Unit 5

Database Management System

https://github.com/sanjeevlcc/notes_2081/tree/main

Er Sanjeev Thapa. BE CE, MTech CSE, MBS. DevOps Eng, CKA, RHCSA, RHCE, RHCSA-Openstack, MTCNA, MTCTCE, USRS, HE IPv6. https://github.com/sanjeevlcc   2024/2081

5.1. Introduction to Databases

What is a Database?

- A **database** is an organized collection of data stored electronically, designed for efficient access, management, and retrieval.

- Data in databases can be stored in various formats such as tables, documents, graphs, or files.

Basic LAB test for UNDERSTANDING CRUD STUDYING DBMS, UN:root PW:passwd DB:mydb

Employees Details

Add New Employee

#	Name	Address	Salary	Action
1	Ram Thapa	Butwal 11	36000	👁 ✏ 🗑
2	Sita Gurung	Kalikanagar	44000	👁 ✏ 🗑

Examples of Databases in Use

- **Banking Systems**: Managing accounts, transactions, and customer data.

- **E-commerce Platforms**: Storing product catalogs, customer orders, and payment records.

- **Healthcare**: Keeping patient records, medical histories, and appointments.

- **Social Media**: Storing user profiles, posts, and interactions.

```
mysql> select * from employees;
+----+-------------+-------------+--------+
| id | name        | address     | salary |
+----+-------------+-------------+--------+
|  1 | Ram Thapa   | Butwal 11   |  36000 |
|  2 | Sita Gurung | Kalikanagar |  44000 |
+----+-------------+-------------+--------+
```

Purpose of Databases

- To organize data for easy access and management.

- To reduce redundancy and maintain consistency.

- To facilitate multi-user access and support concurrent operations.
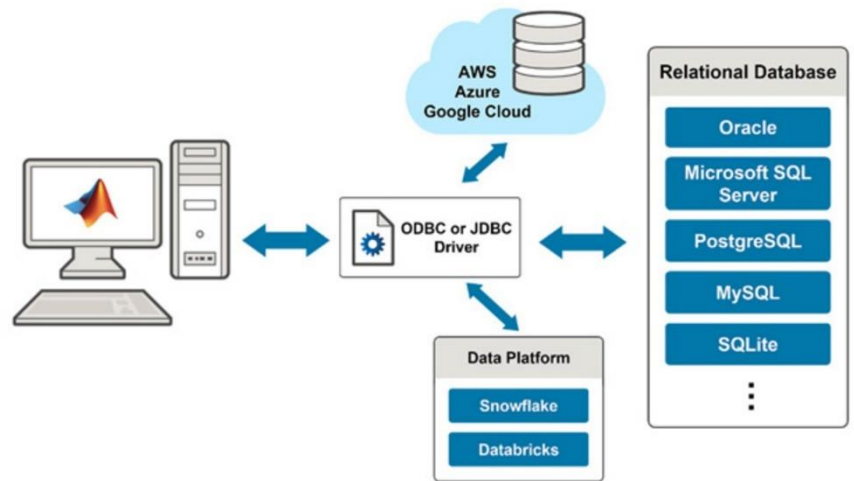
Key Characteristics of Databases

1. **Data Organization**:

    o Structured in a logical way, often using rows, columns, and keys.

2. **Efficient Retrieval**:

    o Optimized for quick data searches and queries.

3. **Data Integrity**:

o Ensures accuracy and consistency of data across the system.

4. **Security**:

o Protects sensitive data from unauthorized access.

Advantages of Using Databases

- **Centralized Data Storage**:

  o Avoids duplication and enables efficient data sharing.

- **Data Consistency**:

  o Keeps data accurate across all users and applications.

- **Scalability**:

  o Can handle growing data needs over time.

- **Data Analysis**:

  o Facilitates generating insights and reports for decision-making.

Components of a Database System

- **Data**: The actual information stored in the database.

- **Hardware**: Physical devices used to store and process data (e.g., servers, hard drives).

- **Software**: Database Management System (DBMS) for interacting with the database.

- **Users**:

  o **End Users**: Interact with the database through applications.

  o **Administrators**: Manage the database and ensure its smooth operation.

5.2. Importance of Data in Business Decision Making

What is Data in Business?

- Data in a business context refers to raw facts and figures that, when processed, provide meaningful information.

- Examples include customer demographics, sales figures, market trends, and operational metrics.

Why is Data Important in Decision Making?

1. **Informed Decisions**:
   - Data-driven insights reduce reliance on intuition and guesswork.
   - Example: Using sales data to determine which products to stock during peak seasons.

2. **Identifying Trends and Patterns**:
   - Helps businesses recognize market trends and customer behavior.
   - Example: Analyzing website traffic to understand peak shopping times.

3. **Measuring Performance**:
   - Data allows businesses to track Key Performance Indicators (KPIs) to assess success.
   - Example: Monitoring monthly revenue to evaluate business growth.

4. **Enhancing Customer Experience**:
   - Personalization is possible through data analysis, improving customer satisfaction.
   - Example: Recommending products based on a customer's purchase history.

5. **Risk Management**:
   - Data enables businesses to predict potential risks and take proactive measures.
   - Example: Using financial data to assess credit risk before approving a loan.

6. **Optimizing Operations**:
   - Data streamlines operations by identifying inefficiencies.
   - Example: Analyzing production data to reduce waste and improve quality.

Applications of Data in Business Decision Making

- **Marketing**:
  - Segmenting customers based on purchasing behavior.
  - Example: Running targeted ad campaigns for different age groups.

- **Finance**:
  - Budgeting and forecasting using historical financial data.
  - Example: Allocating resources for maximum ROI.

- **Human Resources**:
  - Using employee performance data for appraisals and promotions.
  - Example: Identifying skill gaps for training initiatives.

Challenges in Using Data for Decisions

1. **Data Overload**:
   o Too much data can be overwhelming without proper tools.

2. **Data Accuracy**:
   o Poor-quality data leads to inaccurate conclusions.

3. **Data Security**:
   o Safeguarding sensitive data is critical to prevent breaches.

5.3. Types of Data: Structured vs. Unstructured

Aspect	Structured Data	Unstructured Data
Customer Data	- Customer ID: 101	- Customer reviews on a product page
	- Name: John Doe	- Social media comments about a brand
	- Email: john.doe@example.com	- Profile pictures uploaded by users
Sales Data	- Transaction ID: 56789	- Receipt image uploaded for expense claims
	- Date: 2024-12-01	- Sales discussion audio recordings
	- Amount: $120.50	- Video of product demonstrations
Product Data	- Product ID: P123	- Product manual PDF
	- Name: Wireless Headphones	- Promotional videos
	- Price: $150.00	- Customer-generated unboxing videos
Employee Records	- Employee ID: E001	- Performance feedback emails
	- Name: Jane Smith	- Informal chats on workplace messaging apps
	- Department: IT	- Audio recordings of team meetings

➢ Data can be broadly classified into **structured** and **unstructured** categories based on how it is organized and stored.
➢ Understanding the differences is crucial for businesses to manage and utilize their data effectively.

Structured Data

- **Definition**:
    - Data that is organized in a predefined format, such as rows and columns, making it easily searchable and analyzable.

- **Characteristics**:
    - Stored in databases like relational database management systems (RDBMS).
    - Uses schema to define relationships between data entities.
    - Data can be queried using languages like SQL.

- **Examples**:
    - Spreadsheets with rows and columns.
    - Customer databases storing names, emails, and phone numbers.
    - Sales records with timestamps and amounts.

Advantages:

- Easy to store, search, and analyze.
- Suitable for tasks requiring real-time data processing (e.g., banking systems).

Challenges:

- Limited flexibility; requires data to fit predefined formats.

Unstructured Data

- **Definition**:
    - Data that lacks a predefined structure and is stored in its native format, making it harder to organize and analyze.

- **Characteristics**:
    - Often stored in data lakes or NoSQL databases.
    - Requires advanced tools (e.g., AI, machine learning) for processing and analysis.

- **Examples**:

  o Emails and instant messages.

  o Social media posts, comments, and likes.

  o Videos, audio recordings, and images.

  o Sensor data from IoT devices.

Advantages:

- Captures a richer and more diverse range of information.

- Useful for tasks like sentiment analysis or multimedia processing.

Challenges:

- Difficult to process and analyze using traditional methods.

- Requires larger storage capacities and specialized tools.

Key Differences

Feature	Structured Data	Unstructured Data
Format	Predefined (e.g., rows, columns)	Undefined
Storage	RDBMS (e.g., MySQL, Oracle)	NoSQL or data lakes
Processing	Simple (SQL queries)	Complex (AI/ML tools)
Examples	Customer databases, financial data	Social media, videos, images
Scalability	Limited	Highly scalable

Why Understanding Data Types Matters

- Helps in choosing appropriate storage solutions (e.g., databases or data lakes).

- Guides the selection of tools and technologies for processing and analysis.

- Enables businesses to derive maximum value from both types of data.

5.4. Database Management Systems (DBMS)

What is a DBMS?

- A **Database Management System (DBMS)** is software that enables users to create, manage, and interact with databases efficiently.
- It provides a systematic way of organizing, storing, retrieving, and manipulating data.

Functions of a DBMS

1. **Data Storage and Retrieval**:
   - Manages how data is stored on disk and retrieved efficiently.

2. **Data Manipulation**:
   - Allows users to insert, update, delete, and query data.

3. **Data Integrity**:
   - Ensures data accuracy and consistency across the database.

4. **Data Security**:
   - Controls access to data and prevents unauthorized access.

5. **Multi-User Access**:
   - Supports simultaneous access by multiple users.

6. **Backup and Recovery**:
   - Provides tools to back up data and recover it in case of failure.

Types of DBMS

1. **Relational DBMS (RDBMS)**:
   - Organizes data in tables with rows and columns.
   - Example: MySQL, Oracle Database, Microsoft SQL Server.

2. **NoSQL DBMS**:
   - Handles unstructured or semi-structured data.
   - Example: MongoDB, Cassandra.

3. **Hierarchical DBMS**:
   - Data is organized in a tree-like structure.
   - Example: IBM Information Management System (IMS).

4. **Network DBMS**:
   - Data is represented as records connected by links.
   - Example: Integrated Data Store (IDS).

Advantages of a DBMS

1. **Data Consistency**:
   o Avoids data redundancy and ensures consistency across multiple users.

2. **Ease of Data Sharing**:
   o Facilitates data sharing between users and applications.

3. **Improved Data Security**:
   o Enforces user authentication and access controls.

4. **Scalability**:
   o Handles increasing amounts of data effectively.

5. **Support for Transactions**:
   o Ensures that all database operations are completed successfully or rolled back if an error occurs.

Disadvantages of a DBMS

1. **Cost**:
   o High initial investment for licensing and setup.

2. **Complexity**:
   o Requires skilled personnel for administration and maintenance.

3. **Performance Overhead**:
   o Large-scale DBMS can be resource-intensive.

Real-World Applications of DBMS

1. **Banking**:
   o Manages account transactions and customer details.

2. **E-commerce**:
   o Stores product catalogs, customer orders, and payment information.

3. **Healthcare**:
   o Maintains patient records and medication inventories.

4. **Education**:
   o Tracks student performance and course enrollment.

5. **Social Media**:

- o Stores user profiles, posts, and interactions.

Popular DBMS Software

DBMS	Key Features
MySQL	Open-source, widely used for web applications.
Oracle DB	Enterprise-grade, supports complex transactions.
PostgreSQL	Advanced features, supports custom functions.
MongoDB	NoSQL, handles unstructured data.
Microsoft SQL Server	Integration with Microsoft products.

5.5. Relational Databases (Tables, Queries, Reports)

What is a Relational Database?

- A **relational database** is a type of database that organizes data into structured tables, where relationships between the data are defined.

- Each table represents a specific entity (e.g., Customers, Products) and contains rows (records) and columns (fields).

Key Components of Relational Databases

1. **Tables**:

   - o Tables store data in a grid-like structure with rows and columns.

   - o **Rows**: Represent individual records.

   - o **Columns**: Represent attributes or fields of the records.

   - o Example:

CustomerID	Name	Email
1	John Doe	john.doe@example.com
2	Jane Smith	jane.smith@example.com

2. **Primary Key**:

   - o A unique identifier for each record in a table.

   - o Example: **CustomerID** in the table above.

3. **Foreign Key**:

    o   A field in one table that links to the primary key in another table, establishing a relationship.

    o   Example: **OrderID** in an Orders table linking to **CustomerID** in a Customers table.

Customers Table

CustomerID	Name	Email	PhoneNumber	Address
1	Alice Johnson	alice@example.com	1234567890	123 Elm Street
2	Bob Smith	bob@example.com	0987654321	456 Oak Avenue

Orders Table

OrderID	OrderDate	Amount	CustomerID
101	2024-12-20	250.00	1
102	2024-12-21	150.50	2
103	2024-12-22	350.75	1

Querying Data:

To retrieve customer orders, use an INNER JOIN:

SELECT o.OrderID, o.OrderDate, o.Amount, c.Name AS CustomerName

FROM Orders o

JOIN Customers c ON o.CustomerID = c.CustomerID;

OrderID	OrderDate	Amount	CustomerName
101	2024-12-20	250.00	Alice Johnson
102	2024-12-21	150.50	Bob Smith
103	2024-12-22	350.75	Alice Johnson

This structure ensures proper data integrity and establishes a relationship between the two tables using a foreign key.

Queries

- Queries allow users to retrieve, update, or manipulate data in a relational database.

- **SQL (Structured Query Language)** is used to interact with relational databases.

Examples of SQL Queries:

1. Retrieve all customers:

    SELECT * FROM Customers;

2. Find customers from a specific city:

   SELECT Name, Email FROM Customers WHERE City = 'New York';

3. Update customer email:

   UPDATE Customers SET Email = 'new.email@example.com' WHERE CustomerID = 1;

Reports

- Reports present data in a structured and user-friendly format for analysis and decision-making.

- **Features of Reports**:

  - Aggregate data: Summarizing sales, profits, or other metrics.

  - Grouped views: Displaying data grouped by specific criteria (e.g., region, department).

  - Visualizations: Charts, graphs, and dashboards for easier interpretation.

- Example:

  - A sales report showing total revenue by region.

Advantages of Relational Databases

1. **Data Integrity**:

   - Ensures consistency and correctness using constraints (e.g., primary keys).

2. **Flexibility**:

   - Allows complex queries to extract meaningful information.

3. **Data Relationships**:

   - Supports structured relationships between entities (e.g., Customers and Orders).

4. **Ease of Use**:

   - SQL provides a simple and standardized way to interact with the data.

Application of Relational Databases

1. **E-commerce**:

   o Managing product inventories, customer information, and orders.

2. **Finance**:

   o Tracking transactions, accounts, and balances.

3. **Healthcare**:

   o Storing patient records and appointment schedules.

4. **Education**:

   o Maintaining student and course enrollment details.

5.6. Introduction to SQL (Structured Query Language)

What is SQL?

- **SQL (Structured Query Language)** is a programming language designed to manage and manipulate relational databases.

- It allows users to perform various operations such as retrieving, inserting, updating, and deleting data.

Basic SQL Operations

1. **Data Retrieval**:

   o Use the SELECT statement to query data.

   o Example:

     *SELECT * FROM Employees;*

2. **Data Insertion**:

   o Add new records using the INSERT statement.

   o Example:

     INSERT INTO Employees (ID, Name, Department, Salary)

     VALUES (1, 'John Doe', 'IT', 50000);

3. **Data Update**:

   o Modify existing records using the UPDATE statement.

o   Example:

UPDATE Employees SET Salary = 55000 WHERE ID = 1;

4. **Data Deletion**:

o   Remove records using the DELETE statement.

o   Example:

DELETE FROM Employees WHERE ID = 1;

Scenario: Employee Management Database

Table: Employees

ID	Name	Department	Position	Salary
1	John Doe	IT	Developer	50000
2	Jane Smith	HR	Manager	60000
3	Mike Johnson	Finance	Analyst	55000
4	Emily Davis	Marketing	Specialist	45000

Key Operations:

1. **Retrieve all employees in the IT department**:

SELECT Name, Position FROM Employees WHERE Department = 'IT';

Output:

Name	Position
John Doe	Developer

2. **Increase the salary of all employees in the HR department by 10%**:

UPDATE Employees

*SET Salary = Salary * 1.10*

WHERE Department = 'HR';

3. **Add a new employee**:

INSERT INTO Employees (ID, Name, Department, Position, Salary)

VALUES (5, 'Sarah Taylor', 'Finance', 'Consultant', 48000);

4. **Delete employees with a salary less than 45000**:

   DELETE FROM Employees WHERE Salary < 45000;

Why Use SQL?

1. **Ease of Use**:

   o   SQL is user-friendly with simple syntax for powerful operations.

2. **Standardized**:

   o   Supported by all major relational databases like MySQL, PostgreSQL, and SQL Server.

3. **Flexibility**:

   o   Enables complex queries, data analysis, and reports generation.

Key SQL Features

1. **Data Definition Language (DDL)**:

   o   Commands to define and modify database structure (e.g., CREATE, ALTER, DROP).

2. **Data Manipulation Language (DML)**:

   o   Commands to manipulate data (e.g., SELECT, INSERT, UPDATE, DELETE).

3. **Data Control Language (DCL)**:

   o   Commands to control access (e.g., GRANT, REVOKE).

4. **Transaction Control Language (TCL)**:

   o   Commands to manage database transactions (e.g., COMMIT, ROLLBACK).

Applications of SQL

- **Business Analytics:**

  o   Extract insights from customer, sales, and operational data.

- **Web Development:**

  o   Backend database management for e-commerce and social platforms.

- **Healthcare:**

  o   Patient data management and reporting.

- **Finance:**

  o   Transaction tracking and fraud detection.

5.7. Basics of Data Storage and Retrieval

https://github.com/sanjeevlcc/notes_2081/blob/main/BBA_1_Information_Technology_For_Business/Notes_Manaslu/LAB_1_SQL.txt

https://github.com/sanjeevlcc/notes_2081/blob/main/BBA_8_Management_of_Technology/lab%20sql/bba_lab2081.txt

https://github.com/sanjeevlcc/notes_2081/tree/main/DBMS_BIM_BSCIT_BCA/LABS_ON_AIR

5.8. Introduction to Big Data and Its Business Applications

What is Big Data?

- **Big Data** refers to extremely large datasets that cannot be easily processed, stored, or analyzed using traditional database tools or methods due to their volume, velocity, variety, and complexity.

- **Characteristics of Big Data** (often referred to as the "3 Vs"):

  1. **Volume**: Large amounts of data generated every second (e.g., social media posts, transaction logs).

  2. **Velocity**: High speed at which data is generated and needs to be processed (e.g., real-time streaming data).

  3. **Variety**: Data comes in various formats – structured, semi-structured, and unstructured (e.g., text, images, video, logs).

  4. **Veracity**: The quality and accuracy of data.

  5. **Value**: The insights and usefulness that can be derived from the data.

Big Data Technologies

1. **Hadoop**:

- o An open-source framework for storing and processing large datasets using distributed computing.
- o Allows data to be processed across clusters of computers to handle Big Data's size and complexity.

2. **NoSQL Databases**:

- o Databases designed to handle unstructured or semi-structured data, such as MongoDB, Cassandra, and Couchbase.

3. **Data Warehouses**:

- o Centralized repositories for storing large amounts of structured data that can be queried and analyzed (e.g., Amazon Redshift, Google BigQuery).

4. **Cloud Computing**:

- o Provides scalable storage and processing capabilities for Big Data (e.g., AWS, Microsoft Azure, Google Cloud).

5. **Data Lakes**:

- o Storage repositories that can handle structured, semi-structured, and unstructured data in its raw form for further analysis.

Business Applications of Big Data

1. **Customer Insights and Personalization**:

- o By analyzing consumer behavior and preferences, businesses can provide personalized products, services, and marketing.
- o **Example**: E-commerce companies like Amazon and Netflix use Big Data to recommend products and movies based on past behavior.

2. **Predictive Analytics**:

- o Businesses can use Big Data to predict future trends and customer behaviors, enabling them to make proactive decisions.
- o **Example**: Retailers use predictive analytics for inventory management, predicting which products will be in demand and when.

3. **Operational Efficiency**:

- o Big Data helps organizations optimize processes, reduce costs, and improve productivity by identifying inefficiencies.
- o **Example**: Airlines use Big Data to optimize flight schedules, reducing delays and fuel consumption.

4. **Fraud Detection and Risk Management**:

- o Financial institutions and insurance companies use Big Data analytics to detect fraudulent activities and assess risks in real-time.

o **Example**: Banks analyze transaction data to identify unusual patterns and prevent fraud.

5. **Healthcare**:

   o Big Data enables improved patient care through better diagnosis, treatment planning, and personalized medicine.

   o **Example**: Hospitals use data from patient records, medical imaging, and wearable devices to predict health risks and improve treatment outcomes.

6. **Supply Chain and Logistics**:

   o Big Data helps businesses track products in real-time, optimize inventory, and forecast demand, improving supply chain efficiency.

   o **Example**: Amazon tracks inventory levels and shipping routes to ensure timely delivery of goods.

7. **Market Research and Competitive Intelligence**:

   o Companies use Big Data to monitor market trends, analyze customer sentiments, and gain insights into competitors.

   o **Example**: Social media monitoring tools help businesses track consumer sentiment and competitor activity.

Challenges of Big Data

1. **Data Quality**:

   o Handling data with varying levels of accuracy, consistency, and completeness can be challenging.

2. **Data Security and Privacy**:

   o Storing and processing massive amounts of sensitive data raises concerns about cybersecurity and regulatory compliance (e.g., GDPR).

3. **Storage and Processing Power**:

   o Storing large volumes of data and processing them quickly requires powerful infrastructure and advanced technologies.

4. **Skills Gap**:

   o The need for skilled professionals, such as data scientists and analysts, to manage and interpret Big Data.

Future Trends in Big Data

1. **Artificial Intelligence and Machine Learning**:

   o AI and ML will increasingly be used to analyze Big Data, automating decision-making and predictive analytics.

2. **Edge Computing**:

   o Processing data closer to the source (e.g., IoT devices) to reduce latency and improve real-time analysis.

3. **Real-Time Data Analytics**:

   o With the growing importance of real-time decision-making, real-time data analytics tools and techniques will continue to evolve.

4. **Data Democratization**:

   o Making Big Data and analytics tools more accessible to non-technical users, enabling more people to leverage data for decision-making.

5.9. Data Security and Ethical Considerations

What is Data Security?

- **Data Security** refers to the practices, technologies, and processes that protect data from unauthorized access, data breaches, corruption, or theft throughout its lifecycle.

- Ensuring that sensitive data, such as personal, financial, and medical information, remains confidential, accessible only to authorized users, and intact during storage and transmission.

Key Elements of Data Security

1. **Confidentiality**:

   o Ensures that data is only accessible to those who are authorized.

   o **Encryption** is commonly used to protect sensitive data from unauthorized access.

2. **Integrity**:

   o Ensures that data remains accurate and unaltered during storage and transmission.

   o Techniques such as **checksums**, **hashing**, and **digital signatures** help maintain data integrity.

3. **Availability**:

   o Ensures that data is accessible to authorized users when needed.

   o Regular backups and **disaster recovery plans** are critical to ensure availability in case of system failure.

Data Security Measures

1. **Access Control**:

   o Restricts access to data based on roles and responsibilities.

- Role-based access control (RBAC) and mandatory access control (MAC) are common models.

2. **Authentication and Authorization**:

   - **Authentication** verifies the identity of users (e.g., passwords, biometrics).

   - **Authorization** determines what an authenticated user can do with the data (e.g., read, write, delete).

3. **Encryption**:

   - Data is encrypted during storage (at rest) and transmission (in transit) to prevent unauthorized access.

   - **AES (Advanced Encryption Standard)** and **SSL/TLS encryption** are commonly used encryption methods.

4. **Data Masking**:

   - Data is obfuscated or anonymized to protect sensitive information while maintaining its usability for analysis.

   - Example: Masking credit card numbers when displaying customer information.

5. **Firewalls and Intrusion Detection Systems (IDS)**:

   - Firewalls monitor and control incoming and outgoing network traffic to prevent unauthorized access.

   - IDS detects any suspicious activity or potential breaches in the network.

6. **Backup and Recovery**:

   - Regular data backups ensure that data can be recovered in case of corruption, loss, or natural disasters.

Ethical Considerations in Data Management

1. **Data Privacy**:

   - **Privacy** refers to the right of individuals to control their personal data and decide how it is used.

   - Companies must adhere to privacy regulations such as the **GDPR (General Data Protection Regulation)** and **CCPA (California Consumer Privacy Act)**, which protect consumer privacy.

2. **Informed Consent**:

   - Individuals must be informed about what data is being collected, why it is being collected, and how it will be used.

   - **Opt-in** or **opt-out** consent mechanisms should be clear and transparent.

3. **Data Ownership**:

- o Determining who owns the data (individuals or companies) and who has the rights to access, share, and use it.

- o Ethical data use requires transparency and accountability in how data is accessed and utilized.

4. **Bias and Discrimination**:

- o Data can inadvertently reinforce existing biases or discrimination if not properly managed.

- o Ethical data practices involve ensuring fairness and avoiding discriminatory practices in data collection, analysis, and decision-making.

5. **Surveillance**:

- o Ethical considerations around monitoring individuals' activities or behaviors through data collection, especially in areas like social media, government surveillance, and workplace monitoring.

- o Striking a balance between security and individual rights is key.

6. **Data Deletion**:

- o Companies must respect individuals' right to have their personal data deleted (often referred to as the **Right to be Forgotten** under GDPR).

- o Ethical handling of data includes ensuring that data is not kept longer than necessary.

Challenges in Data Security

1. **Cybersecurity Threats**:

- o Increasing threats from hackers, ransomware, and data breaches put businesses at risk.

2. **Insider Threats**:

- o Employees or individuals with authorized access to systems may misuse their privileges or leak data.

3. **Data Encryption Challenges**:

- o Implementing robust encryption can be complex and computationally expensive, especially with large datasets.

4. **Evolving Regulations**:

- o Keeping up with constantly changing privacy laws and regulations requires businesses to remain vigilant and adaptable.

5. **Emerging Technologies**:

- o The use of AI, cloud computing, and the Internet of Things (IoT) introduces new challenges in ensuring the security of data.

Best Practices for Ensuring Data Security and Ethical Compliance

1. **Data Minimization**:

   o Collect only the data you need and ensure that data is kept only for as long as necessary.

2. **Encryption and Secure Data Storage**:

   o Always encrypt sensitive data and ensure it is stored in secure environments.

3. **Regular Security Audits and Updates**:

   o Perform regular security assessments and update systems and software to protect against new vulnerabilities.

4. **Transparency and Clear Privacy Policies**:

   o Ensure users understand how their data is collected, used, and protected. Provide clear privacy policies.

5. **Employee Training**:

   o Regularly train employees on data security policies, ethical considerations, and compliance regulations.

Er Sanjeev Thapa. BE CE, MTech CSE, MBS. DevOps Eng, CKA, RHCSA, RHCE, RHCSA-Openstack, MTCNA, MTCTCE, USRS, HE IPv6. https://github.com/sanjeevlcc   2024/2081

Q/A

Fill in the Blanks

1. A database is an organized collection of _____.

2. _____ is the process of using data to make informed business decisions.

3. Data can be categorized as _____ and _____.

4. A _____ manages and controls access to databases.

5. In relational databases, data is stored in _____.

6. A _____ is used to retrieve specific data from a database.

7. SQL stands for _____ Query Language.

8. Data _____ refers to storing and retrieving data efficiently.

9. _____ refers to massive volumes of data that traditional systems cannot handle.

10. Ensuring that data is not accessed by unauthorized users is part of _____.

Multiple Choice Questions (MCQ)

1. What is the primary purpose of a database? a) Perform calculations b) Store and manage data c) Design websites d) Create graphics

2. Which of the following is a DBMS? a) Python b) MySQL c) Windows d) HTML

3. What type of data is stored in relational databases? a) Unstructured data b) Structured data c) Multimedia files d) Encrypted data

4. SQL is primarily used for _____. a) Creating websites b) Querying and managing data c) Network security d) Graphic design

5. Which of the following is NOT an SQL command? a) SELECT b) INSERT c) DELETE d) EXPORT

6. What is the smallest unit of data in a relational database? a) Table b) Field c) Record d) Column

7. Big Data is characterized by its _____. a) Volume, Variety, Velocity b) Simplicity c) Lack of structure d) Small size

8. Which of the following is an example of unstructured data? a) Excel spreadsheet b) Text documents c) SQL table d) None of the above

9. What is a primary key? a) A unique identifier for records b) A duplicate field c) A type of SQL command d) A backup file

10. Data security primarily focuses on _____. a) Encrypting user accounts b) Protecting data from unauthorized access c) Designing user interfaces d) Testing software

11. Which of the following is a characteristic of Big Data? a) Slow data processing b) Fixed structure c) High velocity of data generation d) Small datasets

12. A foreign key is used to _____. a) Define a unique identifier for a table b) Establish a relationship between tables c) Retrieve data from multiple databases d) Delete records from a table

13. Which SQL command is used to modify data in a table? a) UPDATE b) INSERT c) SELECT d) DROP

14. In a relational database, columns are also called _____. a) Fields b) Records c) Tables d) Queries

15. Which of the following is a benefit of using a DBMS? a) Increased data redundancy b) Centralized data management c) Lack of scalability d) Unsecured data

16. What does the acronym CRUD stand for in database operations? a) Create, Read, Update, Delete b) Copy, Retrieve, Upload, Delete c) Connect, Read, Update, Download d) Create, Retrieve, Upload, Download

17. Which type of database is designed to handle unstructured data? a) Relational database b) NoSQL database c) Distributed database d) Centralized database

18. Which SQL clause is used to filter data? a) WHERE b) SELECT c) ORDER BY d) GROUP BY

19. What is a data warehouse? a) A system for storing backup files b) A central repository for structured and historical data c) A tool for designing user interfaces d) A database for real-time transactions

20. Which of the following best describes data mining? a) Extracting patterns from large datasets b) Storing data in tables c) Encrypting sensitive information d) Designing database schemas

21. The process of ensuring data accuracy and consistency is called _____. a) Data encryption b) Data integrity c) Data warehousing d) Data redundancy

22. Which SQL command is used to delete a table from a database? a) DROP b) DELETE c) REMOVE d) ALTER

23. What is a schema in a database? a) A collection of database objects b) A specific SQL command c) A backup strategy d) A type of data encryption

24. Data visualization is used to _____. a) Encrypt sensitive information b) Represent data graphically c) Perform data backup d) Increase data redundancy

25. Which component of DBMS handles transaction management? a) Query processor b) Transaction manager c) Data dictionary d) User interface

Short Questions

1. Define a database.

2. Why is data important for business decision-making?

3. Differentiate between structured and unstructured data.

4. What is a Database Management System (DBMS)?

5. Name three popular DBMS software.

6. Explain the concept of relational databases.

7. What are tables, and how are they used in databases?

8. Define SQL and its importance.

9. List three basic SQL commands and their purposes.

10. What are the basics of data storage?

11. Explain the term "Big Data."

12. How does Big Data impact businesses?

13. Define data security.

14. What is data retrieval?

15. Give an example of ethical consideration in database management.

16. Explain the concept of a query in databases.

17. How are reports generated in relational databases?

18. Define the term "data redundancy."

19. What is the role of a database administrator?

20. Mention two differences between SQL and NoSQL databases.

21. What is data integrity?

22. Name two challenges in handling unstructured data.

23. Define "data warehouse."

24. Explain the role of encryption in data security.

25. What is a foreign key in relational databases?

Comprehensive Questions

1. Describe the key features and functions of a DBMS.

2. Discuss the importance of relational databases in data organization.

3. How does SQL help in managing large datasets effectively?

4. Explain the process of data storage and retrieval with examples.

5. Discuss the characteristics and applications of Big Data in modern business.

6. Explain the difference between structured and unstructured data with examples.

7. Describe the concept of data security and its challenges.

8. How do tables, queries, and reports interconnect in relational databases?

9. Discuss ethical considerations when handling sensitive data.

10. Explain how database backups ensure data reliability and security.

Answers

Fill in the Blanks

1. *Data*

2. *Data analysis*

3. *Structured, Unstructured*

4. *DBMS (Database Management System)*

5. *Tables*

6. *Query*

7. *Structured*

8. *Storage*

9. *Big Data*

10. *Data security*

Multiple Choice Questions (MCQ)

1. *b) Store and manage data*

2. *b) MySQL*

3. *b) Structured data*

4. *b) Querying and managing data*

5. *d) EXPORT*

6. *b) Field*

7. *a) Volume, Variety, Velocity*

8. *b) Text documents*

9. *a) A unique identifier for records*

10. *b) Protecting data from unauthorized access*

11. *c) High velocity of data generation*

12. b) Establish a relationship between tables

13. a) UPDATE

14. a) Fields

15. b) Centralized data management

16. a) Create, Read, Update, Delete

17. b) NoSQL database

18. a) WHERE

19. b) A central repository for structured and historical data

20. a) Extracting patterns from large datasets

21. b) Data integrity

22. a) DROP

23. a) A collection of database objects

24. b) Represent data graphically

25. b) Transaction manager