

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

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

B.Tech

SEM: VI - 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:-

- (1.1) Please indicate the correct statement. (CO1,K1) 1
- (a) Replication Factor can be configured at a cluster level (Default is set to 3) and also at a file level
 - (b) Block Report from each DataNode contains a list of all the blocks that are stored on that DataNode
 - (c) User data is stored on the local file system of DataNodes
 - (d) DataNode is aware of the files to which the blocks stored on it belong to
- (1.2) Transaction of data of the bank is a type of. (CO1,K3) 1
- (a) STRUTUCTRED
 - (b) UNSTRUCTURED
 - (c) BOTH A & B
 - (d) None of the mentioned
- (1.3) Identify the incorrect big data Technologies. (CO4, K4) 1
- (a) Apache Pytorch
 - (b) Apache Kafka
 - (c) Apache Hadoop
 - (d) Apache Spark
- (1.4) Identify the different features of Big Data Analytics? [CO2,K1] 1
- (a) Open-Source
 - (b) Data recovery
 - (c) Scalability
 - (d) All of the Above

- (1.5) Which Hadoop ecosystem component is used for scripting and high-level data manipulation? (CO3,K2) 1
- (a) HDFS
 - (b) Mahout
 - (c) Pig
 - (d) Hive
- (1.6) Identify which Hadoop component was YARN designed to replace? [CO3,K1] 1
- (a) MapReduce
 - (b) Hive
 - (c) HBase
 - (d) HDFS
- (1.7) The _____ Server assigns regions to the region servers and takes the help of Apache ZooKeeper for this task. (CO4, K1) 1
- (a) Region
 - (b) Master
 - (c) Zookeeper
 - (d) All of the mentioned
- (1.8) On dropping a managed table (CO4, K3) 1
- (a) The schema gets dropped without dropping the data
 - (b) The data gets dropped without dropping the schema
 - (c) An error is thrown
 - (d) Both the schema and the data is dropped
- (1.9) Spark runs on top of _____ a cluster manager system which provides efficient resource isolation across distributed applications. (CO5, K2) 1
- (a) Mesjs
 - (b) Mesos
 - (c) Mesus
 - (d) All of the mentioned
- (1.10) Scala stands for _____. (CO5, K1) 1
- (a) Scalable language
 - (b) Scripted advanced language
 - (c) Sequential language
 - (d) None

2. Attempt all parts:-

- (2.1) Discuss the about top challenges facing big data? (CO1, K2) 2
- (2.2) Differentiate between data analytics and big data analytics? (CO2, K4) 2
- (2.3) List the key features of MapReduce. (CO3, K1) 2
- (2.4) Explain in detail what is the purpose of the "JOIN" command in HiveQL? [CO4,K2] 2
- (2.5) State what is the work of Export in Hadoop sqoop? [CO5,K1] 2

SECTION-B

30

Answer any one of the following:-

- (3.1) Explain “Map Phase” and “Combiner Phase” in MapReduce. (CO1, K2) 6
- (3.2) Explain Big Data in Healthcare, Transportation and Medicine? [CO2,K2] 6

Answer any one of the following:-

- (3.3) Brief out the features of Hadoop HDFS and explain the functions of NameNode and DataNode. (CO2, K2) 6
- (3.4) Explain the role of following nodes in detail: (i) Datanode (ii) NodeManager (iii) Job Tracker (iv) Task Tracker (CO2, K2) 6

Answer any one of the following:-

- (3.5) Describe what is Hbase and how does it differ from other data storage technologies in Hadoop such as HDFS and MapReduce? (CO3,K2) 6
- (3.6) Describe and write and draw anatomy of a file write with HDFS? (CO3,K2) 6

Answer any one of the following:-

- (3.7) Give the comparison between NOSQL & MySQL/RDBMS. (CO4, K4) 6
- (3.8) Discuss what are the Benefits of using zookeeper, also explain it's architecture? (CO4,K2) 6

Answer any one of the following:-

- (3.9) Explain SCALA functions and Closures? (CO5, K2) 6
- (3.10) Write a step-by-step procedure for installation of APACHE SPARK? (CO5, K3) 6

SECTION-C

50

4. Answer any one of the following:-

- (4.1) What are 5 Principles of Big Data Ethics? Compare between analysis and reporting? (CO1, K2) 10
- (4.2) Define Data. Also explain types of digital data. (CO1, K1) 10

5. Answer any one of the following:-

- (5.1) Discuss and Write short notes on? (CO2, K1) 10
- (i) Driver code
- (ii) Mapper code
- (iii) Reducer code
- (iv) Combiner
- (5.2) Discuss and write short notes on Serialization and Byte Writable? [CO2,K2] 10

6. Answer any one of the following:-

- (6.1) Specify the role of job tracker and task tracker in HDFS. (CO3, K2) 10
- (6.2) Describe the concept of NameNode High Availability in Hadoop. Explain the role of Active and Standby NameNode along with failover mechanism. (CO3, K2) 10

7. Answer any one of the following:-

- (7.1) Discuss various data processing operators available in Pig. Explain filtering, grouping, joining, and sorting operations with examples. (CO4,K4) 10
- (7.2) Discuss what are the important components of the Spark ecosystem, Explain how Spark runs applications with the help of its architecture? [CO4,K2] 10

8. Answer any one of the following:-

- | | | |
|-------|---|----|
| (8.1) | Explain Anatomy of SPARK Job run in detail ? [CO5,K2] | 10 |
| (8.2) | Discuss how subset data can be imported in Sqoop. Explain filtering techniques and query-based imports with examples. (CO5, K3) | 10 |

REG_JAN_JUNE_2026