UPPSALA UNIVERSITET

Introduction to the course: Database Design I

George Fakas (Labs)

georgios.fakas@it.uu.se

Khalid Mahmood (only lectures)

UPPSALA UNIVERSITET

Overview of this lecture

- Motivation.
- Overview of the course.
  - Textbooks.
  - Content.
  - Project.
  - Examination.
    - Exams will be on Inspera
- Course preparation

UPPSALA UNIVERSITET

Textbooks

- **Fundamentals of database systems (7E, older versions will be ok)**
  - R. Elmasri, S. Navathe.
  - Chapters 1, 3, 4, 5, 6, 7, 9, 10, 14, 20, 30 (part), 24.
- **Databasteknik**
  - T. Padron-McCarthy, T. Risch.
- Each lecture focuses on one or more chapters.

UPPSALA UNIVERSITET

Course info

- All the information and material is posted on the studium.
- Groups formation,
- Project Submission, etc.

MOTIVATION &
INTENDED LEARNING OUTCOMES

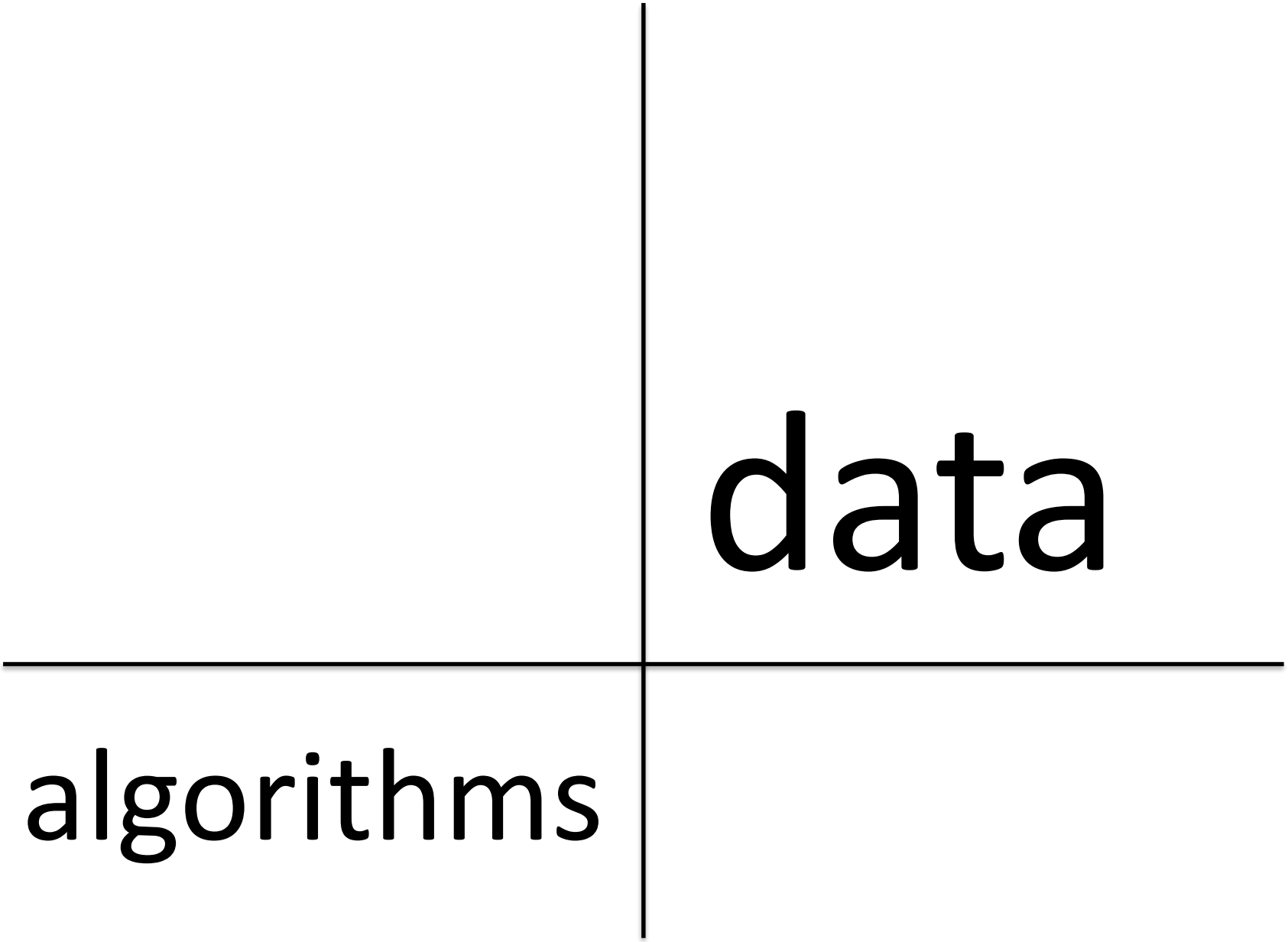
UPPSALA UNIVERSITET

A fundamental asset in nearly all organizations.

Complete, relevant, consistent, and up-to-date data are vital to enable basic operations.

Think of Uppsala University.

data

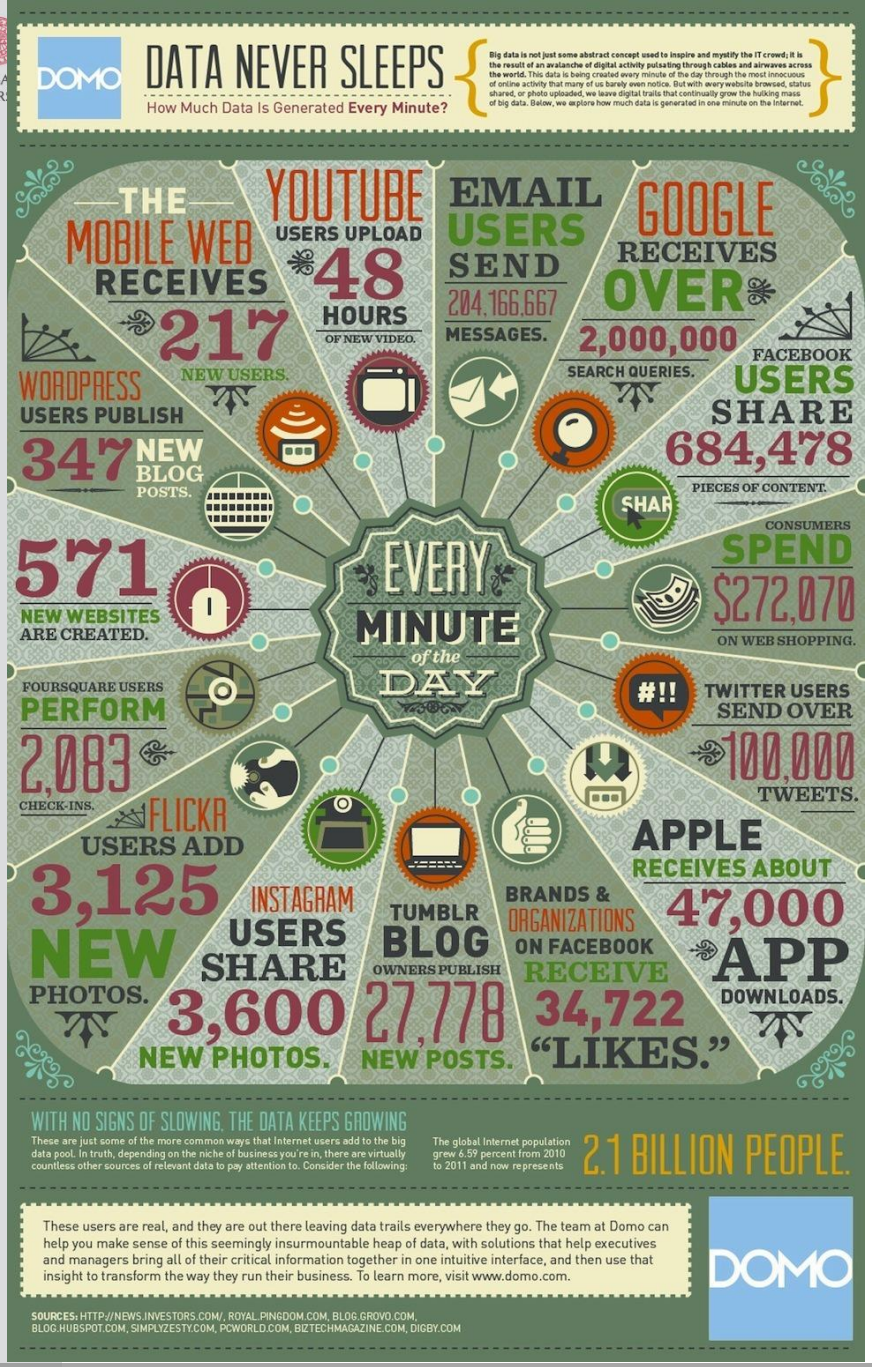

data
algorithms

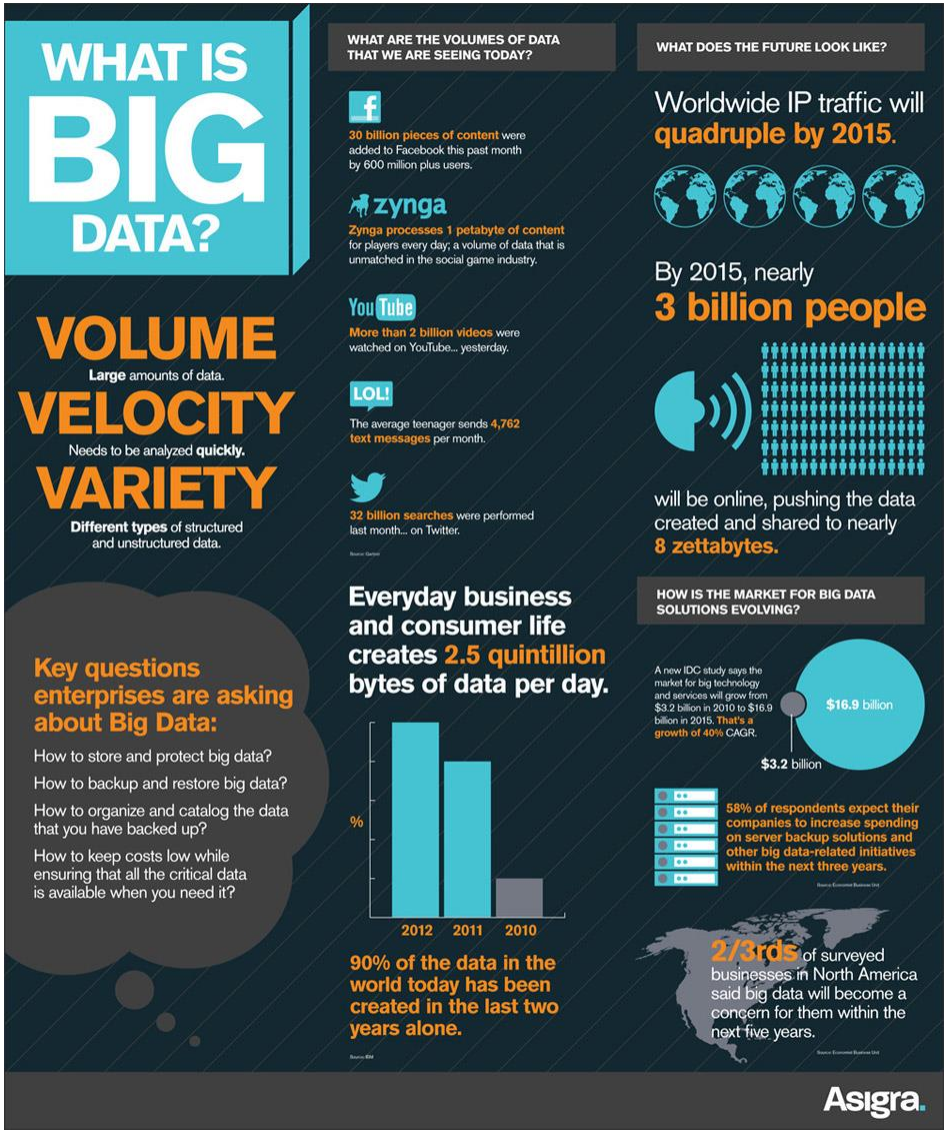
large
(persistent)
(live longer than
applications)

data

databases

querying
(larger set of potential users)


DOMO
DATA NEVER SLEEPS
How Much Data Is Generated Every Minute?
Big data is not just some abstract concept used to inspire and mystify the IT crowd; it is the result of an avalanche of digital activity pulsating through cables and airwaves across the world. This data is being created every minute of the day through the most innocuous of online activity that many of us barely even notice. But with every website browsed, status shared, or photo uploaded, we leave digital trails that continually grow the hulking mass of big data. Below, we explore how much data is generated in one minute on the Internet.
THE MOBILE WEB RECEIVES 217 NEW USERS.
YOUTUBE USERS UPLOAD 48 HOURS OF NEW VIDEO.
EMAIL USERS SEND 204,166,667 MESSAGES.
GOOGLE RECEIVES OVER 2,000,000 SEARCH QUERIES.
FACEBOOK USERS SHARE 684,478 PIECES OF CONTENT.
WORDPRESS USERS PUBLISH 347 NEW BLOG POSTS.
571 NEW WEBSITES ARE CREATED.
EVERY MINUTE of the DAY
CONSUMERS SPEND $272,070 ON WEB SHOPPING.
FOURSQUARE USERS PERFORM 2,083 CHECK-INS.
TWITTER USERS SEND OVER 100,000 TWEETS.
FLICKR USERS ADD 3,125 NEW PHOTOS.
INSTAGRAM USERS SHARE 3,600 NEW PHOTOS.
TUMBLR BLOG OWNERS PUBLISH 27,778 NEW POSTS.
BRANDS & ORGANIZATIONS ON FACEBOOK RECEIVE 34,722 "LIKES."
APPLE RECEIVES ABOUT 47,000 APP DOWNLOADS.
WITH NO SIGNS OF SLOWING, THE DATA KEEPS GROWING
These are just some of the more common ways that Internet users add to the big data pool. In truth, depending on the niche of business you're in, there are virtually countless other sources of relevant data to pay attention to. Consider the following:
The global Internet population grew 6.59 percent from 2010 to 2011 and now represents 2.1 BILLION PEOPLE.
These users are real, and they are out there leaving data trails everywhere they go. The team at Domo can help you make sense of this seemingly insurmountable heap of data, with solutions that help executives and managers bring all of their critical information together in one intuitive interface, and then use that insight to transform the way they run their business. To learn more, visit www.domo.com.
DOMO
SOURCES: HTTP://NEWS.INVESTORS.COM/, ROYAL.PINGDOM.COM, BLOG.GROVO.COM, BLOG.HUBSPOT.COM, SIMPLYZESTY.COM, PCWORLD.COM, BIZTECHMAGAZINE.COM, DIGBY.COM


WHAT IS BIG DATA?
VOLUME
Large amounts of data.
VELOCITY
Needs to be analyzed quickly.
VARIETY
Different types of structured and unstructured data.
WHAT ARE THE VOLUMES OF DATA THAT WE ARE SEEING TODAY?
30 billion pieces of content were added to Facebook this past month by 600 million plus users.
zynga
Zynga processes 1 petabyte of content for players every day; a volume of data that is unmatched in the social game industry.
YouTube
More than 2 billion videos were watched on YouTube... yesterday.
LOL!
The average teenager sends 4,762 text messages per month.
32 billion searches were performed last month... on Twitter.
Everyday business and consumer life creates 2.5 quintillion bytes of data per day.
%
2012 2011 2010
90% of the data in the world today has been created in the last two years alone.
WHAT DOES THE FUTURE LOOK LIKE?
Worldwide IP traffic will quadruple by 2015.
By 2015, nearly 3 billion people
will be online, pushing the data created and shared to nearly 8 zettabytes.
HOW IS THE MARKET FOR BIG DATA SOLUTIONS EVOLVING?
A new IDC study says the market for big technology and services will grow from $3.2 billion in 2010 to $16.9 billion in 2015. That's a growth of 40% CAGR.
$16.9 billion
$3.2 billion
58% of respondents expect their companies to increase spending on server backup solutions and other big data-related initiatives within the next three years.
Key questions enterprises are asking about Big Data:
How to store and protect big data?
How to backup and restore big data?
How to organize and catalog the data that you have backed up?
How to keep costs low while ensuring that all the critical data is available when you need it?
2/3rds of surveyed businesses in North America said big data will become a concern for them within the next five years.
Asigra.

UPPSALA UNIVERSITET

In this course

- Relational Database Management Systems (RDBMS).
- Represent data as mathematical *relations*.
- Corresponding to the user-friendly concept of table.

STUDENT

Name	Student_number	Class	Major
Smith	17	1	CS
Brown	8	2	CS

GRADE_REPORT

Student_number	Section_identifier	Grade
17	112	B
17	119	C
8	85	A
8	92	A
8	102	B
8	135	A

COURSE

Course_name	Course_number	Credit_hours	Department
Intro to Computer Science	CS1310	4	CS
Data Structures	CS3320	4	CS
Discrete Mathematics	MATH2410	3	MATH
Database	CS3380	3	CS

PREREQUISITE

Course_number	Prerequisite_number
CS3380	CS3320
CS3380	MATH2410
CS3320	CS1310

SECTION

Section_identifier	Course_number	Semester	Year	Instructor
85	MATH2410	Fall	07	King
92	CS1310	Fall	07	Anderson
102	CS3320	Spring	08	Knuth
112	MATH2410	Fall	08	Chang
119	CS1310	Fall	08	Anderson
135	CS3380	Fall	08	Stone

Main RDBMSs:

- Oracle Server
- IBM DB2
- MS SQLServer
- SyBase
- IBM Informix
- Teradata

- MySQL
- PostgreSQL
- SQLite

- MS Access

UPPSALA UNIVERSITET

RDBMSs in real life

- E-Commerce (Ebay, Amazon, …)
- Banks, credit cards, …
- Hospitals, …
- Scientific databases (EPDB, …)
- …

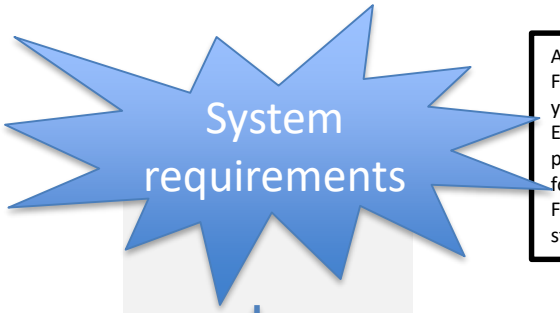
OVERVIEW OF THE COURSE

UPPSALA UNIVERSITET

Intended Learning Outcomes (course)

- **Design and use** relational databases in both theory and practical application.
- Explain basic **differences** between database design and conventional programming.
- **Present and discuss** the course content orally and in writing with proficiency appropriate to the course level.

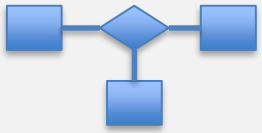
Lectures 1-5: from requirements to the database



A university offers courses to students.
For every student we register name, surname, address, year and student ID.
Every course instance is given on a specific year by a professor,
for whom we store the income and the supervisor.
For each exam we save the date and the grade of the student.

Database Designer

- Identifying the data to be stored
- Choosing appropriate structures to represent and store this data



Entity-Relationship Data Modeling
Relational model

Database Administrator

Data Definition Language (SQL)

- Create the database

UPPSALA UNIVERSITET

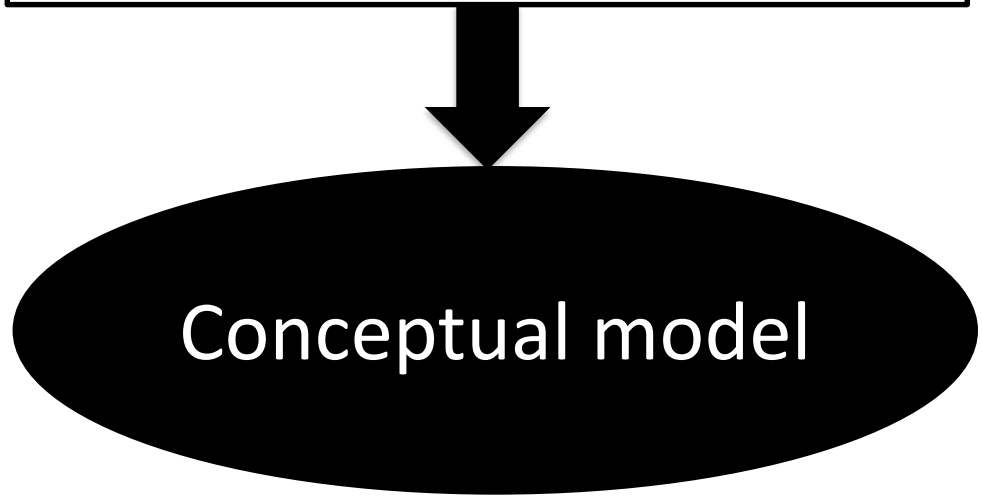
Lectures 2-3: Conceptual data modeling

A university offers courses to students.
For every student we register name, surname, address, year and student ID.
Every course instance is given on a specific year by a professor,
for whom we store the income and the supervisor.
For each exam we save the date and the grade of the student.

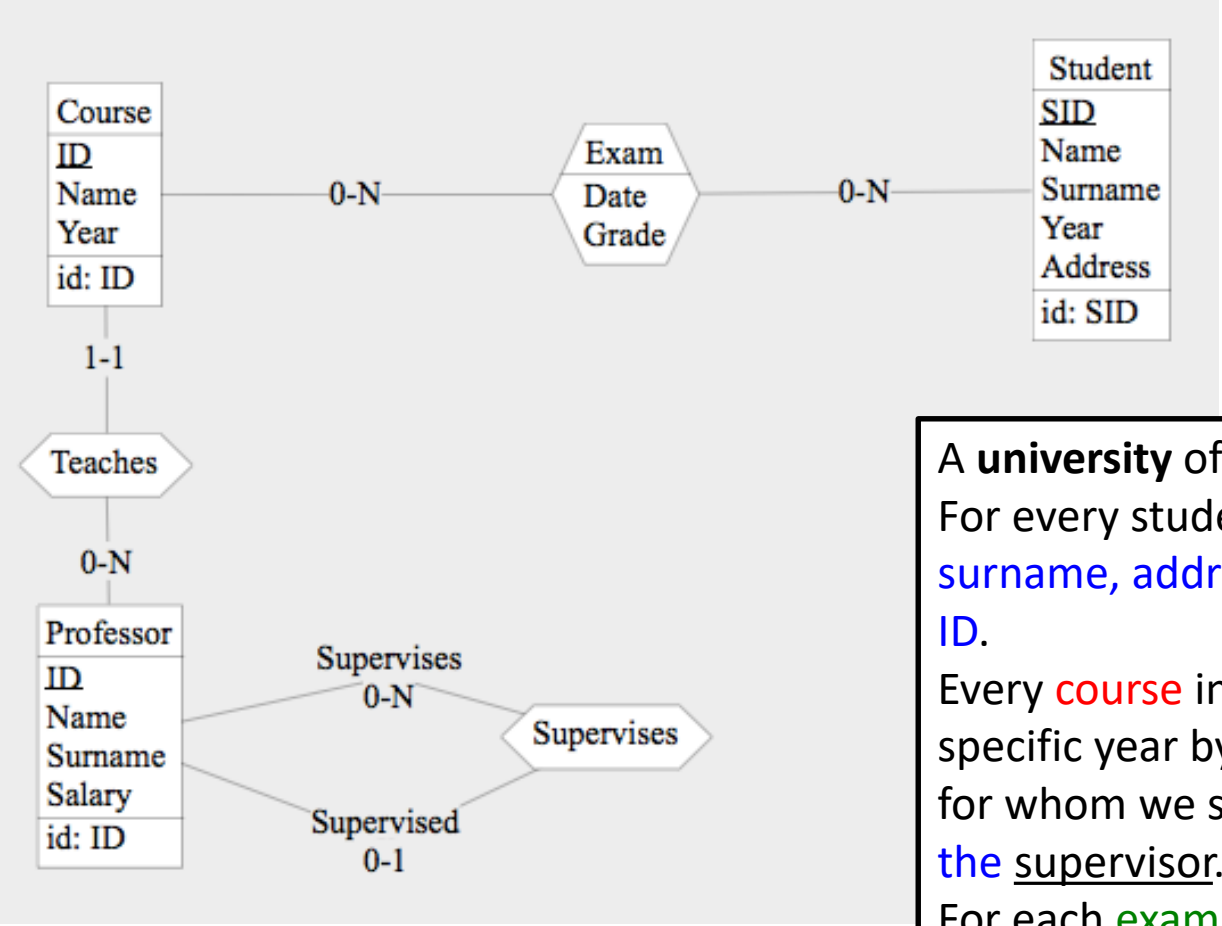
UPPSALA UNIVERSITET

Lectures 2-3: Conceptual data modeling

A **university** offers courses to students.
For every student we register name, surname, address, year and student ID.
Every course instance is given on a specific year by a professor,
for whom we store the income and the supervisor.
For each exam we save the date and the grade of the student.



Lectures 2-3: EER conceptual data modeling

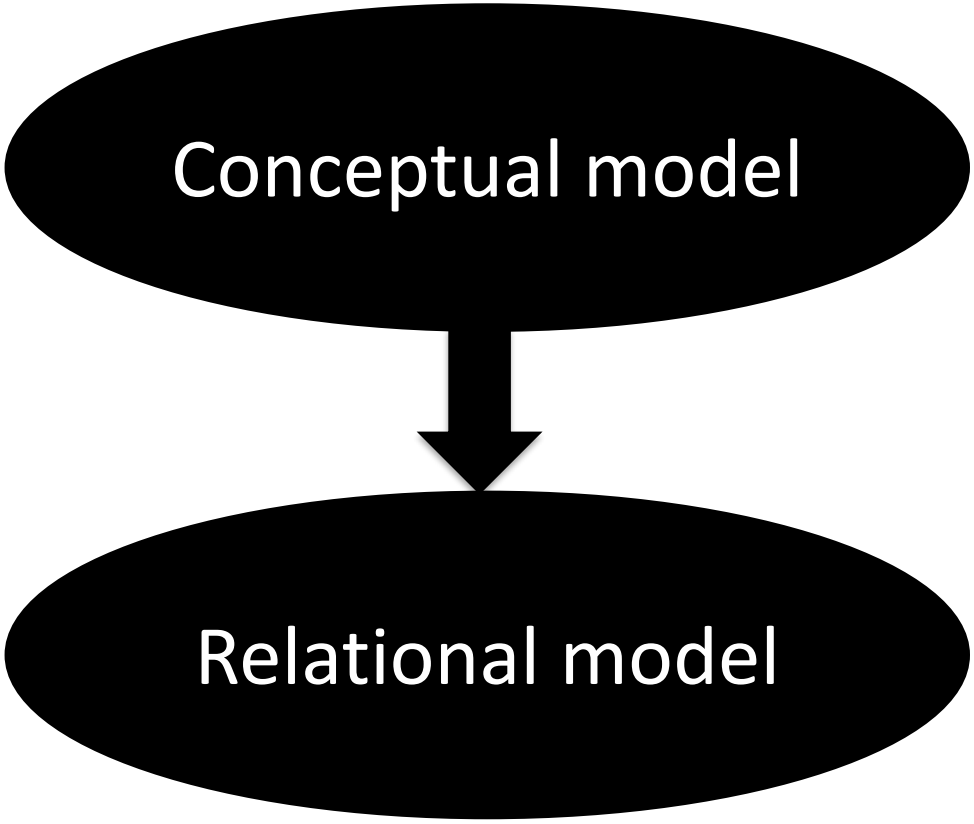


A **university** offers courses to students.
For every student we register name, surname, address, year and student ID.
Every course instance is given on a specific year by a professor,
for whom we store the income and the supervisor.
For each exam we save the date and the grade of the student.

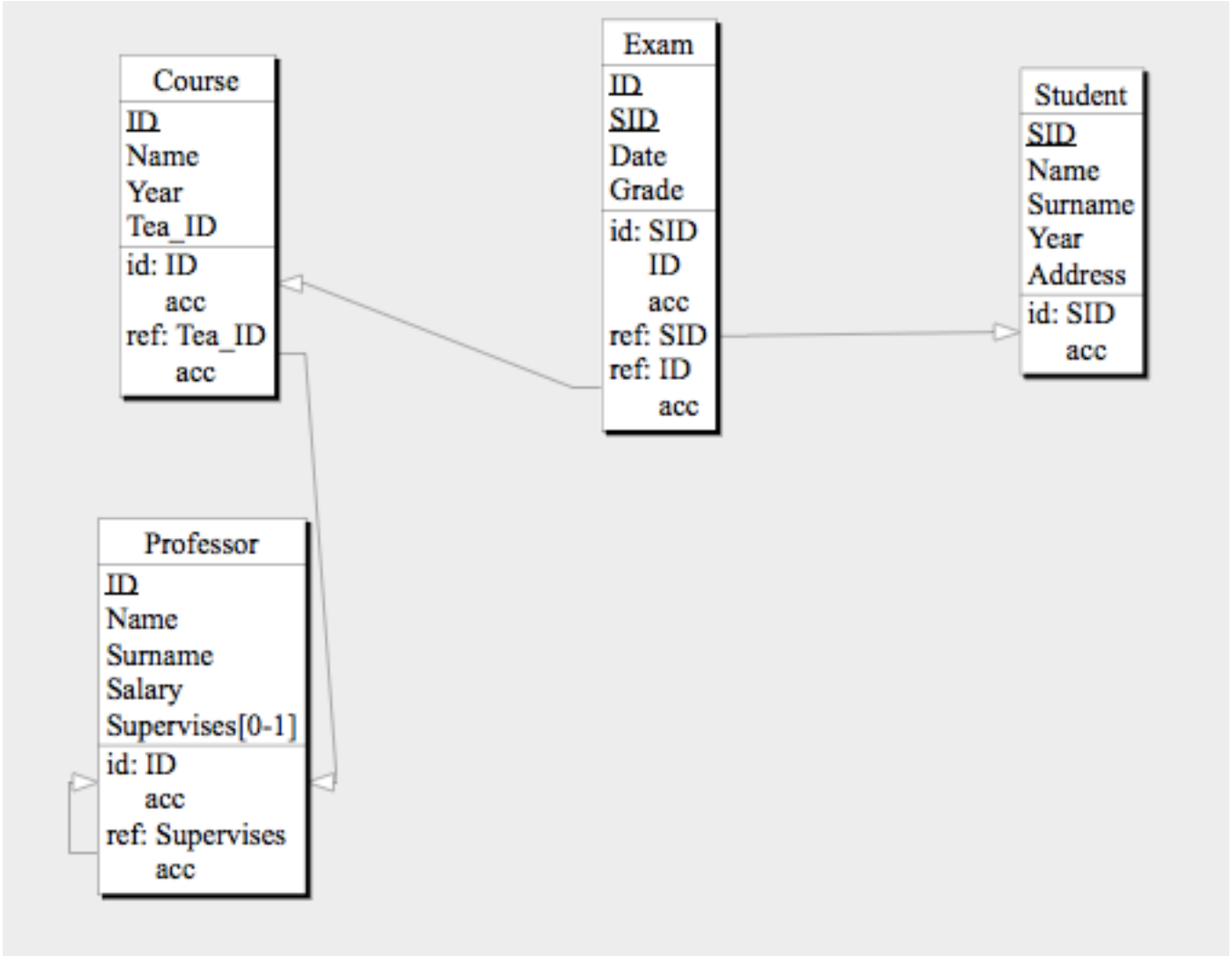
DEMO

UPPSALA UNIVERSITET

Lectures 4-5: Relational model



Lectures 4-5: Relational model (diagram)



Lectures 4-5: Relational model (SQL)

```
12  -- Database Section
13  -- ________________
14
15  create database EER;
16
17
18  -- DBSpace Section
19  -- ________________
20
21
22  -- Tables Section
23  -- _____________
24
25  create table Course (
26       ID varchar(20) not null,
27       Name varchar(100) not null,
28       Year numeric(4) not null,
29       Tea_ID char(1) not null,
30       constraint ID_Course_ID primary key (ID));
31
32  create table Exam (
```

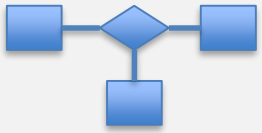
DEMO

Lectures 6-8: SQL as a query language

A university offers courses to students.
For every student we register name, surname, address, year and student ID.
Every course instance is given on a specific year by a professor,
for whom we store the income and the supervisor.
For each exam we save the date and the grade of the student.

Database Designer

- Identifying the data to be stored
- Choosing appropriate structures to represent and store this data



Entity-Relationship Data Modeling
Data normalization
Relational model

Database Administrator

- Create the database

Data Definition Language (SQL)

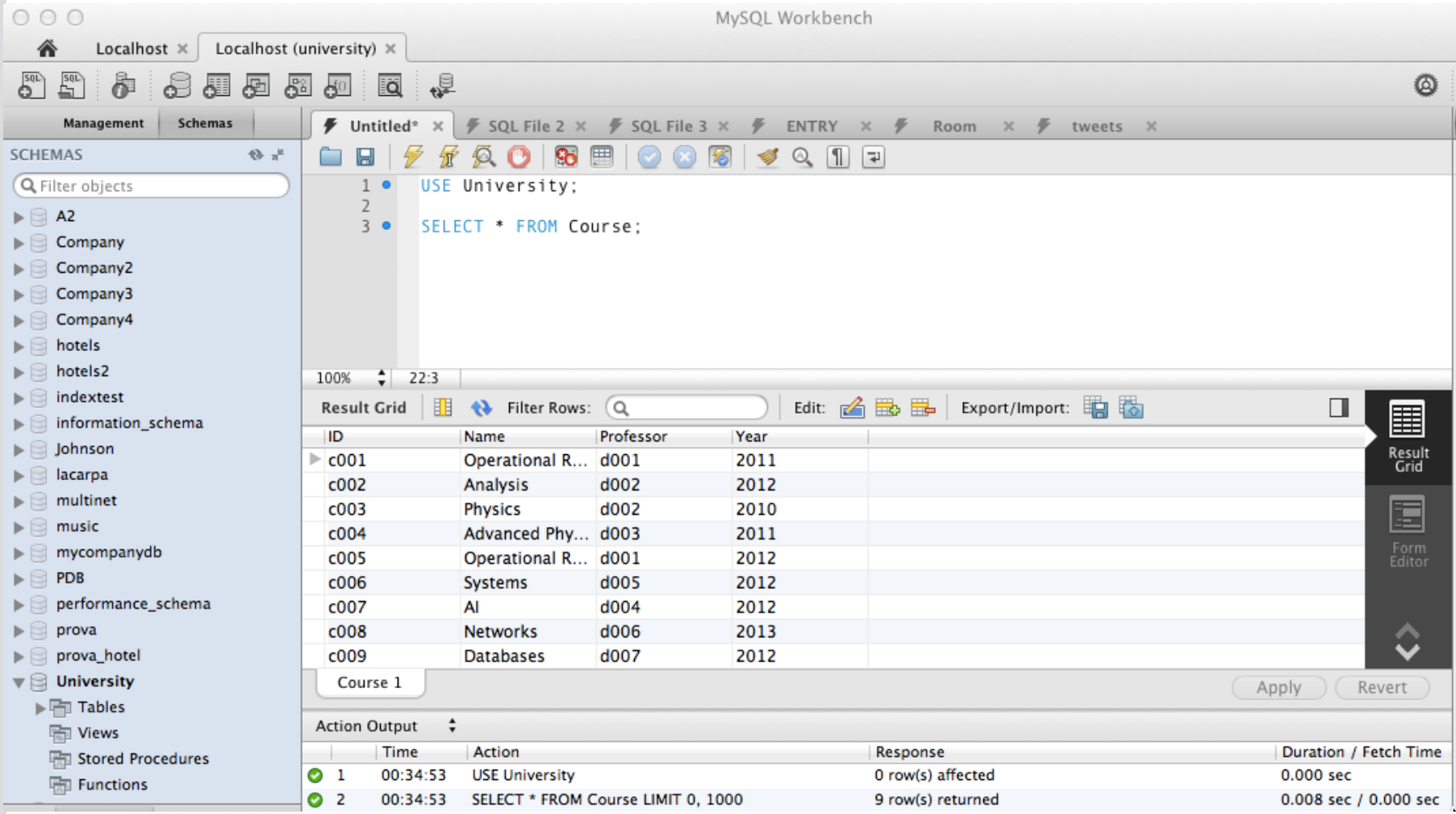
SQL

Database Users

- Populate and query the database

UPPSALA UNIVERSITET

Lectures 6-8: SQL as a query language



DEMO

UPPSALA UNIVERSITET

Lectures 9-10: Data quality and normalization

- If you follow the methodology learned so far, you will probably design a good model.
- Often, data already exists and must be improved.
  - This being an euphemism.
- Theory of normalization.
  - A formal way to check if a database suffers from specific quality issues and fix it.

Essentially, all models are wrong, but some are useful.

George E. P. Box

UPPSALA UNIVERSITET

Two/three lectures

- Database programming.
  - SQL + Java.
- Transactions.
  - Transfer 50$ from Account A to account B
    - Remove 50$ from A account
    - Add 50$ to B account
- Security.
  - Many users accessing various part of the database

UPPSALA UNIVERSITET

Project: A trading database

- Your group has been asked to help with a project for AltOnline AB, a new Swedish company which, inspired
  - by the tremendous success of Amazon.com, Inc., wants to become a market leader in online sales in Sweden.
  - The project is rather large and your part is to design and implement a database system for its online store.

UPPSALA UNIVERSITET

Project: A trading database

- The project gives you the opportunity to practice a relational database design process from A to Z:
  - Reading the customer specification, which might be redundant, vague or even inconsistent;
  - Conceptual modelling using ER diagrams;
  - Creating, populating and querying the database in an RDBMS; SQL queries
  - Normalisation
  - Writing applications accessing and modifying the stored data.

PRACTICAL ISSUES

UPPSALA UNIVERSITET

Examination

- Project (2 credits in total).
  - They are part of the examination,
  - But the objective is to learn through practical applications.
  - Make sure you register on Ladok on time
- Exam (3 credits, theory and practice).
  - The most important part for some of you.
  - The least for me.
  - You will receive a lot of material, but do not assume to find exams with the same structure as the previous ones.

UPPSALA UNIVERSITET

Groups for the Project

- Projects are done by groups of 3-5 students.
- Every group has reserved lab room and a tutor.
  - Feel free to "use" your tutor, but do not ask questions to both me and them.
  - For each Lab, I will make a **doodle** to choose from a number of slots (First come first serve). I will email you instructions (so expect an email)
  - You need to attend only one slot per week (thus 4 in total)

- Form your group on Studium
  - Register autonomously to one group **today**.
  - **End of next week** the group page will be locked.
  - We have created a discussion forum on studium that you can use as to join or form a group

To do list (this week)

1. If you **aren't admitted to the course**
   email: *it-kansli@it.uu.se*.
   - Master students: *studievagledare@it.uu.se*
   - Exchange students: *ulrika.jaresund@it.uu.se*
2. Register-form a group.
3. Install MySQL on your laptop.
   – Details will be provided on the portal **SOON**.

UPPSALA UNIVERSITET

Wrap up

- Importance and role of database systems.
- Basic terminology.
  - Data, database, DBMS, database system.
- Key concepts.
  - Data model – Schema and instance.
  - Data independence.
- Next: database design with the ER model.

UPPSALA UNIVERSITET

Something about yourself

- What is your program?
- What is your current knowledge of database systems?
  - Database what? Beginner? Expert? Guru?
- What do you expect from this course?
- Other questions.