Syllabus

Lecturer

- Werner Nutt, `nutt@inf.unibz.it`, Room TRA 2.01

  Office hours: Thursday, 16:00–18:00

  (If you want to meet up with me, send me an email or catch me after the lectures to make an appointment. In that way we can be sure that you don't have to wait until other students have finished their meetings.)

Teaching Assistants

- Michail Kazimianec, `kazimianiec@inf.unibz.it`, Room SER A 515

  Office hours: to be determined and by appointment

- Damiano Somenzi, `Damiano.Somenzi@unibz.it`

  Office hours: Thursday, 16:00–17.00 and by appointment

Timetable

Day	Time	Event	Room
Monday	10:30–12:30	Lecture	E 420
Thursday	14:00–16:00	Lab (Kazimianec)	E 431
Thursday	14:00–16:00	Lab (Somenzi)	E 531
Thursdaz	10:30–12:30	Lecture	E 420

- We are currently trying to find a more suitable room for the lectures, since E 420 is very small. Students will be informed immediately about such a change. It will also show up in the RIS.
- There will be two lab groups. Each student will be assigned to one group and will attend the lab session of that group.
- Lab sessions will start on 8 October.

Home Page

Various resources for the course will be available on the web under the URL

`www.inf.unibz.it/~nutt/IDBs0910`

This web site will evolve as the lectures proceed.

Aims

Virtually every software system has to manage significant amounts of data. Many tasks regarding the management and the manipulation of data can be executed by specialised software systems, called Database Management Systems (DBMS). To be able to use DBMS's successfully, one has to understand the concepts on which they are based. The aims of this course are

- to introduce you to the *basic concepts* underlying a DBMS;
- to show how they are realized in *specific systems* such as the PostgreSQL DBMS;
- to give you some *hands-on experience* in using the PostgreSQL DBMS.

Course Content

- Fundamental Database Concepts
- The Entity Relationship (= ER) Model
  - Conceptual Database Design
- The Relational Model
  - Relations and Integrity Constraints
    (Keys and Foreign Keys)
- Logical Database Design
  - ER to Relational Schemas
- Relational Algebra: an Algebraic Query Language for the Relational Model
- SQL: Querying and Manipulating Data
  - SQL Data Definition Language
  - Single Block Queries
  - Aggregation
  - Joins and Outer Joins
  - Nesting
  - Negation
- Transaction Management and Concurrency Control
- Database Access from a Programming Language: JDBC

- Data Storage and Indexing
  - File Organisation and Indexes
  - Tree-structured Indexing: B+-trees
  - Indexes in PostgreSQL
- Query Evaluation
  - Sorting
  - Evaluation of Relational Operators
  - Query Optimisation
  - Physical Database Design
- Query Plans in PostgreSQL
- Functional Dependencies and Normalisation

Format of the Course

The course has three main ingredients:

- Lectures
- Exercises
- Group Projects

Lectures

The lectures will introduce *new material.* The lectures will largely follow the structure of the book

> *A First Course in Database Systems*
> by Jeff Ullman and Jennifer Widom.

There are several copies in the library. Students are expected to work with the book to consolidate and revise the material of the lectures.

The sections on data storage, indexing and query execution will be based on

> *Database Management Systems*
> by Raghu Ramakrishnan and Johannes Gehrke.

There will be a sufficient number of copies once we reach this point of the lectures.

As the lectures proceed, copies of the slides will be available from the course home page.

Labs

There will be two kinds of activities during the labs:

- Solving exercises
- Presentations and discussions on the group projects.

During the lectures, a work sheet will be handed out with *exercises* on the subjects of the lectures. Students are expected to prepare for these exercises and may be asked to present their solution at the blackboard. Solutions will be discussed at the tutorials.

Some of the questions of the final exam will be very similar to the lab exercises. Thus, solving the exercises is an excellent preparation for the exam.

Students will also present work on their group projects and receive feedback and support.

Group Projects

The goal of the group projects is to apply the concepts and techniques presented in the lectures to a small database application.

Project Groups: You will work on a project in groups of *three* students. You are free to choose your group as you like.

Individual Database Application (IDA): Each group will develop an Individual Database Application (IDA). You are free to choose the topic of your IDA by yourself. The project will consist of the following steps:

1. Specifying the application and writing up data requirements
2. Designing a conceptual model in the form of an Entity Relationship diagram (= conceptual design)
3. Translating the conceptual model into a relational schema
4. Implementing the relational schema in PostgreSQL and populating the database
5. Querying and modifying the database by suitable SQL statements
6. Optimising the access to data by adding indexes to the relational schema (= physical database design)
7. Writing a Java client that accesses the DB via JDBC, allowing a user to query, insert, delete, and modify data
8. Experimenting with PostgreSQL's concurrency control mechanisms.

Project Home Page: Each group will set up a *project home page* with documents on the project and progress reports. There will be a pointer from the course home page to the project home pages so that you can learn from the work of other groups.

Each group will receive Web space on the faculty servers once they have formed.

Milestones: There will be six *milestones* for the project:

Week 2: Group registered and topic of project defined;

Week 4: Data requirements and conceptual model;

Week 6: Translation into relational schema, implementation of the schema, population of the schema with data

Week 8: SQL queries over the database

Week 10: JDBC client that runs transactions on the database.

Week 12: Physical design, analysis of query execution plans, performance analysis

The *deadline* for each milestone is 10.30pm on Monday of the following week. You submit your work by publishing it at your project home page. You will receive a mark for the work that can be found at that time at your home page. If there is no work, the mark will be 0.

Registration: The only exception to this rule is the registration. Obviously, there will be no mark for the registration. Also, please try to register as soon as possible. To register, send a mail with an attached XML document to

`kazimianec@inf.unibz.it`

containing

- the topic of the project (e.g., "Hotels in South Tyrol")
- a description of the topic comprising one or two sentences (e.g., "A database with hotels in different parts of South Tyrol that contains information about the equipment of the hotels, the tariffs and available holiday packages.")
- the names and email addresses of the group members.

Note: The attached XML document should contain your registration data in the following format:

```
<project>
  <title> Title of the Project </title>
  <shortdescription> Short Description of the Project </shortdescription>
  <people>
    <student>
       <name>  Member name  </name>
       <email> Member's email address </email>
    </student>
    ...

  </people>
</project>
```

You will find this pattern also on the course website from where you can copy it.

Conflicting Topics: All groups have to work on different topics. If two groups choose the same topic, the group that registered second will be informed to choose a different one. Incoming registrations will be published daily at the course home page.

Project Groups and Lab Groups. Project groups will present and discuss their progress in the labs. Therefore, all members of the project should attend the same lab.

Presentations in the Labs: The labs in week 3 will be devoted to the data requirements and conceptual model and the relational schema. Each group will give a short presentation. All participants of the tutorial are encouraged to discuss the project presented. The teaching assistant will ask questions about the planned project so that problems with the design can be identified as soon as possible.

More detailed instructions for the coursework will be given to you later during the term.

Reading

There are a number of good introductory textbooks around and most of them cover the more or less the same material. This list contains a small choice of them. For each book there is a web site that provides additional material.

- *A First Course in Database Systems* by Jeff Ullman and Jennifer Widom.
  This is the book on which by and large the lectures are based.
- *Database Management Systems* by Raghu Ramakrishnan and Johannes Gehrke.
  The sections on data storage, indexing and query execution will be based on this book.
- *Database Systems: The Complete Book* by Hector Garcia-Molina, Jeff Ullman, and Jennifer Widom.
  This an extension of the book by Ullman and Widom by several chapters on the implementation of database management systems
- *Database Systems: Concepts, Languages and Architectures* by Paolo Atzeni, Stefano Ceri, Paolo Paraboschi and Riccardo Torlone.
- *Database System Concepts* by Abraham Silberschatz, Henry F. Korth, and S. Sudarshan.
  This is book gives up to date information about recent technology. The part that discusses how database concepts are realized in various commercial systems is of particular interest.

Assessment

For the project, each group will receive a mark, which will also be the mark of each member. There will also be a written exam for the course. The final mark will be computed as

$$\begin{aligned}\max\ \{&0.3\,\text{project mark} + 0.7\,\text{exam mark},\\ &1.0\,\text{exam mark}\}.\end{aligned}$$

In other words, the final mark is at least as good as the exam mark. If the project mark is better than the exam mark, then the final mark is composed of the mark for the project (contributing 30%) and the mark for the exam (contributing 70%).