SHAMBHUNATH INSTITUTE OF ENGINEERING AND TECHNOLOGY

Subject Code: BCS-061 **Subject: BIG DATA**

Course: B.Tech. **Semester: 6th**

SECOND SESSIONAL EXAMINATION, EVEN SEMESTER, (2024-2025)

Branch: Computer Science & Engineering

Time–1hr **MaximumMarks–15**

NOTE :(Attempt all sections)

1. Attempt ALL questions.

QN	QUESTION	Marks	CO	BL
a.	What are the core components of Hadoop?	1	CO2	L2
b.	What is the purpose of Hadoop Streaming?	1	CO2	L3
c.	What is the default data format in Hadoop?	1	CO2	L1
d.	How does the shuffle and sort phase work in MapReduce?	1	CO2	L2
e	What are some input formats used in MapReduce?	1	CO2	L2

2. Attempt any *ONE* of the following.

QN	QUESTION	Marks	CO	BL
a.	Explain Hadoop Ecosystem in detail.	5	CO2	L3
b.	Explain the Hadoop Distributed File System (HDFS) architecture and its key components.	5	CO2	L4

3. Attempt any ONE of the following

QN	QUESTION	Marks	CO	BL
a.	Discuss the detailed architecture of Map-Reduce?	5	CO2	L3
b.	Explain different failure scenarios in a MapReduce job run. How does Hadoop handle failures at different stages of the job execution?	5	CO2	L3

Bloom's Taxonomy Level(BL): -

Remember(L1), Understanding(L2), Apply(L3), Analyze(L4), Evaluating(L5), Creating(L6)

Roll No. 1 2 0 1 6 2 0 5 2 0 0 8

SHAMBHUNATH INSTITUTE OF ENGINEERING AND TECHNOLOGY,PRAYAGRA

Subject Code: -BCS-061 **Subject: BIG DATA**

Course: B.Tech. **Semester: VIth**

THIRD SESSIONAL EXAMINATION, EVEN SEMESTER, (2024-2025)

Branch: COMPUTER SCIENCE & ENGINEERING

Time – 2Hr **MaximumMarks–45**

1. Attempt any Five questions in brief.

QN	QUESTION	Marks	CO	BL
a.	What is the default block size in HDFS, and why is it large?	2	CO3	L2
b.	What is Hadoop benchmarking, and why is it important?	2	CO3	L1
c.	What is the purpose of Sqoop in Hadoop?	2	CO3	L1
d.	What is data replication in HDFS, and why is it important?	2	CO3	L2
e.	What is HDFS monitoring, and which tools are used for it?	2	CO3	L3
f.	What are the security features in Hadoop?	2	CO3	L3

2. Attempt Any ONE of the following.

QN	QUESTION	Marks	CO	BL
a.	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?	5	CO3	L3
b.	Describe the considerations for deploying Hadoop in a cloud environment. What are the advantage and challenges of running Hadoop clusters on cloud platforms like Amazon Web Services(AWS), Microsoft Azure and Google Cloud Platform(GCP)?	5	CO3	L3
c.	Explain the process of reading and writing files in HDFS. How do Java interfaces and the command-line interface facilitate HDFS operations?	5	CO3	L3

3. Attempt any Five questions in brief.

QN	QUESTION	Marks	CO	BL
a.	What is MRv2 in Hadoop?	2	CO4	L2
b.	What is a capped collection in MongoDB?	2	CO4	L1

c.	What are the characteristics and use cases for schedulers such as Fair scheduler and Capacity scheduler?	2	CO4	L2
d.	Does MongoDB support ACID properties? Justify your answer.	2	CO4	L3
e.	What is HDFS Federation?	2	CO4	L1
f.	Name two components of the Hadoop ecosystem.	2	CO4	L1

. Attempt Any ONE of the following.

QN	QUESTION	Marks	CO	BL
a.	Discuss Resilient Distributed Datasets in Spark. What are RDDs and how do they enable fault-tolerant and distributed data processing in spark applications?	5	CO4	L3
b.	With the help of suitable example explain how CRUD operations are performed in MongoDB	5	CO4	L3
c.	Difference between Java and Scala. Also explain various features of Scala	5	CO4	L3

. Attempt any Five questions in brief.

QN	QUESTION	Marks	CO	BL
a.	What is Grunt in Pig?	2	CO5	L2
b.	Differentiate between Pig and MapReduce?	2	CO5	L2
c.	Discuss meta store in HIVE in brief ?	2	CO5	L2
d.	Discuss the different types of data can be handled with HIVE?	2	CO5	L3
e.	Name any two components of IBM's Big Data Strategy.	2	CO5	L2
f.	What is Big SQL?	2	CO5	L2

. Attempt Any ONE of the following.

QN	QUESTION	Marks	CO	BL
a.	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?	5	CO5	L3
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?	5	CO5	L3
c.	Introduce the concepts of HBase and its role in the Hadoop ecosystem. How does Hbase differ from traditional relational database and what advantage does it offer for storing and accessing large-scale data?	5	CO5	L4

loom's Taxonomy Level (BL):-

Remember(L1), Understanding(L2), Apply(L3), Analyze(L4), Evaluating(L5), Creating(L6)