KAMARAJ COLLEGE (Autonomous)

Accredited with A+ Grade by NAAC

(Affiliated to Manonmaniam Sundaranar University, Tirunelveli)

(3 Pages) **Reg. No:……**

Question Code: 26E03206 **Course Code : 25PMAI22**

PG Degree - End Semester Examinations, April 2026

Second Semester

M.Sc., COMPUTER SCIENCE WITH AI

Big Data Analytics

(For those who joined in June 2025 onwards)

Time : 3Hours **Maximum : 75 Marks**

PART – A (10 × 1 = 10 Marks)

Answer ALL Questions

Choose the correct answer :

CO:1 K:1 1. What is the term used for a collection of large, complex data sets that cannot be processed using traditional data processing tools?

(a) Big Data (b) Small Data

(c) Medium Data (d) Mini Data

CO:1 K:2 2. The total forms of big data is _____

(a) 1 (b) 2

(c) 3 (d) 4

CO:2 K:2 3. NoSQL databases are most often referred to as _____

(a) Relational (b) Distributed

(c) Object-oriented (d) Network

CO:2 K:1 4. What was Hadoop written in?

(a) C (b) C++

(c) JAVA (d) JSP

CO:3 K:2 5. MongoDB achieves replication by the use of _____

(a) Replica set (b) Data set

(c) $first (d) Collections

CO:3 K:2 6. _____ part of the MapReduce is responsible for processing one or more chunks of data and producing the output results.

(a) Maptask (b) Mapper

(c) Reducer (d) Combiner

CO:4 7. Hive is primarily used for _____

K:2

(a) Online Transaction Processing (b) Real-time queries

(c) Batch processing (d) Distributed processing

CO:4 8. We can run Pig in batch mode using _____

K:1

(a) Pig shell command (b) Pig scripts

(c) Pig options (d) Pig Local

CO:5 9. Which function is used to print output to the console in R?

K:1

(a) echo() (b) print()

(c) display() (d) output()

CO:5 10. Which package is commonly used for data manipulation in R?

K:1

(a) ggplot2 (b) shiny

(c) dplyr (d) caret

PART – B (5 X 5 = 25 Marks)

Answer ALL Questions choosing either (a) or (b).

Answer should not exceed 250 words.

CO:1 11. (a) List the Characteristics of Big data.

K:4

(OR)

(b) Examine the responsibilities of a Data scientist in Big data.

CO:2 12. (a) Analyze the anatomy of file Read in HDFS.

K:4

(OR)

(b) Examine the limitations of Hadoop 1.0 architecture.

CO:3 13. (a) Identify the function of Insert method () in Mongo DB Query Language.

K:3

(OR)

(b) Make use of Combiner and Reducer in Map reduce.

CO:4 14. (a) Classify the file formats in Hive.

K:4

(OR)

(b) Inspect the complex data types in Pig.

CO:5 15. (a) List the data types in R.

K:4

(OR)

(b) Analyze the various function in Big Data visualization.

PART – C (5 X 8 = 40 Marks)

Answer ALL Questions choosing either (a) or (b).

Answer should not exceed 600 words.

CO:1 16. (a) Classify the digital data in Big data Analytics.

K:4

(OR)

(b) Examine the CAP theorem in Big Data environments.

CO:2 17. (a) Compare SQL with No SQL.

K:4

(OR)

(b) Analyze how does Map reduce work in Hadoop.

CO:3 18. (a) Interpret the data types in Mongo DB.

K:5

(OR)

(b) Explain about secondary sorting in Map reduce with an example.

CO:4 19. (a) Elaborate the architecture of Hive.

K:6

(OR)

(b) Compile the Eval functions of Pig.

CO:5 20. (a) Discuss the features and strengths of R in analytics.

K:6

(OR)

(b) Predict the libraries and tools for Big Data in R.