

Name:	 UPES UNIVERSITY WITH A PURPOSE
Enrolment No:	

UNIVERSITY OF PETROLEUM AND ENERGY STUDIES
Online End Semester Examination, May 2021

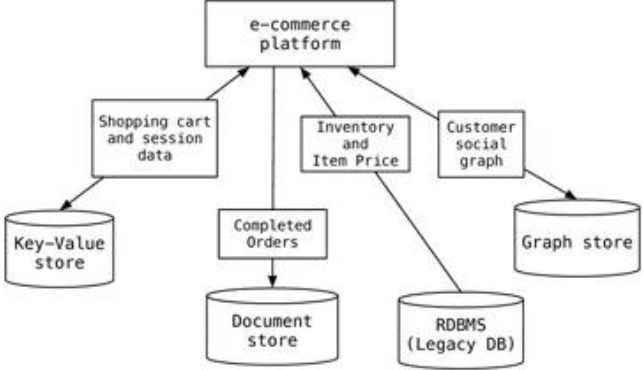
Course: Big Data	Semester: IV
Program: BCA	Time : 03 hrs.
Course Code:	Max. Marks: 100
Instructions: Attempt all questions.	

SECTION A

S. No.		Marks	
Q 1	“Knowing the 5 V's allows data scientists to derive more value from their data while also allowing the scientists' organization to become more customer-centric.” Explain	5	CO1
Q2	It is true that HDFS is used for applications that have large data sets. Why is it not the correct tool to use when there are many small files?	5	CO2
Q3	Write HDFS commands for the following: a) Create directory in HDFS b) Copy a file from one folder to another c) Move a file from one folder to another d) Set replication factor e) Delete directory/file in HDFS	5	CO3
Q4	List the common data challenges. How to organizations overcome data challenges to gain competitive edge.	5	CO4
Q5	Give an account of ‘sharding mechanism’ in NoSQL databases.	5	CO5
Q6	Differentiate between a data swamp and a data lake.	5	CO2

SECTION B

Q 7	Explain the steps to be followed to deploy a Big Data solution.	10	CO1
Q 8	Explain the type of processing used by the following Use Cases: a) Bitcoin b) Multiplayer Online games Also, state the advantages and disadvantages of each type of processing	10	CO2
Q 9	Explain the architecture of the Business data lake. Also, support your explanation with a neat well-labelled figure.	10	CO3

<p>Q 10</p>	 <p>“An e-commerce platform deals with many types of data (i.e. shopping cart, inventory, completed orders, etc). Instead of trying to store all this data in one database, which would require a lot of data conversion to make the format of the data all the same, it stores the data in the database best suited for that type of data” i.e. it means picking the right tool for the right use case.</p> <p>a) Identify the concept presented in the above figure. b) Explain its Pros and Cons.</p>	<p>10</p>	<p>CO4</p>
<p>OR</p>			
<p>Q10</p>	<p>Give a detailed account of Big data ecosystems.</p>		
<p>Q11</p>	<p>Explain “data lineage” process, clearly stating how it is useful for the organizations.</p>	<p>10</p>	<p>CO3</p>
<p>SECTION-C</p>			
<p>Q12</p>	<p>Based in Silicon Valley, @WalmartLabs is a hive of innovation, with hundreds of data specialists who have joined the group through a series of startup acquisitions, the most recent of which was Jet.com and its dynamic pricing platform.</p> <p>As Cognitive Scientist Om Marwah explains in an interview with Forbes: “Traditionally, data science is merely statistics-driven. We also use psychology to make shopping easier, more enjoyable, and more convenient.” Marwah goes on: “Recommendations and targeting are the pillars of our personalization engine.”</p> <p>One of the many innovations developed at the @WalmartLabs is the Social Genome project, a big data analytics solution that analyses millions and billions of Facebook messages, tweets, YouTube videos, blog postings etc., as well as giving customers product information and discounts if they mention a Walmart product on the social media. The group also has a gift-recommending app called Shopycat, which is part of its Facebook page and was launched in 2011.</p> <p>As well as being a key driver for the customer experience for Walmart, big data also enables the retailer to optimize the supply chain. And this is where real-time capabilities are essential, as Naveen Peddamail, who runs the Data Café, explains:</p>	<p>20</p>	<p>CO5</p>

“If you can’t get insights until you’ve analysed your sales for a week or a month, then you’ve lost sales within that time. Our goal is always to get information to our business partners as fast as we can, so they can take action and cut down the turnaround time. It is proactive and reactive analytics.”

Peddmail gives an example of the retailer’s grocery team who were struggling to understand why sales of a particular product were declining. But thanks to data, the Cafe’s analysts were able to attribute the drop in sales to a pricing error. Once this was rectified, sales began to pick up again within days.

Peddmail also recalls monitoring sales of a new range of cookies that were not selling at all in certain outlets. An alert was sent to the team in charge of these stores, who soon realized that the products had not even been put on the shelves. Thanks to real-time data analysis, this oversight was quickly addressed.

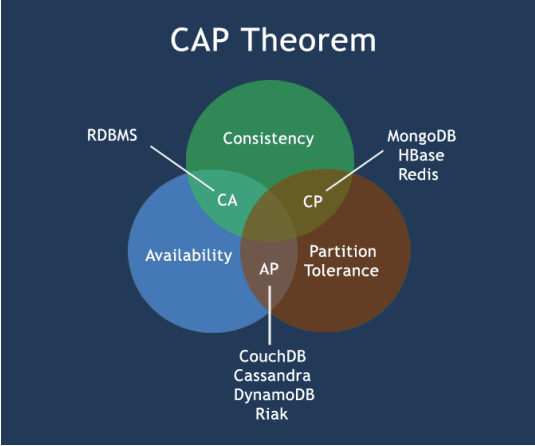
According to Walmart, the Data Café has led to a reduction in the time it takes to resolve problems from an average of two to three weeks to around twenty minutes.

Walmart are precursors in the field, but the retail sector as a whole is on the brink of a revolution thanks to big data, particularly where buying behavior, industry trend analysis, predictive modelling and geographical segmentation are concerned. It is crucial for retailers to seize these opportunities in order to stand out from the competition and offer a unique, innovative customer experience.

- a) Discuss the ways Walmart is using big data to enhance, optimize and customize the shopping experience
- b) How does Walmart improve its operational efficiency?

OR

Q12



With reference to the figure, discuss the cases of CAP theorem- CA, CP and AP for RDBMS, MongoDB and Cassandra respectively.