Reg. No.															

Question Paper Code	14246

M.E. / M.Tech. - DEGREE EXAMINATIONS, NOV / DEC 2025

First Semester

M.E - Big Data Analytics

24PBDPC103 - BIG DATA FRAMEWORK AND TECHNOLOGIES

Regulations - 2024

Duration: 3 Hours | Max. Marks: 100

PART - A (MCQ) (10 × 1 = 10 Marks)

Answer ALL Questions

		Marks	K – Level	CO
1.	NoSQL database is (a) MySQL (b) Oracle (c) MongoDB (d) MS Access	1	K1	CO1
2.	Replication in Big Data ensures: (a) Data loss (b) High availability (c) Reduced redundancy (d) Faster CPU performance	1	K1	CO1
3.	The two main components of YARN are: (a) NameNode and DataNode (b) ResourceManager and NodeManager (c) JobTracker and TaskTracker (d) Master and Worker	1	K1	CO2
4.	Sorting in MapReduce occurs: (a) Before the Map phase (b) Between Map and Reduce phases (c) After Reduce phase (d) During input splitting	1	K1	CO2
5.	The Pig shell used for interactive execution is called: (a) Hog (b) Piglet (c) Bacon (d) Grunt	1	K1	CO3
6.	Relational operator in Pig Latin? (a) LOAD (b) STORE (c) FILTER (d) CONNECT	1	K1	CO3
7.	Bucketing in Hive divides data: (a) Based on ranges (b) Into fixed number of files using hash functions (c) Based on partitions (d) Randomly	1	K1	CO4
8.	Spark supports which programming languages? (a) Java and Python only (b) Java, Python, Scala, and R (c) Only Scala (d) Python and SQL	1	K1	CO4
9.	The Impala Daemon (impalad) is responsible for: (a) Storing metadata (b) Executing queries on local data nodes (c) Scheduling jobs (d) File management	1	K1	CO5
10.	Impala executes queries using: (a) MapReduce framework (b) Its own distributed query engine (c) Spark DAG Scheduler (d) Tez engine	1	K1	CO5

PART - B (12 × 2 = 24 Marks)

Answer ALL Questions

		Marks	K – Level	CO
11.	Differentiate between a File System and a Distributed File System.	2	K2	CO1
12.	Classify different types of data used in Big Data Analytics.	2	K2	CO1
13.	Interpret the significance of Hadoop I/O operations.	2	K2	CO2
14.	Define NameNode and DataNode.	2	K1	CO2
15.	List the scalar data types supported by Pig.	2	K1	CO3
16.	Mention the role of the Pig Execution Engine.	2	K1	CO3
17.	Differentiate between Bucketing and Partitioning in Hive.	2	K2	CO4
18.	Define RDD in Spark.	2	K1	CO4
19.	Compare Impala with Hive in terms of query execution and latency.	2	K2	CO5
20.	How does Impala work with Hive?	2	K1	CO5

21.	State any two advantages of using Pig over traditional MapReduce.		*2*	*K1*	*CO3*
22.	Define pipelining in MapReduce.		*2*	*K1*	*CO2*

PART - C (6 × 11 = 66 Marks)

Answer ALL Questions

23.	a)	(i)	Design a simple Big Data architecture for an e-commerce company.	*6*	*K3*	*CO1*
		(ii)	Illustrae how Sharding and Replication together enhance the performance and reliability of Big Data systems.	*5*	*K2*	*CO1*
			OR			
	b)	(i)	Enumerate the need for Big Data frameworks Hadoop and Spark in handling modern data challenges.	*6*	*K3*	*CO1*
		(ii)	Outline the importance of each "V" in Big Data and explain how they together define the challenges of Big Data management.	*5*	*K2*	*CO1*
24.	a)		Discover a Hadoop-based system architecture for processing sensor data from IoT devices along with the flow from data input to final output using MapReduce.	*11*	*K4*	*CO2*
			OR			
	b)		Analyze the sorting and joining techniques used in MapReduce withtheir contribution to data analysis.	*11*	*K4*	*CO2*
25.	a)		Identify how Pig achieves parallel processing using Hadoop.	*11*	*K3*	*CO3*
			OR			
	b)		Develop a Pig script to load employee data, filter based on salary, and store the results in HDFS.	*11*	*K3*	*CO3*
26.	a)		Describe the Spark cluster design and management process.	*11*	*K2*	*CO4*
			OR			
	b)		Explain HiveQL with suitable examples for DDL, DML, and view creation.	*11*	*K2*	*CO4*
27.	a)		Analyze the process of creating, inserting, and overwriting data in Impala tables with commands.	*11*	*K4*	*CO5*
			OR			
	b)		Examine how Impala and Hive work together for data querying.	*11*	*K4*	*CO5*
28.	a)		Compare and contrast Hadoop and Spark in terms of architecture and performance.	*11*	*K3*	*CO4*
			OR			
	b)		Design a simple Spark application that reads a text file, counts the number of words, and saves the result using RDD transformations and actions.	*11*	*K3*	*CO4*