

Reg No.: _____

Name: _____

APJ ABDUL KALAM TECHNOLOGICAL UNIVERSITY

B.Tech Degree S6 (R,S) Exam April 2025 (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 | Differentiate between Big Data and Conventional Data. | (3) |
| 2 | Explain different types of Big Data. | (3) |
| 3 | Illustrate the stream data architecture. | (3) |
| 4 | Why is conventional data processing insufficient for stream processing? | (3) |
| 5 | Describe Hadoop ecosystem. | (3) |
| 6 | How map-reduce performs parallel processing. | (3) |
| 7 | Identify the ways in which a pig program can be executed. | (3) |
| 8 | Illustrate the interaction of driver with different clients in Hive architecture. | (3) |
| 9 | Write an R program to combine two lists. | (3) |
| 10 | List and explain three R functions used in descriptive statistics. | (3) |

PART B*Answer one question from each module, each carries 14 marks.***Module I**

- | | | |
|----|---|-----|
| 11 | a) What are different phases of the Data Analytics Lifecycle? Explain each in detail. | (7) |
| | b) What are the 5 Vs of Big Data? | (7) |

OR

- | | | |
|----|---|-----|
| 12 | a) Illustrate Big Data Architecture. | (7) |
| | b) Give examples of different Big Data platforms. | (7) |

Module II

- | | | |
|----|---|-----|
| 13 | a) Explain DGIM algorithm with an example. | (8) |
| | b) Suppose our stream consists of the integers 3,1,4,1,5,9,2,6,5. Hash functions are,
1. $2x+1 \bmod 32$ | (6) |

2. $3x+7 \bmod 32$

Treat the result as a 5-bit binary integer and estimate the number of distinct elements in the string.

OR

- 14 a) How to use Flajolet- Martin algorithm to count the distinct elements in a stream. (9)
 b) Calculate the second moment for the stream a,b,c,b,d,a,c,d,a,b,d,c,a,a,b. (5)

Module III

- 15 a) How word count problem can be implemented using MapReduce model? (8)
 b) Discuss the file read and write operations in HDFS. (6)

OR

- 16 a) Explain the Map reduce execution pipeline. (7)
 b) Differentiate Hadoop with RDBMS. (7)

Module IV

- 17 a) Write the syntax to create a table and partition in Hive. (8)
 b) Describe the Data Types and File Formats in Hive. (6)

OR

- 18 a) List out the different data processing operators in the pig. (7)
 b) Illustrate the process of writing and using User-defined functions in Hive. (7)

Module V

- 19 a) Illustrate the operations performed in R vectors. (7)
 b) List the applications of R programming. (7)

OR

- 20 a) Write an R program to find the factorial of a number by reading the input from a file. (7)
 b) Write a program in R to read a dataset 'internals.csv' which consists of id, name and marks of 5 subjects each out of 50. Write code (7)
 i). To Read data frame and display columns
 ii). To add an extra column Total
 iii). To display class topper
 iv). To display student details having total marks above the average
 v). To prepare rank list by sorting rows based on total marks
