



BRAINWARE UNIVERSITY

Term End Examination 2025-2026

Programme – B.Tech.(CSE)-DS-2021/B.Tech.(CSE)-DS-2022/B.Tech.(CSE)-DS-2023

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 stream in the context of stream computing
 - a) A flowing river
 - b) A sequence of data elements over time
 - c) A static dataset
 - d) A graphical representation
- (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) What is a Decaying Window in stream computing ?
 - a) A window that never changes over time
 - b) A window that decreases in size over time
 - c) A window with transparent elements
 - d) A static window with fixed boundaries
- (iv) Select an application area of Real-Time Sentiment Analysis.
 - a) Weather forecasting
 - b) Customer service enhancement
 - c) Historical data analysis
 - d) Sorting algorithms
- (v) Select a characteristic of big data platforms
 - a) Scalability
 - b) Limited data sources
 - c) Single-user access
 - d) Offline functionality
- (vi) Choose a trend in big data analytics
 - a) Increased use of machine learning
 - b) Decreased reliance on cloud computing
 - c) Decreased data volume
 - d) Increased manual processing
- (vii) Explain the importance of analytics in big data platforms
 - a) Extracting insights from data
 - b) Making coffee
 - c) Sending emails
 - d) Printing documents
- (viii) Select tools used for intelligent data analysis
 - a) Python
 - b) R
 - c) Tableau
 - d) All of these

- (ix) Identify a scenario where Hadoop MapReduce performs poorly due to inefficient data serialization.
- a) When using complex serialization formats like XML or JSON b) When using Avro serialization
- c) When using custom partitioning d) When using in-memory caching
- (x) Select the role of the Application Master in the YARN architecture of Hadoop.
- a) Manages lifecycle of applications, Negotiates resources b) Executes MapReduce tasks
- c) Stores metadata of HDFS d) Coordinates node communication
- (xi) Choose the correct option from following statements that is true about Apache Pig.
- a) Pig scripts are executed directly on HDFS. b) Pig uses SQL-like queries to process data.
- c) Pig compiles scripts into MapReduce jobs. d) Pig is optimized for real-time analytics.
- (xii) Select the right option from following statements that is true about Apache Hive.
- a) Hive queries are directly executed on HDFS. b) Hive uses Pig Latin for data processing.
- c) Hive provides a relational database-like interface. d) Hive is optimized for real-time data ingestion.
- (xiii) Choose the proper role of Hive metastore.
- a) Stores metadata about Hive tables and partitions b) Executes HiveQL queries
- c) Manages HDFS storage d) Schedules Hive jobs
- (xiv) Select a key characteristic of the stream data model.
- a) Storage of static, discrete entities b) Continuous flows of data
- c) Batch processing of data d) Limited processing capabilities
- (xv) Select the primary purpose of sampling data in a stream.
- a) Reducing processing time b) Increasing storage efficiency
- c) Improving accuracy of analysis d) Eliminating outliers

Group-B

(Short Answer Type Questions)

3 x 5=15

2. Explain the organizations utilization by deriving the value from unstructured data within big data environments? (3)
3. Define HiveQL and its role in Big Data processing. (3)
4. Explain the job scheduling process in Hadoop. (3)
5. Identify the applications of Real-Time Analytics Platform (RTAP) and provide an example use case. (3)
6. A 12.8 GB file is stored in HDFS using a 128 MB block size with replication factor 3. Assume 10% storage overhead due to metadata. Calculate the total physical storage used. (3)

OR

- A MapReduce job processes 960 GB using 32 map tasks and 8 reduce tasks. Assuming equal data distribution, calculate: a) Input per mapper b) Data per reducer (3)

Group-C

(Long Answer Type Questions)

5 x 6=30

7. Evaluate the key applications of Real-Time Analytics Platforms (RTAP) (5)
8. Analyze the challenges a businesses may face in implementing Real-Time Analytics Platforms, and how can they overcome these challenges (5)
9. Explain the concept of Simple Linear Regression and Multiple Linear Regression. How do we interpret regression coefficients in these models? (5)
10. Illustrate the impact of stream computing on stock market predictions and describe the role of real-time data processing in enhancing accuracy. (5)

11. Identify the fundamental concept behind the stream data model and its difference from traditional batch processing. (5)
12. Develop an interactive visualization dashboard for data analysis and explain the importance of interaction techniques. (5)

OR

Elaborate on the role of visual data analysis techniques in decision-making and how different visualization methods aid interpretation. (5)

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