UNIVERSITY OF PETROLEUM AND ENERGY STUDIES
School of Computer Science

End Semester Examination, December 2022

Course   : **Bigdata Analytics**   Semester   : **VII**
Program : **B.Tech CSE OGI**   Time   : **03 Hours**
Course Code   : **CSBA4001**   Max. Marks : **100**

Instructions   :

SECTION A

	SECTION A 1. Each Question will carry 5 Marks	Marks	
Q1	Differentiate between Pig and hive?	04	CO3
Q2	Defend the statement "Hadoop is not a solution for streaming data applications".	04	CO4
Q3	Discuss the three V's of big data and identify any other possible v's that relate to big data.	04	CO1
Q4	Differentiate between Hadoop and Traditional Database?	04	CO2
Q5	In Hadoop, _____ is algorithm applied to NameNode to decide the way that the blocks and its replicas are placed across DataNodes. Depending on rack definitions network traffic is minimized between DataNodes within the same rack. Example, if we consider replication factor as 3, then two copies will be placed on one rack whereas the third copy in a separate rack.	04	CO2

SECTION B

SECTION B 1. Each question will carry 10 marks 2. Instruction: Write short / brief notes			
Q6	Differentiate between Data at rest and data at motion. Explain their processing also.	5+5	CO2
Q7	Elaborate Big Data lifecycle, explaining each phase and specifying its relevance.	10	CO1
Q9	Write five Jaql core operators with example.	10	CO3
Q10	a. Assume that there are 100 nodes in your Hadoop cluster with a total of 400TB (8TB per node) of raw disk allocated HDFS storage. Assuming Hadoop configuration, how much data will you be able to store? b. Explain the term 'Heartbeat' and 'Block Report' in HDFS.	4+6 Or 10	CO2

	Or Discuss the problem of backup mechanism in Hadoop Gen 1 and how Ha There is no classic backup and recovery functionality in Hadoop.		

SECTION C

SECTION C
1. Each Question carries 10 Marks. 10*2=20
2. Instruction: Write long answer.

Q11	Write the use and syntax of following HDFS commands.
	    a. mkdir
	    b. ls
	    c. cat
	    d. mv
		20	**CO3**

Q12	Discussing JobTracker and TaskTracker elaborate MapReduce architecture. Citing WordCount example draw a MapReduce logical data flow diagram and give details of each logical block.
	
	OR
	
	Explain the term 'Single point of failure'? In Hadoop 2.0 during the data processing phase, if the following failure happens what solution is being provided by Hadoop 2.0
	    a. Mapper fails,
	    b. Reducer Fails,
	    c. NodeManager fails,
	    d. Application Manager fails,
	    e. Resource Manager fails.
		20	**CO2**