



Roll No:

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

BTECH
(SEM VI) THEORY EXAMINATION 2024-25
BIG DATA

TIME: 3 HRS**M.MARKS: 100****Note:** Attempt all Sections. In case of any missing data; choose suitably.**SECTION A****1. Attempt all questions in brief. 2 x 10 = 20**

Q No.	Question	CO	Level
a.	Define Big Data. List and explain its 5 V's.	1	K2
b.	Differentiate between structured, semi-structured, and unstructured data with examples.		
c.	What is HDFS? Mention its key features.		
d.	What are the key components of Hadoop architecture?		
e.	Define MapReduce. What are its main phases?		
f.	What is the role of YARN in Hadoop 2.0 architecture?		
g.	Define NoSQL. How is it different from traditional RDBMS?		
h.	What is an RDD in Apache Spark?		
i.	Write a short note on Pig Latin and its data processing operators.		
j.	What is the role of ZooKeeper in managing distributed systems?		

SECTION B**2. Attempt any three of the following: 10 x 3 = 30**

a.	Explain the architecture of Hadoop and its working with the help of a neat diagram.	1	K2
b.	Describe the design and features of HDFS. How does it ensure fault tolerance?		
c.	Discuss the challenges faced by conventional systems and how Big Data technologies overcome them.		
d.	Explain the working of MapReduce with a suitable real-life example.		
e.	Compare Hive and Pig. Highlight their architecture and suitable use cases.		

SECTION C**3. Attempt any one part of the following: 10 x 1 = 10**

a.	Discuss the evolution and importance of Big Data analytics with examples.	1	K2
b.	What are the ethical issues related to Big Data? How can they be addressed?		

4. Attempt any one part of the following: 10 x 1 = 10

a.	Explain the anatomy of a MapReduce job run with proper steps.	1	K2
b.	What are different input and output formats supported by MapReduce?		

5. Attempt any one part of the following: 10 x 1 = 10

a.	Describe the architecture and working of Apache Spark.	1	K2
b.	What is the execution model of Spark jobs? Explain jobs, stages, and tasks.		

6. Attempt any one part of the following: 10 x 1 = 10

a.	What are the key differences between HBase and RDBMS? Discuss their use cases.	1	K2
b.	Explain the working of HBase with clients and schema design.		

7. Attempt any one part of the following: 10 x 1 = 10

a.	What is IBM BigInsights? Explain IBM's strategy for Big Data with examples.	1	K2
b.	Explain the concept of Big SQL and how it integrates with Hadoop ecosystem.		