BRAINWARE UNIVERSITY

Term End Examination 2022
Programme – B.Tech.(CSE)-2018/B.Tech.(CSE)-2019
Course Name – Big Data Analysis
Course Code - PEC-701A
(Semester VII)

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) Determine the unit of data that flows through a Flume agent is

a) Log b) Row
c) Event d) Record

(ii) Analyze which is the term that is used to describe data that is high volume , highvelocity and /or high variety.

a) Analytics b) Big data
c) Hadoop Data d) Big data analytics

(iii) 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

(iv) 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 have improved that how we look at the data d) All of these

(v) Measure which of the following is a wide-column store?

a) Cassandra b) Riak
c) MongoDB d) Redis

(vi) Define Exabyte

a) 10^15 b) 10^17
c) 10^18 d) 10^19

(vii) Define about big Data approaches which are reducing the

a) costs of data management b) data quality
c) standardization d) None of these

(viii) Identify the Big data which focus on

a) Identify the data we have b) Identify the data we need
c) what you want to achieve d) All of these

(ix) State about HDFS that provides

a) difficult access
b) easier access
c) Both easier and difficult access
d) None of these

(x) Describe about Datanodes that perform

a) read-write operations
b) read operations
c) write operations
d) None of these

(xi) Consider the number of maps which is usually driven by the total size of _____

a) inputs
b) outputs
c) tasks
d) None of the mentioned

(xii) Determine according to analysts, for what can traditional IT systems provide afoundation when theyâ€™re integrated with big data technologies like Hadoop?

a) Big data management and data mining
b) Data warehousing and business intelligence
c) Management of Hadoop clusters
d) Collecting and storing unstructured data

(xiii) Justify which of the following function is used for k-means clustering?

a) k-means
b) k-mean
c) heatmap
d) none of the mentioned

(xiv) Determine all of the following accurately describe Hadoop, EXCEPT _____

a) Open-source
b) Real-time
c) Java-based
d) Distributed computing approach

(xv) Justify which of the following clustering requires merging approach?

a) Partitional
b) Hierarchical
c) Naive Bayes
d) None of the mentioned

Group-B

(Short Answer Type Questions) 3 x 5=15

2. Explain Stream Data Model. (3)
3. Explain HDFS architecture with proper diagram. (3)
4. Explain Bayesian Classifiers with example. (3)
5. Explain why is Hadoop used for Big Data Analytics (3)
6. Express how would you transform unstructured data into structured data. (3)

OR

Prepare a short note about Association Rule (3)

Group-C

(Long Answer Type Questions) 5 x 6=30

7. Define how will you solve a classification problem using Decision Tree? (5)
8. Explain NO SQL Database. (5)
9. Illustrate how E-Commerce is attached with Big Data to improve business in detail? (5)
10. Differentiate RDBMS and Hadoop (5)
11. Explain Stream Data Model and Architecture with the help of a diagram. (5)
12. Explain how can you achieve security in Hadoop? (5)

OR

Explain do you prefer good data or good models? Why? (5)

****