A
00G-22T27

1200ADT302122501

Pages: 2

Reg No.:_____ Name:_____

APJ ABDUL KALAM TECHNOLOGICAL UNIVERSITY

B.Tech Degree S6 (R,S) Examinations April 2026 (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		Identify the five key characteristics (Vs) that define Big Data.	(3)
2		Explain the concept of a Big Data Platform along with its key characteristics and examples.	(3)
3		Illustrate the stream data architecture.	(3)
4		Describe three major benefits of using the DGIM algorithm.	(3)
5		Explain the core components of a Hadoop ecosystem.	(3)
6		Explain the functions of data node and name node.	(3)
7		Describe the different ways Pig scripts can be executed.	(3)
8		Describe the communication process between the driver and clients within Hive architecture.	(3)
9		Explain the features of R programming.	(3)
10		Write an R program that reads a number from the keyboard and finds its factorial.	(3)

PART B

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

Module I

11	a)	With the help of a neat diagram explain big data architecture in detail.	(7)
	b)	Differentiate conventional data and big data.	(7)

OR

12	a)	Describe the evolution of Big Data and its key characteristics.	(7)
	b)	Explain the important features offered by an integrated IT system for handling Big Data.	(7)

Module II

13	a)	Explain DGIM algorithm with an example.	(7)
	b)	Find the number of distinct elements in the given input stream	(7)

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.

OR

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

b) Explain Sampling techniques for efficient stream processing. (7)

Module III

15 a) Write about the anatomy of file read operation in HDFS. (7)

b) How wordcount problem can be implemented using MapReduce model? (7)

OR

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

b) Describe Hadoop architecture with neat sketch. (7)

Module IV

17 a) Explain the Architecture of Hive. (8)

b) List out the different data processing operators in the pig. (6)

OR

18 a) Describe HBase along with its important features and provide an overview of its data model. (7)

b) Describe the different types of operators used for data processing in Pig Latin. (7)

Module V

19 a) Explain the categories of data types and attributes available in R programming. (8)

b) List and explain four R functions used in descriptive statistics. (6)

OR

20 a) Write an R program to perform basic matrix handling operations. The program should: (8)

1. Create two matrices, each of size 2×2.
2. Display both matrices.
3. Perform and display the results of the following operations:
   - Matrix Addition
   - Matrix Multiplication
   - Transpose of a Matrix

b) Illustrate the operations performed in R vectors. (6)

****