



# COS 221 Practical Assignment 4

- Date Issued: **22th April 2020**
- Date Due: **20th May 2020 before 11:30 (in the morning)**
- Submission Procedure: **Upload source program and report to CS web**
- This assignment consists of **6 tasks** for a total of **40 marks**.

## 1 Introduction

In this assignment, you are required to use the Employees MySQL sample database that provides a combination of a large base of data spread over six separate tables and consisting of 4 million records in total. An ER-diagram of the database is given in Figure 1. You are required to execute and optimise high level queries in SQL and Relational Algebra (RA).

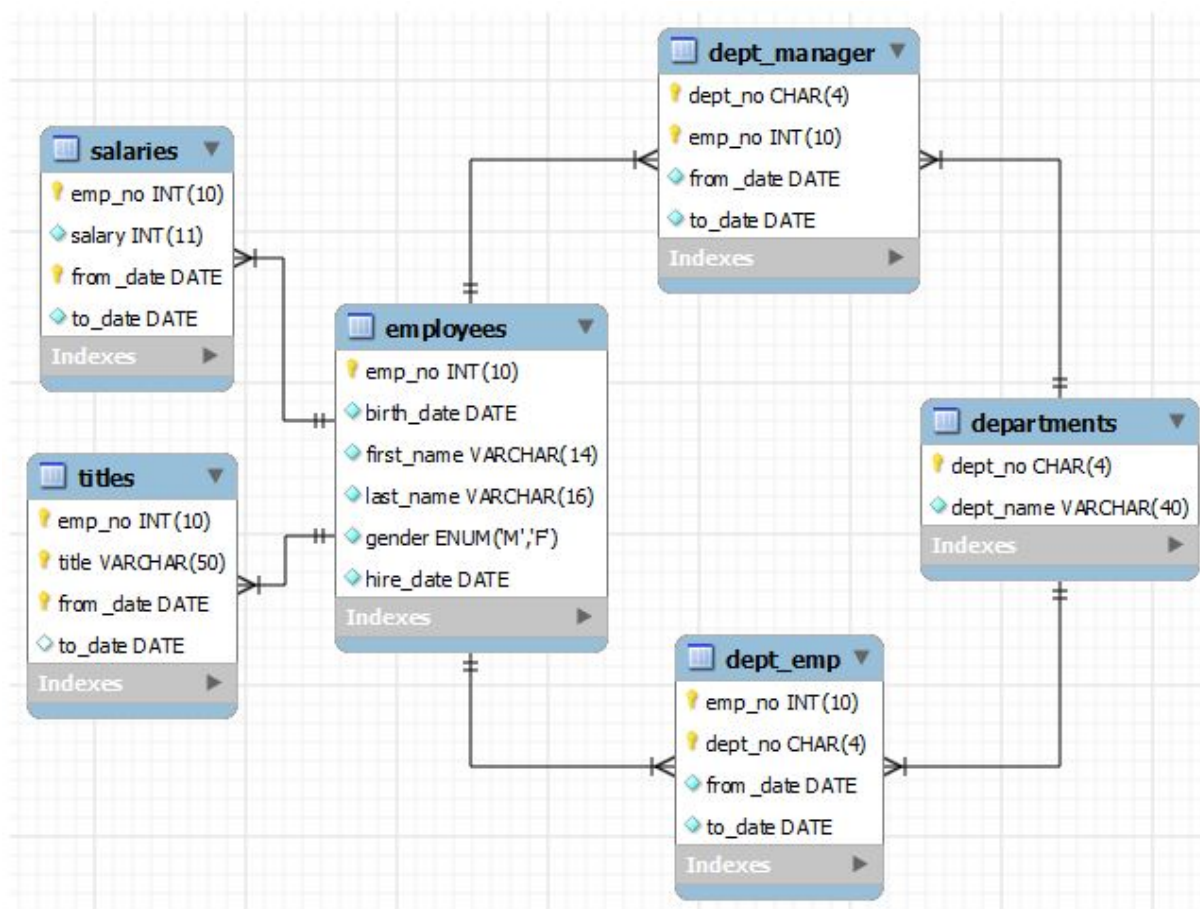


Figure 1: The ERD

After successful completion of this assignment you should be able to:

- import any database
- optimise high level queries,
- use `indexes` to find rows with specific column values quickly,
- use `explain` for examining queries involving partitioned tables

## 2 Constraints

1. You must complete this assignment individually.
2. The SQL scripts will be marked
  - (a) Scripts which run and perform what they are supposed to do get full marks
  - (b) Scripts which run but do not perform as required, will receive partial marks
  - (c) Scripts which do not run will be allocated partial marks based on the functionality they would have exhibited..
3. The Java benchmark program will be marked.
4. Your report will be marked as uploaded to CS-web.
5. You may ask the Teaching Assistants for help but they will not be able to give you the solutions.
6. You may utilise any text editor or IDE, upon an OS of your choice. Whilst at home, you will use either MySQL Workbench or MariaDB to import your database on `localhost server`. All files and scripts must be uploaded and run on `localhost` for demonstration purposes.

### 3 Submission Instructions

You are required to upload all your source files to the Computer Science web-portal along with your report in a single archive. You also need to make sure that `localhost` mirrors what you uploaded to CS web and works on the web server before the deadline. No late submissions will be accepted, so make sure you upload in good time. You will be required to download the files you uploaded to CS web and load them onto `localhost` as part of the assessment of the practical assignment.

### 4 Online resources

employees database install tutorial: [https://stackoverflow.com/questions/36322903/error-1193-when-following-employees-database-install-tutorial-with-mysql-5-7-1?utm\\_medium=organic&utm\\_source=google\\_rich\\_qa&utm\\_campaign=google\\_rich\\_qa](https://stackoverflow.com/questions/36322903/error-1193-when-following-employees-database-install-tutorial-with-mysql-5-7-1?utm_medium=organic&utm_source=google_rich_qa&utm_campaign=google_rich_qa)

Access a free SQL Tutorial at: [https://www.w3schools.com/sql/sql\\_create\\_table.asp](https://www.w3schools.com/sql/sql_create_table.asp) and  
<https://dev.mysql.com/doc/employee/en/employees-installation.html>

Optimization Overview <https://dev.mysql.com/doc/refman/5.7/en/optimize-overview.html>

RA documentation: <https://users.cs.duke.edu/~junyang/radb/basic.html>

**IMPORTANT NOTE:** Make use of your textbook<sup>1</sup> and/or the lecture notes for query processing and optimization.

### 5 Rubric for marking

Importing the dataset	5
Benchmark Java program	5
Timing the execution of queries	5
Using EXPLAIN keyword	5
Query trees	10
Comparing relational algebra expressions	10
<b>Total</b>	<b>40</b>

---

<sup>1</sup>Some of the source code examples and database design techniques are in Edition 6 [1]

## 6 Assignment Instructions

### Task 1: Importing the dataset ..... (5 marks)

You are required to download the sample database using the link:

[https://launchpad.net/test-db/employees-db-1/1.0.6/+download/employees\\_db-full-1.0.6.tar.bz2](https://launchpad.net/test-db/employees-db-1/1.0.6/+download/employees_db-full-1.0.6.tar.bz2)  
and import it to localhost server. Remember to name your schema uXXXXXXXX business, where XXXXXXXX is your student number and business is the name of the database schema.

### Task 2: Benchmark Java program ..... (5 marks)

Write a Java program that connects to the database and accepts an SQL query as input as well as the number of times the query must execute. The Java program will then execute the query the specified number of times taking a time measurement from when each query is submitted to when the query has completed.

### Task 3: Timing the execution of queries ..... (5 marks)

You are required to time how long it takes to execute a query. It would be best to run the query a number of times and store the timing values. The query you need to execute is *to count the number employees with salary between 50000 and 70000, and who have been employed for less than 1 year*. Perform following tests:

- Execute the query 20 times in succession taking the time for each run using your Java benchmark program.
- Add an index for the salary attribute.
- Execute the query 20 times in succession with the index and note the time each run takes.
- Plot the execution times on a graph. X-axis represents the run (i.e. 1 to 20) and Y-axis the time.
- Write a paragraph explaining the results.
- Include the plot and paragraph in your report.

### Task 4: Using EXPLAIN keyword ..... (5 marks)

Write a query to *find the name and salary of the bottom five employees with the lowest salary of all employees* by using an inner join.

- Run the query.
- Use the EXPLAIN statement to display information from the optimiser about the statement execution plan.
- Note how many rows are queried for each table and plot them on a histogram.
- Include the histogram in your report.

### Task 5: Query trees ..... (10 marks)

```
SELECT e.first_name, e.last_name, t.title, s.salary
FROM employees e, titles t, salaries s
WHERE e.emp_no = t.emp_no
AND t.emp_no = s.emp_no
AND t.title = 'engineer'
AND s.salary > 80000
AND s.salary < 100000;
```

- Execute the query and show a physical query plan for this query, assuming there are no indexes and data is not sorted on any attribute.
- Using the knowledge acquired about the EXPLAIN statement and indexes, suggest two indexes and an alternate query plan for the above query.
- Write your answers in a short paragraph in your report.

### Task 6: Comparing relational algebra expressions ..... (10 marks)

Write an SQL query to *retrieve employees who earn 43699*.

- Translate the query to relational algebra and represent this as a query tree. Include this tree in your report.

- Using the optimisation algorithm discussed in class, optimise the initial query tree. Include one of the more optimised query trees in your report.
- Translate the optimised query tree to relational algebra statements for the input into the Relational Algebra Interpreter from a previous practical.
- Using the Relational Algebra Interpreter, execute the statements for the initial query tree and the optimised query tree, noting the differences in execution time and results as you go.
- Write a paragraph in your report stating whether you saw what you expected in terms of execution time and results.

## References

- [1] R. Elmasri and S. Navathe, *Fundamentals of database systems*, 6th ed. Addison-Wesley Publishing Company, 2010.