BRAINWARE UNIVERSITY

Term End Examination 2023-2024
Programme – B.Tech.(CSE)-2019/B.Tech.(CSE)-2020
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) Identify about big data velocity deals with
a) business processes b) sensors
c) mobile devices d) All of these

(ii) Identify about big data application which is directly dependent on
a) Awareness of thearchitecture ofcomputing platform b) Hardware
c) software d) All of these

(iii) Select 422 exabytes=
a) Total digital data created in 2008 b) Total digital data created in 2009
c) Total digital data created in 2010 d) Total digital data created in 2011

(iv) State about the data come from many sources like
a) Facebook b) Google
c) Amazon d) All of these

(v) Express about tabular stores which is/are
a) structured in a tabularform. b) arranges data elements in verticacolumns
c) arranges dataelements in horizontalrows d) All of these

(vi) Express about content-based approach problem is/are
a) Finding the appropriate features is hard b) Recommendations for new users
c) Both Finding the appropriate features is hard and Recommendations for new users d) None of these

(vii) Express about decision tree which can be used for clustering
a) possible b) impossible
c) impossible in some scenario d) None of these

(viii) Express about Hadoop which is

a) open-source
b) Java based programming framework
c) designed to storage and process large datasets
d) All of these

(ix) Express about_____ which maps input key/value pairs to a set of intermediate key/value pairs

a) Mapper
b) Reducer
c) Both Mapper and Reducer
d) None of the above

(x) Express the number of maps is usually driven by the total size of_____ .

a) task
b) output
c) input
d) none

(xi) Express the following is a characteristic of a NoSQL database.

a) Uses JSON
b) Needs a schema
c) Requires JOINs
d) Uses tables for storage

(xii) Predict the consequence betweena node and its predecessors while creating bayesian network.

a) Functionally dependent
b) Dependant
c) Conditionally independent
d) Both Conditionally dependant & Dependant

(xiii) State the following is true about Naive Bayes.

a) Assumes that all the features in a data set are equally important
b) Assumes that all the features in a dataset are independent
c) both
d) None of these

(xiv) State the bayes rule can be used.

a) Solving queries
b) Increasing complexity
c) Decreasing complexity
d) Answering probabilistic query

(xv) Identify from the following that are Decision Tree nodes.

a) Decision Nodes
b) End Nodes
c) Chance Nodes
d) All of Above

Group-B

(Short Answer Type Questions) 3 x 5=15

2. Write about NO SQL Database. (3)
3. Explain Key Value Stores. (3)
4. Discuss storage of big data. (3)
5. Illustrate about datastream. (3)
6. Differentiate RDBMS and Hadoop (3)

OR

Illustrate the applications of Association rules. (3)

Group-C

(Long Answer Type Questions) 5 x 6=30

7. Define how do you approach data preparation for big data analysis. (5)
8. Illustrate the concepts of decision tree. (5)
9. Explain about collaborative recommendation (5)
10. Express about HBASE. (5)
11. Explain about data stream. (5)
12. Summarize Big data with Twitter. (5)

OR (5)

Summarize E-Commerce Using Big Data

****