.

Printed Page:-4 **Subject Code:- AMICA0715**

Roll. No:

NOIDA INSTITUTE OF ENGINEERING AND TECHNOLOGY, GREATER NOIDA
(An Autonomous Institute Affiliated to AKTU, Lucknow)
MCA (Integrated)
SEM: VII - THEORY EXAMINATION (2025 - 2026)
Subject: Big Data Analytics

Time: 3 Hours **Max. Marks: 100**

General Instructions:

IMP: *Verify that you have received the question paper with the correct course, code, branch etc.*

1. *This Question paper comprises of* ***three Sections -A, B, & C.*** *It consists of Multiple Choice Questions (MCQ's) & Subjective type questions.*

2. *Maximum marks for each question are indicated on right -hand side of each question.*

3. *Illustrate your answers with neat sketches wherever necessary.*

4. *Assume suitable data if necessary.*

5. *Preferably, write the answers in sequential order.*

6. *No sheet should be left blank. Any written material after a blank sheet will not be evaluated/checked.*

SECTION-A 20

1. Attempt all parts:-

1-a. Identify the 5 Vs of Big Data. (CO1, K1) 1

(a) Volume, Velocity, Variety, Veracity, Value
(b) Volume, Variety, Validation, Velocity, Value
(c) Volume, Velocity, Variety, Visibility, Value
(d) Volume, Variety, Veracity, Variance, Value

1-b. Recall Volume means in Big Data. (CO1, K1) 1

(a) Amount of data collected and stored
(b) Speed of data generation
(c) Types of data
(d) Data accuracy

1-c. Apache Hadoop is maintained under _____ foundation? (CO2, K2) 1

(a) Linux Foundation
(b) Apache Software Foundation
(c) Mozilla Foundation
(d) Google Foundation

1-d. Pig in the Hadoop ecosystem is used for? (CO2, K3) 1

(a) Batch processing
(b) Data flow scripting
(c) Real-time processing
(d) Data storage

1-e. Map Reduce framework is primarily used for. (CO3, K1) 1

(a) Large-scale data processing
(b) Data visualization
(c) Security testing
(d) Hardware maintenance

1-f. Job scheduler in Map Reduce is responsible for. (CO3, K3) 1

(a) Managing cluster resources
(b) Cleaning data
(c) Networking configuration
(d) Security management

1-g. Hadoop Distributed File System is designed for primary purpose? (CO4, K1) 1

(a) Large data processing
(b) Local file access
(c) Distributed storage for big data
(d) Networking only

1-h. By default, HDFS user files are stored under_____ directory? (CO4, K1) 1

(a) /user directory
(b) /root
(c) /system
(d) /core

1-i. Data processing operators belong to component? (CO5, K3) 1

(a) MapReduce
(b) Pig
(c) Sqoop
(d) Hive

1-j. scheduler in YARN aims for equitable resource sharing? (CO5, K3) 1

(a) Capacity scheduler
(b) Round-robin scheduler
(c) Resource manager
(d) Fair scheduler

2. Attempt all parts:-

2.a. Demonstrate compliance auditing in Big Data by citing an industry example. (CO1, K3) 2

2.b. Illustrate how Hadoop analyzes large data sets efficiently. (K3, CO2) 2

2.c. Analyze how failure recovery is handled during a Map Reduce job run. (CO3, K4) 2

2.d. The process of reading a file from HDFS, highlighting each step. (CO4, K3) 2

2.e. Describe the difference between fair and capacity schedulers in Hadoop. (CO5, K3) 2

SECTION-B 30

3. Attempt all parts:-

3.a. Answer any one of the following:-

3.a.(i) Demonstrate how you would design a solution for real-time sensor data 6

	management using big data technology. (CO1, K3)	
3.a.(ii)	Design a compliance and data protection framework for an enterprise handling global Big Data sets. (CO1, K6)	6
3.b. Answer any one of the following:-		
3.b.(i)	Compare HDFS file storage with traditional file systems, mentioning at least three differences. (CO2, K4)	6
3.b.(ii)	Compare the data format requirements of HDFS and any two Hadoop ecosystem components. (CO2, K4)	6
3.c. Answer any one of the following:-		
3.c.(i)	Compare two common input formats used in Map Reduce and mention typical uses. (CO3, K4)	6
3.c.(ii)	Describe how Map Reduce's fault tolerance ensures job completion even after node failures. (CO3, K4)	6
3.d. Answer any one of the following:-		
3.d.(i)	Describe the importance of file and block sizes in HDFS. (CO4, K3)	6
3.d.(ii)	Describe how block abstraction simplifies data management in HDFS. (CO4, K3)	6
3.e. Answer any one of the following:-		
3.e.(i)	Describe the difference between the fair and capacity schedulers in YARN. (CO5, K4)	6
3.e.(ii)	Describe how MRv1 jobs can run in YARN environment for backward compatibility. (CO5, K4)	6
SECTION-C		50
4. Answer any one of the following:-		
4-a.	Compare and contrast the security and privacy frameworks of two Big Data platforms, providing evidence-based recommendations for enterprise adoption. (CO1, K4)	10
4-b.	Formulate a new Big Data governance model for multinational enterprises operating under diverse regulatory regimes. (CO1, K6)	10
5. Answer any one of the following:-		
5-a.	Compare and contrast Hadoop streaming and Hadoop pipes with examples of when to use each. (CO2, K4)	10
5-b.	Examine how fault tolerance mechanisms in Hadoop contribute to data integrity during cluster failures. (CO2, K4)	10
6. Answer any one of the following:-		
6-a.	Analyze the Map Reduce architecture and explain how its components interact to process large data sets. (CO3, K4)	10
6-b.	Discuss the impact of resource allocation and job prioritization on task scheduling in Map Reduce clusters. (CO3, K5)	10
7. Answer any one of the following:-		
7-a.	Explain the design and architecture of HDFS and how it stores large data files. (CO4, K4)	10
7-b.	Analyze the role of Hadoop benchmarks in capacity planning and performance	10

tuning. (CO4, K5)

8. Answer any one of the following:-

8-a.	Explain the workflow and components involved in using the Hive shell for data management. (CO5, K4)	10
8-b.	Discuss how scheduling strategies in YARN impact cluster utilization and job throughput. (CO5, K4)	10