


Name:	 UPES UNIVERSITY WITH A PURPOSE
Enrolment No:	

UNIVERSITY OF PETROLEUM AND ENERGY STUDIES
Online End Semester Examination, December 2020

Course: Big Data Analytics Program: MBA(BA) Course code: DSBA 8002	Semester: III Time: 03 hrs. Max. Marks: 100
---	--

SECTION A

- 1. Each Question will carry 5 Marks**
- 2. Instruction: Select/write the correct answer(s)**

S. No.	Question	CO
Q1	<p>Select correct option:</p> <p>I. What does commodity Hardware in Hadoop world mean?</p> <ol style="list-style-type: none"> a) Very cheap hardware b) Industry standard hardware c) Discarded hardware d) Low specifications Industry grade hardware <p>II. Which of the following are NOT big data problem(s)?</p> <ol style="list-style-type: none"> a) Parsing 5 MB XML file every 5 minutes b) Processing IPL tweet sentiments c) Processing online bank transactions d) both (a) and (c) <p>III. What does “Velocity” in Big Data mean?</p> <ol style="list-style-type: none"> a) Speed of input data generation b) Speed of individual machine processors c) Speed of ONLY storing data d) Speed of storing and processing data <p>IV. The term Big Data first originated from:</p> <ol style="list-style-type: none"> a) Stock Markets Domain b) Banking and Finance Domain c) Genomics and Astronomy Domain d) Social Media Domain 	CO1

	<p>V. Which of the following are NOT true for Hadoop?</p> <ul style="list-style-type: none"> a) It's a tool for Big Data analysis b) It supports structured and unstructured data analysis c) It aims for vertical scaling out/in scenarios d) Both (a) and (c) 	
<p>Q2</p>	<p>Select correct option:</p> <p>I. Which of the following are the core components of Hadoop?</p> <ul style="list-style-type: none"> a) HDFS b) Map Reduce c) HBase d) Both (a) and (b) <p>II. Hadoop is open source.</p> <ul style="list-style-type: none"> a) ALWAYS True b) True only for Apache Hadoop c) True only for Apache and Cloudera Hadoop d) ALWAYS False <p>III. What is the default HDFS block size?</p> <ul style="list-style-type: none"> a) 32 MB b) 64 KB c) 128 KB d) 64 MB <p>IV. What is the default HDFS replication factor?</p> <ul style="list-style-type: none"> a) 4 b) 1 c) 3 d) 2 <p>V. Which of the following is NOT a type of metadata in NameNode?</p> <ul style="list-style-type: none"> a) List of files b) Block locations of files c) No. of file records d) File access control information 	<p>CO1</p>
<p>Q3</p>	<p>Write query in MongoDB for the following:</p> <ul style="list-style-type: none"> a) Create collection 'mammals' in side 'petshop' database b) Insert two documents in 'mammals': <ul style="list-style-type: none"> name: Polar Bear name: Star Nosed Mole 	<p>CO2</p>

Q4	Write query in MongoDB for the following: a) Create collection by name “people” with below document: name: "dave", email: <u>davey@aol.com</u> Phone: 9423456734 Age:69 b) Display records where name is “dave” and email is <u>davey@aol.com</u>	CO2
Q5	Consider the mycol collection has the following data. <ul style="list-style-type: none"> • { "title": "MongoDB Overview" } • { "title": "NoSQL Overview" } • { "title": "Tutorials Point Overview" } a) Set the new title 'New MongoDB Tutorial' of the documents whose title is 'MongoDB Overview'. b) Remove all the documents whose title is 'MongoDB Overview'.	CO2
Q6	Consider the collection mycol has the following data: <ul style="list-style-type: none"> • { _id: 1, "title": "MongoDB Overview" } • { _id: 2, "title": "NoSQL Overview" } • { _id: 3, "title": "Tutorials Point Overview" } a) Display only the title of the document while querying the document. b) Display Title of only two documents while querying the document.	CO2

SECTION B

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

Q7.	Explain in brief about Name node, Data Node and Secondary Name node in HDFS.	CO2
Q8.	Explain MongoDB CRUD operations with examples.	CO2
Q9.	Differentiate between structured, semi structured and un-structured data with examples.	CO2
Q10.	What is Big Data? Explain the different characteristics of Big Data.	CO2
Q11	Describe the importance of map reduce for Big data with the help of example.	CO2

Section C

1. Each Question carries 20 (5+15) Marks.

2. Instruction: Write long answer.

Q12

A) Write query to insert following records in collection name restaurants:

```
{
  "address": {
    "building": "1007",
    "coord": [ -73.856077, 40.848447 ],
    "street": "Morris Park Ave",
    "zipcode": "10462"
  },
  "borough": "Bronx",
  "cuisine": "Bakery",
  "grades": [
    { "date": { "$date": 1393804800000 }, "grade": "A", "score": 2 },
    { "date": { "$date": 1378857600000 }, "grade": "A", "score": 6 },
    { "date": { "$date": 1358985600000 }, "grade": "A", "score": 10 },
    { "date": { "$date": 1322006400000 }, "grade": "A", "score": 9 },
    { "date": { "$date": 1299715200000 }, "grade": "B", "score": 14 }
  ],
  "name": "Morris Park Bake Shop",
  "restaurant_id": "30075445"
}
```

B) Write the following queries using MongoDB:

- i) To display all the documents in the collection restaurants.
- ii) To display the fields restaurant_id, name, borough and cuisine for all the documents in the collection restaurant.
- iii) To display the fields restaurant_id, name, borough and cuisine, but exclude the field _id for all the documents in the collection restaurant.
- iv) To display all the restaurant which is in the borough Bronx
- v) To display the first 5 restaurant which is in the borough Bronx.
- vi) Write a MongoDB query to find the restaurants who achieved a score more than 90.
- vii) Write a MongoDB query to display the next 5 restaurants after skipping first 5 which are in the borough Bronx.
- viii) To find the restaurants that achieved a score is more than 80 but less than 100.

CO3

ix) To find the restaurant Id, name, borough and cuisine for those restaurants which contain 'Wil' as first three letters for its name.	
---	--

x) To find the restaurant Id, name, borough and cuisine for those restaurants which contain 'ces' as last three letters for its name.	
---	--