Reg No.:_____ Name:_____

**APJ ABDUL KALAM TECHNOLOGICAL UNIVERSITY**

Sixth Semester B.Tech Degree Regular and Supplementary Examination June 2023 (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 | | What is data? Explain about nature of data. | (3) |
| 2 | | Give examples of different Big Data platforms. | (3) |
| 3 | | What are the key components of streaming data architecture? | (3) |
| 4 | | Is conventional data processing sufficient for stream processing. Why? | (3) |
| 5 | | Why do we need Hadoop for Big Data Analytics? | (3) |
| 6 | | Give examples of Hadoop file systems. | (3) |
| 7 | | What are the different execution modes of Pig? | (3) |
| 8 | | Differentiate Hive from traditional databases. | (3) |
| 9 | | What are the important features of R programming? | (3) |
| 10 | | Write an R program to find the factorial of a number by reading the input from keyboard. | (3) |

**PART B**

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

**Module I**

| | | | |
|---|---|---|---|
| 11 | a) | Describe the various big data analytical methods in detail. | (7) |
| | b) | Explain Intelligent Data Analysis in detail. | (7) |

**OR**

| | | | |
|---|---|---|---|
| 12 | a) | Define the term "Big data". Which are the different 5 V's that help to decide whether a given data source contributes to big data. | (7) |
| | b) | Explain in detail about data analytic processes and tools. | (7) |

**Module II**

| | | | |
|---|---|---|---|
| 13 | a) | What is Bloom filter and what are its properties. Illustrate the working of Bloom filter with examples. | (8) |
| | b) | Illustrate Flajolet- Martin algorithm. | (6) |

**OR**

14 a) Explain in detail the algorithm used for counting the number of ones in a window. Consider there is a window of length N=24 on a binary system. Find the total number of ones in the last 20 bits, when 0111 enters into the given stream ....10101100010111010010110.... The new bit enters from the right side and time stamp of first new bit is 100. (9)

b) Explain the term moments of the stream, and list out its importance. (5)

**Module III**

15 a) How wordcount problem can be implemented using MapReduce model? (10)

b) Write about the anatomy of file read operation in HDFS. (4)

**OR**

16 a) Explain map reduce programming model. (7)

b) Which are the core components of Hadoop? Explain in detail. (7)

**Module IV**

17 a) Explain the main components of the Hadoop Pig framework (7)

b) Write about different data processing operators in Pig Latin. (7)

**OR**

18 a) What are the different types of Tables in Hive? Explain. (5)

b) Explain the data model and Read -Write operations in HBase? (9)

**Module V**

19 a) 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 (10)

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

b) Discuss how reading and writing files is handled in R. (4)

**OR**

20 a) Explain Matrix creation and Matrix handling in R. Discuss about applying functions to Matrix rows and columns. (8)

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

****