

# COMP207

## Database Development

2023-24

Rasmus Ibsen-Jensen

# What is this course all about?

- Databases...
- Roughly half is also about concurrency

# Questions?

- If you have a question, pick one:
  - If during a lecture, you can raise your hand and I will let you ask it when I notice you
  - You can ask after/before a lecture
    - Just outside this room however
  - Ask during the labs I attend
    - I intend to attend every second week
  - Ask on the online discussion board
  - **If the question is personal and should not be shared**, you can ask it by email:  
R.Ibsen-Jensen@Liverpool.ac.uk

# Attendance monitoring

- As you might have heard, we are required to monitor attendance in lectures
- It is done as follows:
  - You log in on [timetables.liverpool.ac.uk](http://timetables.liverpool.ac.uk)
    - or download UoL timetables app
  - You find the session you want to register for
    - during the session
  - You click on it
  - Click on “Register your attendance”
  - You then write in a 6-digit code we give you during each session
  - Click “Register”

*Databases are everywhere*

Topics ▾

Current edition

More ▾

Regulating the internet giants

## The world's most valuable resource is no longer oil, but data

*The data economy demands a new approach to antitrust rules*

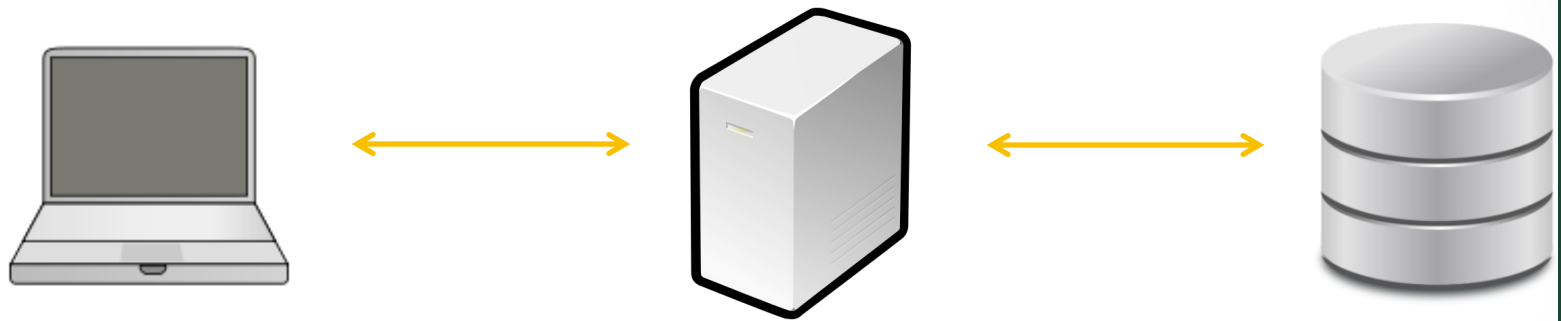


Print edition | Leaders >

May 6th 2017

... and are  
valuable

# Databases Drives the Web



User

Web service

Database

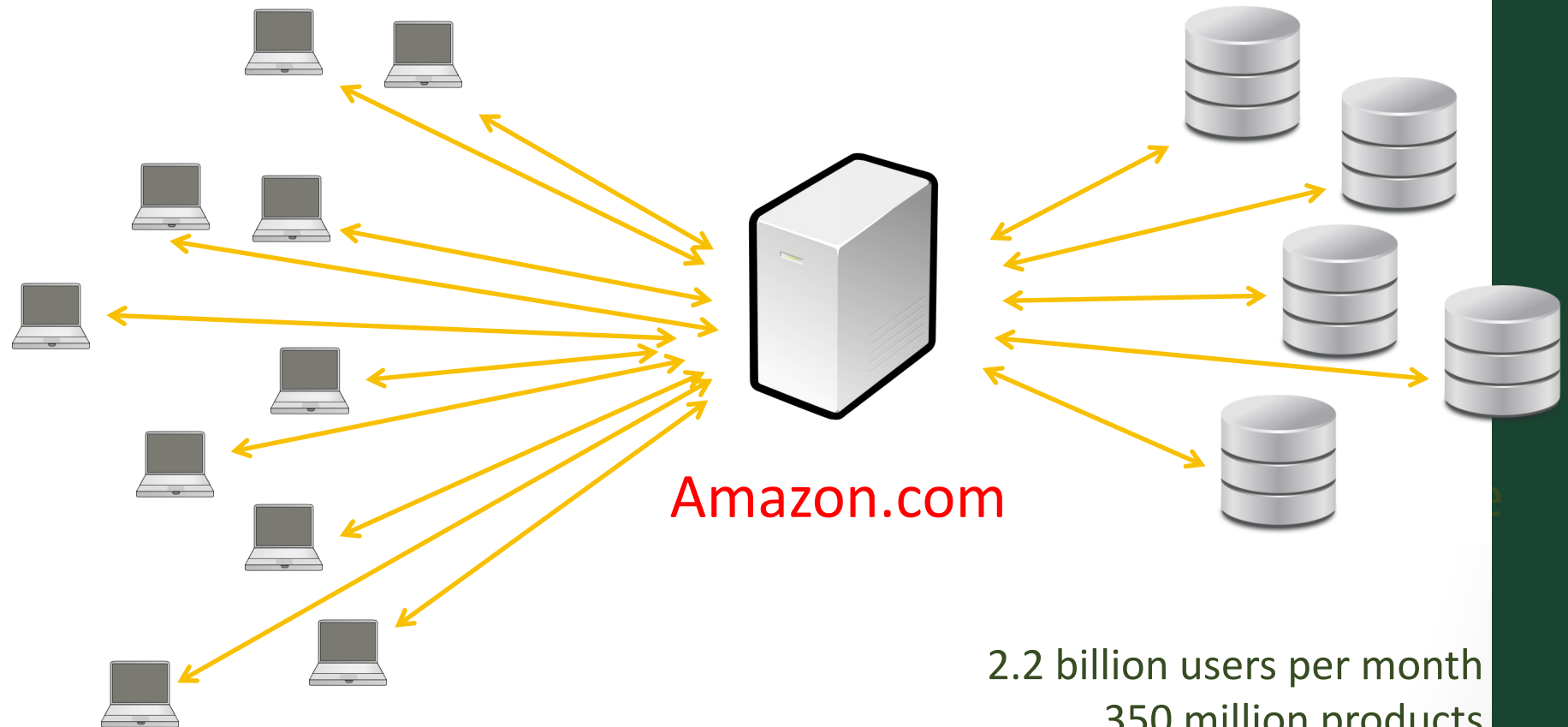
Online store

Social media site

Airline/train/hotel reservation

– many more –

# Databases Drives the Web

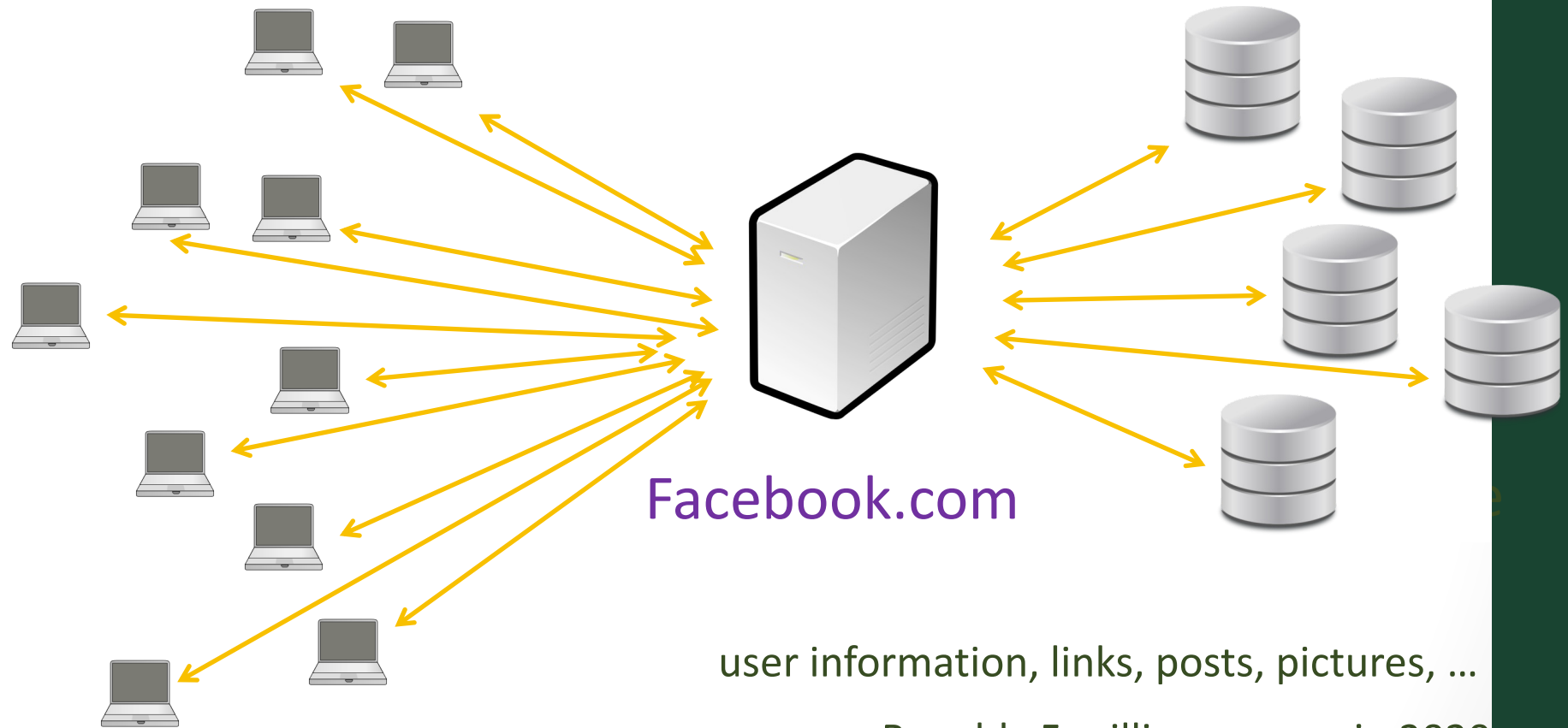


2.2 billion users per month  
350 million products  
spread over a number of databases

millions of users per day  
many at the same time



# Databases Drives the Web



Facebook.com

user information, links, posts, pictures, ...

loading the homepage of a user  
requires accessing 100s of servers

Roughly 1.62 billion users daily

Roughly 5 million servers in 2020  
(from: power usage (7.1 terawatts)  
and avg server (around 1megawatt)  
and a few more details)

# Databases Drives the Web

Requires *extremely efficient* systems for *querying data*, and for *storing and manipulating* data in a *safe* way.

Facebook.com

user information, links, posts, pictures, ...

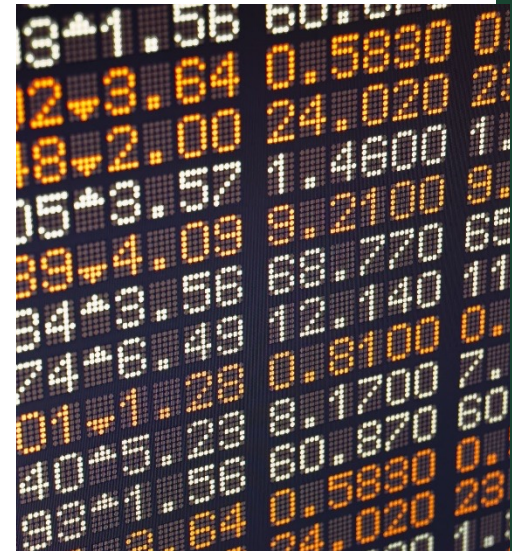
loading the homepage of a user  
requires accessing 100s of servers

Roughly 1.62 billion users daily

Roughly 5 milion servers in 2020  
(from: power usage (7.1 terawatts)  
and avg server (around 1megawatt)  
and a few more details)

# Databases Drives Businesses

- Banking: debit/credit card transactions
  - E.g., in stores/super markets, at ATMs, etc.
- Stock market
- Big companies & organisations:
  - Employees, customers, products, sales, schedules, logistics, clients, etc.
  - Purposes:
    - Some are worthwhile directly
    - Management, data analytics and decision making



# Databases in Science & Public Service

- Science: for data management and data analysis in
  - Astronomy
  - Human Genome
  - Biochemists exploring properties of proteins
  - Biology and life sciences
  - Geology
- Medicine/NHS:
  - Patient records
  - Diseases and their relationships, treatments, etc.
  - Decision making
- Many others

Not always using proper databases!



# NHS example



## Covid: Test error happened' - Hancock

🕒 5 October 2020

Coronavirus pandemic

The technical error was caused by some Microsoft Excel data files exceeding the maximum size after they were sent from NHS Test and Trace to Public Health England.

It meant 15,841 cases between 25 September and 2 October were left out of the UK daily case figures.

PHE said the error itself, discovered overnight on Friday, has been fixed, and outstanding cases had been passed on to tracers by 01:00 BST on Saturday.

# University of Liverpool example



INDEPENDENT

Rugby World Cup >

The Ripple Effect >

NEWS

SPORT

VOICES

CULTURE

LIFESTYLE

TRAVEL

PREMIUM

MORE

Student > News

## Students furious as Liverpool University emails the wrong results

Tom Mendelsohn • Monday 01 July 2013 16:47 • [Comments](#)



# Early Data Management

- Early DBMS of the 1960's were based on this idea
- Disadvantages:
  - difficult to program
  - not very robust, especially when dealing with updates to data by many users in parallel
  - Hard to add fields or new efficient queries

```
class Student {  
    String name;  
    int number;  
    String programme;  
    ...  
}
```

```
Vector<Student>  
students;  
...
```

```
Anna, 20181989, G402  
John, 20184378, G702  
...
```

External data file



# Relational Databases to the Rescue...

- Modern database management systems (DBMSs) are based on 50+ years of database research
  - Very sophisticated tools
  - Can manage very large amounts of data over a long period of time
  - Highly efficient, flexible, robust

ORACLE®



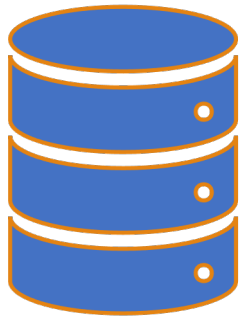


# Relational Model

- Data organised as relations (“tables”)

Items		Employees		
name	price	birthday	first_name	family_name
2L Cola	3.00	1990-11-10	Anne	Smith
Banana	0.10	2000-02-05	David	Jones
...	...	...	...	...

- Data is queried/modified at a high level (e.g. via SQL)  
No need to know how data is stored and where
- Introduced in 1970, most dominant model by 1990



*“Relational DBMS are  
one of the biggest success  
stories in computer  
science.”*

– a colleague, 2012

Also: one of the most complex pieces of software...

# Course overview

- What are we doing in each week?

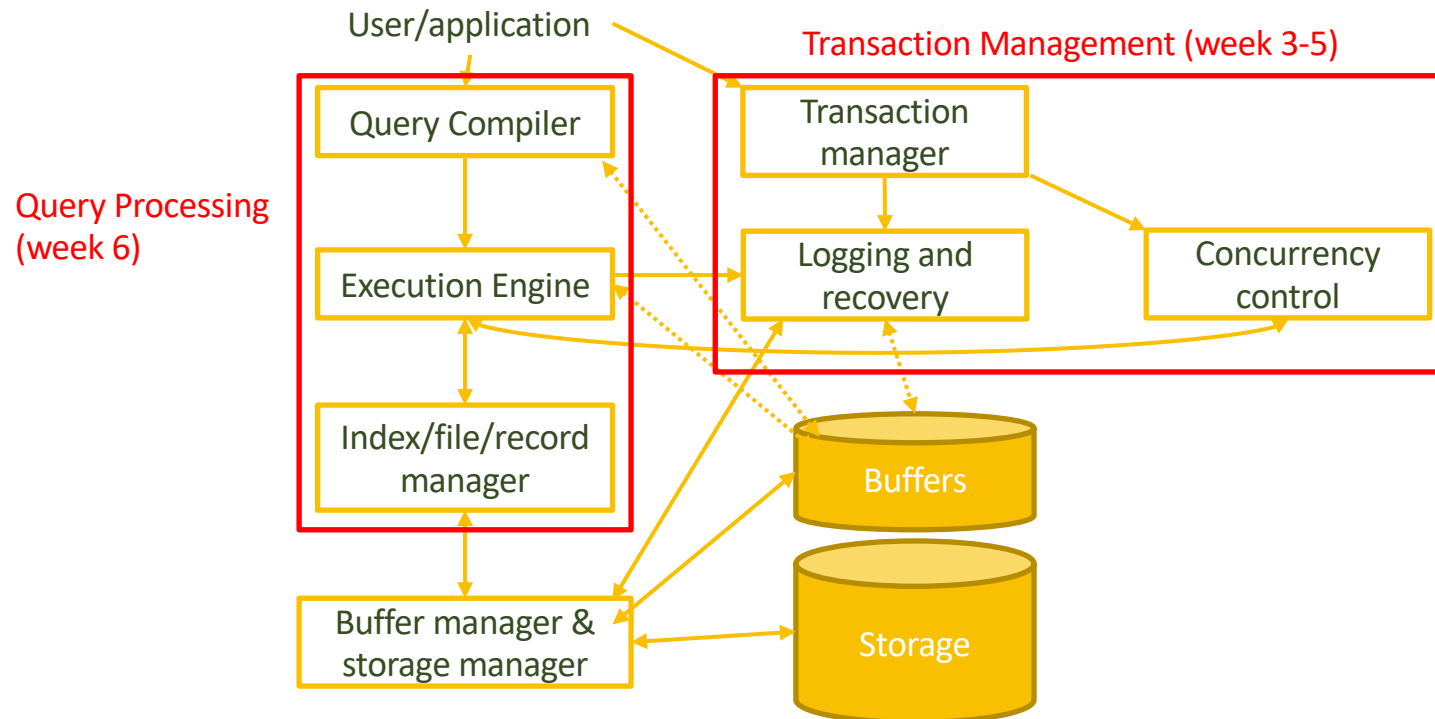
# Basics of SQL and the relational model

- First two weeks (week 1-2):
  - We will show you how to make relational databases
  - How to insert/delete/query them

Student		Module		
name	number	code	name	lecturer
Anna	20181989	COMP207	Database Development	R. Ibsen-Jensen
John	20184378	COMP219	Artificial Intelligence	F. Oliehoek
...	...	...	...	...

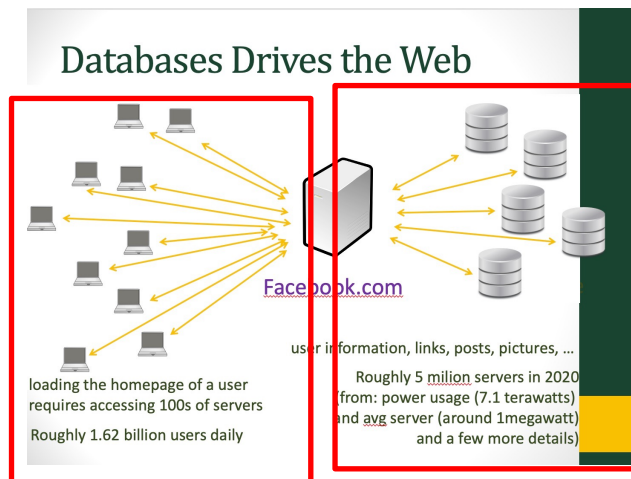
# Relational DBMS Components

(Simplified)



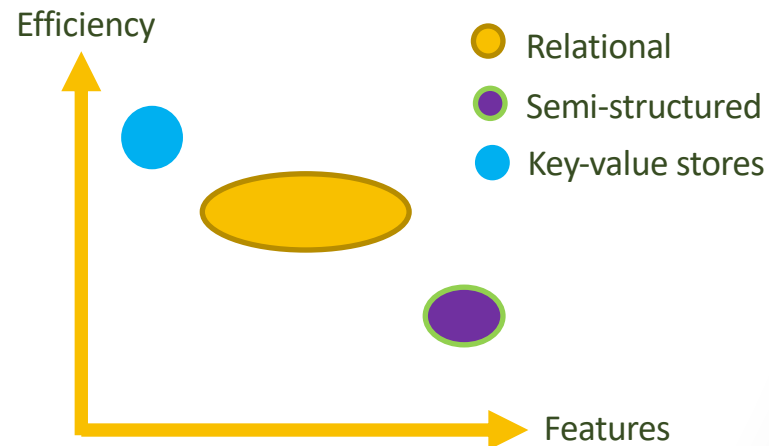
# Beyond Plain Relational DBMS

- Distributed DBMSs (week 7)
- Beyond relational data (week 8-9)
- Data warehouses/analysis/mining (week 10)



Transaction  
Management  
(week 3-5)

Distributed  
DBMSs  
(week 7)



# What you should be able to do at the end

## (Learning Outcomes)

- Basic/advanced SQL:
  - Application in problem solving (focused on in week 1,2 and 6)
- Transaction management:
  - Identify and apply the principles underpinning transaction management within DBMS (focused on in weeks 3,4,5 and 7)
- Web technologies:
  - Illustrate issues related to web technologies as a semi-structured data representation formalism (focused on in weeks 8 and 9)
- Data warehouses and data mining:
  - Interpret the main concepts and security aspects in data warehousing
  - Interpret the main concepts of data mining
  - Focused on in week 10

# Topics We Will Not Cover

- No modelling of databases
- No administration of DBMS
- No implementation of database system components



Some words on logistics...

# Lectures & Labs/Tutorials

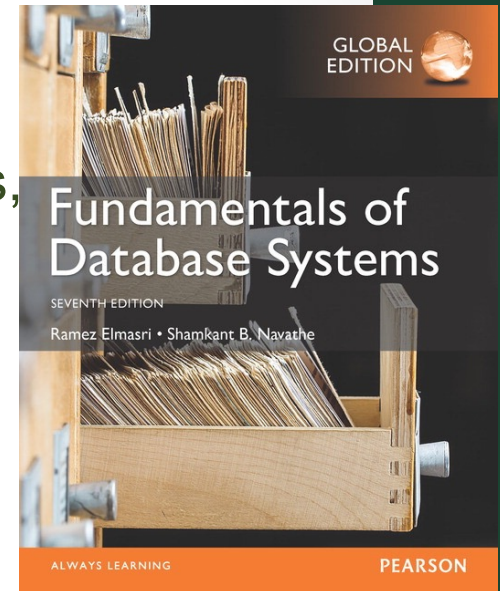
- 20 “lectures” in total (week 1-10)
- 10 weeks of labs (starts in week 2)
  - Each week, you have 1 lab hour
  - You should be assigned to one of the lab sessions
  - *If not: let me know as soon as possible*

# Format for lectures

- Each week we have 2 “lectures”
- The lectures will focus on examples: You **NEED** to watch the online videos for definitions and similar before the Thursday lecture
  - Exception: The first two weeks focus on SQL and should be possible to follow without the videos
    - If you struggle (with or) without watching let me know
  - From week 3 onwards, I very much expect that you won't be able to follow the lectures if you did not watch the videos
    - If you struggle while watching, do let me know
    - If you do manage without the videos and no prior knowledge on the subject, do let me know, because I would be impressed

# Resources

- Course page on Canvas:
  - Announcements, videos, slides, assessments, stream lectures, discussion board, etc.
  - *Let me know if you don't have access*
- Textbooks (any is fine):
  - Garcia-Molina, Ullman and Widom: *Database Systems – The Complete Book*, Pearson Education
    - Free e-book
  - Elmasri and Navathe: *Fundamentals of Database Systems*, Pearson Education
    - Also free e-book, but that might be removed
  - Connolly and Begg: *Database Systems*, Addison Wesley
- Your own notes + the web



# Exam

- The exam is 60% of your grade
- It has 30 MCQs
- Each MCQ has:
  - 5 options and you pick one
  - One is correct
  - Each correct answer gives 1 point (i.e. 2% of the final grade), and each wrong answer 0 point
- You will get more information about it in the last lecture and get some mock exams and similar towards the end

# Main assessment

- There will be a 25% main, practical assessment in SQL with deadline  
Wednesday the 8<sup>th</sup> of November
- Will discuss this main assessment next week

# Weekly assessment

- There will be a weekly assessment, one for each of the 10 weeks
  - Each will have 3 MCQs similar to the exam
  - Each will give 1.5% and you will have 30 mins to do it once started (i.e. each question gives 0.5%)
  - You will get access to it on Tuesday each week
  - You should do it before Monday at 17:00 two weeks later (to let you have the labs first)
  - Means: you got the one for the lectures this week this Monday and have to finish it not next Monday, but the one after that
  - Similar for the remaining weeks

# Weekly survey

- Each week, there will be a survey, asking the following questions:
  - Can you list 1-3 things that worked particularly well this week (and come with feedback explaining why you liked that if possible)?
  - Can you list things that did not work well during this week (1-3, but if there are more, please list those as well) - preferably with some explanation if it is not self-evident.



# Typical weekly lectures

- Tuesday lecture:
  - What I'll do about the weekly survey from previous week (not this week)
  - Walkthrough of just finished weekly assessment
    - (not this or next week though)
  - Examples
- Thursday lecture:
  - A brief summary of that week's videos
  - Examples

# Summary

- Database systems
  - Success story...
- Understanding how they work is important
  - For practitioners and researchers alike
  - This is what this course is about
- **For next lecture:**
  - Familiarise yourself with the **Canvas course page**
  - Keep checking timetables for first few weeks