Nirma University

Institute of Technology

Sessional Examination, October 2023

B.Tech. Semester VII

2CS702 – Big Data Analytics

Roll/Exam No

Supervisor's initial with date 9/10

Time: 2 Hour

Max Marks: 50

Instructions:

1. Attempt all questions.
2. Figures to the right indicate full marks.
3. Assume suitable data wherever necessary and specify them.
4. Draw neat sketches wherever necessary.

Q. 1

1. Explain how social media data are considered as big data. Explain with suitable examples. **5**
2. Explain how relational database systems are suitable for OLTP systems. Is HDFS considered an OLTP system? Explain in detail. **5**

Q. 2

1. Explain in detail the HDFS write operation with illustrations. **5**
2. Discuss and differentiate Hadoop V 1.0 and Hadoop V 2.0 with respect to resource utilization. Explain in detail the role of the job tracker in Hadoop V 1.0. How HDFS architecture will respond in case of job tracker is down? **5**

Q. 3

1. Explain the following HDFS command with a proper example. **3**
   a) CopyToLocal b) rmr c) mv
2. Discuss and differentiate combiner and reducer in MapReduce programming. **4**
3. Explain input split and block in HDFS architecture. **3**

Q. 4 Consider the following dataset in text format,

Product (Product_No, Description, Profit_Percent, Unit_Measure, Qty_on_hand, Reorder_Level, Sell_Price, Cost_Price)

1. Design input and output for mapper tasks and reducer tasks to list the product names whose sell price is between 10,000 and 20,000. **5**
2. Illustrate the execution of the Map and Reduce phase with a proper diagram for the following input data. **5**

P00001	1.44 Floppies	5	Piece	100	20	525	500
P03453	Monitors	6	Piece	10	3	12000	11200
P06734	Mouse	5	Piece	20	5	1050	500
P07865	1.22 Floppies	5	Piece	100	20	525	500
P07868	Keyboards	2	Piece	10	3	3150	3050

Q. 5

1. Consider input text file of size 1TB. Consider the input split size is 2048 KB and HDFS block size is the default size. How many mapper tasks will be generated in MR execution? Is the number of mapper tasks dependent on input file size? Explain your answer. **5**
2. Explain map-only jobs in the MR programming paradigm. What is the optimal number of reducers required to run map-only jobs? **5**