations") of columns and rows, making it easy to see and understand how different data structures relate to each other.

**Examples -** MySQL, PostgreSQL, MariaDB, Microsoft SQL Server, and Oracle Database.

**Components of RDBMS**

**Tables**

A table is a structured set of data organised in rows and columns. Its purpose is to store and manage data efficiently

**Columns (Attributes)**

Columns, also known as attributes, define the structure of a table by specifying the type of data each entry in the table will hold.

**Rows (Records)**

Rows, or records, are individual instances of data stored within a table. Each row represents a single entry in the table.

**Primary Key**

A primary key is a unique identifier for each row in a table. It ensures that no two rows have the same identifier, providing a consistent way to reference specific records.

**Foreign Key**

A foreign key is a column or set of columns in a table that refers to the primary key of another table. It is used to establish relationships between tables, allowing you to associate data across multiple tables in a meaningful way.

**SQL and its features**

Sql is structured query language (SQL) is a programming language for storing and processing information in a relational database. A relational database stores information in tabular form, with rows and columns representing different data attributes and the various relationships between the data values.

## **SQL commands**

Structured query language (SQL) commands are specific keywords or SQL statements that developers use to manipulate the data stored in a relational database. You can categorise SQL commands as follows.

### **1) Data definition language**

Data definition language (DDL) refers to SQL commands that design the database structure. Database engineers use DDL to create and modify database objects based on the business requirements. For example, the database engineer uses the CREATE command to create database objects such as tables, views, and indexes.

### **2) Data query language**

Data query language (DQL) consists of instructions for retrieving data stored in relational databases. Software applications use the SELECT command to filter and return specific results from a SQL table.

### **3) Data manipulation language**

Data manipulation language (DML) statements write new information or modify existing records in a relational database. For example, an application uses the INSERT command to store a new record in the database.

### **4) Data control language**

Database administrators use data control language (DCL) to manage or authorise database access for other users. For example, they can use the GRANT command to permit certain applications to manipulate one or more tables.

### **5) Transaction control language**

The relational engine uses transaction control language (TCL) to automatically make database changes. For example, the database uses the ROLLBACK command to undo an erroneous transaction.

**Features of SQL**

1) SQL can execute queries against a database

2) SQL can retrieve data from a database

3) SQL can insert records in a database

4) SQL can update records in a database

5) SQL can delete records from a database

6) SQL can create new databases

7) SQL can create new tables in a database

8) SQL can create stored procedures in a database

9) SQL can create views in a database

10) SQL can set permissions on tables, procedures, and views