1 Before attempting the problems

This exercise sheet covers the part of the course that gives an overview of relational databases and SQL (first tick and the first four lectures).

When studying databases, remember to think about whether and how concurrent/distributed queries are supported. This is explored in more detail in Concurrent and Distributed Systems, but concurrency is a very important practical concern that motivates some of the design decisions.

2 Problems

1. State and explain the fundamental tradeoff in database design and choice of data representation.
2. What are the ACID properties in the context of databases?
3. Explain how data redundancy and data duplication are distinct concepts in database design and implementation.
4. Explain the issue of implementing many-to-many and one-to-one relations in the relational model.
5. State the formal definitions of the following concepts (and introduce notation when needed):
   (a) natural join;
   (b) superkey;
   (c) key;
   (d) foreign key.
6. Why does the need for weak entities arise?
7. What are data anomalies in the context of relational databases?

8. What does it mean for a database to have referential integrity? Why is it an important concept?

9. The example given in Lecture 4 shows how you can encode several different relations using a single table. Discuss the approach.

10. Suppose you are working on IT tasks for a company that provides catering services for weddings, parties and businesses. One of the requests you received from accounting and management departments is to create a handy database system that will allow them to keep track of all company finances, clients and employees. You are only told that the users of this database should be able to retrieve information about the company performance with a click of a button.

This database should contain details about all employee salaries, all events done for clients, the necessary information about these clients, which employees attended which event, as well as the costs required to organise a particular event (not counting buying all the ingredients through deliveries, which are usually bought in bulk for multiple events). The company policy is to provide a full refund in the case an event is cancelled, with the refund happening as soon as the cancellation is processed.

Note that the answers will depend on how you decide to interpret and implement the requirements.

(a) Draw an ER diagram to show how you would model this scenario.

(b) Give a simple overview of the tables you would have in your database. State all of the constraints.

(c) Write SQL SELECT queries that will perform the following tasks:
   i. list the details of all employees who attended a particular event;
   ii. list the details of all employees whose average salary is greater than the average salary of all company employees;
   iii. compute the total company profit between the given start and end dates;
   iv. compute how much each client has paid for their events (returned as a two column result |CLIENT_NAME|AMOUNT|, with clients sorted by how much they have paid).

(d) Suppose that, after a few months of using the database, you are asked to optionally provide a way to associate some deliveries with a particular event. How would you change your schemas? What are the difficulties you might run into?