BHARATIYA VIDYA BHAVAN'S

SARDAR PATEL INSTITUTE OF TECHNOLOGY

MUNSHI NAGAR, ANDHERI (WEST), MUMBAI – 400 058, India

(Autonomous College Affiliated to University of Mumbai)

Mid Semester Examination

Max. Marks: 60	Duration: 1 Hr
Class: BE Computer	Semester: VII
Course Code: CEE71B	Date: 23/11/20
Subject: Big Data Analytics	Time: 5 pm -7 pm

Instructions: (1) All questions are compulsory.
(2) Use of scientific calculator is allowed.
(3) Assume any necessary data but justify the same.

Q. No.	Questions	Max Marks	CO	BL
Q. 1	What are the drawbacks of traditional frequent pattern mining algorithm? Explain SON algorithm in detail with example? **OR** Using flajolet martin algorithm Find number of distinct elements from the the following sequence of stream 1 ,4,2,1,3.4,4,2,5,3,2,5,1,4,6,6,5,3 using hash function H(a)= (2*x +2) mod 5.	10	CO 3	L2
Q. 2	What are the various features of Big data? explain with the help of example of Facebook and YouTube?	10	CO 1	L3
Q. 3	What is recommendation? what are various types of recommendation? Which recommendation is efficient for big data and why? **OR** Explain DGIM algorithm in detail with the help of example?	10	CO 4	L3
Q. 4	How you create a model to categorize tweets into positive negative and neutral, explain whole process with the help of example with the help of naïve bayes algorithm?	10	CO3	L3
Q. 5	What is No SQL? How NOSQL handles big data problems?	10	CO4	L3
Q. 6	Explain following Distance measures with the help of example? a. Euclidian distance b. Cosine distance c. Jaccard Distance d. Edit Distance	10	CO3	

Mid Semester Examination	
Max. Marks: 20	Duration: 1 Hr
Class: TE Computer	Semester: V
Course Code: CE53	Date: 17/09/19
Subject: Data Warehousing and Mining	Time: 10.00AM
Instructions: (1) All questions are compulsory. (2) Use of scientific calculator is allowed. (3) Assume any necessary data but justify the same.	

Synoptic

Q1. Online transaction processing, or OLTP, refers to a class of systems that facilitate and manage transaction-oriented applications, typically for data entry and retrieval transaction processing

•Current data • Changeability: Frequent data changes • Priority: High availability, High data volume • Database Operation: Online update/insert/delete and read • Normalization is very high • Data Structure: Relational (flat tables) • Integration: Minimal • Data Set: 6-18 months

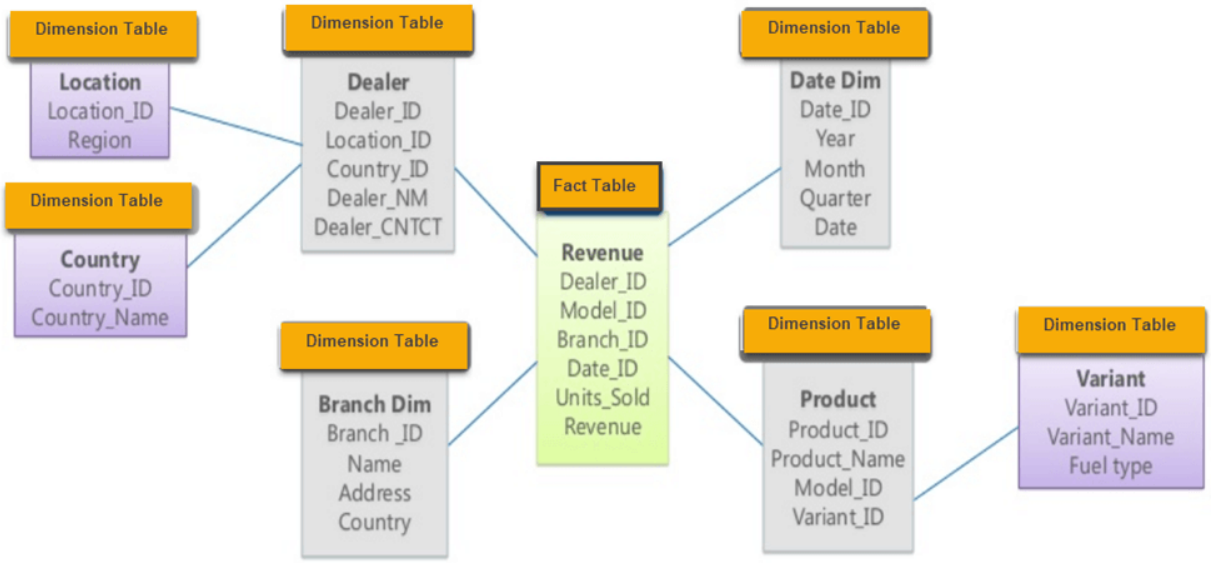
OLAP stands for On Line Analytical Processing, a series of protocols used mainly for business reporting. Using OLAP, businesses can analyze data in all manner of different ways, including budgeting, planning, simulation, data warehouse reporting, and trend analysis

• Current and historical data • Changeability: Data frozen • Priority: Simple to use, flexible access to data • Database Operation: Read • Less Normalization due to data staging and less performance • Data Structure: Multi Dimensional format • Integration: Comprehensive • Data Set: 2-7 years (Min 8 Points) 0.5 Mark for each

OR

Subject Oriented, Integrated Nonvolatile, Time Variant explanation of each feature with example 1 mark each

Q2.



Q3. Correct Frequent Pattern Tree 4 Mark

Correct Frequent Pattern set 2 marks

OR

Correct Frequent Pattern set 4 Mark

Correct Association rules 2 marks

Q4.

Distance Matrix with correct results 4 marks

Correct Dendogram 2 Mark