22SSP645B

PSG COLLEGE OF ARTS & SCIENCE
(AUTONOMOUS)

MSc (SS) DEGREE EXAMINATION DECEMBER 2025
(Sixth Semester)

Branch – **SOFTWARE SYSTEMS (Five Years Integrated)**

MAJOR ELECTIVE COURSE – II : BIG DATA ANALYTICS

Time: Three Hours | Maximum: 50 Marks

SECTION-A (5 Marks)
Answer **ALL** questions
ALL questions carry **EQUAL** marks (5 x 1 = 5)

1. Which of the following best describes Big Data?
   (i) Large volume of data (ii) High velocity of data
   (iii) Variety of data (iv) All of the above

2 What is the primary function of HDFS in Hadoop?
   (i) Data storage (ii) Data processing
   (iii) Data visualization (iv) Data cleaning

3 Which of the following is a NoSQL database?
   (i) MySQL (ii) PostgreSQL
   (iii) MongoDB (iv) SQLite

4 What is the key feature of MapReduce?
   (i) Centralized data processing (ii) Distributed computing
   (iii) Small data handling (iv) Single-thread processing

5 Which industry widely uses Big Data Analytics?
   (i) Healthcare (ii) Finance
   (iii) Retail (iv) All of the above

SECTION - B (15 Marks)
Answer **ALL** Questions
ALL Questions Carry **EQUAL** Marks (5 x 3 = 15)

6 a Explain the characteristics of Big Data.
OR
b Classify different types of Data Sources in Big Data.

7 a Discuss the architecture of the Hadoop framework.
OR
b Illustrate the working of HDFS with an example.

8 a Determine the role of NoSQL in handling Big Data.
OR
b Show how MapReduce performs distributed computing?

9 a Solve a sample problem using Pig for data analysis.
OR
b Apply Hive for querying structured data in Big Data.

10 a Explain how Big Data is applied in healthcare analytics?
OR
b Evaluate the impact of Big Data in social networking.

Cont...

SECTION - C (30 Marks)

Answer **ALL** questions
ALL questions carry **EQUAL** Marks (5 x 6 = 30)

11 a Categorize different types of Big Data applications with examples.

OR

b Develop a solution for processing unstructured data using Hadoop.

12 a Assess the importance of HDFS in Big Data storage.

OR

b Compare Hadoop and Spark frameworks in data processing.

13 a Formulate a query using NoSQL databases for a real-time application.

OR

b Interpret the significance of MapReduce in Big Data computing.

14 a Construct a Pig script for analyzing large datasets.

OR

b Design a Hive architecture for an enterprise data warehouse.

15 a Justify the use of Big Data in personalized recommendations.

OR

b Plan an end-to-end Big Data pipeline for fraud detection.

Z-Z-Z END