UNIVERSITY OF PETROLEUM AND ENERGY STUDIES
End Semester Examination, May 2023

Course: Disk Based Processing   **Semester: IV**
Program: B.Tech CSE spl in Big Data   **Time   : 03 hrs.**
Course Code: CSBD 2008   **Max. Marks: 100**

Instructions: Kindly stick to the question and support every answer with proper explanation.

SECTION A
(5Qx4M=20Marks)

S. No.		Marks	CO
Q 1	List the advantage and disadvantages of "Fire and forget" method.	4	CO1
Q 2	Define data locality principle. Explain how Hadoop facilitate the same	4	CO1
Q 3	State the mechanism adopted by Hadoop to overcome slow running data nodes.	4	CO2
Q 4	List difference between Schema on Read and Schema on write	4	CO2
Q 5	State umbilical protocol.	4	CO4

SECTION B
(4Qx10M= 40 Marks)

Q 6	"Map Reduce is most basic processing performed by Hadoop to process Big data". Demonstrate the benefit of Map Reduce with proper example.	10	CO1
Q 7	Yet Another Resource Negotiator "YARN", reduces the workload of Job tracker while Schedule multiple tasks on the run time. Demonstrate techniques followed by Map Reduce Ver 2 to handle parallel processing.	10	CO4
Q 8	Illustrate the Role of Input Split. Explain how to customize the size of input split. Draw the various stages of Map Reduce to explain the role of Input Split.	10	C02
Q 9	"Map reduce can process data in petabyte". Interpret the Architecture of MapReduce **OR** "Map reduce can process data in petabyte". Interpret the working of Mapper and Reducer in MapReduce.	10	C03

SECTION-C
(2Qx20M=40 Marks)

Q 10	Explain the working of Map reduce with proper java code to processing to Identify the top the Cold and Hot days for the given dataset.	**20**	**CO3**
Q 11	Write detail note to appraise the role of the following: 1. Primary Sorting techniques in MR 2. Partitioner **OR** Write detail note to appraise the role of the following: 1. Secondary Sorting techniques in MR 2. Input and Output Contract	**10+10**	**CO2**