


Name:	 UPES <small>UNIVERSITY WITH A PURPOSE</small>
Enrolment No:	

UNIVERSITY OF PETROLEUM AND ENERGY STUDIES
School of Computer Science

End Semester Examination, December 2020

Course : Bigdata Analytics	Semester : VII
Program : B.Tech CSE BAO/Bigdata/IoTSc	Time : 03 Hours
Course Code : CSBA4001	Max. Marks : 100

SECTION A

SECTION A		Marks	
1. Each Question will carry 5 Marks			
Q1	Differentiate between Pig and hive?	05	CO3
Q2	Defend the statement “Hadoop is not a solution for streaming data applications”.	05	CO4
Q3	Discuss the three V’s of big data and identify any other possible v’s that relate to big data.	05	CO1
Q4	Differentiate between Hadoop and Traditional Database?	05	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.	05	CO2
Q6	Explain the purpose of Sqoop import and export command. Write the commands also.	05	CO3

SECTION B

SECTION B

1. Each question will carry 10 marks
2. Instruction: Write short / brief notes

Q7	Differentiate between data at rest and data at motion. Explain their processing also.	6+4	CO4
Q8	Elaborate Big Data lifecycle, explaining each phase and specifying its relevance.	10	CO1
Q9	Write the use and syntax of following HDFS commands. a. Put b. Get c. copyFromLocal d. copyToLocal	10	CO3

	e. jps		
Q10	Write five Jaql core operators with example.	10	CO3
Q11	<p>a. Assume that there are 50 nodes in your Hadoop cluster with a total of 200TB (4TB per node) of raw disk allocated HDFS storage. Assuming Hadoop configuration, how much data will you be able to store?</p> <p>b. Explain the term ‘Heartbeat’ and ‘Block Report’ in HDFS.</p> <p>Or</p> <p>Discuss the problem of backup mechanism in Hadoop Gen 1 and how Hadoop Gen2 resolve this problem</p>	4+6 Or 10	CO2
SECTION C			
SECTION C			
1. Each Question carries 20 Marks.			
2. Instruction: Write long answer.			
Q12	<p>Discuss JobTracker and TaskTracker and elaborate MapReduce architecture. Citing WordCount example draw a MapReduce logical data flow diagram and give details of each logical block.</p> <p>OR</p> <p>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</p> <p>a. Mapper fails,</p> <p>b. Reducer Fails,</p> <p>c. NodeManager fails,</p> <p>d. Application Manager fails,</p> <p>e. Resource Manager fails.</p>	20	CO2