Name: Enrolment No:	

UNIVERSITY OF PETROLEUM AND ENERGY STUDIES
End Semester Examination, December 2019

Course: Big Data Storage   **Semester: III**
Program: BTech CSE + Big Data   **Time: 03 hrs.**
Course Code: CSBD2004   **Max. Marks: 100**

Instructions: Attempt all questions from section A. There is an internal choice in question 9 and 11

	SECTION A		
S. No.		**Marks**	**CO**
Q 1	Differentiate between HBase and Big Table.	**4**	**CO2**
Q 2	State rack awareness and on what basis is data stored in a rack?	**4**	**CO1**
Q 3	Define column families. What happens if we alter the block size of the column family on an already populated database?	**4**	**CO2**
Q 4	List the different modes in which Hadoop can be installed and what is the use of each mode from application and developer point of view?	**4**	**CO1**
Q 5	Describe the features of HBase.	**4**	**CO2**
	SECTION B		
Q 6	Design HBase architecture. Explain all the components with functionality.	**10**	**CO2**
Q 7	Explain distributed system classification based on the CAP theorem.	**10**	**CO3**
Q 8	Discuss different scenarios and conditions when one has to choose Apache HBase as compared to the HDFS in Big Data projects.	**10**	**CO2**

Q 9	Draw the Hadoop logical architecture and explain its components.
	
	OR
	
	Define the following:
	   1. Distcp
	   2. Rack Awareness
	   3. Zookeeper
	   4. Compaction
		10	**CO4**
			
			CO1

	SECTION-C		
Q 10	Define database sharding with suitable examples. What are the benefits and drawbacks of sharding? Discuss whether one should shard or not.	**20**	**CO3**
Q 11	Sketch the High availability architecture of the Hadoop Distributed File System. Explain all the components used in the architecture. **OR** a) Compare NoSQL with RDBMS. Investigate why and why not to use NoSQL databases. b) Discuss why NoSQL database is required in Big data. What are the various advantages and disadvantages of NoSQL databases?	**20**	**CO4** **CO2**