

**JAMIA HAMDARD**  
**Department of Computer Science & Engineering**  
**School of Engineering Sciences & Technology**  
**New Delhi-110062**  
**Odd Semester Examination-Dec 2024**

**Paper Name: Big Data Analytics**  
**Time: 3 Hours**

**Paper Code: BTCSE DED42/LE**  
**Maximum Marks: 75**

**Section A**

**All Questions are Compulsory in this section. Each question is of 5 Marks**

**7\*5=35 Marks**

| S.No | Question                                                                                                                                       | Marks | CO  | BL |
|------|------------------------------------------------------------------------------------------------------------------------------------------------|-------|-----|----|
| 1.   | Elaborate upon the History of Hadoop                                                                                                           | 5     | CO1 | L1 |
| 2.   | Explain the architecture of HDFS. What is the write mechanism in HDFS                                                                          | 5     | CO2 | L2 |
| 3.   | What is the mechanism behind the working of Map Reduce?                                                                                        | 5     | CO3 | L2 |
| 4.   | What are the various types of NoSQL databases? Explain HBase architecture                                                                      | 5     | CO4 | L1 |
| 5.   | Explain the RHadoop architecture. What are the various packages in R? Which packages of R are used for Hadoop?                                 | 5     | CO5 | L2 |
| 6.   | What data types are supported by R? Write the commands with correct syntax for performing statistical operations of mean, median and mode in R | 5     | CO5 | L3 |
| 7.   | What is the HiveQL? Evaluate the Hive operations of performing Union and Join and write their commands with proper syntax.                     | 5     | CO4 | L4 |

**Section B**

**Attempt any four. Each question is of 10 Marks**

**4\*10=40 Marks**

| S.No | Question                                                                                                                                                                            | Marks | CO  | BL |
|------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------|-----|----|
| 1)   | Explain in detail the IBM big data strategy, Infosphere Big Insights and Big sheets.                                                                                                | 10    | CO1 | L1 |
| 2)   | Write short notes on<br>(a)Hadoop IO (b)Data Serialisation (c)Avro<br>(d)Floom (e)Sqoop                                                                                             | 10    | CO2 | L1 |
| 3)   | How is job scheduling done in Map Reduce? Explain the working of all 3 types of Hadoop Schedulers.                                                                                  | 10    | CO3 | L2 |
| 4)   | Show the use of the following Pig commands with proper syntax<br>(a)Filter (b)Group (c)Join<br>(d)Count (e)Split                                                                    | 10    | CO4 | L4 |
| 5)   | Evaluate the output after writing a program in R for implementing any algorithm of Supervised Machine Learning.                                                                     | 10    | CO5 | L5 |
| 6)   | For the HiveQL write the commands with correct syntax and output for the following<br><br>(a)Create and alter a database (b)Select clause (c)Order by clause<br>(d)Count (e)Average | 10    | CO4 | L4 |

**BL-Bloom's Taxonomy Levels (1-Remembering, 2- Understanding, 3-Applying, 4 -Analysing, 5- Evaluating, 6 - Creating)**