BRAINWARE UNIVERSITY

Term End Examination 2023-2024
Programme – MCA-2022
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) Discuss which of the following is a wrong statement.

   a) The big volume actually represents Big Data   b) Big Data is just about tons of data
   c) The data growth and social media explosion   d)
   have improved that how we look at the   All of these
   data

(ii) Discuss that beyond Volume, velocity and variety are the concerns and problems of big data veracity.

   a) Correct   b) Wrong
   c) Partially wrong   d) Partially correct

(iii) Discuss Which of the big data characteristic is comparatively more concerned with data science?(Probable duplicate)

   a) Variety   b) Velocity
   c) Volume   d) None of these

(iv) Find errors from following options: Which is the type of a NoSQL database?

   a) Cassandra   b) MongoDB
   c) SQL Server   d) None of the above

(v) Focus on Facebook which Tackles Big Data With _____ based on Hadoop.

   a) Project Prism   b) Prism
   c) Project Big   d) Project Data

(vi) Explain which is general-purpose computing model and runtime system for distributed data analytics.

   a) Mapreduce   b) Drill
   c) Oozie   d) Both Drill and Oozie

(vii) Explain which function is responsible for consolidating the results produced by each of the Map() functions/tasks.

   a) Reduce   b) Map

d) All of the mentioned

c) Reducer

(viii) Although the Hadoop framework is implemented in Java, MapReduce applications need not be written in _____

a) Java

b) C

c) C#

d) None of the mentioned

(ix) Demonstrate about veracity

a) Filter

b) Reliable

c) manage data

d) All of these

(x) The primary aspect of Big Data is to provide

a) demanding data rapidly

b) demanding data slowly

c) no demanding data

d) None of these

(xi) Uses of velocity in

a) mobile devices

b) sensors

c) business processes

d) All of these

(xii) Demonstrate about categories of clustering methods

a) Partitioning methods

b) Hierarchical methods

c) Grid based methods

d) All of these

(xiii) Identify among the options below which is general-purpose computing model and runtime system for Distributed Data Analytics.

a) HDFS

b) YARN

c) Map Reduce

d) Utilities

(xiv) Define HDFS is implemented in which programming language?

a) C

b) Java

c) C++

d) Python

(xv) Describe when a_____ serves as the master and there is only one NameNode per cluster.

a) Data Node

b) replication

c) Name node

d) Data block

Group-B
(Short Answer Type Questions)

3 x 5=15

2. Explain the CAP theorem in the context of NoSQL databases and provide an example of how it impacts database design. (3)

3. What are some common use cases for MapReduce? (3)

4. Illustrate the various types of data in big data. (3)

5. Report the difference between clustering and classification. (3)

6. Evaluate the trade-off between computational complexity and interpretability in the context of the Apriori Algorithm. (3)

OR

Examine the impact of parameter tuning on the Apriori Algorithm's mining capabilities. (3)

Group-C
(Long Answer Type Questions)

5 x 6=30

7. Write down the characteristics of Big Data Applications. (5)

8. What is HDFS? (5)

9. Contrast the ramifications of overfitting and underfitting in Decision Tree models, elucidating the potential consequences on predictive outcomes. (5)

10. What are Bayesian Classifiers? (5)

11. What are Recommenders? (5)

12. What is Cluster Analysis? (5)

OR

How will you solve a classification problem using Decision Tree? (5)

****