Assignment 1: Distributed Databases (10%)

Guidelines for Assignment Submission:

1. Submission Deadline: **March 28, 2024, 16:00 (AEST).**
2. For questions requiring queries, kindly include both a screenshot of the query and a screenshot of the execution results.
3. When addressing discussion-based questions, provide comprehensive textual responses, supplemented with suitable visual aids if necessary.
4. The submission should be in PDF format, named as 'A1_s1234567.pdf' (replace '1234567' with your student ID). The submitted file should not exceed 10 MB.
5. All implementations and project components should be finalized within the UQZones environment. Evaluators might assess the assignment based on the established checkpoints within UQZones.
6. All assignment submissions must be submitted exclusively through the UQ Blackboard. Alternative methods of submission will not be acknowledged. Please be mindful that submissions via email will not be entertained under any circumstances.
7. It is imperative to adhere to the stipulated submission deadline to avoid penalties, as explained in the Educational Course Policies (ECP).
8. Ensuring the successful submission of your assignment within the designated timeline is your responsibility.

Introduction

In this assignment, you will be provided with four sql files that corresponding to four tables ('salaries', 'dept_emp', 'dept_manager', 'employees' and 'departments') in the 'EMP' database. Figure 1 shows the conceptual model of 'EMP' database, where red rectangular indicate primary keys, and blue squares indicate the foreign keys. The table 'titles' is stored in a remote server.
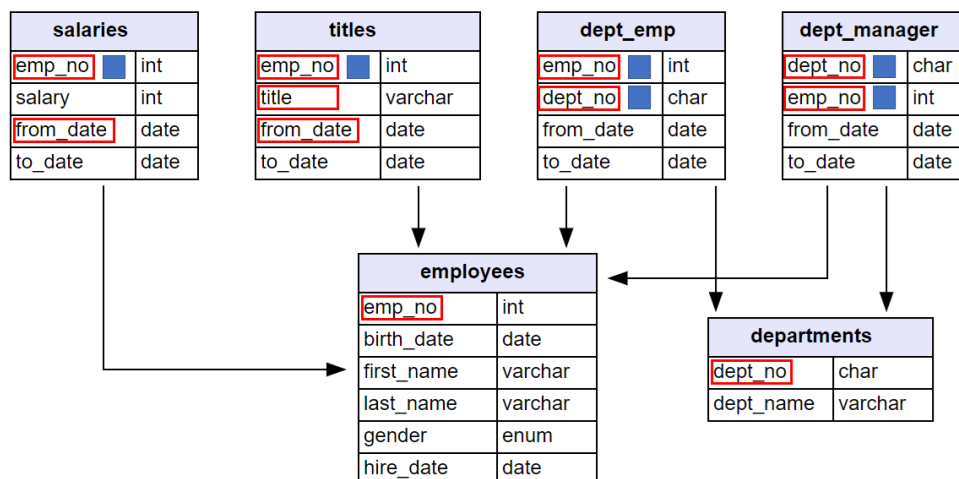

Figure1. The conceptual model of EMP database.

Task 1: Load Database (1 point)

Create a database named 'EMP_s1234567' (1234567 should be replaced by your studentID). Load the database with the provided sql files.

- (0.5) Write sql queries to count the number of entries in table 'employees'.
- (0.5) Write sql queries to count the number of employees from the 'Marketing' department.

Task 2: Database Fragmentation (2.5 points)

- (1) Write sql queries to perform horizontal fragmentation on table 'salaries', based on the following rules:
  - 'from_date' before 1990-01-01
  - 'from_date' no earlier than 1990-01-01 and before 1992-01-01
  - 'from_date' no earlier than 1992-01-01 and before 1994-01-01
  - 'from_date' no earlier than 1994-01-01 and before 1996-01-01
  - 'from_date' no earlier than 1996-01-01 and before 1998-01-01
  - 'from_date' no earlier than 1998-01-01 and before 2000-01-01
  - 'from_date' no earlier than 2000-01-01

- (0.5) Calculate the average employee salary in between 1996-06-30 and 1996-12-31 on the fragmented 'salaries' table. Show the query explanation.

- (1) Write sql queries to perform vertical fragmentation on table 'employee', based on the following rules:
  - 'emp_no', 'first_name', 'last_name', and 'hire_date' should be stored in table 'employees_public'
  - 'emp_no', 'birth_date', and 'gender' should be stored in table 'employees_ confidential'
  - The 'employees_ confidential' table should be stored in a new database called 'EMP_Confidential'.
    - *Hint: you can export data to a sql file.*

Task 3: Database Replication (2.5 points)

Consider the table 'employee' is sliced into 10 fragmentations based on the value of 'emp_no'.

- (1.5) Consider you have 5 different server sites. How to design full replication/ partial replication/ no replication? What could be the pros and cons for each fragmentation strategy?
- (1) Consider your master server has the fragmentation allocation schema. Based on your design, what is the process to update a record with a specific 'emp_no'?

Task 4: Access foreign data with FDW (4 points)

The table 'titles' is stored in a remote database that requires you to use Foreign Data Wrappers (FDW) to access the data. You are provided with the following login details:

- ➢ User Name: sharedb
- ➢ Password: Y3Y7FdqDSM9.3d47XUWg
- ➢ Host: infs3200-sharedb.zones.eait.uq.edu.au
- ➢ Port: 5432
- ➢ Database Name: sharedb

- (1) Write sql queries to establish a FDW to the external DB
- (1) For each unique 'title' in table 'titles', calculate the averaged current salary.
    - o *Hint: current staff has 'to_date' that equals to '9999-01-01'*
- (1) Consider that the current employee table is vertically fragmented and stored on different databases. Perform semi join from 'EMP' database to select the last name and first name of employees that 'birth_date' is no earlier than 1970-01-01 and before 1975-01-01.
    - o *Note: 'birth_date' information can only be accessed from 'EMP_Confidential' database through FDW. Direct read from the original employee table will receive 0 mark for this question.*
- (1) Compared with inner join, will semi join incur higher transmission cost under this scenario?
    - o *Note: Screenshots are required for the comparison.*