UPES
End Semester Examination, May 2024

Course: Big Data Analytics   Semester: VI
Program: B.Tech. CSE+IoT   Time: 03 hrs.
Course Code: CSBA3012   Max. Marks: 100

Instructions: Attempt all the questions. Section B and Section C have an internal choice.

SECTION A
(5Qx4M=20Marks)

S. No.		Marks	CO
Q 1	Identify key technologies that have made big data processing and storage possible.	4	CO1
Q 2	Explain veracity in big data. Briefly explain its implications.	4	CO1
Q 3	Categorize different categories of data that you can process in Hadoop platform.	4	CO2
Q 4	Justify the validity of statement "HDFS clusters do not benefit from using RAID".	4	CO2
Q 5	Discuss the logical data flow in map reduce framework.	4	CO4

SECTION B
(4Qx10M= 40 Marks)

Q 6	Explain map reduce architecture in detail.	10	CO2
Q 7	What are the broad data analysis techniques? Discuss any two techniques in each category. OR How is big data reporting different from big data analysis? Elaborate it considering valid cases.	10	CO1
Q 8	After differentiating batch processing from steam processing conclude stream processing is not possible using Hadoop.	10	CO4
Q 9	Discuss the main types of possible reporting techniques used post big data processing.	10	CO1

SECTION-C
(2Qx20M=40 Marks)

Q 10	Write short notes on: 1. Apache Sqoop 2. SPL 3. JAQL	20	CO3

	4. Apache Pig		

Q 11	Referring Stream Processing Language (SPL), review the following:
	   1. Basics of SPL
	   2. SPL Building blocks
	   3. Structure of program file
	   4. Different available operators
	
	OR
	
	MapReduce is the processing module in Hadoop framework. It involves series of map and reduce operations. In context to this answer the following:
	
	a) Who and how the number of map and reduce operations are decided.
	
	b) Outline map and reduce operation.
	
	c) Explain role of job tracker and task tracker.
	
	d) Explain the whole process preferably using a flow diagram considering word count example
		20	CO4