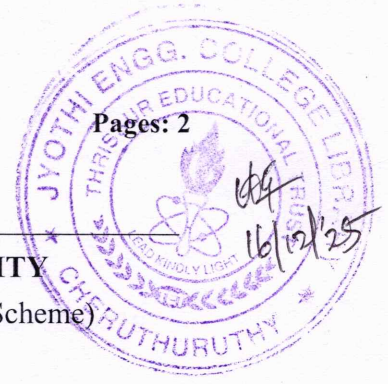


Reg No.: _____

Name: _____

APJ ABDUL KALAM TECHNOLOGICAL UNIVERSITY
B.Tech Degree S6 (S,FE) Examination December 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 | Identify the differences between data analysis and data reporting | (3) |
| 2 | Compare conventional Data and Big Data | (3) |
| 3 | Discuss the issues in stream processing. | (3) |
| 4 | How do moments play a significant role in statistical data analysis? | (3) |
| 5 | Compare the specific file types of HDFS. | (3) |
| 6 | List the different Hadoop commands for basic file system operations.. | (3) |
| 7 | Compare Hbase and RDBMS | (3) |
| 8 | Compare Pig with Databases | (3) |
| 9 | Explain the different categories of attributes and data types in R | (3) |
| 10 | What are the key features of R programming for data analytics? | (3) |

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

- | | | |
|----|---|-----|
| 11 | a) How are big data and Hadoop related to each other? | (7) |
| | b) Summarize Big data analytics life cycle. | (7) |

OR

- | | | |
|----|---|-----|
| 12 | a) What is meant by Nature of data | (7) |
| | b) Describe the role of Intelligent Data Analysis in handling Big Data? | (7) |

Module II

- | | | |
|----|--|-----|
| 13 | a) Some websites check availability of username by searching millions of usernames registered with it. Identify one effective method to filter data as in this type of scenario. | (7) |
| | b) Explain DGIM algorithm with its use cases | (7) |

OR

- 14 a) Explain Flajolet-Martin algorithm with an example (10)
- b) How does the stream data model impact real-time data processing and analytics? (4)

Module III

- 15 a) Write about the anatomy of file read and write operation in HDFS. (7)
- b) Which are the core components of Hadoop? Explain in detail. (7)

OR

- 16 a) Describe the Map Reduce job implementation in the case of Road Enrichment example (7)
- b) Explain the types of Schedulers available in YARN (7)

Module IV

- 17 a) Explain two execution types or modes in PIG (7)
- b) What is HBase and what are its features? Explain HBase Data Model. (7)

OR

- 18 a) What are the different types of Tables in Hive? Explain. Write the syntax to create a table and partition in Hive. (7)
- b) Explain Pig architecture in detail. What are its components? (7)

Module V

- 19 a) **Write an R program to read a dataset 'sales_data.csv' containing columns: Emp_ID, Name, Region, Sales, and Commission. Perform the following tasks** (9)
- i) Delete the "**Region**" column.
- ii) Update the "**Sales**" column by applying a 15% increment.
- iii) Print the details of the **top-performing salesperson (highest sales)**.
- iv) Sort the dataset based on **Sales in descending order**.
- v) Save the updated dataset as '**sales-out.csv**'.
- b) Write about list data structure and its use cases (5)

OR

- 20 a) Write about how filtering operations are performed in data frame. Create a filter to find the missing values in a data frame (7)
- b) Write a R program to create an array, passing in a vector of values and a vector of dimensions. Also provide names for each dimension. (7)
