

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

NOIDA INSTITUTE OF ENGINEERING AND TECHNOLOGY, GREATER NOIDA

(An Autonomous Institute Affiliated to AKTU, Lucknow)

B.Tech

SEM: VII - THEORY EXAMINATION (2024 - 2025)

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. State what is the purpose of the Hadoop Distributed File System (HDFS)? (CO1, K1) 1
- (a) To provide a SQL-like interface to query and analyze data stored in Hadoop
 - (b) To manage the metadata of the files stored in Hadoop
 - (c) To store and manage large datasets in a distributed environment
 - (d) To coordinate the processing of distributed applications
- 1-b. State what is the purpose of YARN (Yet Another Resource Negotiator) in Hadoop? (CO1, K1) 1
- (a) To manage the metadata of the files stored in Hadoop
 - (b) To store and manage large datasets in a distributed environment
 - (c) To coordinate the processing of MapReduce jobs
 - (d) To provide a SQL-like interface to query and analyze data stored in Hadoop
- 1-c. State which of the following are the Benefits of Big Data Processing? (CO2, K1) 1
- (a) Buisness can utilize outside intilligence while taking decisions.
 - (b) Better operational efficiency
 - (c) Improve customer service
 - (d) All of the Above
- 1-d. State Big data analysis does the following except? (CO2, K1) 1
- (a) Spreads data

- (b) Analyze data
 - (c) Organizes data
 - (d) Collect data
- 1-e. Select what is the default block size in HDFS? (CO3, K1) 1
- (a) 64 MB
 - (b) 128 MB
 - (c) 256 MB
 - (d) 512 MB
- 1-f. State which component in HDFS is responsible for storing data blocks on multiple nodes in the cluster? (CO3, K1) 1
- (a) NameNode
 - (b) DataNode
 - (c) JobTracker
 - (d) TaskTracker
- 1-g. List which of the following is an aggregation operation in Pig? (CO4, K1) 1
- (a) FILTER
 - (b) GROUP
 - (c) DISTINCT
 - (d) DUMP
- 1-h. Select which of the following is a Pig function used for splitting data into multiple relations? (CO4, K1) 1
- (a) FILTER
 - (b) SPLIT
 - (c) JOIN
 - (d) GROUP
- 1-i. Select from below how the Scala compiler starts executing from ? (CO5, K1) 1
- (a) The first line of code
 - (b) main() method
 - (c) main class
 - (d) End of code
- 1-j. Select which is not a Scala method for printing text on the screen? (CO5, K1) 1
- (a) print()
 - (b) printf()
 - (c) println()
 - (d) println()

2. Attempt all parts:-

- 2.a. State what is Big Data? How much data do we consider as big data? (CO1, K1) 2
- 2.b. State what are the three modes that Hadoop can execute on effectively? (CO2, K1) 2

- | | | |
|------|--|---|
| 2.c. | Discuss what is commodity hardware? (CO3, K2) | 2 |
| 2.d. | Discuss Big Data in Healthcare, Transportation & Medicine? (CO4, K2) | 2 |
| 2.e. | Explain the work of export in Hadoop SQOOP? (CO5,K2) | 2 |

SECTION-B

30

3. Answer any five of the following:-

- | | | |
|------|---|---|
| 3-a. | Explain Concept of blocks in details, Let incoming data be 300 files and per file is 200 MB. IF block size is 64 MB and RF 03, then how many blocks will be created? (CO1, K1) | 6 |
| 3-b. | Discuss what is big data analytics? Explain Five 'V's of Big data, briefly discuss applications of big data also, State what is Map Reduce, Explain working of various phases of Map Reduce with appropriate example and diagram? (CO1, K2) | 6 |
| 3-c. | Describe what is HBase, and how does it differ from other data storage technologies in Hadoop such as HDFS and MapReduce? (CO2, K2) | 6 |
| 3-d. | Explain the difference between Name Node, Backup Node and Checkpoint Name Node? (CO2, K2) | 6 |
| 3.e. | 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? (CO3, K1) | 6 |
| 3.f. | Explain with Neat sketch explain in detail Apache Hive architecture? (CO4, K2) | 6 |
| 3.g. | State the network topology types and also discuss and write a program in Spark Core Processing RDD to run Word Count program? (CO5, K1) | 6 |

SECTION-C

50

4. Answer any one of the following:-

- | | | |
|------|---|----|
| 4-a. | Explain the role of following nodes in detail: (CO1, K2) | 10 |
| | 1. Resource Manager | |
| | 2. Application Manager | |
| | 3. Job Tracker | |
| | 4. Task Tracker | |
| | 5. Namenode | |
| | 6. Datanode | |
| 4-b. | Differentiate between Hadoop 1.x, Hadoop2.x and Hadoop 3.x in detail? (CO1, K4) | 10 |

5. Answer any one of the following:-

- | | | |
|------|--|----|
| 5-a. | Explain the Concept of “Rack Awareness” & “Locality of Reference”? (CO2, K2) | 10 |
| 5-b. | Write short notes on? (CO2, K1) | 10 |
| | (i) Driver code | |
| | (ii) Mapper code | |
| | (iii) Reducer code | |
| | (iv) Combiner | |

6. Answer any one of the following:-

- | | | |
|------|---|----|
| 6-a. | Draw neat and clean diagram of Hive execution Process and also discuss Apache | 10 |
|------|---|----|

Pig all execution mechanisms in detail ? (CO3, K1)

- 6-b. 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? (CO3, K2) 10

7. Answer any one of the following:-

- 7-a. Discuss what are the key security considerations when moving applications to the cloud and also what is the difference between IaaS, PaaS, and SaaS in cloud computing? (CO4, K2) 10
- 7-b. 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

8. Answer any one of the following:-

- 8-a. State 5 Applications Areas of spark in Industry, draw its architecture and explain in detail. (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-2024