Question Paper

Exam Date & Time: 02-May-2024 (10:00 AM - 01:00 PM)

MANIPAL ACADEMY OF HIGHER EDUCATION

Manipal School of Information Sciences (MSIS), Manipal
Second Semester Master of Engineering - ME (Big Data Analytics) Degree Examination - April / May 2024

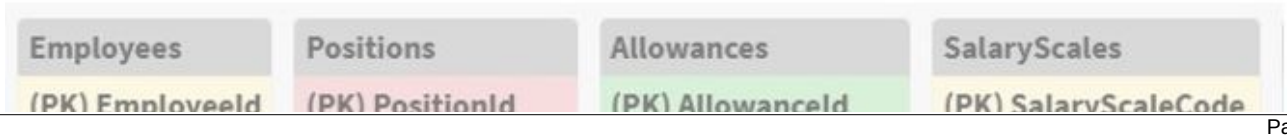
Modern Databases for Big Data [BDA 5202]

Marks: 100 **Duration: 180 mins.**

Thursday, May 02, 2024

Answer all the questions.

1) [L3, CO1] A small accounting firm wants a simple HR application that will help it to keep track of its employees, their positions, allowances and salary scales. The application must keep track of all the positions at the firm, the employees filling these positions, the allowances for these positions and the salary scales for these positions. By referring to the image, design the e-r diagram (you can make few assumptions). Identify primary key, foreign key, composite, dependent and multivalued attribute. Identify the relationship with the entities. Identify cardinality relationship. 10 Marks (10)



SSNumber	**PositionName**	**AllowanceName**	**SalaryScaleName**
LastName	PositionDescription	AllowanceDescription	SalaryScaleDescription
FirstName	Details	Amount	MinimumSalary
MiddleName			MaximumSalary
Gender			
DOB			
Email			
Mobile			
HTel			
AddressLine1			
AddressLine2			
City			
State			
PostCode			

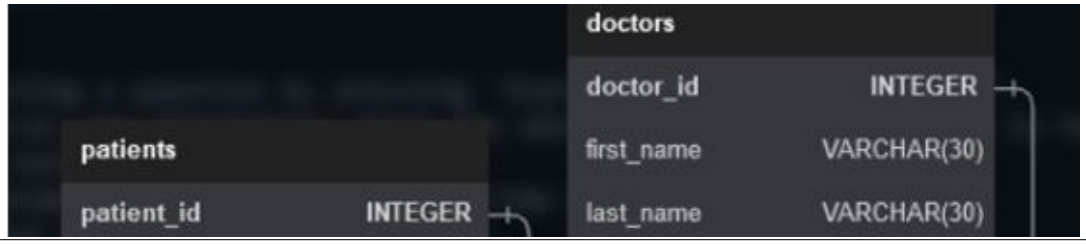
2) [L3, CO1] (10)

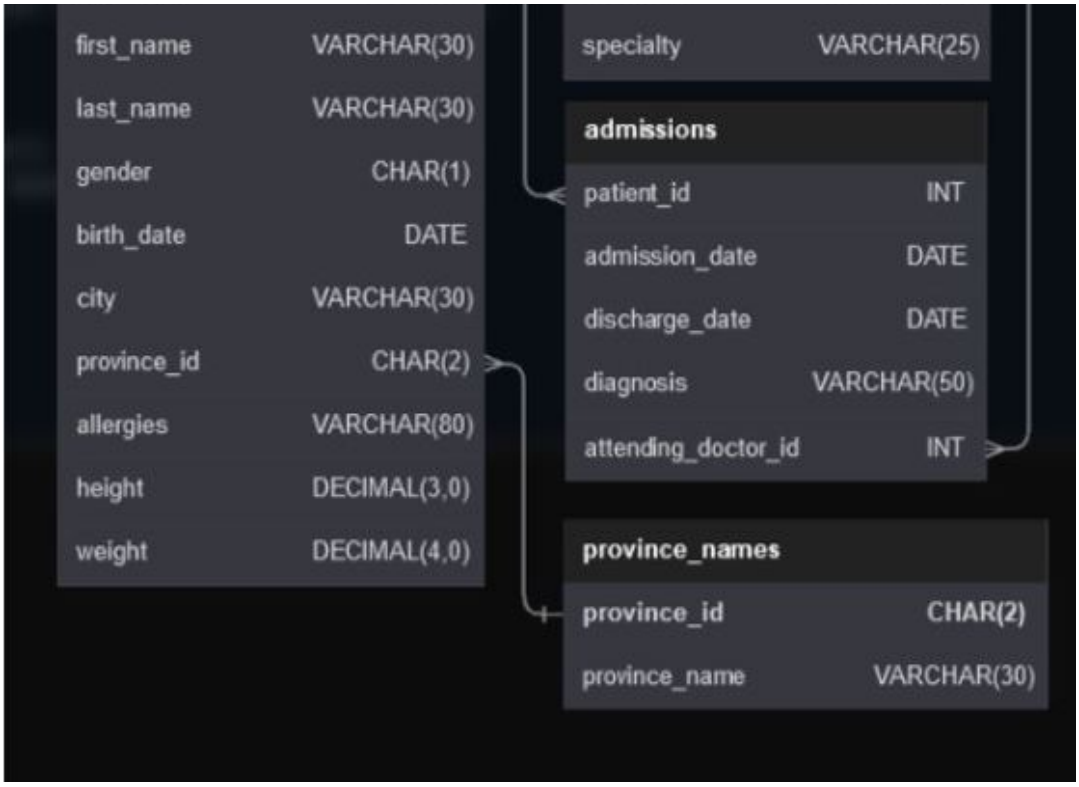
Part A: Refer to Question number 1 and design relational tables by normalizing the entities into 2NF. Specialize the tables to accommodate different types of employees. Ensure that every row can be identified uniquely. 6 Marks

Part B: Illustrate how RDBMS adheres to ACID Properties 4 Marks

3) [L3, CO1[Write SQL statement for the following queries (2x5 = 10 Marks) (10)

A. Show unique birth years from patients and order them by ascending.
B. Count number of patients from each Province.
C. Show first name, last name, and province_name of patients whose attending doctor in "Mr. XYZ"
D. Display number of patients admitted on 12-05-2022.
E. Show the city and the total number of patients in the city. Order from most to least patients and then by city name ascending.





4) [L2, CO2] (10)
Part A. Illustrate data sharding? When do you apply key based, range based, directory based and geo-based sharding? 7 Marks

[L3, CO2] Part B. Which sharding architecture would you choose for the following application for better performance 3 Marks
1. Financial software
2. OTT Platform
3. Online Shopping Platform

5) [L2, CO2] Illustrate are BASE properties? Describe replication in database? Explain any two implementations of Read-your-write consistency. (2+2+4 = 10 Marks) (10)

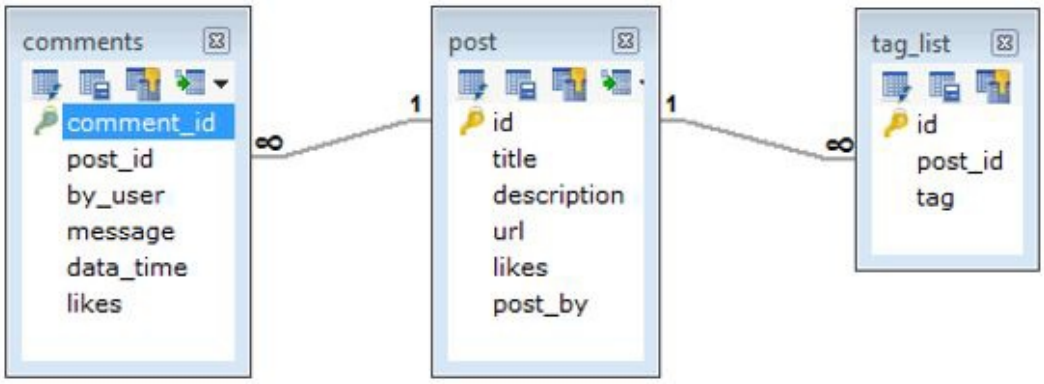
6) [L2, CO2] List out the advantages of using noSQL over tradional database . Explain any to two types of NoSQL database with its advantages. (2+4+4 = 10 Marks) (10)

7) [L2, CO2] Part A. What is MongoDB? What are its features? 4 Marks (10)

[L3, CO2] Part B. Mention which database is suitable for the following example (Mention the type of DB, breifly in a line or two explain why) (6 Marks):

1. To store and update various web page metrics and counter
2. LinkedIn
3. Suggesting reels in Youtube
4. Monitoring traffic from each IP address in a company
5. Displaying product content on online store
6. Monitoring temperature and humidity in different cities every 10 minutes

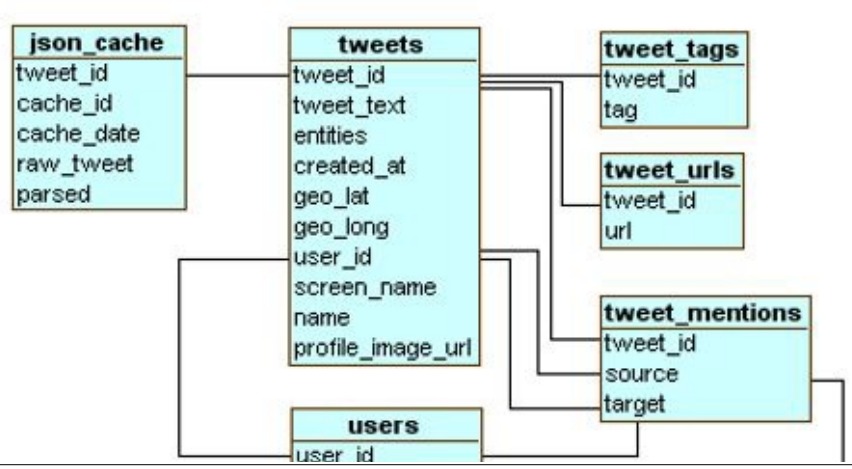
8) [L3, CO3] Part A: Convert the given relational database to document database (Write json document) 5 Marks (10)



[L2, CO3] Part B: Illustrate the role of aggregate pipeline in mongoDB. List out 4 most important stages in aggregate pipeline. 5 Marks

9) [L2, CO3] What is Graph Data Model? List out the applications of GraphDB? What is Cypher query? Explain CRUD operations in Neo4J with an example each. (2+2+6 = 10 Marks) (10)

10) [L3, CO3] Model the following twitter RDBMS model to graph model. (10)




screen_name
name
profile_image_url
location
url
description
created_at
followers_count
friends_count
statuses_count
time_zone
last_update

-----End-----