A **1200ADT302012401** **Pages: 2**

Reg No.:_____ Name:_____

APJ ABDUL KALAM TECHNOLOGICAL UNIVERSITY

Sixth Semester B.Tech Degree (R, S) Examination May 2024 (2019 Scheme)

Course Code: ADT302

Course Name: CONCEPTS IN BIG DATA ANALYTICS

Max. Marks: 100 Duration: 3 Hours

PART A

Answer all questions, each carries 3 marks. Marks

1		Explain the differences between data analysis and data reporting.	(3)
2		What is big data platform? What are its features?	(3)
3		What are the challenges in stream processing?	(3)
4		Describe estimating moments and its types.	(3)
5		Compare name node and data node in HDFS.	(3)
6		Describe the Map Reduce job implementation in the case of Road Enrichment Example.	(3)
7		State the differences between Pig and SQL.	(3)
8		What is Hbase? What are its features?	(3)
9		Illustrate any three R functions used in data analytics.	(3)
10		Write an R program to print Fibonacci series.	(3)

PART B

Answer one question from each module, each carries 14 marks.

Module I

11	a)	Explain big data architecture in detail with the help of a neat diagram.	(7)
	b)	What are the different phases in big data analytics life cycle? Explain in detail.	(7)

OR

12	a)	Explain the Evolution of Big Data and their characteristics. What are the different types of big data?	(7)
	b)	Explain in detail about Nature of Data and its applications. What are the different types of data.	(7)

Module II

13	a)	Illustrate working of Bloom filter with examples. Explain in detail the operations that a Bloom filter supports.	(7)

b) Find the number of distinct elements in the given input stream 1,3,2,1,2,3,4,3,1,2,3,1. The hash function is h(x)=6x+1 mod 5. Explain in detail the algorithm used to count distinct elements in the stream. (7)

OR

14 a) What are the different sampling techniques for efficient stream processing? Explain in detail. (7)

b) Illustrate DGIM algorithm in detail with example. Explain the rules for forming the buckets. (7)

Module III

15 a) Illustrate map reduce job execution flow. (7)

b) Explain in detail the anatomy of file write operation in HDFS with neat diagram. (7)

OR

16 a) Explain how map reduce can be used as a framework for parallel processing. (7)

b) What is Hadoop ecosystem? Explain in detail its components with neat diagram. (7)

Module IV

17 a) Explain Pig architecture in detail. What are its components? (7)

b) Illustrate features and Architecture of Hive with neat diagram. (7)

OR

18 a) Write the syntax to create a table and partition in Hive. (7)

b) What are the different data types used in Pig Latin and describe its various operators. (7)

Module V

19 a) Describe about vectors in R. What are the operations that can be performed on vectors? (7)

b) Explain the different categories of attributes and data types in R. (7)

OR

20 a) What are lists in R? How lists are handled? (7)

b) Describe data frames in R. How it can be created and explain the various operations on frames. (7)

****