

MASTER OF COMPUTER APPLICATIONS (MCA_NEW)

MCA_NEW 3rd Semester Assignments

(January,2026 & July,2026 sessions)

MCS-224, MCS-225, MCS-226, MCS-227

MCSL-228, MCSL-229



**SCHOOL OF COMPUTER AND INFORMATION SCIENCES
INDIRA GANDHI NATIONAL OPEN UNIVERSITY
MAIDAN GARHI, NEW DELHI – 110 068**

CONTENTS

Course Code	Assignment No.	Submission-Schedule		Page No.
		For January-June Session	For July-December Session	
MCS-224	MCA_NEW(III)/224/Assignment/2026	30 th April, 2026	31 st October, 2026	3
MCS-225	MCA_NEW(III)/225/Assignment/2026	30 th April, 2026	31 st October, 2026	6
MCS-226	MCA_NEW(III)/226/Assignment/2026	30 th April, 2026	31 st October, 2026	7
MCS-227	MCA_NEW(III)/227/Assignment/2026	30 th April, 2026	31 st October, 2026	8
MCSL-228	MCA_NEW(III)/L-228/Assignment/2026	30 th April, 2026	31 st October, 2026	10
MCSL-229	MCA_NEW(III)/L-229/Assignment/2026	30 th April, 2026	31 st October, 2026	11

Important Notes

1. Submit your assignments to the Coordinator of your Study Centre on or before the due date. Also, be in touch with concerned IGNOU RC for any updates on Assignments schedule etc.
2. Assignment submission before due dates is compulsory to become eligible for appearing in corresponding Term End Examinations. For further details, please refer to Programme Guide of MCA_NEW.
3. To become eligible for appearing the Term End Practical Examination for the lab courses, it is essential to fulfill the minimum attendance requirements as well as submission of assignments (on or before the due date). For further details, please refer to the Programme Guide of MCA_NEW.
4. The viva voce is compulsory for the assignments. For any course, if a student submitted the assignment and not attended the viva-voce, then the assignment is treated as not successfully completed and would be marked as ZERO.

Course Code	:	MCS-226
Course Title	:	Data Science and Big Data
Assignment Number	:	MCA_NEW(III)/226/Assign/2026
Maximum Marks	:	100
Weightage	:	30%
Last Dates for Submission	:	30th April, 2026 (for January session) 31st October, 2026 (for July session)

This assignment has ten questions of 8 Marks each, answer all questions. Rest 20 marks are for viva voce. You may use illustrations and diagrams to enhance the explanations. Please go through the guidelines regarding assignments given in the Programme Guide for the format of presentation.

- Q1:** Explain the scope and objectives of **Data Science**. Discuss the importance of **data collection and sampling techniques** in data-driven decision making. Illustrate your answer with suitable examples.
- Q2:** Define **measures of central tendency and dispersion**. Explain how **variance and standard deviation** help in understanding data distribution. Discuss the relevance of these measures in Data Science applications.
- Q3:** What is **data preprocessing**? Explain the major steps involved in preprocessing large datasets. Discuss methods used for **noise removal, normalization, and data transformation**.
- Q4:** Explain the concept of **data visualization**. Discuss any four commonly used charts and justify their use for analyzing different types of datasets.
- Q5:** Discuss the need for **Big Data technologies**. Explain the **characteristics of Big Data** and describe how Big Data processing differs from traditional data processing systems.
- Q6:** Describe the architecture of the **Hadoop ecosystem**. Explain the role of **HDFS, YARN, and MapReduce** in handling large-scale data processing.
- Q7:** What is **Apache Spark**? Explain its core components and advantages over traditional Hadoop MapReduce. Discuss one practical application where Spark is effectively used.
- Q8:** Explain the concept of **NoSQL databases**. Compare **key-value stores, column-family databases, and document databases** with suitable examples and use cases.
- Q9:** Explain the techniques used for **similarity measurement** in Big Data analytics. Discuss **Jaccard similarity, cosine similarity**, and their role in recommendation systems.
- Q10:** Write R programs for the following tasks:
 - (a) Perform **simple linear regression** on a sample dataset and interpret the output.
 - (b) Apply a **classification technique** of your choice on a dataset and explain the results obtained.