B.TECH DEGREE EXAMINATIONS: NOV/DEC 2024

(Regulation 2018)

Sixth Semester

INFORMATION TECHNOLOGY

U18ITI6304: Big Data Analytics

COURSE OUTCOMES

CO1: Outline the big data technologies used for storage, analysis and manipulation of data

CO2: Explain Big Data eco system and its components

CO3: Analyze the Big Data stored in HDFS using Hadoop Map Reduce framework

CO4: Understand the Pig scripting and HBase architecture

CO5: Apply the Hive concepts, Hive Data types, loading and querying for Big Data

CO6: Explain the MongoDB architecture and its operations

Time: Three Hours   **Maximum Marks: 100**

Answer all the Questions:-
PART A (10 x 2 = 20 Marks)
(Answer not more than 40 words)

1. List out the best practices of Big Data Analytics	CO1	[K$_2$]
2. Write down the four computing resources of Big Data Storage.	CO1	[K$_2$]
3. Discuss about the map reduce programming model.	CO2	[K$_2$]
4. Suppose there is a file of size 518 MB stored in HDFS (Hadoop 2. x) using default block size configuration and default replication factor. How many blocks will be created in total and what will be the size of each block?	CO2	[K$_3$]
5. Compare Name node and Data node in the HDFS platform.	CO3	[K$_2$]
6. Illustrate the limitation of MapReduce overcome in a future version of Hadoop.	CO3	[K$_2$]
7. Difference Between Hbase and Hive.	CO4	[K$_2$]
8. Explain the use of the FILTER operator in Apache Pig.	CO4	[K$_2$]
9. For the given input data John 28 Alice 35 Bob 22 Maria 40.Use the Filter command in pig to select rows where age is greater than 30	CO5	[K$_3$]
10. Recall the purpose of sharding, which is to distribute large datasets across multiple servers for horizontal scaling.	CO6	[K$_2$]

Answer any FIVE Questions:-
PART B (5 x 16 = 80 Marks)
(Answer not more than 400 words)

11.	a)	Describe the role of Big Data Analytics in transforming business strategies and illustrate its impact with a relevant real-world example	8	CO1	[K$_2$]
	b)	Discuss about the Hadoop ecosystem with example.	8	CO1	[K$_2$]
12.	a)	Describe and illustrate the architecture and operations of HDFS, including a detailed explanation of its components and functionalities, supported by a diagram to demonstrate its structure and workflow.	8	CO2	[K$_2$]
	b)	Explain the working of the MapReduce framework, detailing its data flow and task execution with a clear diagram. Using a text file (sample.txt) containing the words Banana, Mango, Banana, Apple, Mango, Orange, Mango, Banana, perform a word count operation and illustrate the process step by step	8	CO2	[K$_3$]
13.	a)	Summarize the necessity of HDFS federation for efficient block management and its impact on system scalability and performance	8	CO3	[K$_2$]
	b)	Describe single node Hadoop cluster and installation along with the Hadoop configuration files.	8	CO3	[K$_2$]
14.	a)	Discuss the architecture and components of HBase, including an illustrative diagram to demonstrate its structure and functionality	8	CO4	[K$_2$]
	b)	Elaborate the two modes used for running the Pig scripts.	8	CO4	[K$_2$]
15.	a)	i)Write the Hive command to create a table with columns: Product ID, Product Name, Price, Category, and Stock.	2	CO5	[K$_3$]
		ii) Assume the 'products' table is as given below with fields: Product_ID, Product_Name, Price, Category, and Stock. • Generate a query to retrieve all product details ordered by Category. iii) Generate a query to retrieve the number of products in each category. iv) Generate a query to retrieve product details where the Category is 'Electronics' and the Price is greater than 5000.	6		
	b)	Outline the architecture of Hive, including its key components and their roles.	8	CO5	[K$_2$]

Illustrate the architecture with an example to demonstrate how it processes queries and manages data

16. a) Describe the sharding mechanism in MongoDB, detailing its architecture, data distribution, and configuration steps, with appropriate diagrams and real-world examples   8   CO6   [K$_2$]

    b) Explain the CRUD basic operations of MongoDB with relevant examples.   8   CO6   [K$_2$]

****