



Paper ID : 260927

Printed Page: 1 of 1

Subject Code: BOE099

Roll No:

BTECH

(SEM. VIII) THEORY EXAMINATION 2025-26

BIG DATA

TIME: 3 HRS

M.MARKS: 70

Note: 1. Attempt all Sections. If require any missing data; then choose suitably.

SECTION A

1. Attempt all questions in brief.

2x7 = 14

(a)	List the key characteristics of Big Data.	CO-1	K1
(b)	Differentiate between structured, semi-structured, and unstructured data.	CO-1	K2
(c)	State the role of HDFS in Hadoop.	CO-2	K1
(d)	Enumerate the input and output formats in MapReduce.	CO-2	K1
(e)	What is block size in HDFS and why is it important?	CO-3	K1
(f)	List the advantages of scala over java.	CO-4	K1
(g)	What is Apache Hive and its use?	CO-5	K1

SECTION B

2. Attempt any three of the following:

7 x 3 = 21

(a)	Discuss the importance and applications of Big Data in various domains such as healthcare, finance, e-commerce, and social media.	CO-1	K2
(b)	Describe the MapReduce programming model, explaining the roles of Mapper and Reducer, and illustrate the complete workflow including input splitting, mapping, shuffling, sorting, and reducing with a suitable example.	CO-2	K2
(c)	Explain the architecture of HDFS in detail, describing the roles and responsibilities of NameNode, DataNode, and Secondary NameNode, and discuss how metadata is managed in a distributed environment.	CO-3	K2
(d)	Explain the architecture of YARN in detail, describing its components, and discuss how YARN manages resources and scheduling in a Hadoop cluster.	CO-4	K2
(e)	Discuss the advantages and limitations of Hive, Pig, and HBase in big data processing, comparing their performance, usability, and application scenarios.	CO-5	K2

SECTION C

3. Attempt any one part of the following:

7 x 1 = 7

(a)	Explain the architecture of Big Data systems in detail, describing various components, and discuss how these components interact with each other.	CO-1	K2
(b)	Discuss data privacy, security, compliance, and ethics in Big Data. Highlight major challenges and solutions.	CO-1	K2

4. Attempt any one part of the following:

7 x 1 = 7

(a)	What is Hadoop Ecosystem? Explain various components of Hadoop Ecosystem.	CO-2	K2
(b)	Explain the concept of data locality in Hadoop and discuss how it improves performance by minimizing network congestion and enhancing processing efficiency.	CO-2	K2

5. Attempt any one part of the following:

7 x 1 = 7

(a)	Discuss the replication mechanism in HDFS. How the system ensures fault tolerance and data reliability in case of node failures?	CO-3	K2
(b)	Explain a scenario of node failure in HDFS during data write operations. Describe how the system recovers from failures while ensuring no data loss and maintaining consistency.	CO-3	K2

6. Attempt any one part of the following:

7 x 1 = 7

(a)	Explain NoSQL databases in detail and compare them with relational databases. Also, describe the basics of MongoDB including data types, document creation, querying, and indexing.	CO-4	K2
(b)	Explain basic concepts of Scala programming, including classes, objects, and functional programming features, and discuss its importance in Spark development.	CO-4	K2

7. Attempt any one part of the following:

7 x 1 = 7

(a)	Compare HBase with traditional relational databases, focusing on data model, consistency, scalability, and performance. Explain when HBase is preferred over RDBMS.	CO-5	K2
(b)	Discuss the role of Zookeeper in the Hadoop ecosystem, explaining how it helps in coordination, synchronization, and maintaining configuration information in distributed systems	CO-5	K2