Introduction to Databases

Lecture #1

Course Overview

Ljubov Jaanuska
(ljubov.jaanuska@ut.ee)
I think we should build an SQL database.

Uh-oh.

Does he understand what he said or is it something he saw in a trade magazine ad?

What color do you want that database?

I think mauve has the most RAM.
Quiz

• Is there any data you care enough to manage?
• What if you start an e-commerce project?
• Do your favourite apps require a database system?
• How can you tell it?
Importance of the course

- Database systems are at the core of CS
- They are incredibly important to society
Importance of the course
Importance of the course
Importance of the course
Increasingly many companies see themselves as data driven.
Importance of the course

- Database systems are at the core of CS
- They are incredibly important to society
- The topic is intellectually rich
  - Shift from computation to information
- A capstone course for undergrads
- Be a data ninja
THE WORLD IS INCREASINGLY DRIVEN BY DATA...

This class teaches the basics of how to use & manage data.
Introduction to the course

- Introductory course on databases
- Covers both theoretical and practical aspect
  - Principles of data models
  - Relational database
  - SQL structure and syntax
  - Database design and development
  - Database management systems work
- This course is **NOT** how to be a DBA or how to tune Oracle or MySQL
On successful completion of this course, you will be able to

- Use database terms
- Develop simple E-R models
- Compose simple queries using SQL
- Use SAP SQL Anywhere to develop and manage a database
Organization of the course

• Course websites:
  • https://courses.cs.ut.ee/2017/AB/
  • Moodle

• 16 class sessions
  • Theoretical part: main concepts
  • Practical part: 2 databases
  • Exercises for home assignments
Organization of the course

• Project in groups

• Reading
  • Self-study material

• Exam
Workload

• 3 ECTS = 3 ECTS x 26 hours/ECTS = 78 hours
• Lectures and Practical – 32 hours
• Independent work – 46 hours
Attendance

I dislike mandatory attendance… but I noticed…

• People who did not attend did worse 😞
• People who did not attend used more course resources 😞
• People who did not attend were less happy with the course 😞
Rules

• All deadlines are strict
• Study group work should be finalized privately!
Assessment

- To be admitted to the exam, **at least 50% of assignments and 50% project** need to be collected!

- To pass the course at **least 20 points** should be collected!

<table>
<thead>
<tr>
<th>Type</th>
<th>Max</th>
<th>Min</th>
</tr>
</thead>
<tbody>
<tr>
<td>Home Assignments</td>
<td>19</td>
<td>9.5</td>
</tr>
<tr>
<td>Project</td>
<td>10</td>
<td>5</td>
</tr>
<tr>
<td>Exam</td>
<td>10</td>
<td>5</td>
</tr>
<tr>
<td>Bonus</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>TOTAL</strong></td>
<td>39</td>
<td>20</td>
</tr>
</tbody>
</table>
Ungraded elements

• Readings provided to help you!
  • Only items in homework and project are fair game.
Changes are possible!
Questionnaire

- To assess the basic starting point and expectations before the course starts
- https://goo.gl/forms/1hMsUVPlxB0R4hcI2
Lecture 1. What will you learn

1. Definition of data and information
2. Definition of database
3. Definition of database management system
4. Database users
Data

<table>
<thead>
<tr>
<th>Name</th>
<th>Age</th>
<th>Salary</th>
</tr>
</thead>
<tbody>
<tr>
<td>Marko</td>
<td>23</td>
<td>55,000</td>
</tr>
<tr>
<td>Sander</td>
<td>25</td>
<td>45,000</td>
</tr>
<tr>
<td>Mart</td>
<td>24</td>
<td>60,000</td>
</tr>
<tr>
<td>Sofia</td>
<td>26</td>
<td>70,000</td>
</tr>
</tbody>
</table>

- Without the titles or labels associated with the data
- It is not much useful
Information is interpreted data

<table>
<thead>
<tr>
<th>Company: ABC</th>
<th>Dept: IT</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Name</strong></td>
<td><strong>Age</strong></td>
</tr>
<tr>
<td>Marko</td>
<td>23</td>
</tr>
<tr>
<td>Sander</td>
<td>25</td>
</tr>
<tr>
<td>Mart</td>
<td>24</td>
</tr>
<tr>
<td>Sofia</td>
<td>26</td>
</tr>
</tbody>
</table>

- After attaching these labels, it’s bring some meanings to use.
Difference and link between data and Information

• Data is the collection of raw facts collected from any environment for a specific purpose.

• Get desired types of results from the data we transform it into information by applying certain processing on it
Metadata – DB foundation

- Data about data – data description
Database definitions

• Broad def.: A database is a self-describing collection of integrated records.
  - collection of interrelated data + description of data

• A database is a shared collection of logically related data, that store to meet the requirements of different users an organization

• A database models a particular real world system in the computer in the form of data
- **Database** is used to store
- **DBMS** use mechanism to get data from the database
DB Management System (DBMS)

• A collection of small programs to perform certain operation on data and manage the data
  • how data will be stored, structured and accessed in the database

• DBMS also ensure management of users associated with the database
  • authentication
  • user privileges and restriction
Advantages of DBMS

• Data Consistency
• Better data security
• Faster development of new applications
• Economy scale
• Better concurrency control
• Better backup and recovery procedures
Data Consistency

- Data inconsistency leads to a number of problems
- Loss of information
- Incorrect result.
Better data security

- All application programs access data through DBMS
- DBMS very efficiently check which user is performing which action and accessing which part

Data Consistency
Better data security
Faster development of new applications
Economy scale
Better concurrency control
Better backup and recovery procedures
Fast Application Development

- The data needed for the new application already resides in the database.
- The data might not already reside in the database but it could be derived from the data present in the database.

Data Consistency
Better data security
Faster development of new applications
Economy scale
Better concurrency control
Better backup and recovery procedures
Better Concurrency Control

- Access a database from as number of point simultaneously.
- Database access completed correctly and transparently.
Better Backup and Recovery

- Data as a very important resource for organization must be periodic backed up
- Due to disastrous situation, data should be recovered

Data Consistency
Better data security
Faster development of new applications
Economy scale
Better concurrency control
Better backup and recovery procedures
Application programs talk to DBMS and ask for the data required
DB Application

- It is a program or a group of programs, which is used for performing certain operations on the data store in the database:
  - insertion,
  - extraction
  - update
Database designers design the database and install the DBMS for use by the users of the database in any specific organization.
Once database has been installed and is functioning properly in a production environment of an organization the DBA take over the charge and perform specific DBA related activities.

- Such as: Database maintenance, Backup, Grant of rights to database users, Monitoring of running jobs, Ensuring quality service to all users.
- DBA can interact with **DB Designer** during database design
- DBA can also interact with **App. Programmer** during development time
DB-Engines and Rating

http://db-engines.com/en/ranking
Which DB-Engine?

- In most cases, there’s no noticeable difference between MySQL, PostgreSQL, and other types of relational databases.
- The decision may already be made for you depending on what you need the software for.
- There’s really no need to worry about which is better, and once you understand one you’re pretty much set to work with all of them.