15350

BRAINWARE UNIVERSITY

Term End Examination 2024-2025
Programme – B.Tech.(CSE)-DS-2021/B.Tech.(CSE)-DS-2022
Course Name – Big Data and Analytics
Course Code - PEC-CSD601A
(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) Select the primary advantage of stream computing
a) Processing static data
b) Handling large volumes of data efficiently
c) Generating graphical outputs
d) Storing data in databases

(ii) Define sampling in respect to stream computing.
a) Selecting a subset of data for analysis
b) Storing data in a stream
c) Creating visual representations of data
d) Deleting unwanted data

(iii) Select the purpose of filtering streams in stream computing
a) Removing unwanted elements from a stream
b) Adding more data to a stream
c) Increasing the speed of a stream
d) Sorting data in a stream

(iv) Select the main purpose of a Real-Time Analytics Platform (RTAP)
a) Storing historical data
b) Processing static datasets
c) Analyzing data as it arrives in real-time
d) Deleting unwanted data elements

(v) Identify a characteristic of intelligent data analysis
a) Predictive modeling
b) Basic arithmetic operations
c) Copying and pasting data
d) Manual data entry

(vi) Select a feature that distinguishes big data from small data
a) Volume
b) Color
c) Font size
d) Date

(vii) Select a characteristic of intelligent data analysis
a) Predictive modeling
b) Drawing pictures
c) Typing documents
d) Playing video games

(viii) Define reporting in the context of big data
a) Presenting facts
b) Data analysis
c) Predictive modeling
d) Data visualization

(ix) Identify the components of Hadoop.

a) HDFS, MapReduce
b) SQL, NoSQL
c) Python, Ruby
d) HTML, CSS

(x) Select the primary functionality of HDFS.

a) Distributed storage
b) Relational database
c) Client-server model
d) Peer-to-peer network

(xi) Choose an application of MapReduce.

a) WordCount example
b) Sorting algorithms
c) Graph traversal
d) Encryption algorithm

(xii) Select the types and formats applied in MapReduce application

a) Text, Sequence
b) JPEG, MP3
c) PDF, Excel
d) PNG, TIFF

(xiii) Identify the steps of a MapReduce job run.

a) Input, Map, Reduce
b) Compile, Execute
c) Read, Write
d) Loop, Condition

(xiv) Choose the proper language that is primarily used for scripting in Apache Pig.

a) Java
b) Python
c) Pig Latin
d) Ruby

(xv) Choose the correct option for primary purpose of IBM InfoSphere BigInsights.

a) Real-time stream processing
b) Data integration and transformation
c) Graph analytics
d) Predictive modeling

Group-B

(Short Answer Type Questions) 3 x 5=15

2. List the main advantages of using a big data platform. (3)
3. Illustrate some limitations of traditional relational database systems in handling big data. (3)
4. Identify the applications of Real-Time Analytics Platform (RTAP) and provide an example use case. (3)
5. Explain the primary purpose of Pig in Big Data. (3)
6. Explain streaming in Hadoop and its significance in data processing. (3)

OR

Determine the components of Hadoop. (3)

Group-C

(Long Answer Type Questions) 5 x 6=30

7. Define the role data visualization plays in big data analysis. (5)
8. Identify the key challenges in Hadoop's MapReduce framework and discuss how they can be addressed. (5)
9. Distinguish between simple linear regression and multiple linear regression with examples. (5)
10. Discuss the applications of linear regression in various domains and explain how it is used for predictive modeling. (5)
11. Identify the role of HBase and ZooKeeper in Big Data applications and explain their key features and functionalities. (5)
12. Design a framework for handling Big Data analytics, considering data collection, storage, processing, and visualization. (5)

OR

Develop an approach to implement fragmentation techniques in a distributed database to optimize performance and data availability. (5)

****