SEMESTER-UG 4th

MID-SEMESTER EXAMINATION, FEB, 2024

Course Code- **CDCSC11**

Course Title- **Big Data Analytics**

Time- 1.5 Hours
Max. Marks- 15

Note: - Attempt all questions. Missing data/ information (if any), may be suitably assumed & mentioned in the answer.

Q. No.	Question	Marks	CO
1a	Discuss the integration of cloud computing with big data and its advantages.	1.5	2
1b	What are the challenges associated with inter and trans firewall analytics?	1.5	
2a	How does Hadoop enable parallel processing, and what is the significance of data locality?	2	1
2b	Write a pseudocode for implementing word-count program using Map-reduce	1	
3a	Provide a detailed comparison of different types of NoSQL databases	2	2
3b	Provide examples of scenarios where materialized views in NoSQL databases prove beneficial.	1	
4a	How can you insert, update, and delete documents in a MongoDB collection?	2	3
4b	How do aggregate data models address the challenges of handling large datasets in NoSQL databases?	1	
5a	Explain HDFS architecture in detail. How HDFS ensures fault tolerance?	2	2
5b	How does the MapReduce programming model enable parallel processing?	1	

SEMESTER-UG 4th

END-SEMESTER EXAMINATION, APRIL MAY 2024

Course Code- **CDCSC11**

Course Title- **Big Data Analytics**

Time- 3 Hours

Max. Marks- 40

Note: - Attempt all questions in the given order only. Missing data/information (if any), maybe suitably assumed & mentioned in the answer.

Q. No.	Question	Marks	CO
Q1	**Attempt any two parts of the following**		
1a	What drives organizations to transition from traditional data warehouse tools to smarter data hubs rooted in Hadoop ecosystems? Furthermore, what differentiates the regular FileSystem from HDFS?	4	
1b	What is Crowdsourcing and how it helps in big data applications. Explain the role of big data analytics and crowdsourcing in following applications: • Credit Card Risk prediction • Healthcare and Medicine	4	
1c	What are some key trends contributing to the generation and management of unstructured data? How can big data analytics be applied in fraud detection and prevention across industries such as banking and insurance?	4	
Q2	**Attempt any two parts of the following**		
2a	Why Hadoop Explain various Hadoop daemons and their roles in a Hadoop cluster. Write Hadoop commands for following: • Creating a file in HDFS • Copying a file from local file system to HDFS	4	
2b	Does MongoDB support ACID transaction management and locking functionalities? How do I perform the SQL JOIN equivalent in MongoDB?	4	
2c	Where are the two types of metadata that NameNode server stores? If a file of 1 GB has to be written on HDFS, Explain with diagram, how it will be stored in datanodes.	4	
Q3	**Attempt any two parts of the following**		
3a	How do you select among the Different File Formats for Storing and Processing Data? What is Avro Serialization in Hadoop?	4	
3b	State the reason why we can't perform "aggregation" (addition) in mapper? Why do we need the "reducer" for this?	4	

	You have a file that contains 200 billion URLs. How will you find the first unique URL using Hadoop MapReduce?		
3c	Explain the architecture of YARN and how it allocates various resources to applications?	4	
Q4	**Attempt any two parts of the following**		
4a	What is the role of MasterServer and Zookeeper in HBase. What is lazy evaluation in Spark?	4	
4b	What is the difference between logical and physical plans? Explain COGROUP in Pig with example?	4	
4c	Why does Hive not store metadata information in HDFS? Write command to create a table student in Hive and create partitions based on semester and Elective: Student(Name, Roll No., Elective, semester, marks)	4	
Q5	**Attempt any two parts of the following**		
5a	Define Multivariate Analysis and how is it used in big data analytics. Fit a simple linear regression model to predict the price of a house based on its size. Size (in sq. feet): 1000, 1500, 2000, 2500, 3000 *Price (in $1000s)*: 300, 450, 600, 750, 900 Using the regression model, predict the price of a house with a size of 1800 sq. feet.	4	
5b	Explain how a weather prediction problem can be solved using fuzzy logic. Given U={1,2,3,4,5,6,7} A={(3,0.7),(5,1),(6,0.8)} What will be the value of ~A (where ~ -> complement)	4	
	Explain Bayesian Networks with example. Below, given the frequency table, what if the probability, that players will play if it is sunny weather. Frequency Table Weather | No | Yes Overcast | 1 | 4 Rainy | 3 | 2 Sunny | 2 | 3 Total | 5 | 9	4	