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

Course: Big Data Analytics　　　　　　　　　　　　　**Semester: VII**
Program: B.Tech (CSE+BAO)　　　　　　　　　　　　**Time   : 03 hrs.**
Course Code: CSIB385　　　　　　　　　　　　　　　**Max. Marks: 100**

Instructions: All Questions are compulsory. Question 8 and 11 has internal choice.

SECTION A			
S. No.		**Marks**	**CO**
Q1	Differentiate between Input split and HDFS Block.	**4**	**CO2**
Q2	Discuss various characteristics of Big Data.	**4**	**CO1**
Q3	Big data analysis helps businesses in increasing their revenue. Justify the statement with the help of example.	**4**	**CO1**
Q4	Explain the term DStream.	**4**	**CO5**
Q5	Explain the difference between Sqoop and Flume with the help of example.	**4**	**CO3**
SECTION B			
Q6	Explain the ecosystem of Hadoop with the help of diagram. Also, discuss the architecture of HDFS.	**10**	**CO2**
Q7	1. Write the use and syntax of Sqoop import and Export commands. 2. Write the use and syntax of following HDFS commands. 　　a. Put 　　b. Get 　　c. copyFromLocal 　　d. copyToLocal	**6+4**	**CO3**
Q8	Differentiate between Hadoop1.0 and Hadoop2.0. What is the need of zookeeper? **OR** Explain the following terms 　　a. High Availability 　　b. Federation	**6+4** **5+5**	**CO2**
Q9	Let's assume that, you have 100TB of data to store and process with Hadoop. The configuration of each available DataNode is as follows: 8GB RAM 10TB HDD 100 MB/s read-write speed	**4+3+3**	**CO2**

	You have a Hadoop cluster with replication factor 3 and block size 128MB. Based on above scenario, answer the following:
	    a. How many data nodes are required to store the data?
	    b. How much time is required to process the data?
	    c. How many such datanodes you would need to read 100TB data in 15 minutes in your Hadoop cluster?		

SECTION-C

Q10	1. List out the difference between Map Reduce and SPARK. 2. Is it necessary to install spark on all the nodes of a YARN cluster while running Apache Spark on YARN ? 3. Is it necessary to start Hadoop to run any Apache Spark Application? 4. Discuss workflow of any spark program.	**8+1+1 +10**	**CO4, CO5**
Q11	1. How is Hadoop CLASSPATH essential to start or stop Hadoop daemons? 2. Differentiate between FSImage and Editlog. Explain the Back up mechanism in Hadoop1.0 and Hadoop2.0. **OR** 1. Discuss Data processing in Hadoop1.0 and Hadoop2.0 with the help of commands. 2. Explain 'Sort and shuffle', 'Mapper' and 'Reducer' Phase in Mapreduce.	**5+15** **10+10**	**CO2**