PAPER ID-410559

Subject Code: KCS061

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

BTECH
(SEM VI) THEORY EXAMINATION 2023-24
BIG DATA

TIME: 3 HRS **M.MARKS: 100**

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

SECTION A

1. Attempt *all* questions in brief. **2 x 10 = 20**

Q no.	Question	Marks	CO
a.	What are the different types of digital data commonly encountered in Big Data applications? Provide examples of structured, semi-structured, and unstructured data.	02	1
b.	What constitutes a Big Data platform?	02	1
c.	What is Hadoop Streaming?	02	2
d.	Discuss the data formats commonly used in Hadoop environments.	02	2
e.	Describe the concepts of file sizes, block sizes, and block abstraction in HDFS.	02	3
f.	What are the benefits and challenges of using HDFS for distributed storage and processing?	02	3
g.	What are the characteristics and use cases for schedulers such as Fair Scheduler and Capacity Scheduler?	02	4
h.	What is YARN?	02	4
i.	What is Apache Pig?	02	5
j.	Describe the Grunt shell in Apache Pig.	02	5

SECTION B

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

a.	Distinguish between data analysis and reporting in the context of Big Data. How does advanced analytics go beyond traditional reporting to uncover hidden patterns, trends, and correlations in data?	10	1
b.	Explain Apache Hadoop and its role in big data processing. What are the core components of the Apache Hadoop ecosystem, and how do they work together to enable distributed data storage and processing?	10	2
c.	Explain the core concepts of HDFS, including NameNode, DataNode, and the file system namespace. How do these components work together to manage data storage and replication in Hadoop clusters?	10	3
d.	Define NoSQL databases. What are the key characteristics and benefits of NoSQL databases compared to traditional relational databases?	10	4
e.	Provide an overview of Apache Hive architecture and its components. How does Hive translate SQL-like queries into MapReduce jobs for data processing in Hadoop?	10	5

SECTION C

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

a.	Describe the "5 Vs" of Big Data. What do each of these terms represent in the context of Big Data, and why are they essential considerations for data management and analysis?	10	1

PAPER ID-410559

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

BTECH
(SEM VI) THEORY EXAMINATION 2023-24
BIG DATA

TIME: 3 HRS **M.MARKS: 100**

b.	Provide examples of real-world applications where Big Data analytics have been instrumental. How do industries such as healthcare, finance, e-commerce, and transportation leverage Big Data to gain insights and create value?	10	1

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

a.	Describe the Hadoop Distributed File System (HDFS). How does HDFS manage the storage and replication of data across a distributed cluster of machines?	10	2
b.	Discuss the process of developing a MapReduce application. What are the key steps involved in writing, testing, and deploying a MapReduce program?	10	2

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

a.	Explain how HDFS stores, reads, and writes files. Describe the sequence of operations involved in storing a file in HDFS, retrieving data from HDFS, and writing data to HDFS.	10	3
b.	Describe the considerations for deploying Hadoop in a cloud environment. What are the advantages and challenges of running Hadoop clusters on cloud platforms like Amazon Web Services (AWS), Microsoft Azure, and Google Cloud Platform (GCP)?	10	3

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

a.	Explain the operations for creating, updating, and deleting documents in MongoDB. What are the MongoDB CRUD operations, and how are they used to manipulate data in collections?	10	4
b.	Discuss Resilient Distributed Datasets (RDDs) in Spark. What are RDDs, and how do they enable fault-tolerant and distributed data processing in Spark applications?	10	4

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

a.	Introduce the concepts of HBase and its role in the Hadoop ecosystem. How does HBase differ from traditional relational databases, and what advantages does it offer for storing and accessing large-scale data?	10	5
b.	Discuss the HiveQL language used in Apache Hive. How does HiveQL support SQL-like syntax for defining tables, querying data, and performing data manipulation operations?	10	5