

67101-N

M.C.A. (MCA 2 Year Programme)  
w.e.f. 2021-2022, Examination, December-2022

DATA MINING AND BIG DATA ANALYSIS

Paper-21MCA23C1

Time allowed : 3 hours] [Maximum marks : 80

*Note: Question No. 1 is compulsory. In addition to Question No. 1, attempt four more questions, by selecting at least one question from each Unit. All questions carry equal marks.*

1. (a) What do you understand by data characterization as data mining functionality?
- (b) Describe the steps involved in data mining when viewed as a process of knowledge discovery.
- (c) Formulate the Information Gain used as attribute selection method in classification.
- (d) What do you understand by unsupervised learning?
- (e) What do you mean by variability and value characteristics of Big data?
- (f) Differentiate between fsimage and editlog maintained by Namenode.

67101-N-P-4-Q-9 (22)

[P.T.O.]

7. (a) Why and how Hadoop HDFS is called as Fault Tolerant System? List down atleast five measures taken for this purpose.
- (b) Which type of scaling is preferred in Hadoop system? Justify your answer.

#### Unit-IV

8. (a) Write down about the anatomy of MapReduce and Job Scheduling carried out in MapReduce.
- (b) List down any four characteristics of Hive Metastore, Hbase and Zoo Keeper deployed in Hadoop.
9. (a) Explain different components of Pig environment and how Pig coding is converted into MapReduce.
- (b) What important role does collaborative filtering plays in handling Big data in R environment?