

				Printed Page: 1 of Subject Code: KCS06			of 2		
				· ·				6061	
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	02	1
	Big Data applications? Provide examples of structured, semi-structured,		
	and unstructured data.		
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	02	3
	HDFS.		
f.	What are the benefits and challenges of using HDFS for distributed	02	3
	storage and processing?		
g.	What are the characteristics and use cases for schedulers such as Fair	02	4
	Scheduler and Capacity Scheduler?		
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	10	1
	Data. How does advanced analytics go beyond traditional reporting to		
	uncover hidden patterns, trends, and correlations in data?		
b.	Explain Apache Hadoop and its role in big data processing. What are	10	2
	the core components of the Apache Hadoop ecosystem, and how do they		
	work together to enable distributed data storage and processing?		
c.	Explain the core concepts of HDFS, including NameNode, DataNode,	10	3
	and the file system namespace. How do these components work together		
	to manage data storage and replication in Hadoop clusters?		
d.	Define NoSQL databases. What are the key characteristics and benefits	10	4
	of NoSQL databases compared to traditional relational databases?		
e.	Provide an overview of Apache Hive architecture and its components.	10	5
	How does Hive translate SQL-like queries into MapReduce jobs for data		
	processing in Hadoop?		

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	10	1
	in the context of Big Data, and why are they essential considerations for		
	data management and analysis?		



				e				2 of 2	,	
				Printed Page: 2 of 2 Subject Code: KCS061			5061			
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	10	1	
	have been instrumental. How do industries such as healthcare, finance,			
	e-commerce, and transportation leverage Big Data to gain insights and create value?			

4.	Attempt any one part of the following:	1 x 10	= 10
a.	Describe the Hadoop Distributed File System (HDFS). How does HDFS	10	2
	manage the storage and replication of data across a distributed cluster of		
	machines?		
b.	Discuss the process of developing a MapReduce application. What are	10	2
	the key steps involved in writing, testing, and deploying a MapReduce		
	program?		

5.	Attempt any one part of the following:	1 x 10	=10
a.	Explain how HDFS stores, reads, and writes files. Describe the sequence	10	3
	of operations involved in storing a file in HDFS, retrieving data from		
	HDFS, and writing data to HDFS.		
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)?	. 9	3
	O'V	D .	

6.	Attempt any one part of the following:	1 x 10	= 10
a.	Explain the operations for creating, updating, and deleting documents in	10	4
	MongoDB. What are the MongoDB CRUD operations, and how are		
	they used to manipulate data in collections?		
b.	Discuss Resilient Distributed Datasets (RDDs) in Spark. What are	10	4
	RDDs, and how do they enable fault-tolerant and distributed data		
	processing in Spark applications?		

7.	Attempt any <i>one</i> part of the following:	1 x 10	= 10
a.	Introduce the concepts of HBase and its role in the Hadoop ecosystem.	10	5
	How does HBase differ from traditional relational databases, and what		
	advantages does it offer for storing and accessing large-scale data?		
b.	Discuss the HiveQL language used in Apache Hive. How does HiveQL	10	5
	support SQL-like syntax for defining tables, querying data, and		
	performing data manipulation operations?		