

**MASTER OF COMPUTER
APPLICATIONS
(MCA-NEW)**

Term-End Examination

June, 2026

**MCS-226 : DATA SCIENCE
AND BIG DATA**

Time : 3 Hours

Maximum Marks : 100

***Note :** Question No. 1 is compulsory and carries
40 marks. Attempt any **three** questions
from the rest.*

-
-
1. (a) Explain the following types of data in
the context of data science : 8
- (i) Structured Data

- (ii) Semi-Structured Data
 - (iii) Unstructured Data
 - (iv) Data Streams
- (b) What is data curation ? Explain the steps of data curation. 6
- (c) What is Big Data ? Explain the *five* characteristics of big data. 6
- (d) What is a random variable ? Explain the types of random variables. Also, draw the probability distribution graph of a discrete random variable, which represents the number of head in tossing a fair coin three times. 6
- (e) What is meant by link spam ? Explain the strategy to combat the link spam. 6
- (f) Explain the following with the help of an example in context of R programming : 8
- (i) Vectors

(ii) Lists

(iii) Box Plots

(iv) Line Graph

2. (a) What is Distributed File System (DFS) ? Explain the components of DFS.

Also, write the features of DFS in details. 8

(b) What is MapReduce ? Explain the advantages of MapReduce. Also, write about <key-value> pair mechanism in MapReduce. 6

(c) What do you mean by HBase ? Explain the features of HBase. Also briefly write about the Write Ahead Log (WAL) in HBase. 6

3. (a) What is Document similarity ? Explain Jaccard similarity with suitable example. 8
- (b) Explain how data stream differs from database. 6
- (c) What is page rank ? Explain different mechanisms of finding PageRank with suitable example. 6
4. (a) What is Recommendation System ? Explain the process of content based recommendation system. 5
- (b) What are the different types of social media networks ? Explain the utility of graphs to represent the social media networks. 5

(c) What is a Heat Map ? Explain the best practices for Heat Map. 5

(d) What is Data Clearing in Data Science ? Explain the various steps of Data cleaning process. 5

5. Explain and write the steps of how R-Programming can be used for the following :

5×4=20

(i) Time Series Analysis

(ii) Linear regression

(iii) Decision Tree

(iv) Chi-square Test

× × × × ×