UNIT:1

Database Concepts (File System and DBMS),

| **Basis** | **File System** | **DBMS** |
| --- | --- | --- |
| **Structure** | The file system is a way of arranging the files in a storage medium within a computer. | DBMS is software for managing the database. |
| **Data Redundancy** | Redundant data can be present in a file system. | In DBMS there is no redundant data. |
| **Backup and Recovery** | It doesn’t provide backup and recovery of data if it is lost. | It provides backup and recovery of data even if it is lost. |
| **Query processing** | There is no efficient query processing in the file system. | Efficient query processing is there in DBMS. |
| **Consistency** | There is less data consistency in the file system. | There is more data consistency because of the process of normalization. |
| **Complexity** | It is less complex as compared to DBMS. | It has more complexity in handling as compared to the file system. |
| **Security Constraints** | File systems provide less security in comparison to DBMS. | DBMS has more security mechanisms as compared to file systems. |
| **Cost** | It is less expensive than DBMS. | It has a comparatively higher cost than a file system. |
| **Data Independence** | There is no data independence. | In DBMS data independence exists. |
| **User Access** | Only one user can access data at a time. | Multiple users can access data at a time. |
| **Meaning** | The user has to write procedures for managing databases | The user not required to write procedures. |
| **Sharing** | Data is distributed in many files. So, not easy to share data | Due to centralized nature sharing is easy |

Database Storage

Structures (Tablespace, Control files, Data files)

In a database management system (DBMS), there are several important components involved in storing data. Let's explore the key structures involved: tablespace, control files, and data files.

Tablespace: A tablespace is a logical storage container within a database that holds data objects, such as tables, indexes, and partitions. It is used to organize and manage the physical storage of these objects. A database can have multiple tablespaces, and each tablespace consists of one or more data files.

Control files: Control files are essential components of a database that store metadata about the physical and logical structure of the database. They contain information about the database name, data file locations, tablespace details, redo log files, and other important parameters. Control files are crucial for the startup and recovery of the database.

Data files: Data files are physical files that store the actual data of the database. These files contain the table and index data, as well as other database objects. Each tablespace is composed of one or more data files. The data files are typically stored on the file system of the operating system where the database resides.

When a DBMS is installed and a database is created, the administrator configures the tablespace(s) and specifies the control file(s) and data file(s) locations. The control files and data files are managed by the DBMS and are updated during normal database operations such as inserting, updating, and deleting data

Structured and Unstructured data

|  |  |  |
| --- | --- | --- |
| **On the basis of** | **Structured data** | **Unstructured data** |
| **Technology** | It is based on a relational database. | It is based on character and binary data. |
| **Flexibility** | Structured data is less flexible and schema-dependent. | There is an absence of schema, so it is more flexible. |
| **Scalability** | It is hard to scale database schema. | It is more scalable. |
| **Robustness** | It is very robust. | It is less robust. |
| **Performance** | Here, we can perform a structured query that allows complex joining, so the performance is higher. | While in unstructured data, textual queries are possible, the performance is lower than semi-structured and structured data. |
| **Nature** | Structured data is quantitative, i.e., it consists of hard numbers or things that can be counted. | It is qualitative, as it cannot be processed and analyzed using conventional tools. |
| **Format** | It has a predefined format. | It has a variety of formats, i.e., it comes in a variety of shapes and sizes. |
| **Analysis** | It is easy to search. | Searching for unstructured data is more difficult. |

Structured Data

The data which is to the point, factual, and highly organized is referred to as structured data. It is quantitative in nature, i.e., it is related to quantities that means it contains measurable numerical values like numbers, dates, and times.

## Unstructured Data

All the unstructured files, log files, audio files, and image files are included in the unstructured data. Some organizations have much data available, but they did not know how to derive data value since the data is raw.

SQL Commands (DDL, DML & DCL)

Structured Query Language(SQL) as we all know is the database language by the use of which we can perform certain operations on the existing database and also we can use this language to create a database. [SQL](https://www.geeksforgeeks.org/structured-query-language/) uses certain commands like CREATE, DROP, INSERT, etc. to carry out the required tasks.

SQL commands are like instructions to a table. It is used to interact with the database with some operations. It is also used to perform specific tasks, functions, and queries of data. SQL can perform various tasks like creating a table, adding data to tables, dropping the table, modifying the table, set permission for users.

These [SQL](https://www.geeksforgeeks.org/sql-concepts-and-queries/)commands are mainly categorized into five categories:

1. DDL – Data Definition Language
2. DML – Data Manipulation Language
3. DCL – Data Control Language

### **DDL (Data Definition Language)**

[DDL](https://www.geeksforgeeks.org/features-of-structured-query-language-sql/) or Data Definition Language actually consists of the SQL commands that can be used to define the database schema. It simply deals with descriptions of the database schema and is used to create and modify the structure of database objects in the database. DDL is a set of SQL commands used to create, modify, and delete database structures but not data. These commands are normally not used by a general user, who should be accessing the database via an application.

List of DDL commands:

* [**CREATE**](https://www.geeksforgeeks.org/sql-create/): This command is used to create the database or its objects (like table, index, function, views, store procedure, and triggers).
* [**DROP**](https://www.geeksforgeeks.org/sql-drop-truncate/): This command is used to delete objects from the database.
* [**ALTER**](https://www.geeksforgeeks.org/sql-alter-add-drop-modify/)**:**This is used to alter the structure of the database.
* [**TRUNCATE**](https://www.geeksforgeeks.org/sql-drop-truncate/)**:**This is used to remove all records from a table, including all spaces allocated for the records are removed.
* [**COMMENT**](https://www.geeksforgeeks.org/sql-comments/): This is used to add comments to the data dictionary.
* [**RENAME**](https://www.geeksforgeeks.org/sql-alter-rename/)**:**This is used to rename an object existing in the database.

### **DML(Data Manipulation Language)**

The SQL commands that deal with the manipulation of data present in the database belong to DML or Data Manipulation Language and this includes most of the SQL statements. It is the component of the SQL statement that controls access to data and to the database. Basically, DCL statements are grouped with DML statements.

List of DML commands:

* [**INSERT**](https://www.geeksforgeeks.org/sql-insert-statement/): It is used to insert data into a table.
* [**UPDATE**](https://www.geeksforgeeks.org/sql-update-statement/)**:** It is used to update existing data within a table.
* [**DELETE**](https://www.geeksforgeeks.org/sql-delete-statement/): It is used to delete records from a database table.
* [**LOCK:**](https://www.geeksforgeeks.org/sql-lock-table/) Table control concurrency.
* **CALL:**Call a PL/SQL or JAVA subprogram.
* **EXPLAIN PLAN:** It describes the access path to data.

### **DCL (Data Control Language)**

DCL includes commands such as GRANT and REVOKE which mainly deal with the rights, permissions, and other controls of the database system.

List of  DCL commands:

[**GRANT:**](https://www.geeksforgeeks.org/mysql-grant-revoke-privileges/)This commandgives users access privileges to the database.

**Syntax:**

*GRANT SELECT, UPDATE ON MY\_TABLE TO SOME\_USER, ANOTHER\_USER;*

[**REVOKE:**](https://www.geeksforgeeks.org/difference-between-grant-and-revoke/)This command withdraws the user’s access privileges given by using the GRANT command.

**Syntax:**

*REVOKE SELECT, UPDATE ON MY\_TABLE FROM USER1, USER2;*

Data ware Housing concept and tools (ETL tools),

1. ETL stands for Extract, Transform, Load and it is a process used in data warehousing to extract data from various sources, transform it into a format suitable for loading into a data warehouse, and then load it into the warehouse. The process of ETL can be broken down into the following three stages:
2. **Extract**: The first stage in the ETL process is to extract data from various sources such as transactional systems, spreadsheets, and flat files. This step involves reading data from the source systems and storing it in a staging area.
3. **Transform**: In this stage, the extracted data is transformed into a format that is suitable for loading into the data warehouse. This may involve cleaning and validating the data, converting data types, combining data from multiple sources, and creating new data fields.
4. **Load**: After the data is transformed, it is loaded into the data warehouse. This step involves creating the physical data structures and loading the data into the warehouse.

UNIT:-2