

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

UNIVERSITY OF PETROLEUM AND ENERGY STUDIES
School of Computer Science

End Semester Examination, May 2021

Course : Bigdata Analytics
Program : B.Tech CSE BAO
Course Code : CSBA4001

Semester : VI
Time : 03 Hours
Max. Marks : 100

Instructions :

SECTION A

SECTION A		Marks	
1. Each Question will carry 5 Marks			
Q1	Elaborate in detail the various Hadoop installation options and from the very scratch discuss the steps in deploying a Hadoop cluster.	05	CO2
Q2	Differentiate confirmatory data analysis from exploratory data analysis.	05	CO1
Q3	Discuss partitioning and bucketing in HIVE.	05	CO3
Q4	Explain the purpose of Sqoop import and export command. Write the commands also.	05	CO3
Q5	Explain the Rackawareness concept.	05	CO2
Q6	Explain the importance of Volume, velocity and variety of in Big Data?	05	CO1

SECTION B

SECTION B

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

Q7	Explain the following in context to stream processing language: <ul style="list-style-type: none"> a. Source b. Sink c. Beacon d. Custom e. Functor f. Eviction Policy g. Trigger policy 	10	CO4
----	---	-----------	------------

	<ul style="list-style-type: none"> h. List i. Map j. Host control 		
Q8	Elaborate Big Data lifecycle, explaining each phase and specifying its relevance.	10	CO1
Q9	Explain read anatomy and write anatomy in HDFS.	10	CO2
Q10	<ul style="list-style-type: none"> 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? b. Explain the term federation and the working of Zookeeper. 	4+6	CO2
Q11	<ul style="list-style-type: none"> a. Difference between external table and managed table in hive b. Explain the following PIG datatypes <ul style="list-style-type: none"> • Bag • Tuple • Atom <p style="text-align: center;">Or</p> <p>Outline the relevance of Hadoop application layer and also give details of below mentioned applications:</p> <ul style="list-style-type: none"> a) Apache Spark b) Apache Kafka c) Pig 	4+6 Or 10	CO3

SECTION C

SECTION C

1. Each Question carries 20 Marks.

2. Instruction: Write long answer.

Q12	<p>MapReduce is the processing module in Hadoop framework. It involves series of map and reduce operations. In context to this answer the following:</p> <ul style="list-style-type: none"> a) Who and how the number of map and reduce operations are decided. b) Outline map and reduce operation. c) Explain role of job tracker and task tracker. d) Explain the whole process preferably using a flow diagram considering word count example. <p>OR</p>	20	CO2
-----	--	-----------	------------

	Elaborating in good detail the core components of Hadoop framework, namely HDFS, MapReduce and YARN, justify the statement opening statement “Hadoop framework is one of the most widely used solution to store and analyze big data”.		
--	--	--	--