

--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--

NOIDA INSTITUTE OF ENGINEERING AND TECHNOLOGY, GREATER NOIDA
(An Autonomous Institute Affiliated to AKTU, Lucknow)

B.Tech

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. Transaction of data of the bank is a type of. [CO1, K2]

1

- (a) STRUTUCTRED
- (b) UNSTRUCTURED
- (c) BOTH A & B
- (d) None of the mentioned

1-b. In which language is Hadoop written? [CO1, K2]

1

- (a) C++
- (b) JAVA
- (c) RUST
- (d) PYTHON

1-c. On which of the following platforms does Hadoop run? [CO2, K2]

1

- (a) Debian
- (b) Coss-Platform
- (c) Bare Metal
- (d) Unix-Like

1-d. Data in ____ bytes size is called big data [CO2, K2]

1

- (a) Meta
- (b) Giga
- (c) Tera
- (d) Peta

1-e. Which Hadoop ecosystem component is used for real-time stream

1

processing? [CO3, K2]

- (a) HDFS
- (b) MapReduce
- (c) YARN
- (d) SPARK

1-f. Which Hadoop ecosystem component is used for SQL-based querying and analysis of data stored in Hadoop? [CO3,K2] 1

- (a) HDFS
- (b) MapReduce
- (c) YARN
- (d) Hive

1-g. Which of the following is a data ingestion framework in the Hadoop ecosystem? [CO4, K2] 1

- (a) Flame
- (b) Impala
- (c) Hive
- (d) Hbase

1-h. Which of the following is an interactive SQL query engine for Hadoop in the Hadoop ecosystem? [CO4, K2] 1

- (a) Spark
- (b) PIG
- (c) Hive
- (d) All of the mentioned

1-i. Sqoop uses _____ to fetch data from RDBMS and stores that on HDFS. [CO5, K2] 1

- (a) Hive
- (b) Map reduce
- (c) Imphala
- (d) BigTOP

1-j. _____ tool can list all the available database schemas. [CO5, K2] 1

- (a) sqoop-list-tables
- (b) sqoop-list-databases
- (c) sqoop-list-schema
- (d) sqoop-list-columns

2. Attempt all parts:-

2.a. Explain about 5v's? [CO1, K2] 2

2.b. What are the basic parameters of a mapper? [CO2, K3] 2

2.c. Define hadoop distributed file system. [CO3,K2] 2

2.d. What is the syntax of Pig Latin? [CO4,K2] 2

2.e. What is the role of JDBC driver in Sqoop? [CO5, K3] 2

SECTION-B

30

3.a. Answer any <u>one</u> of the following:-	
3.a.(i) What is Bigdata? Describe the main features of a big data in detail? [CO1, K2]	6
3.a.(ii) What is Map Reduce? Explain working of various phases of Map Reduce with appropriate example and diagram. [CO1, K2]	6
3.b. Answer any one of the following:-	
3.b.(i) How node failure can be handled in hadoop? Explain. [CO2, K4]	6
3.b.(ii) What are the core components of hadoop. Explain in brief each of its component. [CO2, K4]	6
3.c. Answer any one of the following:-	
3.c.(i) Explain the difference between NameNode, Backup Node and Checkpoint NameNode? [CO3, K2]	6
3.c.(ii) What is commodity hardware? [CO3, K2]	6
3.d. Answer any one of the following:-	
3.d.(i) Explain with Neat sketch explain in detail Apache Hive architecture? (CO4, K2)	6
3.d.(ii) Write the steps for upgrading HDFS. [CO4, K2]	6
3.e. Answer any one of the following:-	
3.e.(i) State and Explain SCALA and Basic types of operators in SCALA. [CO5,K3]	6
3.e.(ii) Brief about Built-in control structures in SCALA.[CO5,K3]	6
<u>SECTION-C</u>	50
4. Answer any <u>one</u> of the following:-	
4-a. Write about top challenges facing big data? [CO1, K2]	10
4-b. Explain the concept of Basically Available soft state Eventual consistency [CO1, K2]	10
5. Answer any <u>one</u> of the following:-	
5-a. How to do Mapper testing using MRUnit Test? [CO2, K4]	10
5-b. Briefly explain the anatomy of MapReduce. Explain with real world example. [CO2, K4]	10
6. Answer any <u>one</u> of the following:-	
6-a. How google file system differs from the hadoop file system and explains the google file system architecture with a neat sketch. [CO3, K2]	10
6-b. How the pig programs can be packaged and explain the modes of running a pig script with a neat sketch. [CO3, K2]	10
7. Answer any <u>one</u> of the following:-	
7-a. State what is the purpose of the "DESCRIBE" command in Grunt, Explain what are the Hadoop daemons and explain their roles in a Hadoop cluster, state what is the purpose of Pig Latin's "LOAD" command? (CO4, K1)	10
7-b. What are the applications on Pig in Big Data? [CO4, K2]	10
8. Answer any <u>one</u> of the following:-	
8-a. Describe what is Spark Streaming and how does it work, State how does Spark achieve parallelism and how would you use Spark with real-time data processing? (CO5, K2)	10

8-b. Explain in detail what do mean by virtual private cloud (VPC) also state all its types, discuss steps to establish VPC on a private network? (CO5, K2)

10

REG_JULY_DEC_2025