# Question Paper

## MANIPAL ACADEMY OF HIGHER EDUCATION

Manipal School of Information Sciences (MSIS), Manipal

First Semester Master of Engineering - ME (Artificial Intelligence & Machine Learning / Big Data Analytics) Degree Examination - January 2023

**Architecture of Big Data Systems (Elective -1) [BDA 5102]**

**Marks: 100**                                                                                           **Duration: 180 mins.**

**Monday, January 9, 2023**

**Answer all the questions.**

| | | |
|---|---|---|
| 1) | List and explain the characteristics of Big Data. (TLO 1.1) (10 marks) | (10) |
| 2) | Differentiate between re-computation algorithm and increment algorithm. List the requirements and responsibilities of batch layer. (TLO 1.2) (5+5 marks) | (10) |
| 3) | With diagram show how write operation is performed in HDFS. List the steps to recover from failure during HDFS write operation. (TLO 2.1) (6+4 marks) | (10) |
| 4) | With neat diagram, explain the architecture of HDFS. (TLO 2.1) (10 marks) | (10) |
| 5) | List out the responsibilities of Resource Manager in YARN. Explain different schedulers used in YARN. (TLO 2.2) (4+6 marks) | (10) |
| 6) | With neat diagram, explain the architecture of Spark. (TLO 3.1) (10 marks) | (10) |
| 7) | Define RDDs? Explain the properties of RDD. Justify why Spark is called lazy evaluation. (TLO 3.1) | (10) |
| 8) | Explain Continuous Stream Processing method. List its advantages and disadvantages. Write Short notes on Data Sources and Data Sinks of Structured Streaming in Spark. (TLO 3.2) (3+3+4 marks) | (10) |
| 9) | Assume dataset with fields like 'Student Name', 'Institute', 'Program Name', and 'Gender'. Write map and reduce code for the following questions. <br> a. Compute number of students from each Institute. <br> b. Number of 'boy' and 'girl' students from selected Institute. It is assumed that Institute name is passed as command line argument. (TLO 4.1) (4+6 marks) | (10) |
| 10) | One of the column in dataset is 'Date'. Data type of this column is String. Format is dd-MM-yyyy. Add a new column 'New Date' and its data type should be Date. Date in new column should be displayed in the same format as in column 'Date'. Using 'New Date' column create columns like 'Year', 'Month' and 'Week'. Delete original column 'Date'. (TLO 4.2) (5+3+2 marks) | (10) |

-----End-----