

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

**NOIDA INSTITUTE OF ENGINEERING AND TECHNOLOGY, GREATER NOIDA**

**(An Autonomous Institute Affiliated to AKTU, Lucknow)**

**B.Tech**

**SEM: VI - THEORY EXAMINATION (2023 - 2024)**

**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. Hadoop is a framework that works with a variety of related tools. Common cohorts include \_\_\_\_\_ [CO1] 1
- (a) MapReduce, Hive and HBase
  - (b) MapReduce, MySQL and Google Apps
  - (c) MapReduce, Hummer and Iguana
  - (d) MapReduce, Heron and Trumpet
- 1-b. Hive also support custom extensions written in \_\_\_\_\_ . [CO1] 1
- (a) C#
  - (b) Java
  - (c) C
  - (d) C++
- 1-c. Hadoop Common Package contains? CO2 1
- (a) msi files
  - (b) war files
  - (c) exe files
  - (d) jar files
- 1-d. Identify the node which acts as a checkpoint node in HDFS. [CO2] 1
- (a) Secondary Name node
  - (b) Secondary data node

- (c) Name node  
(d) Data node
- 1-e. What is Avro? [CO3] 1  
(a) A file format used for storing structured data  
(b) A programming language used for data processing  
(c) A distributed computing framework  
(d) A database management system
- 1-f. Which of the following is not a benefit of file-based data structures in Hadoop? [CO3] 1  
(a) Efficient data compression  
(b) High data compression ratio  
(c) Fast query processing  
(d) High data redundancy
- 1-g. To see the data types details of only a column not the table we should use the command. [CO4] 1  
(a) DESCRIBE  
(b) DESCRIBE EXTENDED  
(c) DESCRIBE FORMATTED  
(d) DESCRIBE COLUMN
- 1-h. Each database created in hive is stored as[CO4] 1  
(a) a directory  
(b) a file  
(c) a hdfs block  
(d) a jar file
- 1-i. Spark powers a stack of high-level tools including Spark SQL, MLlib for \_\_\_\_\_ [CO5] 1  
(a) regression models  
(b) statistics  
(c) machine learning  
(d) reproductive research
- 1-j. Data processed by Scoop can be used for? [CO5] 1  
(a) Hbase  
(b) HDFS  
(c) Mapreduce  
(d) MahOut

2. Attempt all parts:-

- 2.a. Explain about 5v's of "BIG DATA" and also explain the data size required to be called a data as Big Data? [CO1] 2

- 2.b. What are the Limitations of Hadoop 1.x as compared to Hadoop 2.x ? [CO2] 2
- 2.c. Distinguish between Hadoop and Hadoop Common? [CO3] 2
- 2.d. Explain Zookeeper with examples. CO4 2
- 2.e. What are the features of Apache Spark? [CO5] 2

### **SECTION-B**

30

3. Answer any five of the following:-

- 3-a. Difference between Hive and RDBMS? [CO1] 6
- 3-b. Discuss Big Data in Healthcare, Transportation & Medicine? [CO1] 6
- 3-c. What are the main features and Characteristics of Hadoop which make it the most popular and powerful Big Data tool? [CO2] 6
- 3-d. Explain the following (i) HDFS replication (ii) Rack awareness (iii) Locality reference [CO2] 6
- 3.e. How can you overwrite the replication factors in HDFS? [CO3] 6
- 3.f. What are the relation operations in Pig? Explain any two with examples? [CO4] 6
- 3.g. Explain SCALA functions and Closures? [CO5] 6

### **SECTION-C**

50

4. Answer any one of the following:-

- 4-a. Define Command line interface using HDFS files and give a brief note on Hadoop-specific file system types and HDFS commands? [CO1] 10
- 4-b. Describe the working of Map reduce with a relevant example? [CO1] 10

5. Answer any one of the following:-

- 5-a. How is the DataNode failure handled by NameNode, Explain in detail? [CO2] 10
- 5-b. Is it possible to change the number of mappers to be created in a MapReduce job? [CO2] 10

6. Answer any one of the following:-

- 6-a. Explain Hadoop "Mapred-Site.xml" conf file in detail? [CO3] 10
- 6-b. Differentiate between compression and serialization in detail? [CO3] 10

7. Answer any one of the following:-

- 7-a. Explain Hive Integration & Work Flow steps involved with a diagram. [CO4] 10
- 7-b. What is Apache ZooKeeper and Apache Oozie? Explain in detail? [CO4] 10

8. Answer any one of the following:-

- 8-a. What are the important components of the Spark ecosystem? [CO5] 10
- 8-b. Define "For loop" in scala, How it is different in filters, collections and yield? [CO5] 10