

Big Data Analytics and SQL Exam - Paper 3

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. What type of data visualization is typically used to represent part-to-whole relationships?
 - a. Line graph
 - b. Bar chart
 - c. Pie chart
 - d. Scatter plot
2. Which SQL function calculates the total sum of a numeric column?
 - a. SUM
 - b. TOTAL
 - c. COUNT
 - d. AVG
3. Which Python library is primarily used for data manipulation and analysis?
 - a. NumPy
 - b. Matplotlib
 - c. Pandas
 - d. SciPy
4. What does the term 'Data Lake' refer to?
 - a. A centralized repository that allows you to store all your structured and unstructured data at any scale.

- b. A type of database optimized for handling large amounts of transactional data.
 - c. A data visualization tool used to represent large data sets in graphical format.
 - d. A sophisticated algorithm used for processing Big Data.
5. What is the primary purpose of a database index?
- a. To decrease storage space.
 - b. To increase the speed of data retrieval.
 - c. To enforce data integrity.
 - d. To provide backup capabilities.
6. Which of the following is not a characteristic of Big Data?
- a. Veracity
 - b. Volatility
 - c. Volume
 - d. Variety
7. In machine learning, what is a feature?
- a. A type of algorithm used for data processing.
 - b. An individual measurable property or characteristic of a phenomenon being observed.
 - c. A unique type of data visualization.
 - d. The outcome predicted by a model.
8. What does the SQL clause 'HAVING' used for?
- a. To specify conditions on groups created by GROUP BY.
 - b. To limit the number of rows returned.
 - c. To join tables.
 - d. To exclude rows before grouping.
9. Which statistical test is used to determine whether there is a significant association between two categorical variables?
- a. T-test
 - b. ANOVA
 - c. Chi-square test
 - d. Linear regression
10. In Python, what is the output of 'len("Hello World!")'?

- a. 10
- b. 11
- c. 12
- d. 13

Part B: SQL Query Writing (40 Marks)

Attempt all questions. Each question is worth 10 marks.

1. Write an SQL query to display the third highest salary from the 'Employees' table.
2. Given a 'Customers' table with 'CustomerID', 'Name', and 'State', write an SQL query to find all customers from 'California'.
3. Write an SQL query to count the number of transactions each customer has in the 'Transactions' table.
4. Given a 'Books' table with 'BookID', 'Title', and 'AuthorID', write an SQL query to list all books with more than one author assuming there is a linking table 'BookAuthors'.

Part C: Python Programming (40 Marks)

Attempt all questions. Each question is worth 10 marks.

1. Write a Python script to reverse a string input by the user without using built-in functions.
2. Develop a Python function that detects whether a given IP address is valid.
3. Implement a Python program that simulates rolling a six-sided die until a 6 is rolled, and counts the number of rolls.
4. Write a Python script that calculates the number of vowels and consonants in a string provided by the user.