



UPPSALA
UNIVERSITET

Lecture 1 – Introduction

1DL301 Database Design I

Jan Kudlicka (jan.kudlicka@it.uu.se)

Fall 2019, Term 2



TODAY

- ▶ Teachers
- ▶ Motivation
- ▶ Contents of the course
- ▶ Intended learning outcomes, assessment, literature
- ▶ Lecture plan
- ▶ Project, labs and groups
- ▶ Reminders



Jan Kudlicka

jan.kudlicka@it.uu.se

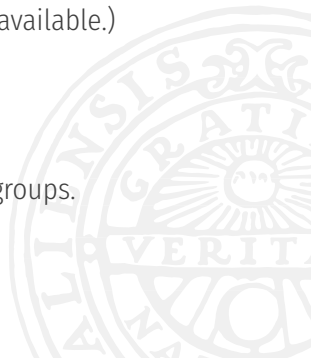
Office: ITC 1136

Feel free to drop by my office if you have questions, comments or suggestions! (Send an email first to check if I am available.)

Khalid Mahmood

khalid.mahmood@it.uu.se

Organization and administration of the labs and groups.



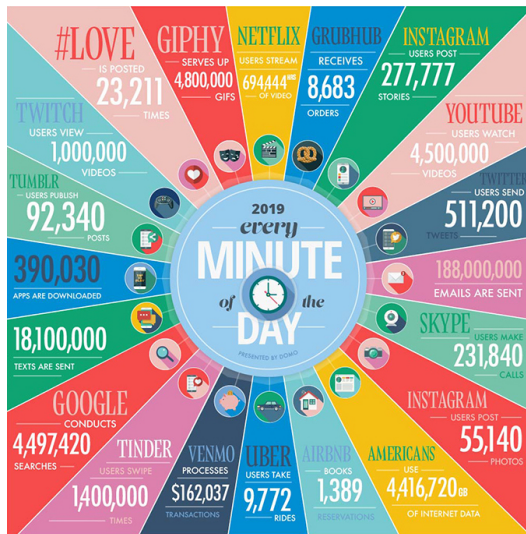
MOTIVATION

Data never sleeps!

How much data do we
generate in 1 minute?

Play video (2017 version)

Source: Domo

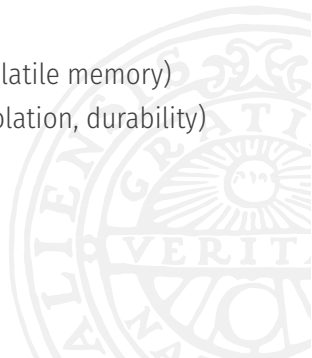


MOTIVATION

With so much data around us we need a systematic approach to process, store, manipulate, search, retrieve data in an “optimal” way.

What does “optimal” mean? There are different aspects (and trade-offs between them):

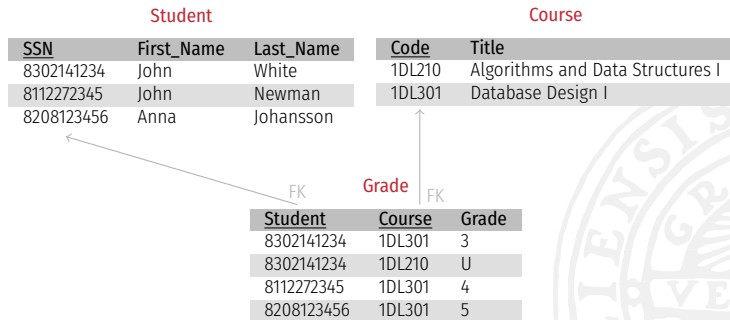
- ▶ Speed
- ▶ Space efficiency (both in volatile and non-volatile memory)
- ▶ Integrity and ACID (atomicity, consistency, isolation, durability)
- ▶ Security
- ▶ Cost



RELATIONAL DATABASES

In this course we will concentrate on relational database management systems (RDBMS).

In relational databases the data are stored in tables (= relations) of columns (= attributes or fields) and rows (= records or tuples).



Some known RDBMS: Oracle Server, IBM DB2, MS SQL Server, SyBase, IBM Informix, Teradata, MySQL, PostgreSQL, SQLite.

There are also non-relational databases (NoSQL = non-SQL, non-relational, not only SQL):

- ▶ Key-value stores
- ▶ Document-oriented databases
- ▶ Graph databases
- ▶ etc.

NoSQL databases are popular in big data and real-time web applications.

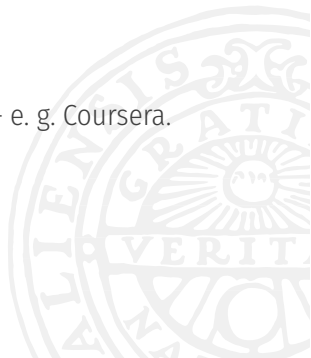


OTHER DATABASE COURSES

There are other database courses you can take after this one:

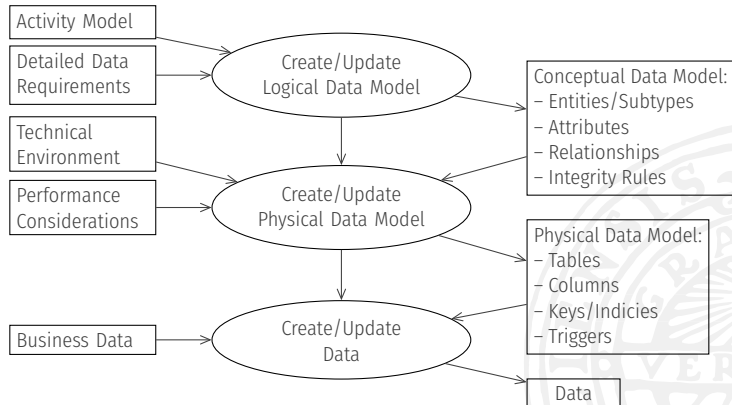
- ▶ Database Design II
- ▶ Data Mining I, II
- ▶ Big Data

There are many MOOC courses (esp. in Big Data) – e. g. Coursera.



CONTENTS OF THE COURSE

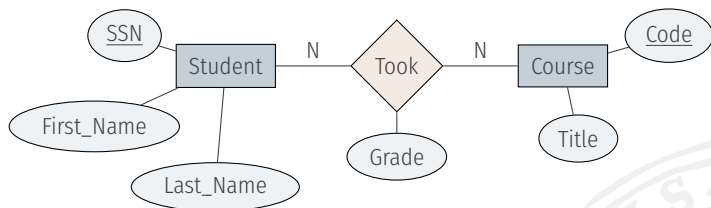
- ▶ Data modeling – creating a data model based on the textual description of the mini world (the part of the real world which we want to model).



Source: EPISTLE

CONTENTS OF THE COURSE, CONT'D

- Data modeling, cont'd
Example of an entity-relationship model:

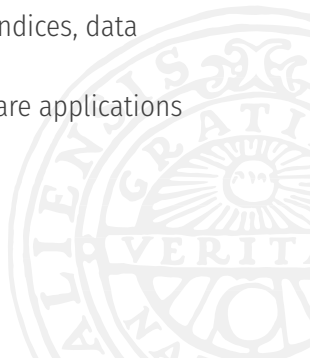


- SQL – “the” language for querying as well as creating and maintaining relational databases.

```
SELECT First_Name, Last_Name, Title, Grade
FROM Student, Course, Grade
WHERE Student.SSN = Grade.SSN
      AND Course.Code = Grade.Course
```

CONTENTS OF THE COURSE, CONT'D

- ▶ Normalization – reducing data redundancy and improving data integrity.
- ▶ Transactions and ACID (atomicity, consistency, isolation, durability).
- ▶ Introduction to physical design and tuning (indices, data de-normalization).
- ▶ Programming with databases – writing software applications using relational databases.
- ▶ Users, roles and security.



INTENDED LEARNING OUTCOMES

On completion of the course the student shall be able to:

- ▶ create conceptual models of relational databases based on requirement specification documents;
- ▶ translate a conceptual model to the relational model;
- ▶ create, modify and query relational databases using the SQL language;
- ▶ apply normalization techniques to reduce data redundancy and improve data integrity;
- ▶ create computer programs which store, modify and query data stored in databases;
- ▶ set up indexes and use other techniques to improve the performance of databases;
- ▶ set up users and their privileges;
- ▶ explain principles of security in database systems and apply protection mechanisms;
- ▶ present and discuss the course content orally and in writing with proficiency appropriate to the course level.

COURSE ASSESSMENT

Oral and written assignments (partial credit 2 hp)

To earn the credit you and your group must successfully do the project (including the written report and the peer review).

Grade scale: fail (U) / pass (G).

Written exam (partial credit 3 hp)

Written exam on January 10, 2020.

Grade scale: fail (U) / 3 / 4 / 5.

More information in December.

To earn 5 hp for the whole course you need to pass both parts.

But remember, the most important is that you learn how to do the database design!

RECOMMENDED LITERATURE

In English:

R. Elmasri, S. Navathe: Fundamentals of database systems (6th or 7th edition)

In Swedish:

T. Padron-McCarthy, T. Risch: Databasteknik



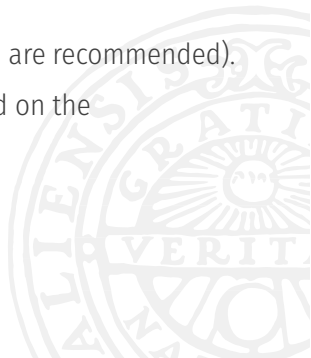
ORGANIZED CLASSES

- ▶ 15 lectures
- ▶ 4 labs (three labs working on the project and one practicing SQL)

Schedule in TimeEdit.

Neither lectures nor labs are mandatory (but both are recommended).

All information and course material will be posted on the Studentportalen.

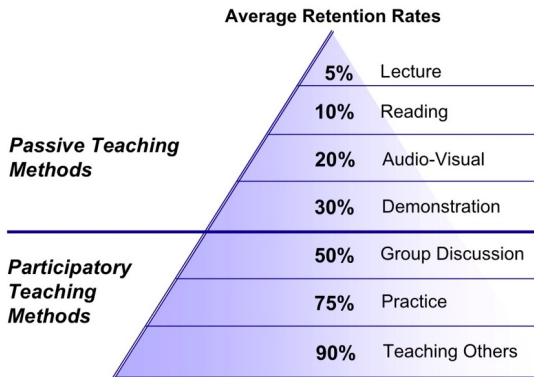


OUTSIDE THE CLASSROOM

- ▶ Learn by discussing and by doing!
- ▶ Work with your group on your project.
- ▶ Solve exercises (feel free to work together).
- ▶ Practice SQL.
- ▶ If you are stuck or have another problem, use Internet:
 - ▶ Google,
 - ▶ Stack Overflow,
 - ▶ other online fora,
 - ▶ W3Schools SQL tutorial.
- ▶ Watch videos online (check the *Lectures* page in the Studentportalen).

I am here for you! Drop by my office.

ACTIVE LEARNING APPROACH

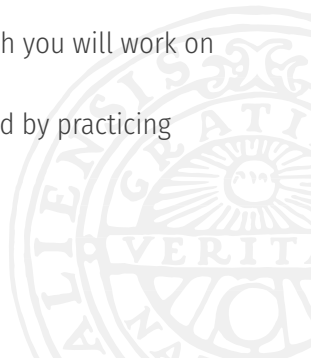


*Adapted from National Training Laboratories. Bethel, Maine

Note: This model is criticized by some scientists, nevertheless you will learn more if you discuss and do/practice.

ACTIVE LEARNING APPROACH

- ▶ Difficult in big lecture halls (but we are going to try anyway!)
- ▶ We will use Vox Aulea – a web-based audience response system (*mentometrar*) – always bring your smart phone, tablet or computer with you.
- ▶ Lectures will include short assignments which you will work on (alone, in pairs or in small groups).
- ▶ Lectures on SQL – short introduction followed by practicing using an online tool.



ACTIVE LEARNING APPROACH, CONT'D

- ▶ For several lectures you will be asked to watch a video or two before coming to the lecture (so that you can be better prepared for solving problems and discussions).
- ▶ Ask questions during the lectures. (You can also ask question via Vox Aulea.)
- ▶ Comments? Concerns? Suggestions? We want to hear them!

PROJECT

During the term you will work on a project (in groups of three).

Design of a database for an online store

- ▶ Conceptual model
- ▶ Relational model
- ▶ Creating and populating a database
- ▶ Analysis of the model
- ▶ Performance optimizations
- ▶ Programs that use the database

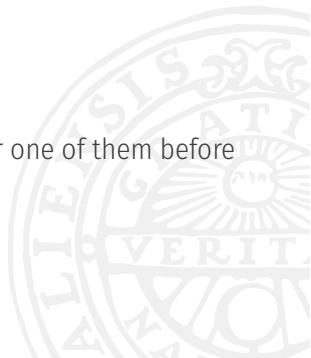
The project description is already published in the Studentportalen.

You will be able to get help with the project during labs.

1. (Project) Conceptual modelling (ER)
2. Practicing SQL queries
3. (Project) Conversion to relational model, creating and populating the database, optimizations
4. (Project) Programming with databases

Labs are not mandatory but recommended.

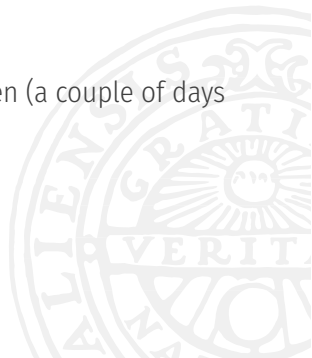
Each lab is offered twice – you need to register for one of them before each lab (mandatory registration via Doodle).



PROJECT MILESTONES

1. Preliminary report including the conceptual model.
2. Preliminary report including the relational model, the database and analysis.
3. Final report.

Deadlines will be published in the Studentportalen (a couple of days after each of the project labs).



GROUPS

Register in a group based on your study program:

- ▶ Sociotechnical Systems Engineering (TST2Y) – **STS**
- ▶ Computer and Information Engineering (TIT2Y) – **IT**
- ▶ Molecular Biotechnology Engineering (TMB2Y) – **X**
- ▶ Bioinformatics (TBK2M/BIOL) – **BI**

Other programs and exchange students – if you study a computer or computational science, math or physics, register in an IT group. Otherwise ask me.

If you are not going to do the project, register yourself in the *Not participating* group!

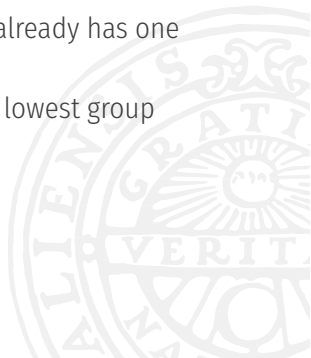
More information in the Studentportalen.

Deadline: November 6, 11:59 pm

GROUPS, CONT'D

After the deadline we will finalize the groups. To minimize the risk that you will be moved to a different group than you have registered:

- ▶ If you are three students, register at the same time.
- ▶ If you are just two, try to join a group which already has one student.
- ▶ Register as soon as possible and choose the lowest group number.

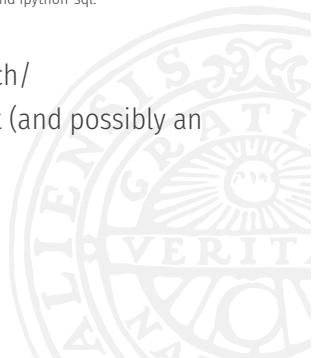


SOFTWARE TO INSTALL

- ▶ Install Anaconda:
<https://www.anaconda.com/distribution/>
(Choose Python 3.7 version.)

Alternatively: install a Python distribution and sqlite3. Use pip to install jupyter and ipython-sql.

- ▶ MySQL Workbench:
<https://dev.mysql.com/downloads/workbench/>
- ▶ If you wish to use Java: Java Development Kit (and possibly an IDE, e. g. NetBeans or Eclipse)



COURSE REGISTRATION

If you are admitted but not registered: Register to the course now!

If you are not admitted:

- ▶ Exchange students: contact Ulrika Jaresund at ulrika.jaresund@it.uu.se
- ▶ Other students (including students who need to be re-registered): contact the Student Office at it-kansli@it.uu.se



TODO LIST FOR THIS WEEK

- ▶ Register at <https://erdplus.com> (before the next lecture).
- ▶ Register in a lab group.
- ▶ Install the software.
- ▶ Read the project description.
- ▶ Make sure you can use the Eduroam wireless network.
(<https://www.uu.se/en/students/it-for-students/connect-to-internet>)



NEXT LECTURE

Conceptual modeling: entity-relationship (ER) model.

- ▶ Register at <https://erdplus.com!>
- ▶ Bring a pen, paper and/or your computer.
- ▶ Before the lecture, watch the following video (24 min):

<https://www.youtube.com/watch?v=ZOJ4J0UxiX0>

(Don't worry if you do not understand everything.)

YOUR TURN



IN THE REMAINING TIME...

Uppsala ancient library has no computer database, but finally sees the need for it. They do not really know what they want and you are asked to guide them through the process of writing requirements.

What does the mini world look like?

- ▶ What are the objects and their properties we need to store?
- ▶ What are the relationships between them?
- ▶ What are the typical tasks in a library? How do they alter the data?

Think about this on your own first and then discuss with people sitting around you.