QUESTIONS BANK

Chapter 1

Q1 .Write short note on: Operational support system.

Q2. Explain Data Warehouse Architecture in detail

Q3. Differentiate Data Warehouse vs. Data Mart

Q4.Differentiate top-down and bottom-up approaches for building data warehouses. Discuss the merits and limitations of each approach.

Q5. What is meant by Metadata in the context of a Data warehouse? Explain the   different types of metadata stored in a data warehouse. Illustrate with a suitable example.

Q6. What is metadata?Why do we need metadata when search engines like google seem so effective?

Q7 Explain the steps in dimensional modeling and its advantages of dimensional modeling

Q8. Explain with examples Factless Fact Table and Fact constellation.

Q9. In real-world data, tuples with missing values for some attributes are a common occurrence. Describe various methods for handling this problem

Q10. Describe the process of Extraction, Transformation and Loading (ETL)with a neat and labeled diagram.

Q11. Discuss various OLAP models and their architecture.

Q12. Describe about the following OLAP operations on a cube:

   i) Rollup ii) Drill down iii) Slice iv) Dice v) Pivot.

Q13. **All Design Star Schema case study problems**

Chapter 2

1)Explain the process of Knowledge discovery in databases

2)For the given data apply smoothing by a) bin- mean, b) bin- boundaries

   8,9,15,30,16,24,26,27,21,21,30,34

3) For the given data values 8,9,15,30,16,24,26,27,21,21,30,34
 lFind Mean, Median, Mode, Midrange,Q1,Q2,Q3, five number summary, also draw box plot

4) Short notes on data visualization techniques for data warehousing and data mining

5) Explain steps involved in data preprocessing.

Numerical

Use k-means algorithm to create two clusters (Use Euclidean distance)
A(2,2) B(3,2) C(1,1) D(3,1)E(1.5,0.5)

Solution

Algorithm 8.1 Basic K-means algorithm.
1: Select K points as initial centroids.
2: **repeat**
3:    Form K clusters by assigning each point to its closest centroid.
4:    Recompute the centroid of each cluster.
5: **until** Centroids do not change.

Use k-means algorithm to create two clusters (Use Euclidean distance)

A(2,2) B(3,2) C(1,1) D(3,1)E(1.5,0.5)

Objects		
A	2	2
B	3	2
C	1	1
D	3	1
E	1.5	0.5

Solution:
Select randomly two objects as centroids eg A(2,2) and C(1,1). Compute distance of every object to each centroids
Euclidean distance between A(2,2) and c(1,1)
$(|2-1|^2 + |2-1|^2)^{1/2} = (1+1)^{1/2} = 1.41$

Iteration 1

		A (2, 2)	C (1,1)	
	Point	Dist Mean 1	Dist Mean 2	Cluster
A	(2, 2)	0	1.41	1

B	(3,2)	1	2.24	1
C	(1,1)	1.41	0	2
D	(3,1)	1.41	2	1
E	(1.5,0.5)	1.58	0.71	2

Clusters are

Cluster1	Cluster2
A(2, 2) B (3,2) D(3,1)	C(1,1) E(1.5,0.5)
Calculate centroid	
C1=((2+3+2)/3 , (2+2+1)/3) (2.67,1.67)	C2=(1+1.5)/2 , (1+.5)/2 =(1.25,0.75)

Iteration 2

		C1 (2.67, 1.67)	C2 (1.25,0.75)	
	Point	Dist Mean 1	Dist Mean 2	Cluster
A	(2, 2)	0.75	1.45	1
B	(3,2)	0.47	2.15	1
C	(1,1)	1.79	0.32	2
D	(3,1)	0.75	1.76	1
E	(1.5,0.5)	1.65	0.36	2

Clusters are

Cluster1	Cluster2
A(2, 2) B (3,2) D(3,1)	C(1,1) E(1.5,0.5)
Calculate centroid	
C1=((2+3+2)/3 , (2+2+1)/3) =(2.67,1.67)	C2=(1+1.5)/2 , (1+.5)/2 =(1.25,0.75)