UNIVERSITY OF PETROLEUM AND ENERGY STUDIES
End Semester Examination, December 2021

Course: B.Tech CSE(Big Data Specialization)　　　　　　　Semester: 3
Program: Big Data Overview　　　　　　　　　　　　　　Time: 03 hrs.
Course Code: CSBD2007　　　　　　　　　　　　　　　　Max. Marks: 100

Instructions:

1. **Section A** shall have 5 Questions of 4 marks each. . All the questions shall be compulsory.

2. **Section B** This section shall have 4 Questions of 10 marks each, out of which 3 Questions shall be compulsory and 1 Questions may have internal choice to attempt any one.

3. **Section C** shall have 2 Questions of 20 marks each, out of which 1 Question shall be compulsory and 1 Question shall have internal choice to attempt any one. This section may be further subdivided as per the requirement of course. These Questions shall be of long answer type.

	SECTION A　　　　　　　　　(5Qx 4M = 20 Marks)		
S. No.		Marks	CO
Q 1.	Enumerate the two key differences between database and file management system .	4	CO1
Q2.	Explain the different types of Data produced from various data sources.	4	CO3
Q3.	Big Data is different from Data .Support this statement by throwing light on the characteristics of Big Data .	4	CO2
Q4.	Why is NoSQL database preferred for Big Data? Elaborate the various types of NoSQL database.	4	CO4
Q5.	Illustrate the various components of Data Lake.	4	CO4
	SECTION B　　　　　　　　　(4Qx10M = 40 Marks)		
Q 1.	HDFS architecture allows storage of Big Data .Elaborate its architecture	10	CO4
Q2.	"Improved Customer Service" and "Reduced operational costs" are among a few benefits of Big Data Analytics. Elaborate.	10	CO1
Q3.	Taking the example of Netflix explain "data explosion ". Or	10	CO2

	Explain how Facebook exploited the benefits of Big Data Analytics.		
Q4.	Analyze the challenges for start-ups and SMEs in Migrating to Big Data Platform.	**10**	**CO3**

	SECTION-C (2Qx 20M= 40 Marks)		
Q 1.	The No. of cases of Dengue are on a steep rise in Dehradun. Following data has been extracted from a dataset that was collected in a small residential area near Raipur (most affected area) area .Apply map-reduce computation and find out the number of Affected and Not-affected people in that area. Affected, Not-Affected, Affected, No- affected, affected, not-affected, Not-affected, Affected , Affected, Not-Affected, Affected	**20**	**CO3**
Q2.	Analyze the challenges of Distributed System. Assess how Hadoop overcame them. OR Illustrate the concept of Polyglot Persistence with the help of an example.	**20**	**CO4**