

Comprehensive Exam Paper: Big Data Analytics

Yashfeen Institute

Instructions

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

Part A: Multiple Choice Questions (10 Marks)

Choose the correct answer from the provided options. Each question is worth 1 marks.

1. What does SQL stand for?
 - a. Standard Query Language
 - b. Structured Query Language
 - c. Simplified Query Language
 - d. None of the above
2. Which command is used to delete a table in SQL?
 - a. DELETE TABLE
 - b. DROP TABLE
 - c. REMOVE TABLE
 - d. CLEAR TABLE
3. In Python, which keyword is used to define a function?
 - a. func
 - b. define
 - c. def
 - d. function
4. Which of the following is a mutable data type in Python?

- a. String
 - b. Tuple
 - c. List
 - d. None of the above
5. Which of the following statements about Big Data is **true**?
- a. Big Data analytics is primarily used for generating small datasets from larger ones.
 - b. The 3Vs of Big Data are Velocity, Variety, and Volume.
 - c. Big Data can only be processed using traditional database systems.
 - d. All Big Data problems can be solved with more computing power.
6. Considering database transactions, which of the following phenomena must be prevented to ensure isolation in a database system?
- a. Phantom Read
 - b. Dirty Read
 - c. Non-repeatable Read
 - d. All of the above
7. Which Python code snippet correctly implements a decorator that times the execution of a function?
- a. `def timer(func): import time; start = time.time(); func(); print(time.time() - start)`
 - b. `def timer(func): import time; def wrapper(): start = time.time(); func(); print(time.time() - start); return wrapper`
 - c. `def timer(func): import time; def wrapper(): start = time.time(); result = func(); print(time.time() - start); return result; return wrapper`
 - d. `def timer(func): import time; start = time.time(); result = func(); print(time.time() - start); return result`
8. What does the "MapReduce" programming model primarily aim to accomplish in Big Data processing?
- a. Reduce the complexity of data to 2D maps for easier analysis.
 - b. Perform in-memory computations on large datasets for speed.
 - c. Simplify the processing of large datasets across distributed systems.
 - d. Directly map data inputs to outputs without reducing steps.
9. Which of the following is a common measure of central tendency in statistics?

- a. Standard deviation
 - b. Median
 - c. Variance
 - d. Mode
10. In the context of data analytics, what does the term 'outlier' refer to?
- a. A value that lies an abnormal distance from other values in a random sample from a population.
 - b. A predictive modeling technique used in machine learning.
 - c. A graphical representation of data points in a coordinate system.
 - d. The process of cleaning data from redundant information.

Part B: Short Answer Questions attempt all(30 Marks)

Answer the following questions briefly. Each question is worth 5 marks.

1. Describe how Big Data Analytics can impact decision-making in businesses.
2. Explain the difference between INNER JOIN and OUTER JOIN in SQL.
3. Explain the difference between supervised and unsupervised learning in the context of machine learning algorithms. Provide one example of each.
4. Describe the concept of "normalization" in databases. Why is it important?
5. What is a "lambda function" in Python? Provide an example of how it might be used in data analysis.
6. Discuss the significance of the "pivot table" function in data analysis. How can it be implemented in Python using pandas?

Part C: SQL Query Writing attempt 3 only(30 Marks)

1. Write an SQL query to find the name and salary of the highest-paid employee in the 'Employees' table.
2. Given a table 'Orders' with columns 'OrderID', 'CustomerID', and 'OrderDate', write an SQL query to list all orders made in the last 30 days.
3. Write an SQL query to list all products with a price higher than the average price of all products in the 'Products' table.
4. Given a table 'Employees' with columns 'EmployeeID', 'Name', 'Department', and 'Salary', write an SQL query to display each department's total salary expenditure.

Part D: Python Programming attempt 3 only(30 Marks)

1. Write a Python function to check if a given number is prime.
2. Create a Python script that asks the user for a list of numbers (separated by commas) and prints the sum of those numbers.
3. Write a Python function that accepts a string as input and returns 'True' if the string is a palindrome, 'False' otherwise.
4. Given a list of integers, write a Python script to find the second-highest number in the list.