22XX405 - DATABASE MANAGEMENT SYSTEM

UNIT 1 & LP 1- UNDERSTANDING DATA AND INFORMATION, DATABASE VS INFORMATION

1. DATA vs INFORMATION:

DATA:

- Data is an individual unit that contains raw materials which do not carry any specific meaning.
- Data doesn't depend on information.

INFORMATION:

- Information is a group of data that collectively carries a logical meaning
- Information depends on data.

DATA	INFORMATION
Derived from Latin word 'Datum'	Derived from word 'informare'
Data is raw fact	Processed form of data
May or may not be meaningful	Always meaningful
Input to any system may be treated as data.	Output after processing system is information.
Understanding is difficult.	Understanding is easy.
Data may not be in order	Information should be in order.
Example: Survey data	Example: census report

2. TYPES OF DATA:

2.1 STRUCTURED DATA:

Structured data is preformatted clean data neatly arranged into memory blocks. The format of structured data is pre-defined in rows and columns and stored in relational database systems (RDBMS) or Microsoft Excel. The data is known as "schema on write, " representing data for a large database schema or blueprint. It is highly scalable and secure and acquires less management.

EXAMPLE OF STRUCTURED DATA

Structured data is used in multiple consumer-oriented business databases or ERPs, such as:

- E-commerce: Review data, pricing data, and SKU number of commodities
- Healthcare: hospital administration, pharmacy, and patient data and medical history of patients.
- Banking: Financial transaction details like name of beneficiary, account details, sender or receiver information and bank details.
- Customer relationship management (CRM) software: lead acquisition data, source, activity and so on of leads in the CRM database.
- Travel industry: Passenger data, flight information, and travel transactions.

BENEFITS OF STRUCTURED DATA

- It is easy to store, retrieve and manage structured data as it has an organized backend mechanism. Using structured data in business can result in the following benefits.
- Structured data can be easily fed into machine learning models as input datasets without any or trimming.
- It does not require any AI or ML expertise to work with structured data. A person with good product information and basic data science knowledge can work with it.
- Structured data is stored evenly in data warehouses or spreadsheets. The specific and organized nature makes it easy to manipulate and query it.
- Structured data predates unstructured data, so more analytics tools are available to measure and analyse it.
- The data is of higher quality, consistency, and usability than unstructured data.
- There are fallback mechanisms to adapt if user encounters and error while managing structured data.
- It is also known as quantitative data, as businesses use their metrics to forecast business trends and strategic impact.
- It is maintained in a stable, centralized repository that improves the flow of business processes and decision-making to optimize ROI.

STRUCTURED DATA TOOLS

Apart from using a structured query language(SQL) or Microsoft Excel to manage structured data manipulations, there are a few more tool extensions you can use.

- PL SQL: Procedural Query Language or PL SQL is an existing version of SQL that deals with **work** transactions. The common transactional queries are "commit" or "rollback."
- Postgre SQL: Postgre SQL is an open-source relational database management system that handles large data volumes. It also supports SQL and JSON querying along with high-level languages.
- SQLite: It is a high-tier, self-contained, and serverless database that software developers use to extract structured data for business app integrations,
- My SQL is a standard integrated data environment that uses user authentication to enter data records through queries in a mass-deployed database.
- OLAP: It encompasses a broader category of database management comprising data mining, report mining, and business intelligence.

2.2 UNSTRUCTURED DATA:

Unstructured data is highly complex, qualitative, and unorganized. It is also referred to as big data, which does not conform to any one particular standard. This data can be numerical, alphabetical, Boolean or a mix of all of them. It is stored using NoSQL database. It cannot be stored in relational database or RDBMS since data strings have mixed datatypes which cannot fit into either a row or a column. Common type of unstructured data is clickstream data, social media data, text and multimedia.

Unstructured data is often categorized as qualitative and cannot be processed and analysed using conventional data tools and methods. It is also known as "schema independent" or "schema on read" data. Examples of unstructured data include text, video files, audio files, mobile activity, social media posts, satellite imagery, surveillance imagery – the list goes on and on.

BENEFITS OF UNSTRUCTURED DATA

Unstructured data, also known as big data nowadays, is free-flowing and native to each specific company. It is schema independent and is known as "schema on read." Customizing this data per your business strategies can give you a competitive edge over competitors still stuck in traditional decision-making. And here is why.

- Unstructured data is easily available and has enough insights businesses can collect to learn about their product response.
- Unstructured data is schema-independent. Hence minor alterations to the database do not impact cost, time, or resources.
- Unstructured data can be stored on shared or hybrid cloud servers with minimal expenditure on database management.
- Unstructured data is in its native format, so data scientists or engineers do not define it until needed. It opens the expandability of file formats, as it is available in different formats like .mp3,

opus, .pdf, .png, and so on.

- Data lakes come with "pay-as-you-use" pricing, which helps businesses cut their costs and resource consumption.

2.3 SEMI STRUCTURED DATA:

- **semi-structured data** is a type of structured data that lies midway between structured and unstructured data. It doesn't have a specific relational or tabular data model but includes tags and semantic markers that scale data into records and fields in a dataset.
- Common examples of semi-structured data are JSON and XML. Semi-structured data is more complex than structured data but less complex than unstructured data. It's also relatively easier to store than unstructured data, bridging the gap between the two data types.
- An XML sitemap contains page information for a website. It embeds URLs, domain scores, do-follow pages, and meta tags.

3. PURPOSE OF DATABASE SYSTEMS:

The purpose of database systems is to make the database user-friendly and do easy operations. Users can easily insert, update, and delete. Actually, the main purpose is to have more control of the data.

The purpose of database systems is to manage the following insecurities:

- data redundancy and inconsistency,
- difficulty in accessing data,
- data isolation,
- atomicity of updates,
- concurrent access,
- security problems, and
- supports multiple views of data.

- **Avoid data redundancy and inconsistency:**
  If there are multiple copies of the same data, it just avoids it. It just maintains data in a single repository. Also, the purpose of database systems is to make the database consistent.
- **Difficulty in accessing data:**
  A database system can easily manage to access data. Through different queries, it can access data from the database.
- **Data isolation:**
  Data are isolated in several fields in the same database.
- **Atomicity of updates:**
  In case of power failure, the database might lose data. So, this feature will automatically prevent data loss.
- **Concurrent access:**
  Users can have multiple access to the database at the same time.
- **Security problems:**

Database systems will make the restricted access. So, the data will not be vulnerable.

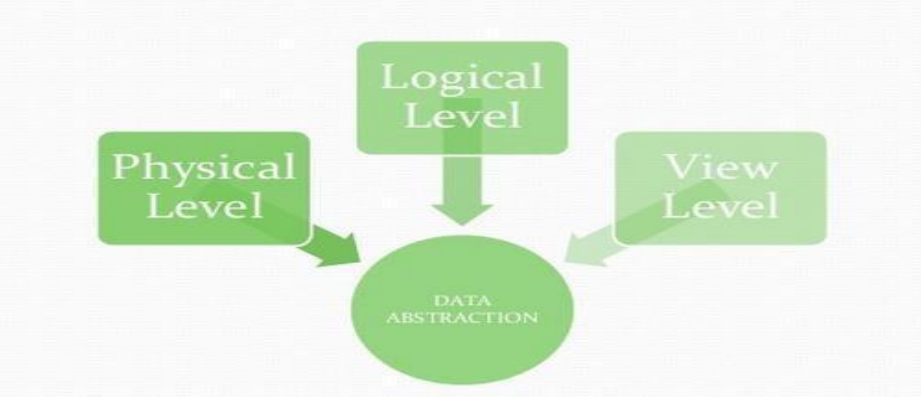
- **Supports multiple views of data:**
  It can support multiple views of data to give the required view as their needs. Only database admins can have a complete view of the database. We cannot allow the end-users to have a view of developers.

4. VIEW OF DATA IN DBMS

Abstraction is one of the main features of database systems. Hiding irrelevant details from user and providing abstract view of data to users, helps in easy and efficient **user-database** interaction. In the previous tutorial, we discussed the three level of DBMS architecture, The top level of that architecture is "view level". The view level provides the "**view of data**" to the users and hides the irrelevant details such as data relationship, database schema, constraints, security etc. from the user.

To fully understand the view of data, you must have a basic knowledge of data abstraction and instance & schema. Refer these two tutorials to learn them in detail.

1. abstraction: Database systems are made-up of complex data structures. To ease the user interaction with database, the developers hide internal irrelevant details from users. This process of hiding irrelevant details from user is called data abstraction.
2. Instance and schema: Design of a database is called the schema. Schema is of three types: Physical schema, logical schema and view schema. The data stored in database at a particular moment of time is called instance of database. Database schema defines the variable declarations in tables that belong to a particular database; the value of these variables at a moment of time is called the instance of that database.



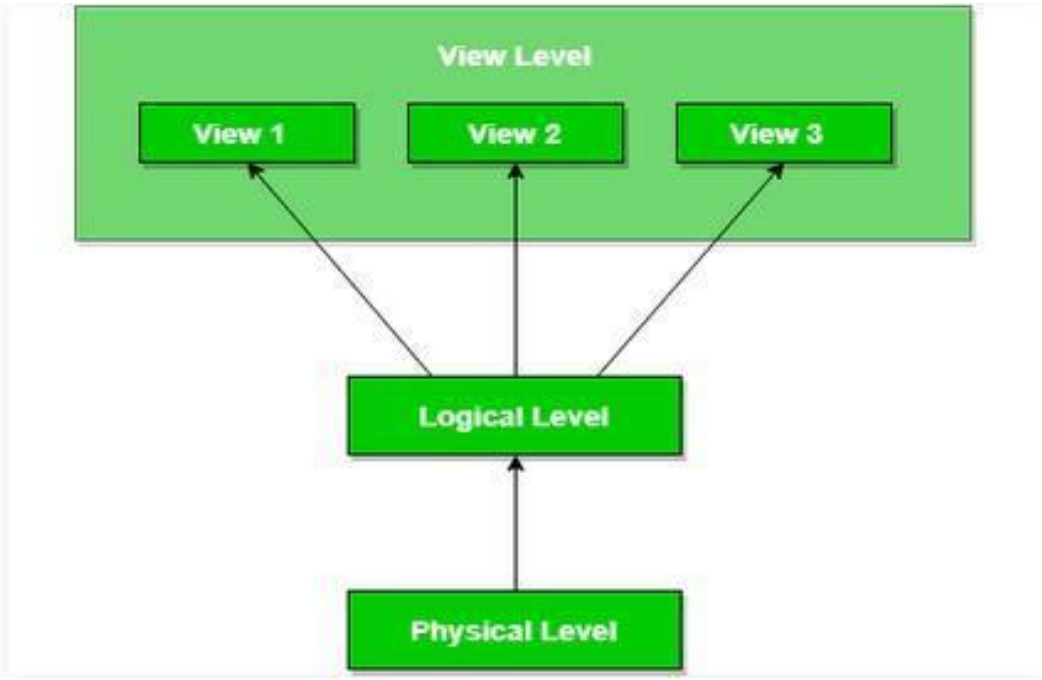
DATA ABSTRACTION IN DBMS

Database systems are made-up of complex data structures. To ease the user interaction with database, the developers hide internal irrelevant details from users. This process of hiding irrelevant details from user is called **data abstraction**. The term "irrelevant" used here with respect to the user, it doesn't mean that the hidden data is not relevant with regard to the whole database. It just means that the **user is not concerned about that data**.

EXAMPLE:

When you are booking a train ticket, you are not concerned how data is processing at the back end when you click “book ticket”, what processes are happening when you are doing online payments. **You are just concerned about the message that pops up when your ticket is successfully booked**. This doesn’t mean that the process happening at the back end is not relevant, it just means that you as a user are not concerned what is happening in the database.

THREE LEVELS OF ABSTRACTION



- **Physical level**: This is the lowest level of data abstraction. It describes how data is actually stored in database. You can get the complex data structure details at this level.

- **Logical level**: This is the middle level of 3-level data abstraction architecture. It describes what data is stored in database.

- **View level**: Highest level of data abstraction. This level describes the user interaction with database system.

Example:

Let’s say we are storing customer information in a customer table.

- At **physical level** these records can be described as blocks of storage (bytes, gigabytes, terabytes etc.) in memory. These details are often hidden from the programmers.

- At the **logical level** these records can be described as fields and attributes along with their data types, their relationship among each other can be logically implemented. The programmers generally work at this level because they are aware of such things about database systems.

- At **view level**, user just interact with system with the help of GUI and enter the details at the screen, they are not aware of how the data is stored and what data is stored; such details are hidden from them.

DBMS SCHEMA

Definition of schema: Design of a database is called the schema. For example: An **employee** table in database exists with the following attributes:

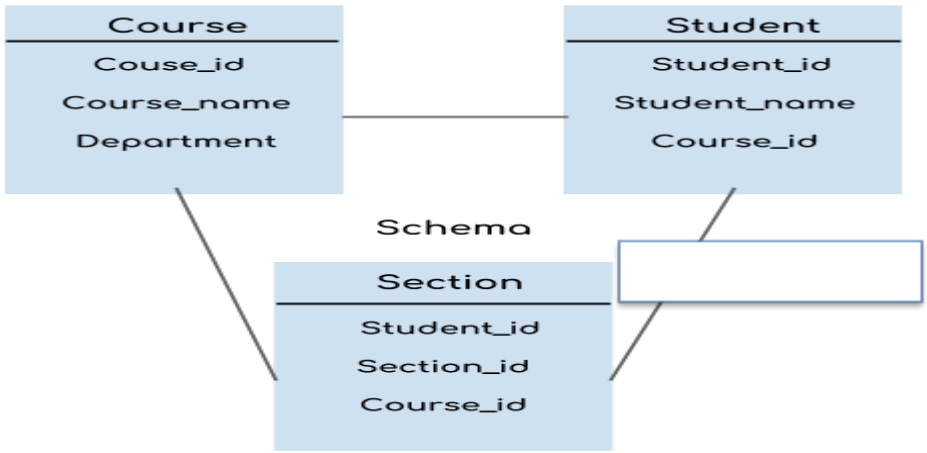
```
EMP_NAME   EMP_ID    EMP_ADDRESS   EMP_CONTACT

--------   ------    -----------   -----------
```

This is the schema of the **employee** table. Schema defines the attributes of tables in the database. **Schema is of three types: Physical schema, logical schema and view schema.**

- Schema represents the **logical view** of the database. It helps you understand what data needs to go where.
- Schema can be represented by a diagram as shown below.
- Schema **helps the database users to understand the relationship between data**. This helps in efficiently performing operations on database such as insert, update, delete, search etc.

In the following diagram, we have a schema that shows the relationship between **three tables: Course, Student and Section**. The diagram only shows the design of the database; it doesn't show the data present in those tables. Schema is only a structural view(design) of a database as shown in the diagram below



- The design of a database at physical level is called **physical schema**, how the data stored in blocks of storage is described at this level.

- Design of database at logical level is called **logical schema**, programmers and database administrators work at this level, at this level data can be described as certain types of data records

gets stored in data structures, however the internal details such as implementation of data structure is hidden at this level (available at physical level).

- Design of database at view level is called **view schema**. This generally describes end user interaction with database systems.

DBMS INSTANCE

Definition of instance: The data stored in database at a particular moment of time is called instance of database. Database schema defines the attributes in tables that belong to a particular database. The value of these attributes at a moment of time is called the instance of that database.

EXAMPLE 1:

we have seen the schema of table "employee" above. Let's see the table with the data now. At this moment the table contains two rows (records). This is the current instance of the table "employee" because this is the data that is stored in this table at this particular moment of time.

```
EMP_NAME    EMP_ID   EMP_ADDRESS   EMP_CONTACT
-------     ------   -----------   -----------
Chaitanya   101      Noida         95********

Ajeet       102      Delhi         99********
```

5. APPLICATIONS OF DATABASE:

There are different fields where a database management system is utilized. Following are a few applications that utilize the information base administration framework.

Applications of DBMS

1. <u>Railway Reservation System</u>

In the rail route reservation framework, the information base is needed to store the record or information of ticket appointments, status of train's appearance, and flight. Additionally, if trains get late, individuals become acquainted with it through the information base update.

2. <u>Library Management System</u>

There are many books in the library so; it is difficult to store the record of the relative multitude of books in a register or duplicate. Along these lines, the data set administration framework (DBMS) is utilized to keep up all the data identified with the name of the book, issue date, accessibility of the book, and its writer.

3. <u>Banking</u>

Database the executive's framework is utilized to store the exchange data of the client in the information base.

4. <u>Education Sector</u>

Presently, assessments are led online by numerous schools and colleges. They deal with all assessment information through the data set administration framework (DBMS). In spite of that understudy's enlistments subtleties, grades, courses, expense, participation, results, and so forth all the data is put away in the information base.

5. <u>Credit card exchanges</u>

The database Management framework is utilized for buying on charge cards and age of month to month proclamations.

6. <u>Social Media Sites</u>

We all utilization of online media sites to associate with companions and to impart our perspectives to the world. Every day, many people group pursue these online media accounts like Pinterest, Facebook, Twitter, and Google in addition to. By the utilization of the data set administration framework, all the data of clients are put away in the information base and, we become ready to interface with others.

7. Broadcast communications

Without DBMS any media transmission organization can't think. The Database the executive's framework is fundamental for these organizations to store the call subtleties and month to month post-paid bills in the information base.

8. Accounting and Finance

The information base administration framework is utilized for putting away data about deals, holding and acquisition of monetary instruments, for example, stocks and bonds in a data set.

9. E-Commerce Websites

These days, web-based shopping has become a major pattern. Nobody needs to visit the shop and burn through their time. Everybody needs to shop through web based shopping sites, (for example, Amazon, Flipkart, Snapdeal) from home. So all the items are sold and added uniquely with the assistance of the information base administration framework (DBMS). Receipt charges, instalments, buy data these are finished with the assistance of DBMS.

10. Human Resource Management

Big firms or organizations have numerous specialists or representatives working under them. They store data about worker's compensation, assessment, and work with the assistance of an information base administration framework (DBMS).

11. Manufacturing

Manufacturing organizations make various kinds of items and deal them consistently. To keep the data about their items like bills, acquisition of the item, amount, inventory network the executives, information base administration framework (DBMS) is utilized.

12. Airline Reservation System

This framework is equivalent to the railroad reservation framework. This framework additionally utilizes an information base administration framework to store the records of flight take off, appearance, and defer status.

13. Healthcare System

DBMS is used in healthcare to manage patient data, medical records, and billing information.

14. Security

DBMS provides security features to ensure that only authorized users have access to the data.

15. Telecommunication

Database Management Systems (DBMS) are essential to the telecommunications industry because they manage enormous volumes of data on billing, customer information, and network optimization.

6. DATABASE vs DBMS

S. No.	Category	Database	DBMS
1	**Definition**	A database is a collection of connected information about people, locations, or things.	A database management system (DBMS) is a collection of programs that allow you to create, manage, and operate a database.
2.	**Storage**	Besides computers, databases can even be maintained in physical ledgers, books, or papers.	In a database management system (DBMS), all the records are maintained only on a computer.
3.	**Data Retrieval**	The retrieval of information from the databases can be done manually, through queries, or by	We can retrieve the data from the database management system through queries written in SQL.

		using programs (C, C++, Java, etc.).	
4.	**Speed**	As databases can be handled manually or via computers, when SQL is not used to retrieve information, it can be very slow.	As a computer system is involved in a database management system, the retrieval of information is very quick.
5.	**Access**	The databases are not designed for a large number of people who can access data at the same time, rather it is designed for a very small number of people (preferably few people) who access data at different times.	The database management system is designed for a large number of people who can access the data at the same time.
6.	**Data**	Data is stored in databases.	A database management system (DBMS) manages and manipulates data.
7.	**Data Manipulation**	In the case of the databases, very less information can be modified at a time.	In the database management system (DBMS), a lot of information can be changed at one time (as it can have many users using it at the same time).