

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

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

B.Tech

SEM: IV - 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) State, In Hadoop, the NameNode is responsible for? [CO1,K1]

1

- (a) Storing actual data blocks
- (b) Managing metadata of files
- (c) Executing MapReduce tasks
- (d) Managing network security

(1.2) State, the default block size in HDFS is generally? [CO1,K1]

1

- (a) 1 MB
- (b) 16 MB
- (c) 64 MB / 128 MB
- (d) Data structure

(1.3) State, the task execution in MapReduce1.0 is handled by? [CO2,K1]

1

- (a) NameNode
- (b) DataNode
- (c) TaskTracker/NodeManager
- (d) Client only

(1.4) Identify, a Combiner in MapReduce is used to? [CO2,K1]

1

- (a) Replace reducer
- (b) Delete duplicates
- (c) Manage cluster
- (d) Perform local aggregation

(1.5) State, the zookeeper is used for? [CO3,K1]

1

- (a) Storage
 - (b) Coordination
 - (c) Processing
 - (d) Query
- (1.6) State, Sqoop transfers data between? [CO3,K1] 1
- (a) Hadoop & RDBMS
 - (b) Hadoop & cloud
 - (c) Hadoop & Python
 - (d) Hadoop & Hive
- (1.7) State, the federated services improve? [CO4,K1] 1
- (a) Interoperability
 - (b) Latency
 - (c) Isolation
 - (d) Complexity
- (1.8) State, A cloud service where only the application is controlled by the user and everything else is managed by the provider is? [CO4,K1] 1
- (a) PaaS
 - (b) SaaS
 - (c) IaaS
 - (d) Hybrid Cloud
- (1.9) State, frequently accessed data stored with low latency is best suited for? [CO5,K1] 1
- (a) Standard storage class
 - (b) Coldline storage
 - (c) Backup storage
 - (d) Archive storage
- (1.10) State, communication between a virtual machine and the network is established using? [CO5,K1] 1
- (a) Gateway
 - (b) Firewall
 - (c) NIC
 - (d) Router

2. Attempt all parts:-

- (2.1) Discuss how does Hadoop ensure fault tolerance through replication? [CO1,K2] 2
- (2.2) Explain the role of NodeManager in YARN? [CO2,K2] 2
- (2.3) Describe the concept of serialization in Hadoop? [CO3,K2] 2
- (2.4) Discuss the role of compression techniques in Hadoop I/O and their impact on performance? [CO4,K2] 2
- (2.5) Explain how data durability and availability are ensured in cloud storage systems? [CO5,K2] 2

SECTION-B

30

Answer any one of the following:-

- (3.1) Differentiate between Hadoop 1.x, Hadoop2.x and Hadoop 3.x in detail? [CO1,K4] 6
- (3.2) Discuss the importance of fault tolerance and load balancing in cloud systems? [CO1,K2] 6

Answer any one of the following:-

- (3.3) Discuss failure recovery in YARN and analyze how high availability is achieved? [CO2,K2] 6
- (3.4) Explain the integration of Hadoop ecosystem tools such as Hive, Pig, and HBase with YARN and analyze their significance? [CO2,K2] 6

Answer any one of the following:-

- (3.5) Explain Flume architecture and its working in detail? [CO3,K2] 6
- (3.6) Describe Spark architecture and its advantages over MapReduce? [CO3,K2] 6

Answer any one of the following:-

- (3.7) Explain the Concept of “Rack Awareness” & “Locality of Reference” in Hadoop architecture detail? [CO1,K2] 6
- (3.8) Explain Hadoop in cloud computing environments and discuss how GCP supports big data applications? [CO4,K2] 6

Answer any one of the following:-

- (3.9) Explain cloud firewall architecture and discuss different types of firewall rules? [CO5,K2] 6
- (3.10) Explain the architecture and working of BigQuery? [CO5,K2] 6

SECTION-C 50

4. Answer any one of the following:-

- (4.1) Explain all the challenges occur during Big Data Processing and also differentiate between Analysis vs reporting? [CO1,K2] 10
- (4.2) 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

5. Answer any one of the following:-

- (5.1) Discuss what is Hadoop streaming and Hadoop Pipes and how are they used in data processing also explain what is data ingestion in Hadoop? [CO2,K2] 10
- (5.2) Describe the working of the MapReduce programming model with a suitable example. Explain Mapper, Reducer, Combiner, and Partitioner phases and also Compare traditional distributed computing with cloud-based Hadoop architecture. Explain the advantages of Hadoop in cloud environments? [CO2,K2] 10

6. Answer any one of the following:-

- (6.1) Explain how ZooKeeper can be used to monitor a distributed cluster and also discuss the consistency model used in ZooKeeper? [CO3,K2] 10
- (6.2) Explain what is the role of a shuffle operation in Hadoop's Spark framework, and how does it contribute to sorting and aggregating data also explain the role of 10

partitioning in Spark's performance & discuss how do you decide the number of partitions for an RDD? [CO3, K2]

7. Answer any one of the following:-

- (7.1) Explain listed below cloud deployment models and evaluate their advantages, limitations, and security implications? [CO4,K2] 10
- (i) Public Cloud
 - (ii) Private Cloud
 - (iii) Hybrid Cloud
- (7.2) Explain the architecture and working of Google App Engine with suitable diagram, also discuss the programming environment for Google App Engine and explain deployment and execution of applications on App Engine? [CO4,K2] 10

8. Answer any one of the following:-

- (8.1) Explain different network topology options used in cloud computing with diagrams? [CO5,K2] 10
- (i) Star topology
 - (ii) Bus topology
 - (iii) Ring topology
 - (iv) Mesh topology
 - (v) Hybrid topology
- (8.2) Explain and critically examine the role of routing policies and traffic flow management in ensuring efficient cloud network performance? [CO5,K2] 10