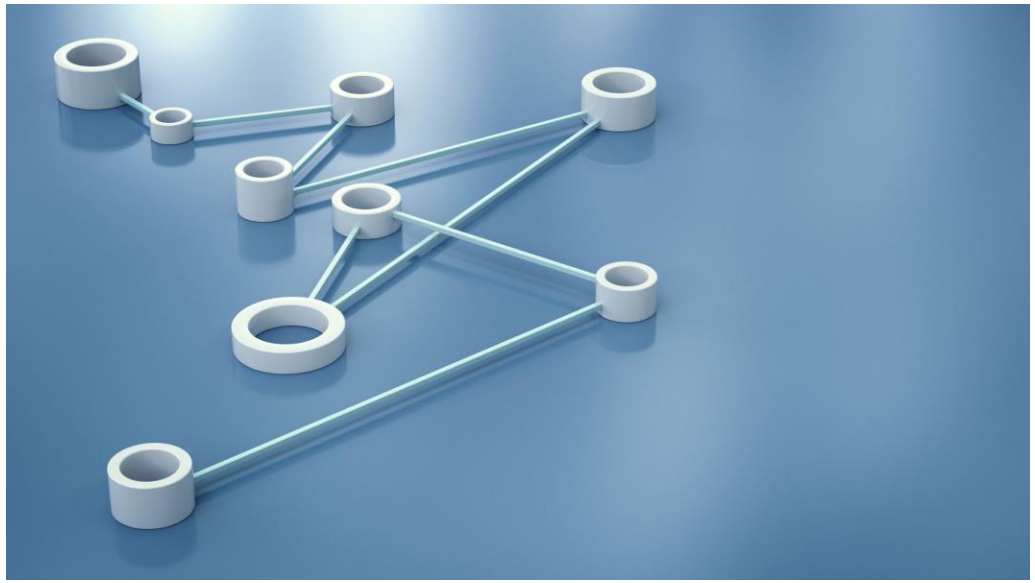
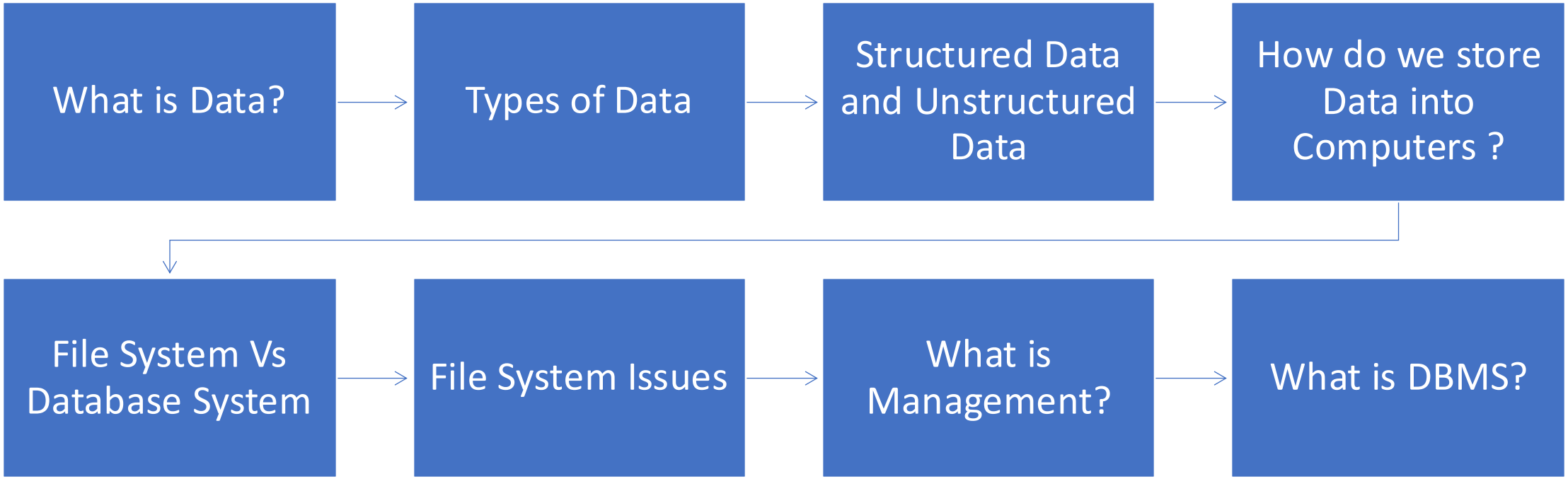


Introduction to Database Management System

By
Dr. Manish Sharma

Agenda


What is Data?
Types of Data
Structured Data and Unstructured Data
How do we store Data into Computers ?
File System Vs Database System
File System Issues
What is Management?
What is DBMS?

What is Data?

Dictionary definition :

Factual information (such as measurements or statistics) used as a basis for reasoning, discussion, or calculation.

Wikipedia definition :

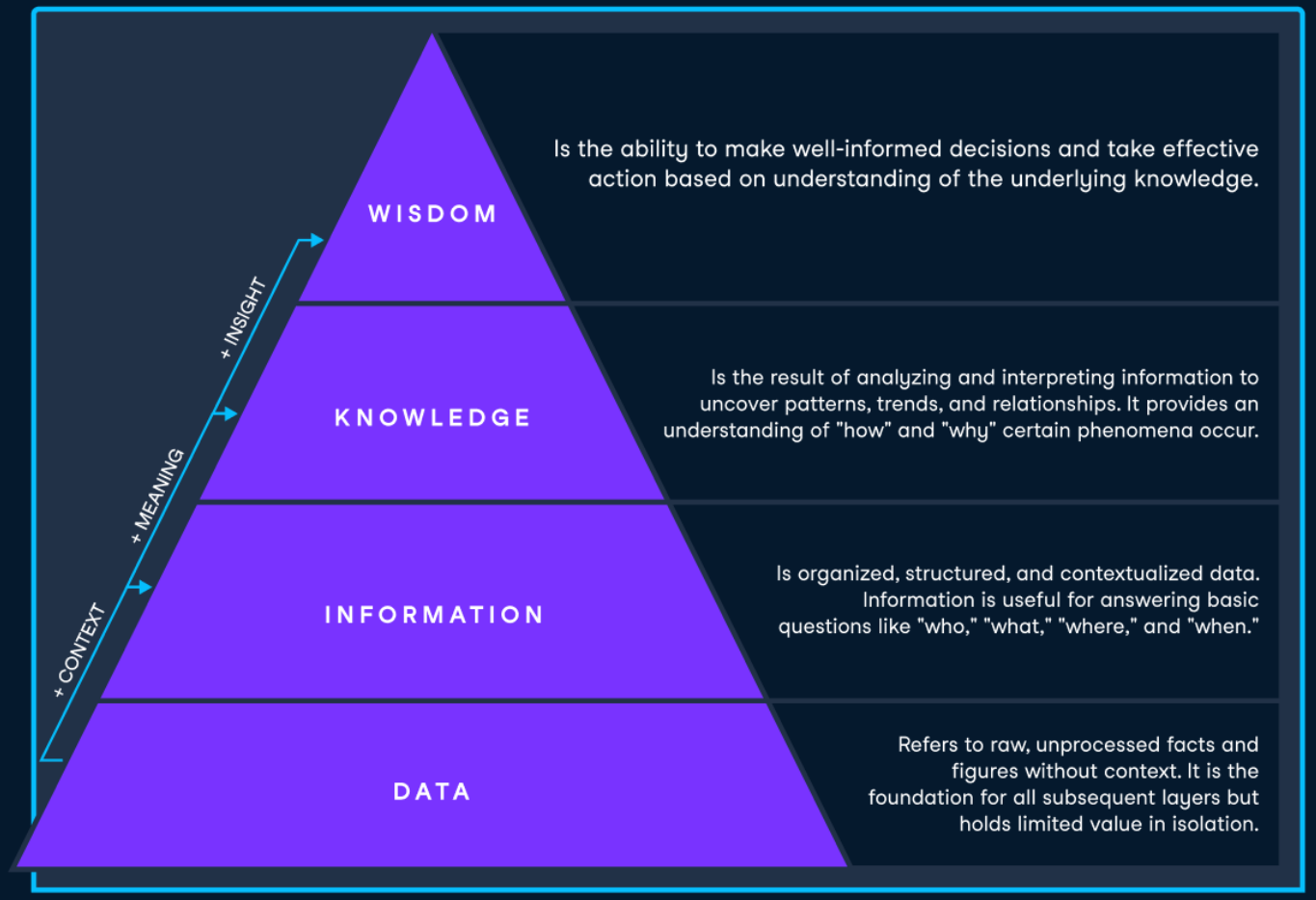
- **Data** are characteristics , usually numerical, that are collected through observation. In a more technical sense, data is a set of values of qualitative or quantitative variables about one or more persons or objects, while a **datum** (singular of data) is a single value of a single variable.

Example: Name, Address, Zip, SSN are characteristics or information about a person.

Data-Information-Knowledge-Wisdom Pyramid

datacamp

The Data-Information-Knowledge-Wisdom (DIKW) pyramid illustrates the progression of raw data to valuable insights. It gives you a framework to discuss the level of meaning and utility within data. Each level of the pyramid builds on lower levels, and to effectively make data-driven decisions, you need all four levels.



Types of Data

Unstructured-Data

Structured – Information (interpreted data – data supplied with semantics)

Structured and Unstructured Data

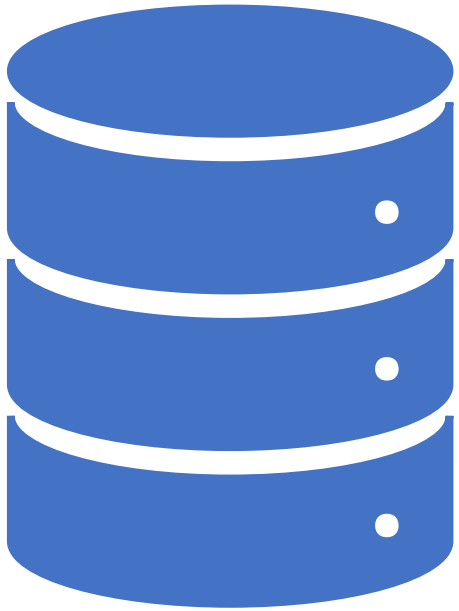
Structured data is most often categorized as quantitative data, and it's the type of data most of us are used to working with. Think of data that fits neatly within fixed fields and columns in **relational databases** and spreadsheets.

Examples of structured data include names, dates, addresses, credit card numbers, stock information, geolocation, and more.

Unstructured data is most often categorized as qualitative data, and it cannot be processed and analyzed using conventional tools and methods.

Examples of unstructured data include text, video, audio, mobile activity, social media activity, satellite imagery, surveillance imagery – the list goes on and on.

Unstructured data is difficult to deconstruct because it has no pre-defined model, meaning it cannot be organized in relational databases. Instead, non-relational, or **NoSQL databases**, are best fit for managing unstructured data.

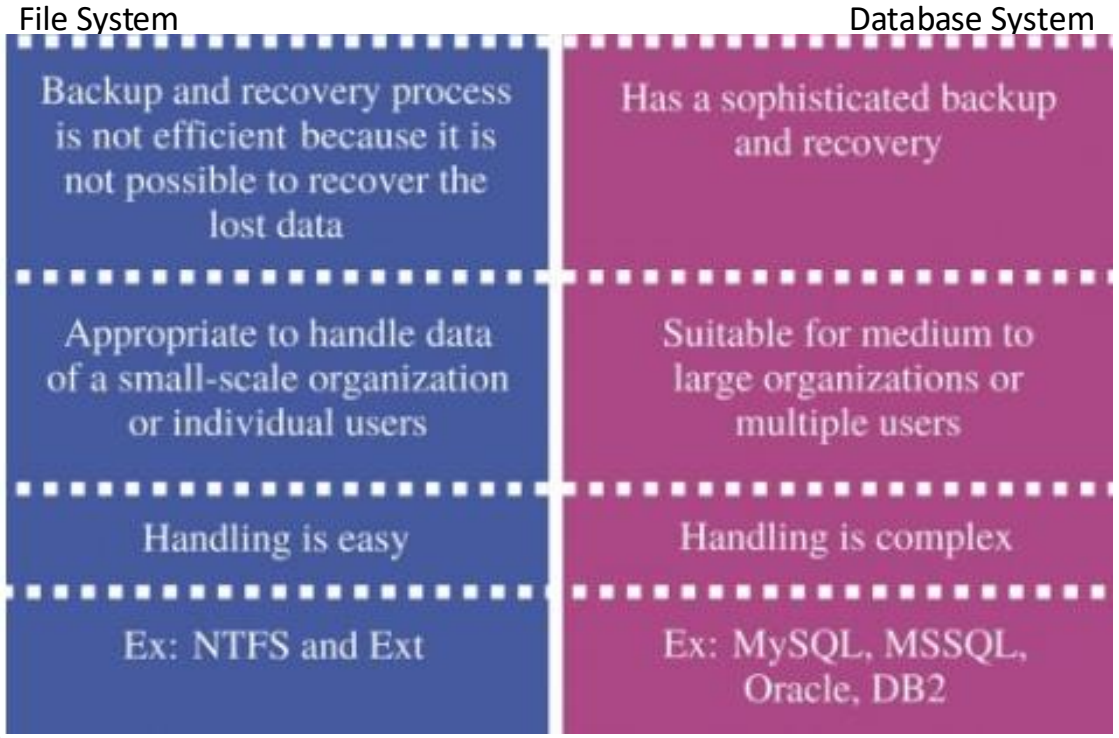


How do we store Data into Computers?.

- Data could be stored computers in File System and or Database Management and or content management systems.
- Let us focus on File System and Database approaches . Each one has its own advantages and disadvantages.

File System Vs Database System

File System Vs Database System Cont…



File System Issues

1

Redundancy of data: Data is said to be redundant if same data is copied at many places. If a student wants to change Phone number, he has to get it updated at various sections. Similarly, old records must be deleted from all sections representing that student.

2

Inconsistency of Data: Data is said to be inconsistent if multiple copies of same data does not match with each other. If Phone number is different in Accounts Section and Academics Section, it will be inconsistent. Inconsistency may be because of typing errors or not updating all copies of same data.

3

Difficult Data Access: A user should know the exact location of file to access data, so the process is very cumbersome and tedious. If user wants to search student hostel allotment number of a student from 10000 unsorted students' records, how difficult it can be.

4

Unauthorized Access: File System may lead to unauthorized access to data. If a student gets access to file having his marks, he can change it in unauthorized way.

5

No Concurrent Access: The access of same data by multiple users at same time is known as concurrency. File system does not allow concurrency as data can be accessed by only one user at a time.

6

No Backup and Recovery: File system does not incorporate any backup and recovery of data if a file is lost or corrupted.

What is Management?

- Generally Management refers create, Retrieve, update and delete.
- For example: Money Management refers earning the money, querying money, distribute the money and spend the money.

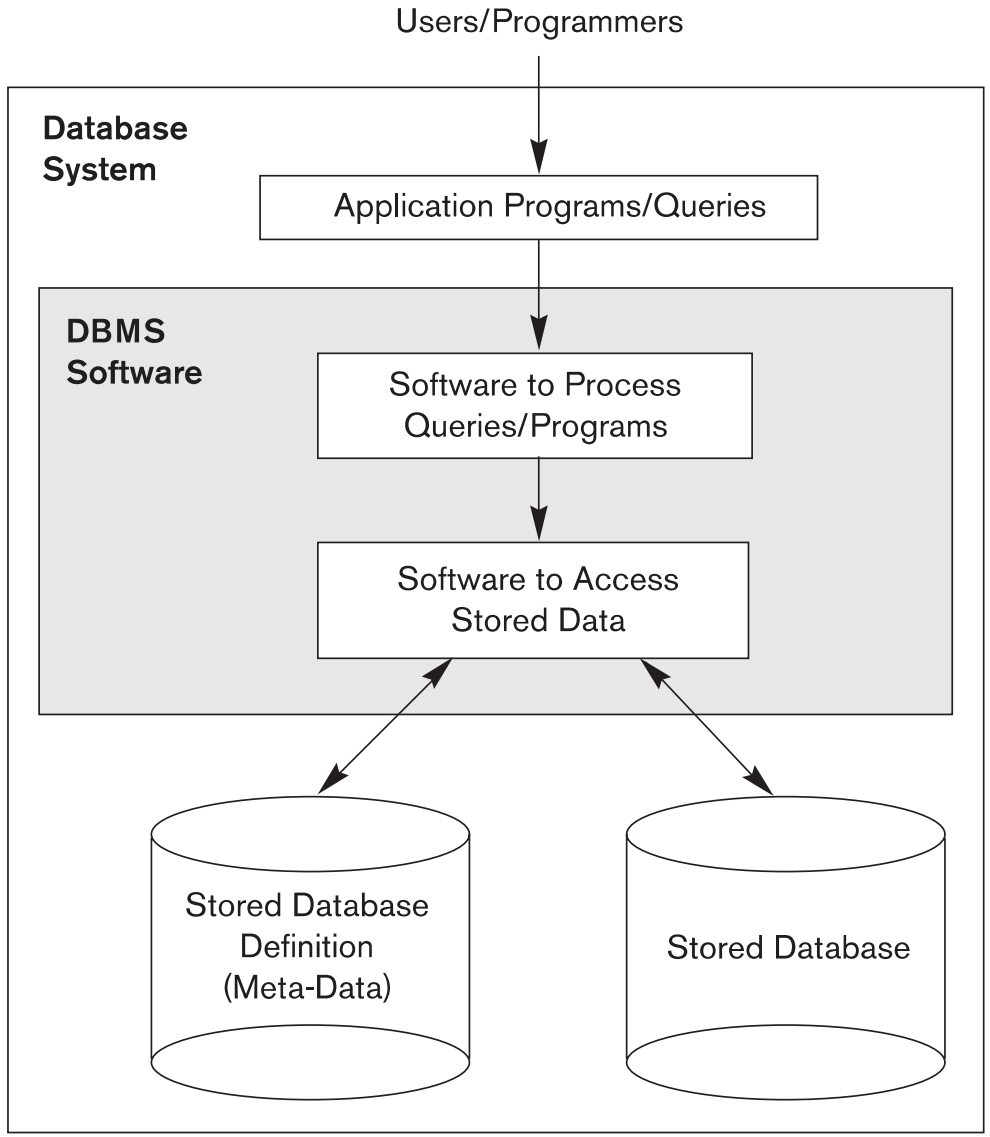
What is Database?

- **Database:** Database is a collection of inter-related data which helps in efficient retrieval, insertion and deletion of data from database and organizes the data in the form of tables, views, schemas, reports etc.
- For Example, university database organizes the data about students, faculty, and admin staff etc. which helps in efficient retrieval, insertion and deletion of data from it.
- Now Database Management talks about
  - Create Database
  - Retrieve(View) database
  - Update database
  - Delete Database

What is DBMS?

- **Database Management System:** The software which is used to manage database is called Database Management System (DBMS). For Example, MySQL, Oracle etc. are popular commercial DBMS used in different applications. DBMS allows users the following tasks:
- **Data Definition:** It helps in creation, modification and removal of definitions that define the organization of data in database.
- **Data Updation:** It helps in insertion, modification and deletion of the actual data in the database.
- **Data Retrieval:** It helps in retrieval of data from the database which can be used by applications for various purposes.
- **User Administration:** It helps in registering and monitoring users, enforcing data security, monitoring performance, maintaining data integrity, dealing with concurrency control and recovering information corrupted by unexpected failure.

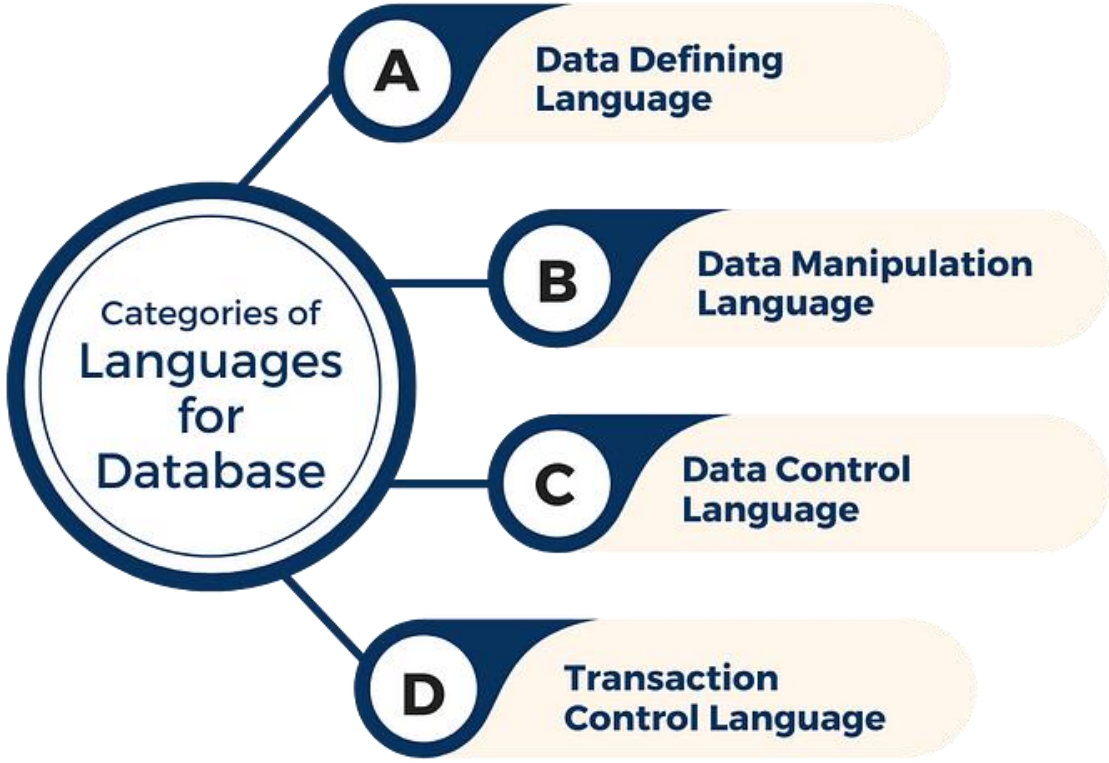
A simplified database system environment.

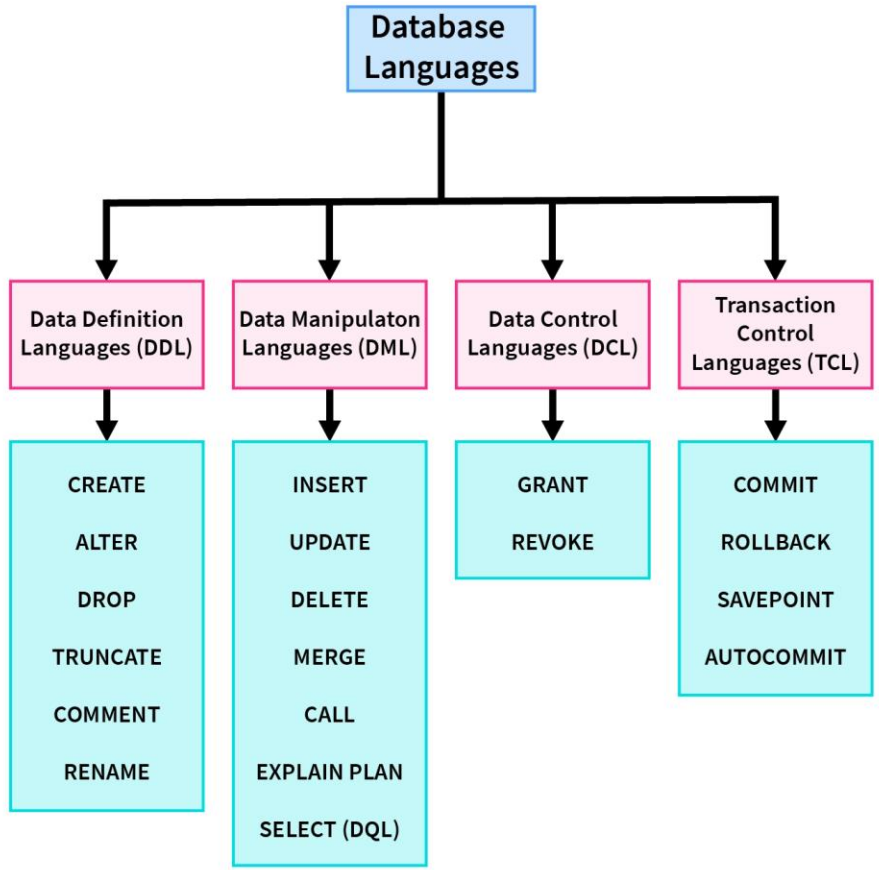


Database Languages

A database system provides a **data definition language** to specify the database schema and a **data manipulation language** to express database queries and updates.

In practice, the data definition and data manipulation languages are not two separate languages; instead they simply form parts of a single database language, such as the widely used SQL language


Categories of
Languages
for
Database
A
Data Defining
Language
B
Data Manipulation
Language
C
Data Control
Language
D
Transaction
Control Language



SCALER Topics

Data Definition Language(DDL)

We specify a database schema by a set of definitions expressed by a special language called a **data-definition language** (**DDL**).

For instance, the following statement in the SQL language defines the *account* table:

create table *account (account-number* ***char****(10), balance* ***integer****)*

Execution of the above DDL statement creates the *account* table. In addition, it updates a special set of tables called the **data dictionary** or **data directory**.

A data dictionary contains **metadata**—that is, data about data. The schema of a table is an example of metadata. A database system consults the data dictionary before reading or modifying actual data.

We specify the storage structure and access methods used by the database system by a set of statements in a special type of DDL called a **data storage and definition** language. These statements define the implementation details of the database schemas, which are usually hidden from the users.

The data values stored in the database must satisfy certain **consistency constraints**. For example, suppose the balance on an account should not fall below $100. The DDL provides facilities to specify such constraints. The database systems check these constraints every time the database is updated.

Data Manipulation Language(DML)

Data manipulation is

The retrieval of information stored in the database

The insertion of new information into the database

The deletion of information from the database

The modification of information stored in the database

A **data-manipulation language (DML)** is a language that enables users to access or manipulate data as organized by the appropriate data model. There are basically two types:

Procedural DMLs require a user to specify *what* data are needed and *how* to get those data.

Declarative DMLs (also referred to as **nonprocedural** DMLs) require a user to specify *what* data are needed *without* specifying how to get those data.

Declarative DMLs are usually easier to learn and use than are procedural DMLs. However, since a user does not have to specify how to get the data, the database system has to figure out an efficient means of accessing data. The DML component of the SQL language is nonprocedural.

DML Continuation…

A **query** is a statement requesting the retrieval of information. The portion of a DML that involves information retrieval is called a **query language**. Although technically incorrect, it is common practice to use the terms *query language* and *data manipulation language* synonymously.

This query in the SQL language finds the name of the customer whose customer-id is 192-83-7465:

```
select customer.customer-name
from customer
where customer.customer-id = 192-83-7465
```

The query specifies that those rows *from* the table *customer* *where* the *customer-id* is 192-83-7465 must be retrieved, and the *customer-name* attribute of these rows must be displayed.

Queries may involve information from more than one table. For instance, the following query finds the balance of all accounts owned by the customer with customerid 192-83-7465.

DML Continuation…

select *account.balance*

from *depositor*, *account*

where *depositor.customer-id* = 192-83-7465 **and**

depositor.account-number = *account.account-number*

There are a number of database query languages in use, either commercially or experimentally.

The levels of abstraction apply not only to defining or structuring data, but also to manipulating data. At the physical level, we must define algorithms that allow efficient access to data. At higher levels of abstraction, we emphasize ease of use. The goal is to allow humans to interact efficiently with the system. The query processor component of the database system translates DML queries into sequences of actions at the physical level of the database system.

Questions?