

CS 210: Data Management for Data Science

Homework 2: SQL, Data Collection, and Data Storage

1 Theory Questions (20 points)

1. Describe the purpose of the following SQL commands, and write down the syntax for each command: SELECT, INSERT, UPDATE, DELETE, JOIN.
2. Describe the differences between CSV and JSON data formats. When might you choose one format over the other?

2 Experiment (30 points)

2.1 SQL Queries

2.1.1 Database Setup

- Create a relational database using MySQL or PostgreSQL. Name it `company_db`.
- Create the following tables:
 - `employees` (`id`, `first_name`, `last_name`, `department`, `salary`)
 - `departments` (`id`, `department_name`)
- Create and insert some data by yourself.

2.1.2 Basic SQL Queries

1. Write a SQL query to retrieve all employees from the `employees` table.
2. Write a SQL query to insert a new employee into the `employees` table.
3. Write a SQL query to update the salary of an employee with a specific `id`.
4. Write a SQL query to delete an employee with a specific `id`.

2.1.3 Advanced SQL Queries

1. Write a SQL query to find the highest salary in each department.
2. Write a SQL query to list employees along with their department names using a JOIN statement.

Please write down the SQL query code of each task and include all screenshots of your experiments in your submitted PDF document.

3 Submission Requirements

1. **Submission Format:** All files should be submitted via Canvas. Submit the report in PDF format, and attach source code files mentioned in the tasks.
2. **Note:** Make sure to test all your scripts thoroughly before submission.