8E1909

B.Tech. VIII-Sem. (Main/Back) Examination, April/May - 2025

Computer Science and Engineering (A.I)

8CA14-01 Big Data Analytics

CS, CAI

Time : 3 Hours

Maximum Marks : 70

Instructions to Candidates:

Attempt all Ten questions from Part A, Five questions out of seven questions from Part B and Three questions out of five questions from Part C.

Schematic diagrams must be shown wherever necessary. Any data you feel missing suitably be assumed and stated clearly. Units of quantities used/calculated must be stated clearly.

Use of following supporting material is permitted during examination. (Mentioned in form No. 205).

PART - A

(Answer should be given up to 25 words only)

All questions are compulsory.   (10×2=20)

1. What are the key features of Big Data?
2. Define the 3V's of Big Data with examples.
3. What is the Google File System (GFS) and its purpose?
4. Explain the role of NameNode and DataNode in Hadoop.
5. What is combiner in MapReduce framework?
6. Define Writable comparable in Hadoop I/O.
7. What are Hive clients? List different types of Hive clients.
8. What is Pig Latin? Mention its importance in Big Data processing.
9. What is the difference between Mapper and Reducer in MapReduce?
10. What is Record Reader in Hadoop?

PART - B

(Analytical / Problem Solving Questions)

Attempt any Five questions.   (5×4=20)

1. Explain the HDFS architecture with a neat diagram.
2. Describe the workflow of MapReduce program with an example dataset.
3. What is Writable class hierarchy in Hadoop? Explain its importance in Big Data processing.
4. Discuss the Pig architecture with its main components.

5. Explain how MapReduce handles data partitioning and the role of the partitioner.
6. Write a Hive QL query to create a table, insert data, and perform basic operations like filtering and grouping.
7. Describe the Hadoop I/O operations and explain how data serialization works.

PART - C
(Descriptive / Analytical / Problem Solving / Design Questions)
Attempt any Three questions.   **(3×10=30)**

1. Explain the complete building blocks of Hadoop. Draw the block diagram and explain the roles of NameNode, DataNode, Job Tracker, and Task Tracker.
2. What is MapReduce? Describe its working principle with a suitable example, including the Mapper, Reducer, and Combiner phases.
3. Write and explain a Pig Script to analyze a weather dataset. Perform operations like filtering, grouping, and sorting.
4. Explain Hive data types and querying techniques. Write Hive QL queries to create tables, load data, and perform basic data analysis operations.
5. Difference between the following:
   a) Hive and Pig.
   b) HDFS and traditional file systems.
   c) MapReduce and Traditional data processing techniques.