



Library
Brainware University
398, Ramkrishnapur Road, Barasat
Kolkata, West Bengal-700125

BRAINWARE UNIVERSITY

Term End Examination 2024-2025
Programme – MCA-2022/MCA-2023
Course Name – Big Data Analysis
Course Code - MCA305B
(Semester III)

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) Define Exabyte as _____.
 - a) 10^{15}
 - b) 10^{17}
 - c) 10^{18}
 - d) 10^{19}
- (ii) Data approaches is defined for reducing the _____.
 - a) costs of data management
 - b) data quality
 - c) standardization
 - d) None of these
- (iii) HDFS provides file for execution _____.
 - a) permissions
 - b) authentication
 - c) Both permissions and authentication
 - d) None of these
- (iv) Define Big Data as:
 - a) structured
 - b) unstructured
 - c) semi-structured
 - d) All of these
- (v) Indicate which of the following emphasises on the discovery of earlier properties that are not known on the data.
 - a) Machine Learning
 - b) Big Data
 - c) Data wrangling
 - d) Data mining
- (vi) Define how many V's exists in Big Data.
 - a) 2
 - b) 3
 - c) 4
 - d) 5
- (vii) Determine what was Hadoop named after.
 - a) MapReduce
 - b) HDFS
 - c) YARN
 - d) All of these
- (viii) Explain Real time data is _____.
 - a) Field
 - b) Primary Key
 - c) unique
 - d) record
- (ix) Select one which has the world's largest Hadoop cluster.

- a) Apple
c) Facebook
(x) Name the type of analytics that answers the question "What has happened?"
a) Descriptive analytics
c) Prescriptive analytics
(xi) Predict the minimum amount of data that a disk can read or write in HDFS.
a) Byte size
c) Heap
(xii) Select the correct option: A loop in a graph in big data analytics is
a) When there is a edge from A->B, there is also on edge from B->A
c) Where there is an edge from a node to itself
(xiii) Choose The correct answer for the blank: Most NoSQL databases support automatic _____ meaning that you get high availability and disaster recovery.
a) Processing
c) Replication
(xiv) Predict the characteristic of a NoSQL database.
a) Uses JSON
c) Requires JOINS
(xv) Select the language which is best for the statistical environment in big data.
a) C
c) Java
b) Datamatics
d) None of these
b) Predictive analytics
d) None
b) Block size
d) both 1 and 2
b) Where there is a path in some way from a node, through 1 or more other nodes - back to original node
d) When there is a edge from A->B->C, there is also on edge from C->B->A
b) Scalability
d) All of these
b) Needs a schema
d) Uses tables for storage
b) R
d) Python

Group-B

(Short Answer Type Questions)

3 x 5=15

2. State the primary objective of the K-means clustering algorithm. (3)
3. How does the K-means algorithm initialize cluster centroids, and what can be the impact of the initial centroids on the clustering results? (3)
4. Explain the two main phases of a MapReduce job. (3)
5. Analyze the core component of Hadoop that is responsible for distributed storage and processing of data. (3)
6. Write down Big Data eCommerce and advantages of it. (3)

OR

Illustrate the features of NoSQL. (3)

Group-C

(Long Answer Type Questions)

5 x 6=30

7. Write down the trade-offs between schema flexibility and data consistency in NoSQL databases, providing practical examples to illustrate the decision-making process. (5)
8. Discuss the best practices of Big Data Analytics. (5)
9. State the four computing resources of Big Data Storage. (5)
10. Analyse the implications of hyperparameter choices on the model's interpretability and predictive accuracy. (5)
11. Consider the concept of classification. (5)
12. Evaluate the paradigm shift introduced by NoSQL databases and justify its relevance in modern data management. (5)

OR

Evaluate the role of NoSQL databases in supporting real-time analytics and provide examples of scenarios where their usages are most advantageous. (5)