Reg. No.																

Question Paper Code	13097

B.E. / B.Tech. - DEGREE EXAMINATIONS, NOV / DEC 2024

Seventh Semester

Computer Science and Engineering

20CSPC701 - BIG DATA ANALYTICS

Regulations - 2020

Duration: 3 Hours | Max. Marks: 100

PART - A (MCQ) (20 × 1 = 20 Marks)

Answer ALL Questions

		Marks	K – Level	CO
1.	Which one is not true about Traditional Business Intelligence (BI). (a) Faster and accurate reporting and analysis (b) Data is stored on a distributed file system (c) BI Solutions are more towards the structured data. (d) BI Solutions carry the data to the processing functions	1	K1	CO1
2.	What does the term "volume" refer to in the context of big data analytics? (a) The accuracy and reliability of data (b) The diverse types of data, including structured and unstructured data (c) The size or amount of data being processed (d) The speed at which data is generated and processed	1	K1	CO1
3.	Find one advantage of using Hadoop over traditional Data Warehouses. (a) Hadoop is designed for smaller data sets (b) Hadoop supports both structured and unstructured data (c) Hadoop requires higher computational resources (d) Data Warehouses are open-source	1	K1	CO1
4.	Hadoop Distributed File System provides (a) performance through distribution of data (b) fault tolerance through replication (c) both performance through distribution of data & fault tolerance through replication (d) hierarchical configuration	1	K1	CO2
5.	What is the default replication factor for HDFS files? (a) 3 (b) 2 (c)1 (d) 4	1	K1	CO2
6.	Which component is part of the Hadoop ecosystem and is used for SQL-like querying on Hadoop data? (a) Apache Hive (b) Apache HBase (c) Apache Spark (d) Apache Flume	1	K1	CO2
7.	Which component of Hive translates HQL queries into MapReduce jobs? (a) Hive Metastore (b) Hive Driver (c) Hive Compiler (d) Hive Execution Engine	1	K1	CO3
8.	Which of the following is a primitive data type in Hive? (a) Array (b) Map (c) String (d) Struct	1	K1	CO3
9.	What is Pig Latin? (a) A programming language for machine learning (b) A query language for data processing in Pig (c) A language for system administration (d) A visualization tool for data	1	K1	CO3
10.	Which method is used to query and retrieve documents from MongoDB? (a) get() (b) find() (c) search() (d) fetch()	1	K1	CO4
11.	How does MongoDB handle NULL values in documents? (a) They are automatically removed (b) NULL values are stored as empty fields (c) NULL values are not supported (d) NULL values are replaced with default values	1	K1	CO4

12.	Which protocol is used to connect applications to MongoDB? (a) HTTP (b) JDBC (c) MongoDB Wire Protocol (d) FTP	*1*	*K1*	*CO4*
13.	Which feature of HBase provides the ability to store and retrieve large amounts of data across a distributed system? (a) Column-Oriented Storage (b) Indexing (c) Join Operations (d) Normalization	*1*	*K1*	*CO5*
14.	Which characteristic differentiates HBase from traditional RDBMS? (a) Strict ACID compliance (b) Schema-less data model (c) Support for SQL queries (d) Fixed schema structure	*1*	*K1*	*CO5*
15.	What type of workflow does Oozie support for defining complex processing pipelines? (a) Simple Workflow (b) Complex Workflow (c) Coordinated Workflow (d) Directed Acyclic Graph (DAG)	*1*	*K1*	*CO5*
16.	Which Oozie component manages the scheduling of workflows? (a) Coordinator (b) Workflow Engine (c) Bundle (d) Action	*1*	*K1*	*CO5*
17.	Which Flume component is used to collect data from Twitter streams? (a) Source (b) Sink (c) Channel (d) Collector	*1*	*K1*	*CO6*
18.	Which component of Sqoop is responsible for transferring data between Hadoop and relational databases? (a) Sqoop Client (b) Sqoop Server (c) Sqoop Connector (d) Sqoop Importer	*1*	*K1*	*CO6*
19.	What distinguishes NewSQL databases from traditional SQL and NoSQL databases? (a) They use NoSQL data models for scalability. (b) They provide SQL-like querying with horizontal scalability and high performance. (c) They do not support ACID transactions. (d) They are designed solely for key-value storage.	*1*	*K1*	*CO6*
20.	Which of the following is NOT a characteristic of NoSQL databases compared to SQL databases? (a) Schema-less data storage (b) Support for complex joins (c) Horizontal scalability (d) Flexible data models	*1*	*K1*	*CO6*

PART - B (10 × 2 = 20 Marks)

Answer ALL Questions

21.	Define big data.	*2*	*K1*	*CO1*
22.	State about CAP theorem.	*2*	*K1*	*CO1*
23.	Compare replication and sharding.	*2*	*K2*	*CO2*
24.	Summarize the function of Mapper and Reducer.	*2*	*K2*	*CO2*
25.	Define Partitioning.	*2*	*K1*	*CO3*
26.	Outline the features of Pig Latin Statements.	*2*	*K2*	*CO3*
27.	Infer the features of MongoDB.	*2*	*K2*	*CO4*
28.	Summarize the features of jasper report.	*2*	*K2*	*CO4*
29.	What is HBase?	*2*	*K1*	*CO5*
30.	Why do we need ZooKeeper in the Hadoop?	*2*	*K1*	*CO6*

PART - C (6 × 10 = 60 Marks)

Answer ALL Questions

31.	a)	Explain the classification of data in detail.	*10*	*K2*	*CO1*
		OR			
	b) i)	Show the responsibilities of Data Scientist.	*5*	*K2*	*CO1*
	ii)	Discuss in detail about Soft state eventual consistency.	*5*	*K2*	*CO1*
32.	a)	Explain in detail about the Map reduce programming architecture with an example.	*10*	*K2*	*CO2*

OR

	b)	Illustrate the major blocks in HDFS architecture.	*10*	*K2*	*CO2*
33.	a)	Describe in detail about HIVE architecture with neat diagram.	*10*	*K2*	*CO3*
		OR			
	b)	Summarize Queries for following relational operators in pig: (i) FILTER (ii) FOREACH (iii) GROUP (iv) DISTINCT (v) LIMIT	*10*	*K2*	*CO3*
34.	a)	Construct queries and explain the following using MongoDB. (i) Insert (ii) Save (iii) Update (iv) Remove (v) Find	*10*	*K3*	*CO4*
		OR			
	b)	Identify the process of connecting with MongoDB NoSQL database using Jasper soft.	*10*	*K3*	*CO4*
35.	a)	Discuss in detail about HBase Architecture.	*10*	*K2*	*CO5*
		OR			
	b)	Summarize about different types of data visualization Techniques.	*10*	*K2*	*CO5*
36.	a)	Build workflow architecture of Oozie for a student data processing application and also explain how does it work?	*10*	*K3*	*CO6*
		OR			
	b)	Using sqoop architecture, explain in detail about data import and export operation.	*10*	*K3*	*CO6*