SREENIVASA INSTITUTE OF TECHNOGY AND MANAGEMENT STUDIES

(AUTONOMOUS)

Murukambattu, Chittoor

MCA DEPARTMENT

QUESTION BANK

For

24MCA213 -Big Data Analytics

Regulation – 24

Academic Year 2025 – 26

Prepared by

Mr. N. P. Gangadhar, Assistant Professor

ISREENIVASA INSTITUTE OF TECHNOLOGY AND MANAGEMENT STUDIES.
(AUTONOMOUS)
MCA DEPARTMENT
QUESTION BANK

SUBJECT NAME :Big Data Analytics **SUBJECT CODE : 24MCA213**

YEAR &SEM : II & I **Academic Year : 2025-26**

UNIT I - Understanding Big Data		
Datasets, Data Analysis, Data Analytics-Descriptive Analysis, Diagnostics Analytics, Predictive Analytics, Prescriptive Analytics, Big Data Characteristics – volume, velocity, variety, veracity, value, Different Types of Data – Structured Data, Unstructured Data, Semi-Structured Data		
PART -A		
Q.No.	**Questions**	**Blooms Taxonomy Level**
1.	Define Dataset	Understanding
2.	Distinguish between Data Analysis and Data Analytics	Analyzing
3.	Mention Different types of Data Analytics	Remembering
4.	Explain Descriptive Analytics with an example	Understanding
5.	Distinguish between Predictive and Prescriptive Analytics	Analyzing
6.	Define Big Data	Understanding
7.	List out the Big Data Characteristics	Remembering
8.	Mention different types of Data	Remembering
9.	Mention some of Big Data Examples	Remembering
10.	Give an example for structured and unstructured data	Understanding
PART –B		
1.	Explain Different types of Data Analytics with examples	Understanding
2.	What is BigData and Discuss different characteristics of Big Data	Understanding
3.	Illustrate different types of Data	Applying
4.	Differentiate Descriptive and Diagnostics analytics with example	Analyzing
5.	Differentiate Structured, unstructured and semi structured data in detail	Analyzing
6.	Explain Dataset, Data Analysis and Data analytics in detail	Understanding
UNIT II - Hadoop Basics		
Brief history of hadoop, Apache hadoop and the hadoop ecosystem. A weather dataset, analyzing the data with unix tools, analyzing the data with hadoop , Understanding different Hadoop modes, understanding Hadoop Features-Understanding HDFS, Understanding MapReduce, Learning the HDFS and Mapreduce Architecture-Understanding the HDFS architecture, Understanding the MapReduce Architecture, Understanding the HDFS and MapReduce architecture by plot.		
PART -A		
1.	What is Hadoop	Understanding
2.	List out various Hadoop ecosystem	Remembering
3.	Mention different Hadoop Modes	Remembering
4.	List out the Hadoop Features	Remembering
5.	Mention two important components of Hadoop	Remembering
6.	What is HDFS	Understanding
7.	What is MapReduce	Understanding
8.	What is Name Node	Understanding
9.	What is Data Node	Understanding
10.	List out Map Reduce Model Phases	Remembering
PART -B		
1.	Write about Brief history of Hadoop	Understanding
2.	Explain Apache Hadoop and the Hadoop Ecosystem	Understanding
3.	List and explain various Features of Hadoop	Remembering
4.	Differentiate how to analyze the weather dataset in unix and hadoop	Analyzing

SUBJECT NAME :Big Data Analytics **SUBJECT CODE : 24MCA213**

YEAR &SEM : II & I **Academic Year : 2025-26**

5.	Explain HDFS and Map Reduce in detail	Understanding
6.	Write about HDFS and Map Reduce Architecture	Understanding
7.	Write about Hadoop History and different Hadoop Modes	Understanding
8.	Illustrate Map Reduce Model with word count example	Understanding
9.	Explain Map Reduce Model with real time example	Understanding
10.	Discuss the need of HDFS with an example	Creating
UNIT III – Working with Pig and Hive		
Pig -Execution Types, An Example, Pig Latin-Structure, Statements,Types, Schemas, Functions, Data Processing Operators.Hive – An example, Tables –Managed Tables and External tables, Partitions and Buckets, Importing data, Altering Data, Dropping Tables, Querying Data-Sorting and Aggregating, Mapreduce Scripts.		
PART – A		
1.	What is Pig	Understanding
2.	List out the Pig Execution Types	Remembering
3.	Mention the Language used in pig	Remembering
4.	List out the Pig Data Types	Remembering
5.	List out two eval functions of pig	Remembering
6.	Purpose of LOAD operator	Understanding
7.	What is hive	Understanding
8.	Mention 2 types of Hive Tables	Remembering
9.	Write Two ways to import data into hive tables	Remembering
10.	Write about Alter statement to add columns	Remembering
11.	Which clause is used to partition Tables	Remembering
12.	Which clause is used to sort the table data	Remembering
13.	Which clause is used to do bucketization	Remembering
14.	Explain the main advantage of partitions and buckets	Understanding
15.	Write the syntax to drop a table in hive	Understanding
PART – B		
1.	Write a Pig Script to analyze weather data set	Understanding
2.	Discuss the Eval and string functions of Pig	Creating
3.	Write about Load and Store operators	Understanding
4.	Explain Math and String functions in pig	Understanding
5.	Discuss Pig Data Processing Operators with example	Creating
6.	Analyze weather data using Hive statements	Analyzing
7.	Create an Internal Table, populate it with data and fetch the data	Creating
8.	Discuss about partitions and Buckets	Creating
9.	Write about alter and drop statements of hive	Understanding
10.	Discuss hive tables with examples	Understanding
11.	Explain the hive partitions with a detailed example	Understanding
12.	Discuss about importing data in hive	Creating
UNIT IV- INTRODUCTION TO APACHE SPARK		
What is Apache Spark, Characteristics – Speed, Easy of Use, Modularity, Extensibility, Apache Spark Components as a Unified Stack, Spark Basic Data Types, Spark Core and SQL- Dynamic Partition Pruning, SQL Operations, Schemas and Creating Data Frames		

ISREENIVASA INSTITUTE OF TECHNOLOGY AND MANAGEMENT STUDIES.
(AUTONOMOUS)
MCA DEPARTMENT
QUESTION BANK

SUBJECT NAME :Big Data Analytics **SUBJECT CODE : 24MCA213**

YEAR &SEM : II & I **Academic Year : 2025-26**

PART – A		
1.	What is Apache Spark?	Remember
2.	What are the main advantages of Apache Spark over Hadoop MapReduce?	Understand
3.	Define Resilient Distributed Dataset (RDD).	Remember
4.	What are DataFrames and Datasets in Spark?	Understand
5.	Why is Spark considered easy to use?	Understand
6.	What is modularity in Apache Spark?	Understand
7.	How does Apache Spark achieve high speed in data processing?	Analyze
8.	Name the major components of Spark's unified stack.	Remember
9.	What is the function of Spark Core?	Understand
10.	What is Spark SQL used for?	Understand
11.	What is the purpose of Spark Streaming?	Understand
12.	List any four basic data types supported in Spark.	Remember
13	What is Dynamic Partition Pruning (DPP) in Spark SQL?	Understand
14	Why is Dynamic Partition Pruning important?	Analyze
15	What is a schema in Spark?	Understand
16	How can you create a DataFrame with an explicit schema?	Apply
PART – B		
1.	Explain Apache Spark. How is it different from Hadoop MapReduce?	Understand
2.	Describe the architecture of Apache Spark with its key components.	Understand
3.	What are RDDs, DataFrames, and Datasets? Compare them with examples.	Analyze
4.	Explain the different components of Apache Spark as a unified stack.	Understand
5.	Discuss the role of Spark Core in managing distributed data processing.	Understand
6.	Describe the basic data types supported in Spark SQL.	Remember
7.	Explain the concept of Dynamic Partition Pruning in Spark SQL.	Understand
8.	Compare Static Partition Pruning and Dynamic Partition Pruning in Spark.	Analyze
9.	Explain the process of running SQL queries on Spark DataFrames.	Apply
10.	Discuss how Spark integrates SQL queries with structured and semi-structured data sources.	Analyze
11.	What is a schema in Spark? Why is schema enforcement important?	Understand
12.	Explain different ways to create DataFrames in Spark (RDDs, external files, schema definition).	Apply/Create
UNIT V – HBase, Zookeeper		
HBase Overview – Limitations of Hadoop, what is HBase,HBase and HDFS, Storage Mechansim in HBase,Features of HBase, Applications of HBase. ZooKeeper Overview – what is ZooKeeper, Distributed Application, Benefits of Distributed Applications,Challenges of Distributed Applications, What is Apache Zookeeper meant for, Benefits of ZooKeeper.		
PART – A		
1.	What is Hbase	Understanding
2.	What is zookeeper	Understanding
3.	What is sqoop	Understanding
4.	List out the limitations of Hadoop	Remembering
5.	Mention the features of HBase	Remembering
6.	Define distributed application	Remembering

SUBJECT NAME :Big Data Analytics **SUBJECT CODE : 24MCA213**

YEAR &SEM : II & I **Academic Year : 2025-26**

7.	Mention the challenges of distributed applications	Remembering
8.	Differentiate Hbase and HDFS	Analyzing
9.	List out the Benefits of Zookeeper	Remembering
	PART – B	
1.	Discuss the Limitations of Hadoop and how they are overcomed with HBase	Understanding
2.	Write about the Hbase and HDFS in detail	Understanding
3.	Describe the storage mechanism of Hbase in detail	Understanding
4.	Explain the features and Applications of HBase	Understanding
5.	Write about distributed application and benefits of distributed applications	Understanding
6.	What is apache zookeeper meant for	Analyzing
7.	Describe the benefits and challenges of zookeeper	Understanding
8.	Write about limitations, applications and storage mechanism of Apache hbase	Understanding