



BRAINWARE UNIVERSITY

Term End Examination 2025-2026

Programme – BCA-Hons-2023

Course Name – Big Data Analytics

Course Code - BCA60206

(Semester VI)

Full Marks : 60

Time : 2:30 Hours

[The figure in the margin indicates full marks. Candidates are required to give their answers in their own words as far as practicable.]

Group-A

(Multiple Choice Type Question)

1 x 15=15

1. *Choose the correct alternative from the following :*

- (i) What does the term Big Data mainly refer to?
 - a) Small amount of data
 - b) Traditional data only
 - c) Very large and complex data
 - d) Only structured data
- (ii) Which of the following is an example of unstructured data?
 - a) Excel sheet
 - b) Database table
 - c) Text messages
 - d) CSV file
- (iii) Which of the following is NOT a Big Data characteristic?
 - a) Volume
 - b) Velocity
 - c) Variety
 - d) Accuracy
- (iv) Which data type includes videos and images?
 - a) Structured
 - b) Text data
 - c) Unstructured
 - d) Numeric
- (v) Classify JSON and XML files into the correct data type.
 - a) Structured
 - b) Semi-structured
 - c) Unstructured
 - d) Binary
- (vi) Predict the Big Data challenge when data size increases rapidly.
 - a) Visualization
 - b) Scalability
 - c) Data entry
 - d) Formatting
- (vii) The Hadoop ecosystem tool best classified as a data warehousing and SQL querying system is
 - a) Pig
 - b) Hive
 - c) Mahout
 - d) Flume
- (viii) A developer wants Map tasks to process larger chunks of data. What should be increased?
 - a) Input split size
 - b) Reducer count

- c) Job priority
d) Replication factor
- (ix) If sorted output is required across all reducers, what should be used?
a) Combiner
b) Partitioning
c) Total Order Sort
d) Speculative execution
- (x) What does skewness measure?
a) Central value of data
b) Symmetry of the data distribution
c) Closeness of data points
d) Average of all values
- (xi) Which of the following is a discrete probability distribution?
a) Normal
b) Poisson
c) Uniform
d) Exponential
- (xii) Which measure of dispersion is expressed as a percentage of the mean?
a) Range
b) Coefficient of Variation
c) Standard Deviation
d) Variance
- (xiii) What does a correlation coefficient of -1 signify?
a) No relation
b) Perfect negative relation
c) Weak positive relation
d) Weak negative relation
- (xiv) Which type of data is measured using nominal scale?
a) Height
b) Gender
c) Weight
d) Temperature
- (xv) What is the probability of an impossible event?
a) 1
b) 0
c) 0.5
d) Undefined

Group-B

(Short Answer Type Questions)

3 x 5=15

2. Define Big Data and mention any two characteristics of Big Data. (3)
3. Explain the 3V characteristics of Big Data. (3)
4. Explain the importance of Big Data analytics. (3)
5. The marks of 5 students are: 40, 50, 60, 70, 80. Find the standard deviation. (3)
6. Examine how the characteristics of Big Data (Volume, Velocity, Variety) influence the design of Big Data storage systems. (3)

OR

Analyze the differences between structured, semi-structured, and unstructured digital data types in the context of Big Data sources, providing one example challenge for each. (3)

Group-C

(Long Answer Type Questions)

5 x 6=30

7. Find the standard deviation and variance for the data: 10, 12, 15, 20, 25. (5)
8. Analyze the role of the Shuffle and Sort phase in ensuring correct MapReduce output. (5)
9. Examine how the Apriori algorithm can be applied to an E-commerce transaction database to identify frequent itemsets, explaining the process and its significance in discovering purchase patterns. (5)
10. Analyze the steps involved in a MapReduce job run and explain how mapping and reducing stages interact during execution. (5)
11. Illustrate the working of gradient descent while training a machine learning model. (5)
12. Evaluate the historical evolution of visualization and justify how technological advancements have shaped modern visualization techniques. (5)

OR

Critique the limitations of Google Charts versus Tableau in handling Big Data visualization scalability (5)