Code No: **R204112D** **R20** **Set No. 1**

IV B.Tech I Semester Regular Examinations, January – 2024

ADVANCED DATABASES

(Information Technology)

Time: 3 hours **Max. Marks: 70**

Answer any FIVE Questions

ONE Question from Each unit

All Questions Carry Equal Marks

*****

UNIT - I

1 a) Consider the universal relation R = {A, B, C, D, E, F, G, H, I, J} and the set of functional dependencies F={ {A,B} → {C}, {A} → {D,E},{B}→{F},{F}→{G,H},{D} →{I,J}}. What is the key for R? Decompose R into 2NF, then 3NF relations. [8]

b) Discuss the main categories of data models. What are the basic differences between the Relational model and the object model? [6]

(OR)

2 a) State and explain BCNF with example. And how does it differ from 3NF? [7]

b) Consider the following relational schemas. Write the following queries in relational algebra, tuple relational calculus.
Suppliers(sid: integer, sname: string, address: string)
Parts(pid: integer, pname: string, color: string)
Catalog(sid: integer, pid: integer, cost: real)

i) Find the sids of suppliers who supply some red part or are at 221 Packer Street.

ii) Find the sids of suppliers who supply every red or green part.

iii) Find the pids of the most expensive parts supplied by suppliers named Yosemite Sham. [7]

UNIT - II

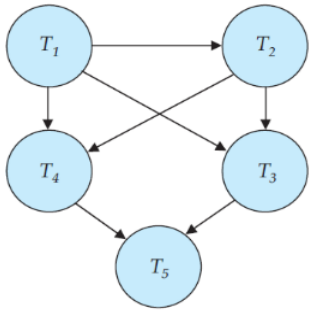
3 a) Why concurrency control is needed? Explain the problems that would arise when concurrency control is not provided by the database system? [7]

b) Discuss the rules for transformation of query trees and identify when each rule should be applied during optimization. [7]

(OR)

4 a) Briefly discuss Online Analytical Processing. [7]

b) Consider the precedence graph of adjacent figure. Is the corresponding schedule conflict serializable? Explain your answer.



[7]

Code No: **R204112D** **R20** **Set No. 1**

UNIT - III

5 a) Suppose a hospital tested the age and body fat data for 18 randomly selected adults with the following result

age	23	23	27	27	39	41	47	49	50
%fat	9.5	26.5	7.8	17.8	31.4	25.9	27.4	27.2	31.2
age	52	54	54	56	57	58	58	60	61
%fat	34.6	42.5	28.8	33.4	30.2	34.1	32.9	41.2	35.7

i) Calculate the mean, median and standard deviation of age and %fat.
ii) Draw the boxplots for age and %fat.
iii) Normalize the two variables based on z-score normalization. [7]

b) In real world data, tuples with missing values for some attributes are common occurrence. Describe various methods for handling this problem. [7]

(OR)

6 a) Describe various schemes used for the design of multi dimensional data model. [7]

b) Use these methods to normalize the following group of data:
200, 300, 400, 600,1000.
i) min-max normalization by setting min = 0 and max = 1
ii) z-score normalization
iii) z-score normalization using the mean absolute deviation instead of standard deviation
iv) normalization by decimal scaling [7]

UNIT - IV

7 a) Find the frequent pattern using FP-Growth algorithm.

T1D	Items
T1	{HotDogs, Buns, Ketchup}
T2	{HotDogs, Buns}
T3	{HotDogs, Coke, Chips}
T4	{Chips, Coke}
T5	{Chips, Ketchup}
T6	{HotDogs, Coke, Chips}

[9]

b) Discuss briefly compact representation of frequent item sets. [5]

(OR)

8 a) Suppose that the data for analysis includes the attribute age. The age values for the data tuples are (in increasing order) 13, 15, 16, 16, 19, 20, 20, 21, 22, 22, 25, 25, 25, 25, 30, 33, 33, 35, 35, 35, 35, 36, 40, 45, 46, 52, 70.
i) What is the mean of the data? What is the median?
ii) Can you find (roughly) the first quartile (Q1) and the third quartile (Q3) of the data?
iii) Show a boxplot of the data. [7]

b) Describe rule generation in Apriori algorithm with example. [7]

UNIT - V

9 a) Discuss web usage mining and its essential components. [7]

b) Briefly explain decision tree algorithm. [7]

(OR)

10 a) What is cluster analysis? Suppose that the data mining task is to cluster points (with (x, y) representing location) into three clusters, where the points are A1(2,10), A2(2,5), A3(8,4), B1(5,8), B2(7,5), B3(6,4), C1(1,2), C2(4,9). The proximity measure is Euclidean distance. Find clusters by applying single linkage, complete linkage algorithms. [9]

b) Briefly explain instance based classification methods. [5]