**Course Code:** 20CS0538 **R20**

**SIDDHARTH INSTITUTE OF ENGINEERING & TECHNOLOGY:: PUTTUR**
**(AUTONOMOUS)**
Siddharth Nagar, Narayanavanam Road – 517583

**QUESTION BANK (DESCRIPTIVE)**

Subject with Code: **BIG DATA(20CS0538)** Course & Branch: **B.Tech – CSM**

Regulation: **R20** Year &Sem: **IV-B.Tech & I-Sem**

## UNIT –I

## INTRODUCTION TO BIG DATA AND HADOOP

| | | | | |
|---|---|---|---|---|
| **1** | **a)** | Examine the different types of digital data with examples. | **[L4][CO1]** | **[6M]** |
| | **b)** | What is Big Data? Explain the sources of Big Data | **[L2][CO1]** | **[6M]** |
| **2** | **a)** | Explain Big Data characteristic in terms of 5 V. | **[L2][CO1]** | **[6M]** |
| | **b)** | Discuss the reason behind the need of bigdata in organizations. | **[L4][CO1]** | **[6M]** |
| **3** | **a)** | Define big data analytics. Explain about the Classification of Analytics. | **[L2][CO1]** | **[6M]** |
| | **b)** | Illustrate how does big data analytics work | **[L3][CO1]** | **[6M]** |
| **4** | | Explain the following<br>a) Benefits of Big Data Analytics<br>b) Challenges of Big Data Analytics<br>c) Usage of Big Data Analytics | **[L2][CO1]** | **[12M]** |
| **5** | | Discuss in detail about Apache Hadoop and History of Hadoop. | **[L2][CO1]** | **[12M]** |
| **6** | | Describe about the Analysis of data through Unix tools with example | **[L2][CO1]** | **[12M]** |
| **7** | **a)** | Explain the process involved in Analyzing data with Hadoop | **[L2][CO1]** | **[6M]** |
| | **b)** | Illustrate in detail about Hadoop streaming | **[L3][CO1]** | **[6M]** |
| **8** | | Analyze the evolution of Hadoop ecosystem with neat diagram. | **[L4][CO1]** | **[12M]** |
| **9** | | Compare the implementation of IBM Big Data Strategy with the list of tools used. | **[L4][CO1]** | **[12M]** |
| **10** | **a)** | Explain in detail about Infosphere Big Insights with its key features. | **[L2][CO1]** | **[6M]** |
| | **b)** | Discuss in detail about Big sheets with its key features | **[L2][CO1]** | **[6M]** |

**Course Code:** 20CS0538 **R20**

## UNIT –II

## HDFS(HADOOP DISTRIBUTED FILE SYSTEM)

| | | | | |
|---|---|---|---|---|
| **1** | **a)** | Illustrate the modules of HDFS. | **[L3][CO2]** | **[6M]** |
| | **b)** | List the advantages, features and goals of Hadoop. | **[L1][CO2]** | **[6M]** |
| **2** | | Explain Hadoop Architecture and its Components with neat diagram. | **[L2][CO2]** | **[12M]** |
| **3** | **a)** | Discuss in detail about HDFS concepts. | **[L2][CO2]** | **[6M]** |
| | **b)** | Describe the dataflow process in Hadoop Distributed file System | **[L2][CO2]** | **[6M]** |
| **4** | | Determine the basic commands in Hadoop command line interface | **[L3][CO2]** | **[12M]** |
| **5** | | Explain in detail about File read and File write operations in HDFS with neat diagrams | **[L2][CO2]** | **[12M]** |
| **6** | **a)** | Explain in detail what is Apache Sqoop and its working with neat diagram. | **[L2][CO2]** | **[10M]** |
| | **b)** | List the features of Apache Sqoop. | **[L1][CO2]** | **[2M]** |
| **7** | **a)** | Discuss about the data ingest operation using Apache Flume with its features | **[L2][CO2]** | **[6M]** |
| | **b)** | List the difference between Apache Sqoop and Flume. | **[L4][CO2]** | **[6M]** |
| **8** | **a)** | What is Hadoop Archives and explain how to create and lookup files in archives. | **[L2][CO2]** | **[6M]** |
| | **b)** | Differentiate the compression and serialization operation in Hadoop I/O. | **[L4][CO2]** | **[6M]** |
| **9** | **a)** | Illustrate the AVRO file format with diagram along with its workflow | **[L3][CO2]** | **[8M]** |
| | **b)** | List the features of Avro. | **[L1][CO2]** | **[4M]** |
| **10** | | Analyze the different types of File Based Data structures with their advantages and disadvantages | **[L4][CO2]** | **[12M]** |

**Course Code:** 20CS0538 **R20**

## UNIT –III
## MAP REDUCE

| | | | | |
|---|---|---|---|---|
| **1** | **a)** | What is Mapreduce? Describe the stages of Mapreduce program execution. | **[L2][CO3]** | **[4M]** |
| | **b)** | Interpret the components of mapreduce architecture with diagram. | **[L3][CO3]** | **[8M]** |
| **2** | **a)** | Illustrate the phases involved in mapreduce data flow. | **[L3][CO3]** | **[6M]** |
| | **b)** | How hadoop mapreduce works? | **[L2][CO3]** | **[6M]** |
| **3** | | Analyse different types of failures in mapreduce with their reason for occurrence and methodology to overcome them. | **[L4][CO3]** | **[12M]** |
| **4** | | Explain the following<br>a) Task Failure<br>b) TaskTracker Failure<br>c) JobTracker Failure | **[L2][CO3]** | **[12M]** |
| **5** | | Describe the types of schedulers in Hadoop with their merits and demerits. | **[L2][CO3]** | **[12M]** |
| **6** | **a)** | Discuss the following<br>i) Steps in Map Reduce<br>ii) Job Scheduling | **[L2][CO3]** | **[6M]** |
| | **b)** | Discriminate the following input to perform word count in MapReduce<br>*Input: "The quick brown fox jumps over the lazy dog"* | **[L5][CO3]** | **[6M]** |
| **7** | **a)** | Explain sort and shuffle in mapreduce with example. | **[L2][CO3]** | **[6M]** |
| | **b)** | Discuss Task execution in mapreduce. | **[L2][CO3]** | **[6M]** |
| **8** | **a)** | Explain the types of mapreduce jobs. | **[L2][CO3]** | **[6M]** |
| | **b)** | Difference between big data and cloud computing. | **[L4][CO3]** | **[6M]** |
| **9** | | Describe different mapreduce formats available in Hadoop. | **[L2][CO3]** | **[12M]** |
| **10** | **a)** | Analyze the different types of InputFormat in MapReduce | **[L4][CO3]** | **[6M]** |
| | **b)** | List the features and use cases for mapreduce. | **[L1][CO3]** | **[6M]** |

**Course Code:** 20CS0538 **R20**

## UNIT –IV

## HADOOP ECO SYSTEM-PIG

| | | | | |
|---|---|---|---|---|
| **1** | **a)** | What is Pig? Explain the need and evolution of PIG. | **[L2][CO4]** | **[6M]** |
| | **b)** | Analyze the difference between Apache Pig and Mapreduce | **[L4][CO4]** | **[6M]** |
| **2** | | Discuss the following<br>a) Features of Apache Pig<br>b) Applications of Apache Pig<br>c) Types of Data Models in Apache Pig | **[L2][CO4]** | **[12M]** |
| **3** | | Explain the different execution modes available in Pig along with its execution mechanisms | **[L2][CO4]** | **[12M]** |
| **4** | | Discuss the different types of Apache Pig Execution Mechanisms | **[L2][CO4]** | **[12M]** |
| **5** | | How does Apache Pig differ from traditional databases, compare and explain in detail. | **[L2][CO4]** | **[12M]** |
| **6** | | Illustrate the concept of grunt with its supporting shell commands. | **[L3][CO4]** | **[12M]** |
| **7** | **a)** | What is Pig Latin? Describe it data types. | **[L1][CO4]** | **[4M]** |
| | **b)** | How do Pig Latin's arithmetic operators work, describe their use with example? | **[L2][CO4]** | **[4M]** |
| | **c)** | Illustrate how to apply comparison operators in Pig Latin to filter or manipulate data? | **[L3][CO4]** | **[4M]** |
| **8** | **a)** | What are the differences types of construction operators available in Pig Latin | **[L1][CO4]** | **[4M]** |
| | **b)** | Discuss the Pig Latin relational operations with example. | **[L2][CO4]** | **[8M]** |
| **9** | | Discriminate User Define Functions in Pig Latin with its detailed implementation procedure. | **[L5][CO4]** | **[12M]** |
| **10** | | What are the main data processing operators in Pig Latin, and describe how it is used in the context of data transformations? | **[L2][CO4]** | **[12M]** |

**Course Code:** 20CS0538 **R20**

## UNIT –V

## HIVE, HBASE, BIG SQL

| | | | | |
|---|---|---|---|---|
| **1** | | What is Hive? Illustrate step-by-step job execution flow in Hive with Hadoop with neat diagram. | **[L3][CO5]** | **[12M]** |
| **2** | **a)** | Discuss the advantages and disadvantages of Hive Architecture | **[L2][CO5]** | **[6M]** |
| | **b)** | Describe in detail about Hive shell command line interface. | **[L2][CO5]** | **[6M]** |
| **3** | | Explain in detail about the various services offered by Hive. | **[L2][CO5]** | **[12M]** |
| **4** | **a)** | What are the key components of the Hive Metastore, and how does it interact with external systems like MySQL and Spark? | **[L2][CO5]** | **[6M]** |
| | **b)** | How does the Hive Metastore manage metadata for Hive tables, partitions, and schemas, and what are the benefits of using it in a distributed environment? | **[L2][CO5]** | **[6M]** |
| **5** | **a)** | Analyze the difference between RDBMS and Hive | **[L4][CO5]** | **[6M]** |
| | **b)** | Explain the types of Tables in Hive and show their creation and deletion with queries. | **[L2][CO5]** | **[6M]** |
| **6** | **a)** | List and explain how basic and advanced querying is performed in Hive | **[L2][CO5]** | **[6M]** |
| | **b)** | Describe the following<br>i) User-Defined functions in Hive<br>ii) Built-in functions in Hive | **[L2][CO5]** | **[6M]** |
| **7** | **a)** | Explain with a neat diagram the architecture of HBase. | **[L2][CO5]** | **[8M]** |
| | **b)** | List the difference between HBase and HDFS | **[L4][CO5]** | **[4M]** |
| **8** | **a)** | List and Explain basic operations performed in HBase | **[L2][CO5]** | **[6M]** |
| | **b)** | Describe the key operations provided by the HBase client API | **[L2][CO5]** | **[6M]** |
| **9** | | List the difference between RDBMS and HBase | **[L4][CO5]** | **[12M]** |
| **10** | | Explain IBM Big SQL with its features along with its working in the Hadoop Ecosystem. | **[L2][CO6]** | **[12M]** |