Data Modelling Final Class Assessment

1. List some popular DBMS.

- MySQL
- PostgreSQL
- Microsoft SQL Server
- Oracle Database
- MongoDB

2. List types of keys in databases and explain any four.

- Primary Key: The primary key is the first key used to uniquely identify one instance of an entity in a database. It must be unique for each record in the table.
- Foreign Key: A foreign key is a column of a table used to point to the primary key of another table. It creates a relationship between two tables and maintains data integrity.
- Super Key: A super key is an attribute set that can uniquely identify a tuple in a database table. It is a superset of a candidate key.
- Composite Key: Whenever a primary key consists of more than one attribute, it is known as a composite key. This key is also known as a Concatenated Key.

Reference: Week 3 lecture material

3. What is a data repository, data warehouse, data lake, data marts, and data cubes?

- Data Repository: A data repository is a centralized location where data is stored, organized, and managed. It serves as a single source of truth for various types of data within an organization. Data repositories can include databases, file systems, or other storage systems.
- Data Warehouse: A data warehouse is a specialized type of data repository designed for storing and analyzing large volumes of structured data from multiple sources. It integrates data from different operational systems and formats it into a consistent schema for reporting and analysis. Data warehouses typically support business intelligence (BI) and analytics applications.
- Data Lake: A data lake is a storage repository that holds a vast amount of raw data in its native format until it is needed. Unlike a data warehouse, which requires structured data, a data lake can store structured, semi-structured, and unstructured data. It provides a flexible and scalable solution for storing data without the need for upfront structuring or schema definition.
- Data Mart: A data mart is a subset of a data warehouse that is focused on a specific business function or department within an organization. It contains a smaller, more tailored set of data that is optimized for the needs of a particular group of users. Data marts are often created to improve performance and accessibility for specific analytical purposes.

- Data Cube: A data cube is a multidimensional representation of data that allows for the analysis of data along multiple dimensions. It enables users to explore and analyze data from different perspectives by summarizing data into a structured format with dimensions such as time, geography, product, and customer. Data cubes are commonly used in online analytical processing (OLAP) systems for interactive analysis and reporting.

4. Explain different types of relationships in a DB.

- One-to-One (1:1) Relationship: In this type of relationship, one record in a table is associated with exactly one record in another table, and vice versa. For example, each person has exactly one passport, and each passport belongs to exactly one person.
- One-to-Many (1:N) Relationship: Here, one record in a table can be associated with multiple records in another table, but each record in the second table is associated with only one record in the first table. An example would be one department having many employees, but each employee belongs to only one department.
- Many-to-One (N:1) Relationship: This is the opposite of a one-to-many relationship. Many records in one table can be associated with a single record in another table. For instance, many employees work in one department, but each department has only one manager.
- Many-to-Many (N:M) Relationship: In this type of relationship, multiple records in one table can be associated with multiple records in another table, and vice versa. This relationship requires a junction table, also known as an associative or linking table, to connect the two tables. For example, students can enroll in multiple courses, and each course can have multiple students.

5. Explain different types of relationships in a DB.

- First Normal Form (1NF): This is the most basic level of normalization. In 1NF, each table cell should contain only a single value, and each column should have a unique name. 1NF helps to eliminate duplicate data and simplify queries.
- Second Normal Form (2NF): 2NF eliminates redundant data by requiring that each non-key attribute be dependent on the primary key. Each column should be directly related to the primary key, and not to other columns. Ensures data integrity by removing partial dependencies.
- Third Normal Form (3NF): 3NF builds on 2NF by requiring that all non-key attributes are independent of each other. Each column should be directly related to the primary key, and not to any other columns in the same table.
- Boyce-Codd Normal Form (BCNF): BCNF is a stricter form of 3NF that ensures that each determinant in a table is a candidate key. Each non-key attribute is dependent only on the candidate key. Prevents anomalies by eliminating all possible dependencies on non-candidate keys.
- Fourth Normal Form (4NF) and Fifth Normal Form (5NF): 4NF ensures that a table does not contain any multi-valued dependencies, while 5NF involves decomposing a table into smaller tables to remove data redundancy and improve data integrity. Further refine the database design to enhance data consistency and integrity.

6. Explain the two kinds of data independence.

- Physical Level Data Independence: Physical Data Independence can be defined as the ability to change the physical level without affecting the logical or Conceptual level. Physical data independence gives us the freedom to modify the - Storage device, File structure, location of the database, etc. without changing the definition of conceptual or view level.
- Logical Level Data Independence: Logical Data Independence is a property of a database that can be used to change the logic behind the logical level without affecting the other layers of the database. Logical data independence is usually required for changing the conceptual schema without having to change the external schema or application programs. It allows us to make changes in a conceptual structure like adding, modifying, or deleting an attribute in the database.

7. Given a Flight reservation system, identify 4 key Entities and their 3 key attributes.

Table 1: Flight

Flight		
FlightID	DepartureAirport	ArrivalAirport

Table 2: Passenger

Passenger		
PassengerID	Name	PhoneNumber

Table 3: Reservation

Reservation		
ReservationID	FlightID	PassengerID

Table 4: Seat

Seat		
SeatID	Class	Availability

8. List the 2 relationships between the entities identified

- Relationship: One-to-Many
  - Table 1: Flight ---(FlightID)--- Table 3: Reservation
  - Table 2: Passenger---(PassengerID)--- Table 3: Reservation

9. With examples in each case, explain the ACID Properties of a Database (8 Marks).

- Atomicity: Atomicity ensures that each transaction related to flight reservations happens all at once or not at all.

For example, when a passenger books a flight, the system ensures that either the entire booking process completes successfully, including deducting the fare from the passenger's account and reserving the seat, or none of it happens. If any part fails (like if the payment fails), the entire transaction is rolled back to maintain consistency.

- Consistency: Consistency ensures that the flight reservation system remains in a valid state before and after each transaction.

  For instance, if a passenger books a seat on a flight, the system ensures that the seat availability is updated accurately and immediately reflects the booked seat. This prevents situations where two passengers book the same seat simultaneously or where the number of available seats becomes negative.

- Isolation: Isolation ensures that multiple transactions can occur simultaneously without interfering with each other.

  In a flight reservation system, if two passengers are booking seats on the same flight at the same time, isolation ensures that each transaction is processed independently. This means that one passenger's booking process does not impact or affect the other passenger's booking process, even if they are happening concurrently.

- Durability: Durability guarantees that once a flight reservation transaction is completed and confirmed (like when a passenger successfully books a seat), the changes made to the database persist even in the face of system failures.

  For example, if the flight reservation system crashes after a passenger books a seat, durability ensures that the booked seat information remains intact and is not lost. When the system restarts, the booking information is still available and can be retrieved, ensuring that the booking process is not lost or repeated.

10. What is the difference between phishing, ransomware attacks, denial of service attack, and malware?

- Phishing: Phishing is a type of cyber attack where attackers attempt to trick individuals into divulging sensitive information such as usernames, passwords, credit card numbers, or other personal data. This is typically done through deceptive emails, messages, or websites that impersonate legitimate organizations or individuals.
- Ransomware Attacks: Ransomware attacks involve malicious software (malware) that encrypts files on a victim's computer or network, rendering them inaccessible. The attackers then demand a ransom payment, usually in cryptocurrency, in exchange for providing the decryption key to unlock the files.
- Denial-of-Service (DoS) Attack: A denial-of-service (DoS) attack aims to disrupt or disable the services of a network, website, or server by overwhelming it with a flood of illegitimate traffic or requests. The goal is to make the targeted system unavailable to its intended users.
- Malware: Malware is a broad term that refers to malicious software designed to disrupt, damage, or gain unauthorized access to computer systems or data. Malware includes various types such as viruses, worms, Trojans, spyware, and ransomware.

11. What is the difference between data privacy and data security?

- Data Privacy: Data privacy refers to the appropriate handling and protection of personal information, ensuring that individuals have control over how their data is collected, used, and shared. It involves ensuring that sensitive data is only accessed by authorized individuals and used for its intended purpose.
  Example: An organization implementing data privacy measures may obtain explicit consent from individuals before collecting their personal information, limit access to sensitive data to only those who need it for their job roles, and comply with regulations such as GDPR or CCPA that govern the processing of personal data.
- Data Security: Data security focuses on safeguarding data from unauthorized access, use, disclosure, modification, or destruction. It encompasses technical, administrative, and physical controls and measures implemented to protect data against cybersecurity threats and vulnerabilities.
  Example: Data security measures include encrypting sensitive data both in transit and at rest, implementing firewalls and intrusion detection systems to prevent unauthorized access to networks, enforcing strong authentication mechanisms like multi-factor authentication, regularly updating software and systems to patch known vulnerabilities, and conducting security audits and assessments to identify and mitigate risks.

Referenced:

1. Data Independence in DBMS: unstop.com
2. Types of Cyberattacks: economictimes.indiatimes.com
3. Lecture material: Week 1, 2, 3 ,4, 6