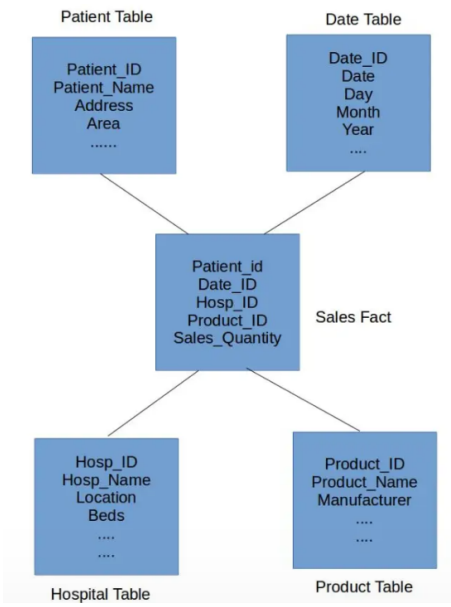
TE AI&DS(2022-23)

DWM : Sample problem statements for Practical exam

1. Suppose that a data warehouse for Big_University consist of the four dimensions student, course, semester and instructor, and two measures count and avg_grade, where Count is the number of students and avg_grade is the course grade of the student.
   Perform following tasks:
   a. Design the start schema for the Big_University
   b. Create dimension tables and fact table

2. Suppose that the data a data warehouse consists of the four dimensions date, spectator, location and game, and two measures count and charge where charge is the fare that a spectator pays when watching on the given date. Spectator may be students, adults or seniors, with each category having its own charge rate.
   a. Design the start schema for the Big_University
   b. Create dimension tables and fact table

3. For given star schema. Create database and apply any 3 OLAP operations



4. For the given dataset calculate the information gain for rank attribute.

NAME	RANK	YEARS	TENURED
Mike	Assistant Prof	3	no
Mary	Assistant Prof	7	yes
Bill	Professor	2	yes
Jim	Associate Prof	7	yes
Dave	Assistant Prof	6	no
Anne	Associate Prof	3	no

5. For the given dataset calculate the information gain for food attribute.

Age	Food	Animal
1	Meat	Dangerous Tiger
3	Meat	Dangerous Tiger
7	Grass	Zebra
10	Meat	Tiger
3	Grass	Zebra
9	Grass	Zebra
5	Meat	Tiger
6	Grass	Zebra

6. For the given dataset WAC to calculate posteriori probability for zebra class for given tuple X(2,"grass")

Age	Food	Animal
1	Meat	Dangerous Tiger
3	Meat	Dangerous Tiger
7	Grass	Zebra
10	Meat	Tiger
3	Grass	Zebra
9	Grass	Zebra
5	Meat	Tiger
6	Grass	Zebra

7. Write a code to handle missing values.

Stud_id	Stud_branch	Marks
1	AI&DS	17
2	COMP	
3	COMP	15
4	COMP	16

5	AI&DS	18
6	AI&DS	15
7	AI&DS	

8. Implement min max normalization on following dataset in python

Emp_id	Salary
101	120000
102	145000
103	320000
104	45000
105	90000
106	134000
107	200000

9. WAC to apply data discretization and concept hierarchy for given dataset.

Age	Food	Animal
1	Meat	Dangerous Tiger
3	Meat	Dangerous Tiger
7	Grass	Zebra
10	Meat	Tiger
3	Grass	Zebra
9	Grass	Zebra
5	Meat	Tiger
6	Grass	Zebra

10. Write a code to handle mission values.

Stud_id	Stud_branch	Marks
1	AI&DS	17
2	COMP	
3	COMP	15
4	COMP	16
5	AI&DS	18
6	AI&DS	15
7	AI&DS	

11. Write a code to apply standardization (z score) on salary attribute using python.

Emp_id	Salary

101	120000
102	145000
103	320000
104	45000
105	90000
106	134000
107	200000

12. Calculate the posterior probability for “yes” class for given tuple P(Ci/X).

age	income	student	credit_rating	buys_computer
<=30	high	no	fair	no
<=30	high	no	excellent	no
31…40	high	no	fair	yes
>40	medium	no	fair	yes
>40	low	yes	fair	yes
>40	low	yes	excellent	no
31…40	low	yes	excellent	yes
<=30	medium	no	fair	no
<=30	low	yes	fair	yes
>40	medium	yes	fair	yes
<=30	medium	yes	excellent	yes
31…40	medium	no	excellent	yes
31…40	high	yes	fair	yes
>40	medium	no	excellent	no

Tuple=(<=30, high, no,fair)

13. Implement k-mean clustering upto 2 iterations on following dataset. Print the iteration data.
2,5,8,12,14,16,22,24,30,32,35

14. Write a code to display dissimilarity matrix for the given dataset.

X	Y
2	3
2	4
4	1
3	2

15. Write a code to display dissimilarity matrix for the given dataset.

X	Y
1	2
2	3
4	1
5	2

16. Write a code to find frequent 2 itemset from the given transactional database.

Support =70%

TID	Items
100	1,3,4,2
200	2,3,5,1
300	1,2,3,4
400	2,5,3
500	1,2,3,4

17. Write a code to find frequent 2 itemset from the given transactional database. Support=60%

TID	Items
100	a,c,b
200	a,c,d
300	a,c
400	b,d,c
500	a,b,d

18. Write a code to find frequent 3 itemset from the given transactional database. Support=60%

TID	Items
100	a,c,b,d
200	a,c,d
300	a,c
400	b,d,c
500	a,,b,d

19. Write a code in python to describe data (dataset is uploaded on desktop)
    a. Display any 5 tuples
    b. To display attribute names and information
    c. Display mean, median, Quartile 1, Quartile 2 and Quartile 3
    d. Correlation between 3 attributes

20. Write a code to implement binning in python. (create 3 bins)

Id	Salary
1	120000
2	23000
3	34000
4	45000
5	50000
6	110000
7	80000
8	60000
9	150000

10	125000

21. WAC to apply data discretization and concept hierarchy in python.

Id	Salary
1	120000
2	23000
3	34000
4	45000
5	50000
6	110000
7	80000
8	60000
9	150000
10	125000

22. WAC to apply data discretization and concept hierarchy in python.

Id	Age
1	27
2	23
3	34
4	45
5	49
6	69
7	62
8	55
9	21
10	18

23. Implement data smoothing technique (smoothing by bin mean) on following dataset.

12,4,5,21,7,14,15,18,20

24. Implement data smoothing technique (smoothing by bin median) on following dataset.

12,4,5,21,7,14,15,18,20

25. Implement data smoothing technique (smoothing by bin boundary) on following dataset.

12,4,5,21,7,14,15,18,20