SREENIVASA INSTITUTE OF TECHNOLOGY AND MANAGEMENT STUDIES
(AUTONOMOUS)

QUESTION BANK

Year / Semester: IV B.Tech VII Semester **Regulation: R23**

Subject and Code: BIG DATA ANALYTICS(23CSD471T)

UNIT–1: INTRRODUCTION TO BIG DATA

Introduction to Big Data platform- What is Big Data? Big Data Sources-Acquisition-Nuts and Bolts of Big data-Features of Big Data-Security - Compliance - auditing and protection- Evolution of Big Data- Best practices for Big Data Analytics-Big Data characteristics- Volume

- Veracity - Velocity - Variety- Structure of Big Data- Exploring the opportunities with Big Data.

UNIT–2: APACHE HADOOP

Introduction to Hadoop- A brief history of Hadoop - Apache Hadoop and The Hadoop Ecosystem AND Hadoop Features - fault tolerance , with data replication, Distributed, parallel processing , High availability, Data locality - The Design of HDFS-HDFS concepts.

UNIT–3: INTRODUCTION TO HIVE AND PIG

Introduction to Hive, data types and file formats, HiveQL data definition- Table creation- Internal table -External table, HiveQL data manipulation, Logical joins, Window functions- rank-dense-rank- row number- Table partitioning-static partitioning -dynamic partitioning

- Bucketing.

Pig- Pig Latin-Statements-Expressions-Types- Pig Architecture-Data processing Operators- Loading and storing Data-Filtering Data-Grouping and Joining Data-Sorting Data-Combining and Splitting Data

UNIT–4: HBASE- ZOOKEEPER – SQOOP

HBasics – Concepts – Example-HBase Versus RDBMS HBase, Hbase architecture, data model, operations—Introduction to Zookeeper-Zookeeper Services-Building applications with Zookeeper - Introduction to Sqoop-Database Imports-Working with Imported data- Importing large objects-performing exports.

UNIT–5: APACHE SPARK

Apache spark- Advantages over Hadoop, lazy evaluation, In memory processing, DAG, Spark context, Spark Session, RDD, Transformations- Narrow and Wide, Actions, Data frames

,RDD to Data frames, Catalyst optimizer, Data Frame Transformations, Working with Dates and Timestamps, Working with Nulls in Data, Working with Complex Types, Working with JSON, Grouping, Window Functions, Joins, Data Sources, Broadcast Variables, Accumulators.

Prepared by Mr.G.YUVARAJU

SREENIVASA INSTITUTE OF TECHNOLOGY AND MANAGEMENT STUDIES
(AUTONOMOUS)

QUESTION BANK

Year / Semester: IV B.Tech VII Semester **Regulation: R23**

Subject and Code: BIG DATA ANALYTICS(23CSD471T)

Max Marks: 10

S.No.	CO	Questions	BT
		Unit I:	
1	**1**	Explain the history and evolution of Hadoop. Discuss the need for Hadoop in Big Data processing.	L2
2	**1**	Describe the Hadoop ecosystem and explain the role of its major components.	L3
3	**1**	Explain the features of Hadoop, including fault tolerance, data replication, and high availability.	L4
4	**1**	Explain distributed and parallel processing in Hadoop with suitable examples.	L4
5	**1**	Describe the concept of data locality in Hadoop and explain its advantages.	L3
6	**1**	Explain the architecture and design of Hadoop Distributed File System (HDFS).	L4
7	**1**	Explain the HDFS architecture and discuss the roles of NameNode, DataNode, and Secondary NameNode.	L4
8	**1**	Discuss the HDFS concepts such as blocks, replication, rack awareness, and fault tolerance.	L5
9	**1**	Compare traditional file systems with HDFS and justify why HDFS is suitable for Big Data applications.	L5
10	**1**	Analyze the working of HDFS during file read and write operations with neat diagrams.	L6
S.No.	**CO**	**Questions**	**BT**
		Unit II	
1	**2**	Explain the history and evolution of Hadoop. Discuss the need for Hadoop in Big Data processing.	L2

SREENIVASA INSTITUTE OF TECHNOLOGY AND MANAGEMENT STUDIES
(AUTONOMOUS)

QUESTION BANK

Year / Semester: IV B.Tech VII Semester **Regulation: R23**

Subject and Code: BIG DATA ANALYTICS(23CSD471T)

2	**2**	Describe the Hadoop ecosystem and explain the role of its major components.	L3
3	**2**	Explain the features of Hadoop, including fault tolerance, data replication, and high availability.	L4
4	**2**	Explain distributed and parallel processing in Hadoop with suitable examples.	L4
5	**2**	Describe the concept of data locality in Hadoop and explain its advantages.	L3
6	**2**	Explain the architecture and design of Hadoop Distributed File System (HDFS).	L4
7	**2**	Explain the HDFS architecture and discuss the roles of NameNode, DataNode, and Secondary NameNode.	L4
8	**2**	Discuss the HDFS concepts such as blocks, replication, rack awareness, and fault tolerance.	L5
9	**2**	Compare traditional file systems with HDFS and justify why HDFS is suitable for Big Data applications.	L5
10	**2**	Analyze the working of HDFS during file read and write operations with neat diagrams.	L6
S. No.	**CO**	**Questions**	**BT**
Unit III			
1	**3**	Explain the architecture of Hive and discuss its advantages in Big Data analytics.	L3
2	**3**	Explain Hive data types and file formats with suitable examples.	L3
3	**3**	Explain HiveQL Data Definition Language (DDL) and demonstrate table creation using internal and external tables.	L4
4	**3**	Explain HiveQL Data Manipulation Language (DML) with suitable examples.	L4
5	**3**	Explain different types of joins in Hive with suitable examples.	L4
6	**3**	Explain Hive window functions (RANK(), DENSE_RANK(), ROW_NUMBER()) with suitable examples.	L5

7	**3**	Explain static partitioning, dynamic partitioning, and bucketing in Hive with examples.	L5
8	**3**	Explain the architecture of Apache Pig and discuss the features of Pig Latin.	L3
9	**3**	Explain Pig Latin statements, expressions, and data processing operators with suitable examples.	L4
10	**3**	Explain loading, filtering, grouping, joining, sorting, combining, and splitting data in Pig with suitable examples.	L6

S.No.	CO	Questions	BT
Unit IV:			
1	**4**	Explain the basic concepts and architecture of HBase with a neat diagram.	L3
2	**4**	Compare HBase and RDBMS based on architecture, storage model, and applications.	L5
3	**4**	Explain the HBase data model and discuss CRUD operations with suitable examples.	L4
4	**4**	Explain the architecture and components of HBase.	L4
5	**4**	Explain the role of ZooKeeper in distributed systems and discuss its services.	L3
6	**4**	Describe how applications are built using ZooKeeper with suitable examples.	L4
7	**4**	Explain the architecture and working of Apache Sqoop.	L3
8	**4**	Explain database import and export operations using Sqoop with suitable examples.	L4
9	**4**	Explain how Sqoop imports large objects and manages imported data efficiently.	L5
10	**4**	Analyze the integration of HBase, ZooKeeper, and Sqoop in a Big Data ecosystem.	L6

SREENIVASA INSTITUTE OF TECHNOLOGY AND MANAGEMENT STUDIES
(AUTONOMOUS)

QUESTION BANK

Year / Semester: IV B.Tech VII Semester **Regulation: R23**

Subject and Code: BIG DATA ANALYTICS(23CSD471T)

S.No.	CO	Questions	BT
		Unit V:	
1	**5**	Explain Apache Spark architecture and discuss its advantages over Hadoop MapReduce.	L3
2	**5**	Explain lazy evaluation and in-memory processing in Apache Spark with suitable examples.	L4
3	**5**	Explain Directed Acyclic Graph (DAG), SparkContext, and SparkSession in Spark.	L4
4	**5**	Explain Resilient Distributed Datasets (RDDs) and discuss different types of transformations and actions.	L4
5	**5**	Differentiate between narrow and wide transformations in Spark with suitable examples.	L5
6	**5**	Explain DataFrames in Spark and demonstrate the conversion of RDDs into DataFrames.	L4
7	**5**	Explain the Catalyst Optimizer and discuss its role in improving Spark SQL performance.	L5
8	**5**	Explain DataFrame transformations involving dates, timestamps, null values, complex data types, and JSON data.	L4
9	**5**	Explain grouping, window functions, joins, and different data sources in Apache Spark with examples.	L5
10	**5**	Explain the concepts of Broadcast Variables and Accumulators in Spark with suitable examples.	L6

Note: L1-Remembering, L2-Understanding, L3-Applying, L4-Analyzing, L5-Evaluating, and L6-Creating

SREENIVASA INSTITUTE OF TECHNOLOGY AND MANAGEMENT STUDIES
(AUTONOMOUS)

QUESTION BANK

Year / Semester: IV B.Tech VII Semester **Regulation: R23**

Subject and Code: BIG DATA ANALYTICS(23CSD471T)

Instruction to Faculty Members:

The Six Levels of Bloom's Taxonomy:

1. **Remembering:** Retrieving, recognizing, and recalling relevant knowledge from long-term memory (e.g., list, define, name, locate).
2. **Understanding:** Constructing meaning, explaining ideas, or concepts (e.g., summarize, interpret, classify, compare).
3. **Applying:** Using information in new situations or implementing procedures to solve problems (e.g., solve, use, demonstrate, implement).
4. **Analyzing:** Breaking material into constituent parts, determining how the parts relate to one another and to an overall structure (e.g., contrast, categorize, distinguish, diagram).
5. **Evaluating:** Making judgments based on criteria and standards through checking and critiquing (e.g., judge, critique, justify, defend, argue).
6. **Creating:** Putting elements together to form a coherent or functional whole; reorganizing elements into a new pattern or structure (e.g., design, construct, develop, formulate).