Data Analysis and Integration

Lab 1: Review of SQL

Note: This lab assumes that you are using the provided virtual machine. If you launch Firefox inside the VM, it will redirect you to a webpage that contains all the necessary files.

Creating the employees database

1. Download and extract **employees.zip** to your Downloads folder (/home/aid/Downloads).

2. Take a moment to inspect the contents of the **employees.sql** script.
   - Locate the CREATE DATABASE statement.
   - Locate all CREATE TABLE statements.
   - Check the columns and data types for each table.
   - Check the primary and foreign keys for each table.
   - Locate all SOURCE instructions to load data into these tables.

3. Take a moment to inspect the contents of each data file:
   - load_departments.dump
   - load_employees.dump
   - load_dept_emp.dump
   - load_dept_manager.dump
   - load_titles.dump
   - load_salaries*.dump

4. Open a terminal and navigate to the folder where the files are located.

5. Execute the following command to login to the local MySQL server: **mysql -u aid -p**
   Password: **aid**

6. On the MySQL prompt, execute the following command to create the database:
   source employees.sql

7. Execute the following command to change to the employees database: **use employees**

Querying the employees database

8. Take a moment to inspect the contents of each table:
   select * **from** employees **limit** 10;
   select * **from** departments;
   select * **from** dept_emp **limit** 10;
   select * **from** dept_manager **limit** 10;
   select * **from** titles **limit** 10;
   select * **from** salaries **limit** 10;

9. Execute the following simple query:

```
select *
from salaries
where salary > 80000
limit 10;
```

10. Narrow down the results to a particular date:

```
select *
from salaries
where salary > 80000
    and from_date <= '2000-01-01'
    and to_date >= '2000-01-01'
limit 10;
```

11. Run the following query to retrieve the current date:

```
select current_date;
```

12. Reformulate the previous query with the current date:

```
select *
from salaries
where salary > 80000
    and from_date <= current_date
    and to_date >= current_date
limit 10;
```

13. Adjust the selected columns:

```
select emp_no, salary
from salaries
where salary > 80000
    and from_date <= current_date
    and to_date >= current_date
limit 10;
```

14. Join the tables **salaries** and **employees**:

```
select *
from salaries as a, employees as b
where a.emp_no = b.emp_no
limit 10;
```

15. Narrow down the results to the current date:

```
select *
from salaries as a, employees as b
where a.emp_no = b.emp_no
    and a.from_date <= current_date
    and a.to_date >= current_date
limit 10;
```

16. Adjust the selected columns:

```
select a.emp_no, a.salary, b.first_name, b.last_name
from salaries as a, employees as b
where a.emp_no = b.emp_no
    and a.from_date <= current_date
    and a.to_date >= current_date
limit 10;
```

17. Join the tables **employees**, **dept_emp** and **departments**:

```
select a.first_name, a.last_name, b.from_date, b.to_date, c.dept_name
from employees as a, dept_emp as b, departments as c
where a.emp_no = b.emp_no and b.dept_no = c.dept_no
limit 10;
```

18. Narrow down the results to the current date:

```
select a.first_name, a.last_name, b.from_date, b.to_date, c.dept_name
from employees as a, dept_emp as b, departments as c
where a.emp_no = b.emp_no and b.dept_no = c.dept_no
    and b.from_date <= current_date and b.to_date >= current_date
limit 10;
```

19. Adjust the selected columns:

```
select a.first_name, a.last_name, c.dept_name
from employees as a, dept_emp as b, departments as c
where a.emp_no = b.emp_no and b.dept_no = c.dept_no
    and b.from_date <= current_date and b.to_date >= current_date
limit 10;
```

Number of employees by department

20. Take a moment to inspect the **dept_emp** table:

```
select *
from dept_emp
limit 10;
```

21. Narrow down the results to the current date:

```
select *
from dept_emp
where from_date <= current_date
    and to_date >= current_date
limit 10;
```

22. Sort by department number:

```
select *
from dept_emp
where from_date <= current_date
    and to_date >= current_date
order by dept_no
limit 10;
```

23. Group by department number and count the employees:

```
select dept_no, count(emp_no)
from dept_emp
where from_date <= current_date
    and to_date >= current_date
group by dept_no;
```

24. Rename the count column:

```
select dept_no, count(emp_no) as count_emp_no
from dept_emp
where from_date <= current_date
    and to_date >= current_date
group by dept_no;
```

Number of employees by department (with department name)

25. Join the tables **dept_emp** and **departments**:

```
select *
from dept_emp as a, departments as b
where a.dept_no = b.dept_no
limit 10;
```

26. Narrow down the results to the current date:

```
select *
from dept_emp as a, departments as b
where a.dept_no = b.dept_no
    and a.from_date <= current_date
    and a.to_date >= current_date
limit 10;
```

27. Group by department number and department name, and count the employees:

```
select a.dept_no, b.dept_name, count(a.emp_no)
from dept_emp as a, departments as b
where a.dept_no = b.dept_no
    and a.from_date <= current_date
    and a.to_date >= current_date
group by a.dept_no, b.dept_name;
```

28. Rename the count column:

```
select a.dept_no, b.dept_name, count(a.emp_no) as count_emp_no
from dept_emp as a, departments as b
where a.dept_no = b.dept_no
    and a.from_date <= current_date
    and a.to_date >= current_date
group by a.dept_no, b.dept_name;
```

29. Sort by department name:

```
select a.dept_no, b.dept_name, count(a.emp_no) as count_emp_no
from dept_emp as a, departments as b
where a.dept_no = b.dept_no
    and a.from_date <= current_date
    and a.to_date >= current_date
group by a.dept_no, b.dept_name
order by b.dept_name;
```

30. Sort by the count of employees:

```
select a.dept_no, b.dept_name, count(a.emp_no) as count_emp_no
from dept_emp as a, departments as b
where a.dept_no = b.dept_no
    and a.from_date <= current_date
    and a.to_date >= current_date
group by a.dept_no, b.dept_name
order by count_emp_no;
```

31. Sort by the count of employees in descending order:

```
select a.dept_no, b.dept_name, count(a.emp_no) as count_emp_no
from dept_emp as a, departments as b
where a.dept_no = b.dept_no
    and a.from_date <= current_date
    and a.to_date >= current_date
group by a.dept_no, b.dept_name
order by count_emp_no desc;
```

32. Select the departments having at least 40 employees:

```
select a.dept_no, b.dept_name, count(a.emp_no) as count_emp_no
from dept_emp as a, departments as b
where a.dept_no = b.dept_no
    and a.from_date <= current_date
    and a.to_date >= current_date
group by a.dept_no, b.dept_name
having count_emp_no >= 40;
```

33. Sort the results by department name in ascending order:

```
select a.dept_no, b.dept_name, count(a.emp_no) as count_emp_no
from dept_emp as a, departments as b
where a.dept_no = b.dept_no
    and a.from_date <= current_date
    and a.to_date >= current_date
group by a.dept_no, b.dept_name
having count_emp_no >= 40
order by b.dept_name asc;
```

Exercise

34. Find the sum of salaries by department. Order the results by the sum of salaries, the output should look like this:

```
+---------+--------------------+------------+
| dept_no | dept_name          | sum_salary |
+---------+--------------------+------------+
| d005    | Development        |    4434974 |
| d007    | Sales              |    3715959 |
| d004    | Production         |    2928341 |
| d009    | Customer Service   |    1914195 |
| d002    | Finance            |    1492870 |
| d001    | Marketing          |    1249477 |
| d006    | Quality Management |    1212103 |
| d008    | Research           |    1064935 |
| d003    | Human Resources    |     643182 |
+---------+--------------------+------------+
```

Take a screenshot of your query and results and submit it on Fénix for lab credit.