SREENIVASA INSTITUTE OF TECHNOGY AND MANAGEMENT STUDIES

(AUTONOMOUS)

Murukambattu, Chittoor

CSE-DATA SCIENCE DEPARTMENT

QUESTION BANK

For

20CSE361 -Big Data Analytics

Regulation – 2020

Academic Year 2024 – 25

Prepared by

Mr.G.YUVARAJU Assistant Professor

ISREENIVASA INSTITUTE OF TECHNOLOGY AND MANAGEMENT STUDIES.
(AUTONOMOUS)
CSE-DS DEPARTMENT
QUESTION BANK

SUBJECT NAME :Big Data Analytics **SUBJECT CODE : 20CSE361**

YEAR &SEM : III & I **Academic Year : 2024-25**

UNIT I - INTRODUCTION 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

PART –A

Q.No.	Questions	Blooms Taxonomy Level
1.	Define Dataset	Understanding
2.	Define Big Data	Analyzing
3.	List out the Big Data Characteristics	Remembering
4.	Mention different types of Data	Understanding
5.	Mention some of Big Data Examples	Analyzing
6.	Give an example for structured and unstructured data	Understanding
7.	What is semi-structured data? Give an example.	Remembering
8.	What is data compliance in the context of Big Data?	Remembering
9.	What is the role of distributed storage in Big Data?	Remembering
10.	Why Big Data important in today's world?	Understanding

PART –B

1.	Explain 5v's in Bigdata	
2.	What is BigData and Discuss different characteristics of Big Data	Understanding
3.	Explore the opportunities of big data	Analyzing
4.	Explain Big Data Sources-Aquisition-Nuts and Bolts of Big data	Analyzing
5.	Differentiate Structured, unstructured and semi structured data in detail	Analyzing
6.	Explain about Evolution of Bigdata.	Understanding
7	Explain about Features of Bigdata.	Understanding
8	Explain Security - Compliance - auditing and protection in big data	Understanding

UNIT II - HADOOP ECOSYSTEM AND YARN

Introduction to Hadoop-Data Storage and Analysis- Comparison with Other Systems - A brief history of Hadoop - Apache Hadoop and The Hadoop Ecosystem - The Hadoop Distributed File System - The Design of HDFS-HDFS concepts - The Command Line Interface.Hadoop ecosystem components - , Hadoop File systems -Schedulers - Fair and Capacity – Hadoop New Features- NameNode High Availability, HDFS Federation, MRv2, YARN, Running MRv1 in YARN.

PART –A

1.	What is Hadoop	Understanding
2.	List out various Hadoop ecosystem Components	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 YARN	Understanding
8.	What is NameNode	Understanding
9.	Describe HDFS Federation	Understanding
10.	List out types of Schedulers in Hadoop	Remembering

SUBJECT NAME :Big Data Analytics **SUBJECT CODE : 20CSE361**

YEAR &SEM : III & I **Academic Year : 2024-25**

11	List out Hadoop new Features	Remembering
12	What is the role of **MRv2 (MapReduce Version 2)** in YARN.	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.	Write about Hadoop History and different Hadoop Modes	Analyzing

5.	Discuss the need of HDFS with an example	Understanding
6.	Explain types of schedulers in Hadoop	Understanding
7.	Explain Ecosystem components in Hadoop	Understanding
8.	Explain NameNode High Availability, HDFS Federation in Hadoop	Understanding
9.	Explain Yarn Architecture in Hadoop	Understanding
10.	Explain the role of **MRv2 ,MRv1** in YARN.	Understanding
	UNIT III - MAPREDUCE PROGRAMMING	
	Developing a Map Reduce Application - How Map Reduce Works - Anatomy of a Map Reduce Job run- Failures-Job Scheduling-Shuffle and Sort - Map Reduce Types and Formats - Map Reduce Features.	
	PART – A	
1.	What is Map Reduce	Understanding
2.	Mention Hadoop Map Reduce Entities	Remembering
3.	List out the Map Reduce Limitations	Remembering
4.	List out the steps to run a Map Reduce job	Remembering
5.	Mention major components of Map Reduce	Remembering
6.	Mention various steps in Hadoop MapReduce Scenario	Remembering
7.	List out the companies that use Map Reduce	Remembering
8.	Write about Map Reduce Objects	Understanding
9	Types of failures in MapReduce	Remembering
10	List out types of formats in MapReduce	Remembering
	PART –B	
1.	Understand the basics of Map Reduce	Understanding
2.	Explain Hadoop Map Reduce Scenario in detail	Understanding
3.	Explain Anatomy of Map Reduce in Detail	Understanding
4.	Illustrate Map Reduce Model with word count example	Analyzing
5	Explain the steps to run the MapReduce Job	Understanding
5.	Illustrate Map Reduce Model HealthCare dataset example	Analyzing
6.	Write about the Hadoop Map Reduce Data Processing stages in detail	Understanding
7.	Discuss the Map Reduce Example in detail	Understanding
8.	Describe one of the major Component of Hadoop called Map Reduce in detail	Understanding
9	Explain the limitations of map Reduce with example	Understanding
10.	Discuss Map Reduce entities with example	Understanding
	UNIT IV – Working with Pig and Hive	

SUBJECT NAME :Big Data Analytics **SUBJECT CODE : 20CSE361**

YEAR &SEM : III & I **Academic Year : 2024-25**

Installing and running pig- An Example- Comparison with Databases- Pig Latin Scripts-User defined functions-Data processing Operators-Pig in Practice.Installing Hive-Running Hive-Comparison with Traditional Databases – HiveQL – Tables-Querying Data

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 about Load and Store operators	Understanding
2.	Discuss Pig Data Processing Operators with example	nderstanding
3.	Create an Internal Table, populate it with data and fetch the data	Understanding
4.	Create an External Table, populate it with data and fetch the data	Understanding
5.	Write about alter and drop statements of hive	Creating
6.	Discuss hive tables with examples	Analyzing
7.	Compare Traditional databases in hive	Analyzing
8.	Explain user defined functions in hive	Understanding
9	Explain types of tables in hive ,different types of load data in hive	Analyzing

UNIT V – HBASE- ZOOKEEPER – SQOOP

HBasics – Concepts – Example-HBase Versus RDBMS-Praxis – Zookeeper-Installing and Running Zookeeper – Example-Zookeeper Services-Building applications with Zookeeper - Introduction to Sqoop-Database Imports-Working with Imported data-Importing large objects-performing exports.Case Study: Mahout, Spark MLLib, Apache Oozie and Apache Flume.

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
7.	Mention the challenges of distributed applications	Remembering
8.	Differentiate Hbase and HDFS	Analyzing
9.	List out the Benefits of Zookeeper	Remembering

ISREENIVASA INSTITUTE OF TECHNOLOGY AND MANAGEMENT STUDIES.
(AUTONOMOUS)
CSE-DS DEPARTMENT
QUESTION BANK

SUBJECT NAME :Big Data Analytics **SUBJECT CODE : 20CSE361**

YEAR &SEM : III & I **Academic Year : 2024-25**

PART – B		
1.	Explain the architecture and working of **HBase**.	Understanding
2.	How does it differ from traditional **RDBMS** in terms of data storage and retrieval	Understanding
3.	What is **Zookeeper**, and why is it used in distributed systems?	Understanding
4.	Explain its services and how it helps in **coordination and synchronization**.	Understanding
5.	Describe the steps involved in installing and running **Zookeeper**. Explain how a **Zookeeper ensemble** works.	Understanding
6.	What is **Sqoop**? Explain its architecture and how it is used to import data from relational databases to Hadoop	Analyzing
7.	Compare **Apache Mahout** and **Spark MLlib** in terms of **machine learning capabilities, performance, and scalability**. Which one is preferred for real-time analytics?	Understanding
8.	What are **Apache Oozie** and **Apache Flume**?	Understanding
9	How do they complement each other in **Big Data workflows and data ingestion pipelines**?	Analyzing
10	Differentiate RDMS and Hbase	Analyzing