Paper id: 250624

Roll No: | | | | | | | | | | | | |

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

TIME: 3 HRS   **M.MARKS: 70**

Note: Attempt all Sections. In case of any missing data; choose suitably.

SECTION A

1. **Attempt *all* questions in brief.**   02 x 7 = 14

Q no.	Question
a.	Define Big Data. What are its key characteristics.
b.	Define the Hadoop Distributed File System (HDFS). How is it different from traditional file systems?
c.	What do you mean by MapReduce? List its main phases.
d.	Describe how compression techniques and file formats impact the performance of big data processing systems
e.	What is the role of file system interfaces in big data storage environments
f.	Explain briefly the role of YARN in the Hadoop ecosystem.
g.	What are the key differences between Hive and traditional SQL databases

SECTION B

2. **Attempt any *three* of the following:**   7 x 3 = 21

a.	Describe its Big Data's evolution over time. Discuss the key drivers behind the rise of Big Data.
b.	What is the role of Big Data analytics in intelligent data analysis.
c.	Explain the architecture and use cases of Apache Flume and Apache Sqoop in Big Data systems. How do these tools support efficient data ingestion from structured and unstructured sources
d.	Evaluate the trade-offs between real-time and batch processing in distributed systems. How do technologies like YARN and Spark support both paradigms
e.	Evaluate the use of Hive, Pig, and HBase for different Big Data application scenarios. When would you choose one over the others?

SECTION C

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

a.	Describe the importance and applications of Big Data in different industries such as healthcare, finance, and e-commerce. How does Big Data contribute to decision-making in these sectors?
b.	What are the major concerns related to Big Data privacy and ethics? Discuss how auditing and compliance play a role in addressing these concerns.

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

a.	Discuss the process of job scheduling and task execution in Hadoop. How does Hadoop ensure fault tolerance during task execution?
b.	What are HDFS block size and block abstraction? How do they contribute to performance in Big Data systems

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

a.	Design a data ingestion and transformation pipeline for a real-time social media analytics system using tools like Flume, Sqoop, and Avro.
b.	Explain the architecture and use cases of Apache Flume and Apache Sqoop in Big Data systems. How do these tools support efficient data ingestion from structured and unstructured sources.

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

a.	Evaluate the use of NoSQL databases in real-time big data applications. Discuss their strengths and weaknesses compared to SQL databases.

Paper id: 250624

Roll No: | | | | | | | | | | | | |

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

TIME: 3 HRS   **M.MARKS: 70**

b.	Evaluate the performance benefits and limitations of Spark over traditional MapReduce in big data processing.

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

a.	Describe HBase and its data model. How does HBase differ from traditional relational databases, and what types of applications benefit most from using HBase
b.	Explain the process of data loading and querying in Hive. How does HiveQL differ from standard SQL, and what are its advantages