DBMS – 4

Functions of a DBMS

1. Data storage, retrieval, and update
2. A user-accessible catalog

* A key feature of the ANSI-SPARC architecture is the recognition of an integrated system catalog to hold data about the schemas, users, applications, and so on.
* Accessible to users.
* **Metadata** is catalog/ data dictionary/ repository of information.
* **E.g of what system catalog stores:**
* names, types, and sizes of data items;
* names of relationships;
* integrity constraints on the data;
* names of authorized users who have access to the data;
* the data items that each user can access and the types of access allowed; for example, insert, update, delete, or read access;
* external, conceptual, and internal schemas and the mappings between the schemas
* usage statistics, such as the frequencies of transactions and counts on the number of accesses made to objects in the database.

**Advantages of system catalog :-**

1. Information about data can be collected and stored centrally. This helps to maintain control over the data as a resource.
2. The meaning of data can be defined, which will help other users understand the purpose of the data.
3. Communication is simplified, since exact meanings are stored. The system catalog may also identify the user or users who own or access the data.
4. Redundancy and inconsistencies can be identified more easily since the data is centralized.
5. Changes to the database can be recorded.
6. The impact of a change can be determined before it is implemented, since the system catalog records each data item, all its relationships, and all its users.
7. Security can be enforced.
8. Integrity can be ensured.
9. Audit information can be provided.
10. Transaction support

* Ensure either that all the updates corresponding to a given transaction are made or that none of them is made.

1. Concurrency control services

* Ensures it’s updated correctly.
* Enable many users to share the same data concurrently without interferences or inconsistencies.

1. Recovery services
2. Authorization services
3. Support for data communication

* Must integrated with communication software.
* Users access the database from workstations: a) either directly connected to computer hosting DBMS or b) over a network.
* In either case, DBMS receives requests as communications messages and responds in a similar way.
* All such transmissions are handled by a Data Communication Manager (DCM).
* DCM is not part of DBMS.

1. Integrity services

* Ensures that both data in db and changes made follow certain rules.
* Database integrity refers to the correctness and consistency of stored data: it can be considered as another type of database protection.
* Connected to quality of data itself.
* Integrity is usually expressed in terms of constraints, which are consistency rules that the database is not permitted to violate.

1. Services to promote data independence

* Data independence is normally achieved through a view or subschema mechanism.
* Data independence is normally achieved through a view or subschema mechanism.
* Physical data independence is easier to achieve: there are usually several types of change that can be made to the physical characteristics of the database without affecting the views.

1. Utility services

Examples of utilities of the latter kind are:

* import facilities, to load the database from flat files, and export facilities, to unload the database to flat files;
* monitoring facilities, to monitor database usage and operation;
* statistical analysis programs, to examine performance or usage statistics;
* index reorganization facilities, to reorganize indexes and their overflows;
* garbage collection and reallocation, to remove deleted records physically from the storage devices, to consolidate the space released, and to reallocate it where it is needed.

Database Languages

1. The Data Definition Language (DDL)  
   - used to specify the database schema or modify an existing one.  
   - Is a language that allows the DBA or user to describe and name the entities, attributes, and relationships required for the application, together with any associated integrity and security constraints.  
   - Stored in system catalog.  
   - The system catalog integrates the metadata, that is data that describes objects in the database and makes it easier for those objects to be accessed or manipulated.  
   - The metadata contains definitions of records, data items, DBMS normally consults the system catalog before the actual data is accessed in the database.  
   - there is one comprehensive DDL that allows specification of at least the external and conceptual schemas.
2. The Data Manipulation Language (DML)  
   - used to both read and update the database.

Data manipulation operations usually include the following:

* insertion of new data into the database;
* modification of data stored in the database;
* retrieval of data contained in the database;
* deletion of data from the database.

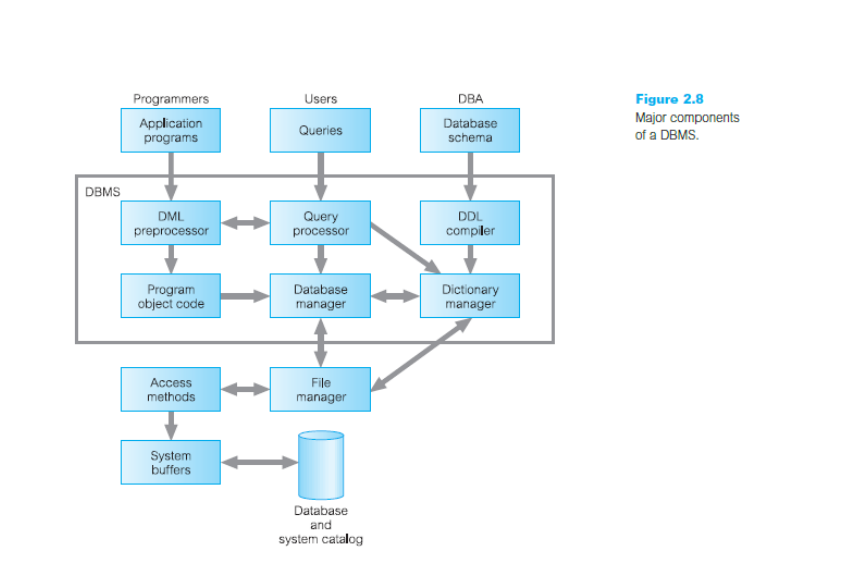
NOTE: DDL and DML are called sublanguages because they do not include constructs for all computing needs such as conditional or iterative statements which are provided by HLL.

1. Fourth-Generation Languages (4GLs)

* Presentation language: such as query languages and report generators;
* Specialty Language: spreadsheets and database languages;
* Application generators: s that define, insert, update, and retrieve data to build apps.
* HLL

Components of a DBMS

Note: DBMS built on top of OS.



* Query processor: transforms queries into a series of low-level instructions directed to the database manager
* Database manager (DM): The DM accepts queries from user- submitted application program queries & examines the external and conceptual schemas to determine what conceptual records are required to satisfy the request. The DM then places a call to the file manager to perform the request.
* File manager: manipulates the underlying storage files and manages the allocation of storage space on disk. It establishes and maintains the list of structures and indexes defined in the internal schema. If hashed files are used it calls on the hashing functions to generate record addresses.

Note: However, the file manager does not directly manage the physical input and output of data. Rather it passes the requests on to the appropriate access methods, which either read data from or write data into the system buffer (or cache).

* DML preprocessor: This module converts DML statements embedded in an application program into standard function calls in the host language. The DML preprocessor must interact with the query processor to generate the appropriate code.
* DDL compiler The DDL compiler converts DDL statements into a set of tables containing metadata. These tables are then stored in the system catalog while control information is stored in data file headers.
* Catalog manager: manages access to and maintains the system catalog. The system catalog is accessed by most DBMS components.

The major software components for the database manager are as follows:

* Authorization control
* Command processor
* Integrity checker
* Query optimizer
* Transaction manager
* Scheduler
* Recovery manager
* Buffer manager (Cache manager): transfer data between main memory and secondary memory.
* Data manager: Buffer + recovery manager.