Reg. No.															

Question Paper Code	13627

B.E. / B.Tech./ M.Tech. - DEGREE EXAMINATIONS, APRIL / MAY 2025

Seventh Semester

Computer Science and Engineering

(Common to Eight Semester - M.Tech. - Computer Science and Engineering (5 Years Integrated))

20CSPC701 - BIG DATA ANALYTICS

Regulations - 2020

Duration: 3 Hours Max. Marks: 100

PART - A (MCQ) (10 × 1 = 10 Marks)

Answer ALL Questions

		Marks	K – Level	CO
1.	Which of the following tasks is most closely associated with the veracity characteristic of data in big data analytics? (a) Processing real-time streaming data from IoT devices (b) Cleaning and validating data to remove errors and inconsistencies (c) Extracting insights from unstructured text data (d) Integrating data from multiple sources into a unified platform	1	K1	CO1
2.	In the BASE model, what does "Soft State" mean? (a) The state of the system is always consistent (b) The state of the system may change over time (c) The system always stores data in a permanent state (d) The system avoids any state changes	1	K1	CO1
3.	Hadoop runs the jobs by dividing them into_____ (a) maps (b) individual files (c) tasks (d) None of the mentioned	1	K1	CO2
4.	Which of the following is a major Hadoop distributor providing enterprise solutions? (a) Apache Software Foundation (b) Hortonworks (c) ElasticSearch (d) MongoDB	1	K1	CO2
5.	Which command is used to execute a Pig script? (a) pig -x local script.pig (b) pig -run script.pig (c) pig -execute script.pig (d) pig -start script.pig	1	K1	CO3
6.	What is the purpose of serialization and deserialization in Hive? (a) To compress and decompress data (b) To convert data between formats for storage and retrieval (c) To encrypt and decrypt data (d) To partition and bucket data	1	K1	CO3
7.	What is Jaspersoft used for in conjunction with MongoDB? (a) Real-time data streaming (b) Business intelligence and reporting (c) Database administration (d) Data warehousing	1	K1	CO4
8.	What is required to connect a MongoDB client application to a MongoDB database? (a) Connection URL (b) API Key (c) ODBC Driver (d) SQL Query	1	K1	CO4
9.	Which command starts ZooKeeper in a typical Hadoop environment? (a) zkServer.sh start (b) zookeeper start (c) zkStart.sh (d) startZooKeeper.sh	1	K1	CO5
10.	What is the primary use case for Apache Flume? (a) Real-time data processing (b) Log collection and aggregation (c) SQL querying (d) Data visualization	1	K1	CO6

PART - B (12 × 2 = 24 Marks)

Answer ALL Questions

11.	How is traditional BI environment different from the big data environment?	2	K2	CO1
12.	Define predictive and prescriptive analytics.	2	K1	CO1
13.	Justify why Hadoop's HDFS is optimized for large files rather than small files.	2	K2	CO2
14.	State the role of YARN in the Hadoop ecosystem.	2	K1	CO2

15.	Define RCFile with an example.		*2*	*K1*	*CO3*
16.	Define pig and state the execution modes of pig.		*2*	*K1*	*CO3*
17.	Write the query for MongoImport.		*2*	*K2*	*CO4*
18.	Summarize the features of jasper report.		*2*	*K2*	*CO4*
19.	Compare HbaseVs RDBMS.		*2*	*K2*	*CO5*
20.	Define data model.		*2*	*K1*	*CO5*
21.	Summarize the advantages of NoSQL.		*2*	*K2*	*CO6*
22.	Why do we need ZooKeeper in the Hadoop?		*2*	*K1*	*CO6*

PART - C (6 × 11 = 66 Marks)

Answer ALL Questions

23.	a)	Explain the terminologies used in big data environment.	*11*	*K2*	*CO1*
		OR			
	b) (i)	Why big data analytics is important? Justify with example.	*6*	*K2*	*CO1*
	(ii)	Summarize the challenges of big data.	*5*	*K2*	*CO1*
24.	a)	Illustrate the key components of the HDFS architecture and their functions.	*11*	*K2*	*CO2*
		OR			
	b)	Discuss in detail about Mapper, Reducer, Combiner and Partitioner with example.	*11*	*K2*	*CO2*
25.	a)	Illustrate Queries for following relational operators in pig: FILTER FOREACH GROUP DISTINCT LIMIT	*11*	*K2*	*CO3*
		OR			
	b)	Demonstrate the concept of Partitioning and bucketing in hive with suitable example.	*11*	*K2*	*CO3*
26.	a)	Explain Map Reduce programming in MongoDB with a suitable example.	*11*	*K2*	*CO4*
		OR			
	b)	Illustrate queries and explain the Mongo DB Query language. 1.Insert 2.Save 3.Update 4.Remove 5.Find	*11*	*K2*	*CO4*
27.	a)	Explain in detail about Visual Data Analysis Techniques.	*11*	*K2*	*CO5*
		OR			
	b)	Explain in detail about HBase Architecture.	*11*	*K2*	*CO5*
28.	a)	Demonstrate the sqoop architecture and explain detail about import and export data.	*11*	*K2*	*CO6*
		OR			
	b)	What is lambda architecture? Explain the Lambda Architecture with Spark.	*11*	*K2*	*CO6*