Q. P. Code: 20CSE363 HALL TICKET NO.: \_\_\_\_\_\_\_\_\_\_\_

B.TECH VI SEMESTER (R20)

REGULAR / SUPPLEMENTARY EXAMINATIONS - JUN - 2024

BIG DATA ANALYTICS

(Common to CSE, CSM, CAI and CSD)

Time: 3 Hours Max. Marks: 70

---------------------------------------------------------------------

## PART – A

Answer ALL questions. Each question carries 2 marks.

\*\*B. Tech VI Semester (CO: R20 Computer Networks Exam\*\*, BT: -)

\*\*Time:\*\* 3 hours \*\*Max Marks:\*\* 70. (CO: -, BT: -)

\*\*Part A. (CO: 20 Marks\*\*, BT: -)

\*\*Answer all questions. Each question carries 2 marks.\*\* (CO: -, BT: -)

1. Define Big Data and list three key characteristics. (CO: CO1, BT: BT1)

2. What are the advantages of using Hadoop Distributed File System (CO: HDFS over traditional file systems?, BT: -)

3. Explain the role of the NameNode and DataNodes in HDFS. (CO: CO2, BT: BT1)

4. Briefly describe the Map and Reduce phases in a MapReduce job. (CO: CO3, BT: BT1)

5. What is the purpose of the `shuffle and sort` phase in MapReduce? (CO: CO3, BT: BT2)

6. Differentiate between Pig Latin and HiveQL. (CO: CO4, BT: BT2)

7. What is the primary function of ZooKeeper in the Hadoop ecosystem? (CO: CO5, BT: BT1)

8. Mention two advantages of using Sqoop for data transfer. (CO: CO5, BT: BT1)

9. What is the significance of data veracity in Big Data? (CO: CO1, BT: BT2)

10. Briefly explain the concept of YARN. (CO: CO2, BT: BT1)

\*\*Part B. (CO: 50 Marks\*\*, BT: -)

\*\*Answer all questions. Each question carries 10 marks. Choose either (CO: a or, BT: -)

\*\*Unit 1: Introduction to Big Data\*\*. (CO: -, BT: -)

1. (CO: a Elaborate on the challenges posed by Big Data and discuss best practices for addressing Big Data security concerns., BT: -)

. (CO: b Discuss the different types of Big Data sources and explain the process of Big Data acquisition., BT: -)

\*\*Unit 2: Hadoop Ecosystem and YARN\*\*. (CO: -, BT: -)

2. (CO: a Explain the architecture of Hadoop and describe the key components of the Hadoop ecosystem., BT: -)

. (CO: b Compare and contrast the functionalities of MRv1 and YARN. Explain the advantages of using YARN., BT: -)

\*\*Unit 3: MapReduce Programming\*\*. (CO: -, BT: -)

3. (CO: a Develop a MapReduce algorithm to count the occurrences of each word in a large text file. Explain each step of the process., BT: -)

. (CO: b Explain the concept of Job Scheduling in MapReduce and describe the different types of MapReduce jobs., BT: -)

\*\*Unit 4: Working with Pig and Hive\*\*. (CO: -, BT: -)

4. (CO: a Describe the advantages of using Pig over writing raw MapReduce code. Give an example of a Pig Latin script., BT: -)

. (CO: b Explain how Hive interacts with traditional databases. Compare and contrast HiveQL and SQL., BT: -)

\*\*Unit 5: HBase, ZooKeeper, Sqoop\*\*. (CO: -, BT: -)

5. (CO: a Describe the architecture of HBase and explain how it differs from a traditional Relational Database Management System, BT: -)

. (CO: b Explain the role of ZooKeeper in managing HBase. Discuss how Sqoop can be used to import and export data between Hadoop and relational databases., BT: -)

\*\*Note:\*\* CO refers to Course Outcome, and BT refers to Bloom's Taxonomy level. (CO: BT1: Remembering, BT: BT2: Understanding)

## PART – B

Answer ONE question from each UNIT – Each question carries 10 marks.