

Name:

Enrolment No:



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

**Course: Big Data Analytics**

**Program: B. Tech CSE+IOTSc**

**Course Code: CSEG485**

**Semester: VII**

**Time : 03 hrs.**

**Max. Marks: 100**

**Instructions:** All questions are compulsory. Questions 9 and 11 has internal choices.

**SECTION A**

S. No.		Marks	CO
Q 1	Differentiate between Relational Database Management System (RDBMS) and Hadoop Distributed File System (HDFS).	4	CO1
Q 2	Explain the procedure involved in resolving DataNode failures in HDFS.	4	CO2
Q 3	Define hadoop streaming.	4	CO5
Q 4	Discuss the relationship between big data and Hadoop.	4	CO1
Q 5	Illustrate the procedure of accessing the same file by two clients in HDFS.	4	CO2

**SECTION B**

Q 6	Identify and discuss the steps required to deploy a Big Data solution.	10	CO1
Q 7	Illustrate various Hadoop daemons and their roles in a Hadoop cluster.	10	CO3
Q 8	Discuss the benefits of Apache Pig over MapReduce.	10	CO4
Q 9	Define Sqoop and explain five important Sqoop commands. <b>OR</b> Define “SerDe” in “Hive”. Can the default “Hive Metastore” be used by multiple users (processes) at the same time?	[5+5]	CO4

**SECTION-C**

Q 10	Explain partitioning, shuffle and sort phases of hadoop architecture using an example. Discuss the importance of combiner in hadoop.	[15+5]	CO2, CO5
Q 11	Explain MapReduce wordcount process using Mapper and Reducer program. <b>OR</b> a) Differentiate between Structured and Unstructured data. b) Discuss most commonly defined input formats in Hadoop. c) Justify what happens when a user submits a Hadoop job and the Job Tracker is down. Does the job get in to hold or does it fail? d) Write port number for NameNode, Task Tracker and Job Tracker.	10+10  [5+5+5+5]	CO1, CO3