Q. P. Code: 17089

Time: 3 Hours   **Marks: 80**

Note: 1. Question 1 is compulsory

     2. Answer any three out of remaining questions.

Q1 A) What is dimensional modelling? Design the data warehouse for wholesale furniture [10] Company. The data warehouse has to allow analysing the company's situation at least with respect to the Furniture, Customer and Time. More ever, the company needs to analyse: The furniture with respect to its type, category and material. The customers with respect to their spatial location, by considering at least cities, regions and states. The company is interested in learning the quantity, income and discount of its sales.

B) Discuss different steps involved in Data Pre-processing.   [10]

Q2 A) The college wants to record the Marks for the courses completed by students using [10] the dimensions: i) Course, ii) Student, iii) Time & a measure Aggregate marks. Create a Cube and describe following OLAP operations:
(i) Slice (ii) Dice (iii) Roll up (iv) Drill down (v) Pivot

B) Apply the Naive Bayes classifier algorithm for buys computer classification and [10] classify the tuple X=(age="young", income="medium", student="yes" and credit-rating="fair")

Id	Age	Income	Student	Credit-rating	buys computer
1	young	high	no	fair	no
2	young	high	no	good	no
3	middle	high	no	fair	yes
4	old	medium	no	fair	yes
5	old	low	yes	fair	yes
6	old	low	yes	good	no
7	middle	low	yes	good	yes
8	young	medium	no	fair	no
9	young	low	yes	fair	yes
10	old	medium	yes	fair	yes
11	young	medium	yes	good	yes
12	middle	medium	no	good	yes
13	middle	high	yes	fair	yes
14	old	medium	no	good	no

Q3 A) Explain ETL of data warehousing in details?   [10]

B) Explain types of attributes and data visualization for data exploration.   [10]

DF85B2CAF8DDC703193679392ACDE71E

Q4 A) Illustrate the architecture of Data Warehouse system. Differentiate Data warehouse [10] and Data Mart

B) Explain K-Means clustering algorithm? Apply K-Means algorithms for the [10] following Data set with two clusters.
Data Set = { 15,15,16,19,19,20,20,21,22,28,35,40,41,42,43,44,60,61,65}

Q5 A) Explain Updates to dimension tables in detail. [10]

B) A database has ten transactions. Let minimum support = 30% and minimum [10] Confidence = 70%
i] Find all frequent patterns using Apriori Algorithm.
ii] List strong association rules.

Transaction _Id	Items
01	A,B,C,D
02	A,B,C,D,E,G
03	A,C,G,H,K
04	B,C,D,E,K
05	D,E,F,H,L
06	A,B,C,D,L
07	B,I,E,K,L
08	A,B,D,E,K
09	A,E,F,H,L
10	B,C,D,F

Q6 Write short note on the following (Answer any **FOUR**) [20]

a) Major issues in Data Mining

b) Metadata in Data Warehouse

c) FP Tree

d) DBSCAN

e) Hierarchical Clustering

QP Code : 16278

(3 Hours) [Total Marks : 80

N.B. : (1) Question No. **1** is **compulsory**.
    (2) Solve any **three** questions from remaining questions **2** to questions **6**.
    (3) Please specify you answers with **neat** sketch and examples wherever **necessary**.
    (4) Assume **suitable** information to support your answers and specify the **same**.

1. Answer any **four** questions from the following :—   20
    (a) What are the three categories of the users ? How they can be provided with facilities to ensure feasibility in operation.
    (b) What are the factors that are considered to choose colours ?
    (c) What do you mean by keyboard accelerator and keyboard equivalent ?
    (d) How images and graphics are important in Design.
    (e) Provide brief information on Qualitative and Quantitative Research.

2. (a) Explain following with respect to Handle response Time   12
      (i) Progress Indicator
      (ii) Elapsed Time Message
      (iii) Hourglass Pointer.
  (b) What are three levels of processing and seven stages of Action ? How they are interrelated   8
to each other ?

3. (a) Petroleum Company want to establish unmanned pertrol Pumps at major locations,   12
Where Vehicle owners can fill the petrol on their own and the payment will be either
by cashless in the form of Debit Card or Credit Card or at some points there may be
Bank Note payment (Cash) by automatic machines. Following are some functional
requirements :
There will be automatic gate that will not allow owners to take their vehicles if they didn't
make the payment. Provide detailed system analysis and possible interaction design that
will help to execute the same for pertroleum companies. Your design should also contain
suitbale diagrams.
  (b) List general principles of user interface design, Explain any 4 in detail.   8

4. (a) What are different presentation styles of windows ? State advantages and disadvantages   10
of each style of window.
  (b) Provide different device based and screen based control for following.   10
    (i) Filling up Online Application for Admission to Engineering Course.
    (ii) Online payment of Utility Bills such as Telephone, Electricity and Water charges
    etc.

[TURN OVER

2

5. (a) What do you mean by Persona ? mention steps to design persona.   10
   (b) Explain six behavioral patterns in detail.   10

6. Write short note on any **four** :–   20
   (a) Goal Directed Design
   (b) Gestalt principles
   (c) Menus
   (d) Feedback and Guidance
   (e) Learning.

Duration: 3 Hours   **[Total Marks -80]**

N.B.   (i) **Q. No. 1** is compulsory

(ii) Attempt any **three** questions out of the remaining **five** questions

1	(a)	What are the common issues with which the designer of a heterogeneous distributed system must deal?	05
	(b)	State and prove Amdahl's Law to compute speedup of parallel computers. From experiment it was verified that 70% of execution time was spent on parallel execution. What is the maximum speedup that can be obtained with 16 processors?	05
	(c)	Explain the concept of Processing Element in SIMD architecture.	05
	(d)	Explain stream oriented communication with an example.	05

2	(a)	Discuss Raymond's Tree based algorithm of token based distributed mutual exclusion.	10
	(b)	How pipeline hazards are classified? Discuss data hazard in detail and list the techniques used to eliminate data hazard.	10

3	(a)	Discuss and differentiate various client-centric consistency models.	10
	(b)	Illustrate the parallel Algorithm for matrix multiplication and compare the performance of this algorithm with sequential matrix multiplication algorithm.	10

4	(a)	Describe code migration issues in detail.	10
	(b)	What is a logic clock? Why are logic clocks required in distributed systems? How does Lamport synchronize logical clocks? Which events are said to be concurrent in Lamport timestamps.	10

5	(a)	What is the requirement of Election algorithm in Distributed Systems? Describe any one Election algorithm in detail with an example.	10
	(b)	Define a Remote Procedure Call. Explain the working of RPC in detail	10

6	(a)	Describe File-Cashing schemes.	10
	(b)	Differentiate between Job scheduling and load balancing. Discuss the issues in designing Load Balancing Algorithm.	10

---------------- X ----------------

QP Code : 725401

(3 Hours) [Total Marks :80

N.B. : (1) Question no. **1** is **compulsory**
     (2) Attempt any three from the remaining.
     (3) Assume suitable data.

1. (a) What is Big Data? What is Hadoop? How Big Data and Hadoop are linked?   5
    (b) Explain Page Rank with Example. Can a Website's Page rank Ever Increase? What   5
       are its chances of Decreasing?
    (c) Explain Hubs and Authorities with neat diagram.   5
    (d) With respect to data stream querying, give example of   5
         (a) One Time queries
         (b) Continuous Queries
         (c) Pre-defined queries
         (d) Ad-hoc queries

2. (a) Explain Hadoop Ecosystem with core components, Explain its Physical architecture.   10
      State Limitations of Hadoop.
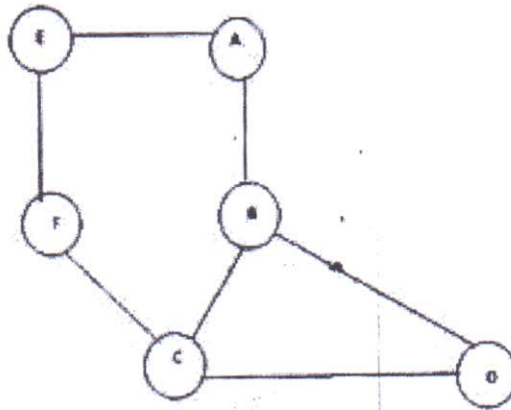
    (b) What is MapReduce ? Explain How Map and Reduce Work? What is Shuffling in   10
      MapReduce?

3 (a) For the Graph given below use betweenness factor and find all communities.   10
    (b) How would you get the features of the document in a content -based system? Explain   5
      document similarity.
    (c) What is triangular matrix? How it is used for main memory counting?   5

[TURN OVER]

4 (a) Explain Collaborative Filtering based recommendation System. How it is different from content based recommendation systems ?   10

   (b) What are Combiners? When Should one use combiner in mapreduce job?   5

   (c) How to count distinct elements in a stream? Explain Flajolet-Martin Algorithm.   5

5 (a) Given a 1 Dim Dataset {1,5,8,10,2} Use the agglomerative clustering algorithm with Euclidean distance to establish hierarchical grouping relationship. Draw the dendrogram .   10

   (b) Consider a Portion of Web Graph Shown below: 10
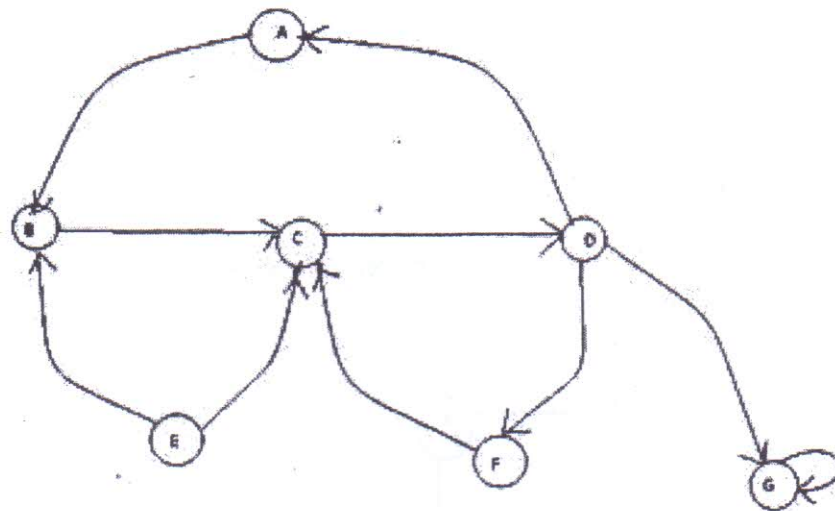
   (a) Compute the hub and authority scores for all the nodes.

   (b) Does this graph contains spider traps? Dead ends? If so, which nodes?

   (c) Compute the page Rank of the nodes with teleportation β =0.8? (Show two iterations only)

6. (a) What is NoSQL? What are the business drivers for NoSQL? Discuss any two architectural patterns of NoSQL.   10

   (b) What is a Data Stream Management System? Explain with Block Diagram   10

[Time: 3 Hours] [Marks:80]

Please check whether you have got the right question paper.

N.B:
1. Question no 1 is compulsory.
2. Attempt any three questions out of remaining five questions.
3. Assume any suitable data wherever required but justify the same.

Q.1	A.	What is machine learning? Explain how supervised learning is different from unsupervised learning.	05
	B.	Explain Bayes theorem.	05
	C.	What are the elements of reinforcement learning?	05
	D.	Describe the two methods for reducing dimensionality.	05

Q.2 A. The following table shows the midterm and final exam grades obtained for students in a database   10
course.

Midterm exam (x)	Final exam (y)
72	84
50	63
81	77
74	78
94	90
86	75
59	49
83	79
65	77
33	52
88	74
81	90

Use the method of least squares using regression to predict the final exam grade of a student who received 86 on the midterm exam.

B. Explain the steps in developing a machine learning application.   10

Q.3 A. For a SunBurn dataset given below, construct a decision tree.   10

Name	Hair	Height	Weight	Location	Class
Sunita	Blonde	Average	Light	No	Yes
Anita	Blonde	Tall	Average	Yes	No
Kavita	Brown	Short	Average	Yes	No
Sushma	Blonde	Short	Average	No	Yes
Xavier	Red	Average	Heavy	No	Yes
Balaji	Brown	Tall	Heavy	No	No
Ramesh	Brown	Average	Heavy	No	No
Swetha	Blonde	Short	Light	Yes	No

1

Turn over

B. What is Support Vector Machine (SVM)? How to compute the margin?   10

Q.4   A. Explain Hidden Markov Models.   10

B. Use Principal Component analysis (PCA) to arrive at the transformed matrix for the given matrix A.   10
A^T = 2 1 0 -1
        4 3 1 0.5

Q.5   A. Explain how Back Propagation algorithm helps in classification.   10

B. For the given set of points identify clusters using complete link and average link using agglomerative   10
clustering.

	A	B
P1	1	1
P2	1.5	1.5
P3	5	5
P4	3	4
P5	4	4
P6	3	3.5

Q.6   Write short notes on any two:-
     A. Temporal difference learning   20
     B. Logistic regression
     C. Machine learning Applications

2