Register Number:……..

KUMARAGURU
college of technology
character is life

B.TECH DEGREE EXAMINATIONS: APRIL /MAY 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.	Write down the four computing resources of Big Data Storage.	CO1	[K_1]
2.	How can be big data analytics used to improve business sales for an online bookstore Justify your answer. Mention some of the big data applications.	CO1	[K_1]
3.	Outline the need for YARN in the Hadoop ecosystem.	CO2	[K_2]
4.	Difference between Hadoop 2.X and Hadoop 3.X. Suppose there is file of size 1028 MB stored in HDFS (Hadoop 3.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?	CO3	[K_3]
5.	List out the five basic operations of the Map reduce programming Model.	CO3	[K_1]
6.	List out the features of HBase.	CO4	[K_1]
7.	Describe the usage or operation of the LOAD keyword in Pig.	CO4	[K_2]
8.	Difference between Hive and HBase.	CO5	[K_2]
9.	Explain the “Sharding” operation in MongoDB.	CO6	[K_2]
10.	Illustrate the need for object identifiers in MongoDB.	CO6	[K_2]

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

11.	a)	What are the advantages of Hadoop? Explain Hadoop Architecture and its Components with a proper diagram.	8	CO1	[K_2]
	b)	Explain the seven Vs of Big Data in detail. Briefly discuss the types of data handled by Hadoop.	8	CO1	[K_2]
12.	a)	Explain in detail about HDFS with an example.	8	CO2	[K_2]
	b)	Brief about Map Reduce data flow with map-reduce tasks with neat diagram. We have a text file with data Orange, Apple, grapes, Orange, Orange, Pineapple, Pear, Jackfruit, and Pear. Operate on a word count on the sample.txt using Map Reduce.	8	CO2	[K_3]
13.	a)	Briefly explain HDFS Name node federation Gateway snapshots, checkups, and backups.	10	CO3	[K_2]
	b)	Describe the need and operations of the block awareness algorithm in HDFS operation.	6	CO3	[K_2]
14.	a)	Narrate the operations and analyze how many tweets are stored per user, in the given tweet tables for the below scenario. (Using Pig) Twitter had both semi-structured data like Twitter Apache logs, Twitter search logs, Twitter MySQL query logs, application logs and structured data like tweets, users, block notifications, phones, favorites, saved searches, re-tweets, authentications, SMS usage, user followings, etc. which can be easily processed by Apache Pig. Twitter dumps all its archived data on HDFS. It has two tables i.e. user data and tweets data. User data contains information about the users like username, followers, followings, number of tweets etc. While Tweet data contains tweet, its owner, number of re-tweets, number of likes etc. Now, twitter uses this data to analyze their customer's behaviors and improve their past experiences.	12	CO4	[K_3]
	b)	Summarize the complex and scalar data types in pig with relevant examples.	4	CO4	[K_2]

15.	a)	i) Write the Hive command to create a table with columns: Employee id, Name, Salary, department, and Designation.	4	CO5	[K_3]
		ii) Assume the employee table as given below, with the fields named Id, Name, Salary, Designation, and Dept. -Generate a query to retrieve the employee details in order by using the Department name. -Generate a query to retrieve the number of employees in each department. -A query is used to retrieve employee details whose department is TP and Salary is more than Rs 40000.	6		
	b)	Outline how a Zookeeper helps in monitoring a cluster.	6	CO5	[K_2]
16.	a)	Explain the basic CRUD operations of MongoDB with relevant examples	10	CO6	[K_2]
	b)	Illustrate the integration of MongoDB with Hadoop with relevant examples.	6	CO6	[K_2]

****