

Big Data Analytics and SQL Exam - Paper 2

Instructions

- This exam paper is worth 100 marks in total.
- Answer all questions.
- For MCQs, circle the correct answer.
- For coding questions, ensure your code is legible and correctly solves the given problem.

Part A: Multiple Choice Questions (20 Marks)

Choose the correct answer from the provided options. Each question is worth 2 marks.

1. Which SQL keyword is used to remove duplicates from the results of a query?
 - a. SELECT DISTINCT
 - b. UNIQUE
 - c. REMOVE DUPLICATES
 - d. FILTER DISTINCT
2. What is a key feature of NoSQL databases?
 - a. Fixed schema
 - b. High vertical scalability
 - c. Support for complex joins
 - d. Flexibility to handle unstructured data
3. Which statistical method is commonly used to predict categorical outcomes?
 - a. Regression Analysis
 - b. Clustering
 - c. Logistic Regression
 - d. Correlation
4. What does the 'commit' command do in SQL?
 - a. Reverts the last change
 - b. Saves changes made by a transaction

- c. Starts a new transaction
 - d. Checks the transaction status
5. In Python, which module is primarily used for statistical operations?
- a. NumPy
 - b. SciPy
 - c. Pandas
 - d. Statistics
6. What is the result of the SQL query 'SELECT COUNT(*) FROM Users'?
- a. The number of tables in the users database
 - b. The sum of all values in the Users table
 - c. The number of rows in the Users table
 - d. The number of unique users in the database
7. Which concept is crucial for understanding the distribution of data within a dataset?
- a. Polymorphism
 - b. Encapsulation
 - c. Standard Deviation
 - d. Inheritance
8. Which Python function is used to read a CSV file?
- a. `pandas.read_csv()`
 - b. `pandas.load_csv()`
 - c. `pandas.open_csv()`
 - d. `pandas.get_csv()`
9. What does ACID stand for in the context of databases?
- a. Atomicity, Consistency, Isolation, Durability
 - b. Association, Consistency, Isolation, Data Integrity
 - c. Atomicity, Consistency, Integrity, Durability
 - d. Authentication, Consistency, Isolation, Durability
10. What type of machine learning algorithm is used for grouping data into a predetermined number of groups?
- a. Supervised learning
 - b. Unsupervised learning
 - c. Reinforcement learning
 - d. Semi-supervised learning

Part B: SQL Query Writing (40 Marks)

Attempt all questions. Each question is worth 10 marks.

1. Write an SQL query to find the most recent login date for each user in the 'Logins' table, which includes columns 'UserID' and 'LoginDate'.
2. Given a table 'Products' with columns 'ProductID', 'ProductName', and 'Price', write an SQL query to find all products cheaper than 10.
3. Write an SQL query that returns the total revenue, which is the sum of all 'Order-Amount' from the 'Orders' table.
4. Given a table 'Employees' and 'Departments' with a common column 'DepartmentID', write an SQL query to list all employees and their corresponding department names.

Part C: Python Programming (40 Marks)

Attempt all questions. Each question is worth 10 marks.

1. Write a Python function to calculate the factorial of a number provided by the user.
2. Create a Python script that converts a list of temperatures from Celsius to Fahrenheit.
3. Write a Python function that takes a list and returns a new list with unique elements of the first list.
4. Develop a Python script that uses regular expressions to check if an email address entered by the user is valid.