

Name:

Enrolment No:



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

Course: Big Data Analysis
Program: B.Tech. CSE Big Data
Course Code: CSBD 2003

Semester: IV
Time 03 hrs.
Max. Marks: 100

Instructions:

SECTION A

S. No.		Marks	CO
Q 1	Define big data and write any four sources of big data.	4	CO1
Q 2	How Hive is different from traditional databases.	4	CO1
Q 3	Why Hive does not store metadata information in HDFS?	4	CO2
Q 4	Explain any scenario where HIVE bucket is preferred over Hive partition.	4	CO4
Q 5	What is the difference between local and remote metastore?	4	CO3

SECTION B

Q 6	Discuss any five domains in which big data analytics can be used with applications.	10	CO1
Q 7	Explain use of partitions in Hive. Discuss various types of partitions.	10	CO3
Q 8	Discuss various components of Hadoop Eco System.	10	CO2
Q 9	What is Hive indexing? Explain features of Hive Indexing.	10	CO4

OR

	Explain and differentiate different types of tables used in Hive.	10	CO4
--	---	----	-----

SECTION-C

Q 10	Write the important phases while executing Hive Query. Explain important components of HIVE with the help of diagram.	10 +10	CO2
Q 11	Assuming dataset of ration distribution system of various centers in Uttar Pradesh ration_table(distributer_id STRING, distributer_name STRING, amount_in float, amount_sold float , year INT, location STRING) Write Hive Query for the given statements: a) Create table for the ration distribution system and store it in a text file.	5*4	CO4

	b) Display distributor details who has sold maximum amount of ration in year 2018. c) Display distributor details of all centers in Agra. d) Display total amount of ration sold year wise. e) Display count of total number of distributors in year 2018.		
	OR		
	Assume that you are starting a business of hotel chain in various cities of India. With your own assumptions: <ol style="list-style-type: none"> 1) Write the schema of important tables. 2) Discuss any three new strategies that you may use at the time of starting business. 3) How data analytics using Hive can be done to design various strategies in your business. Explain with help of examples. 4) Write any three challenges that you may face in the business. 	4*5	CO4

Name:	 UPES UNIVERSITY WITH A PURPOSE
Enrolment No:	

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

Course: Big Data Analysis Program: B.Tech. CSE Big Data Course Code: CSBD 2003	Semester: IV Time 03 hrs. Max. Marks: 100
---	--

Instructions:

SECTION A

S. No.		Marks	CO
Q 1	Define different types of big data with example.	4	CO1
Q 2	What is dynamic partitioning and when is it used?	4	CO4
Q 3	Is Hive suitable to be used for OLTP systems? Why?	4	CO2
Q 4	Write any four limitations of HIVE.	4	CO2
Q 5	Write advantages and disadvantages of using HIVE partitions with example.	4	CO4

SECTION B

Q 6	Discuss various HIVE services with the help of diagram	10	CO1
Q 7	Explain the steps involved in execution of HIVE Query.	10	CO2
Q 8	What is the use of metastore in HIVE. Explain and compare different types of HIVE meta store.	4+6	CO3
Q 9	Explain Hive Buckets and its features. How buckets can be used in Hive? Explain with help of example.	10	CO4

OR

	<p>Suppose, I create a table that contains details of all the transactions done by the customers of year 2018: <code>CREATE TABLE t_details (cust_id INT, amount FLOAT, month STRING, country STRING) ROW FORMAT DELIMITED FIELDS TERMINATED BY ',';</code></p> <p>Now, after inserting 30,000 rows in this table, I want to know the total revenue generated for each month. But, Hive is taking too much time in processing this query. How will you solve this problem and list the steps that you will be taking in order to do so?</p>	10	CO4
--	---	----	-----

SECTION-C

Q 10	Explain following concepts of HIVE with example: a) CTAS b) Views c) Indexing d) User Defined Functions	4 * 5	CO3
Q 11	Assuming Olympics Dataset for all the Olympics held till date is stored as:	5*4	CO4

	<p>(AthleteName STRING, Age INT, Country STRING, Year INT, ClosingDate STRING, Sport STRING, Gold Medals INT, Silver Medals INT, Bronze Medals INT, Total Medals INT)</p> <p>Write Hive query for the given statements:</p> <ol style="list-style-type: none"> Create table and load data from Olympics.csv file. List the total number of medals won by each country in badminton. Display total medals won by India year wise. Find total number of medals won by each country in 2016 Olympics. Display country name with maximum medals in year 2016. 		
	<p>OR</p>		
	<p>Assume that you are starting a business of selling old books in various cities of India. With your own assumptions:</p> <ol style="list-style-type: none"> Write the schema of important tables. Discuss any three new strategies that you may use at the time of starting business. How Hive can be used in your business. Explain with help of examples. Write any three challenges that you may face in the business. 	<p>4 * 5</p>	<p>CO4</p>