COMP207
Database Development

2019-20

Rasmus Ibsen-Jensen

Acknowledgments

These slides are based on André Hernich slides,
who taught this module two years ago.

Questions during lectures

- If you have a question during a lecture, pick one:
  1. Put up your hand
  2. Write it in the thread for the lecture on the forum
  3. Ask the question after the lecture

Attendance checking

- Will be in the lab sessions only (starts in week 3)!

Kortext trial

- Running a trial
- In essence, you get two free e-book in this course
  - Access to *Database Systems – The Complete Book* this year
  - Access to *Fundamentals of Database Systems* might be removed

What is a database?

- Reflects an *updated* aspect of the real world
- Consists of a logical collection of data
- Has been build for a specific purpose

Databases are everywhere

… and are valuable

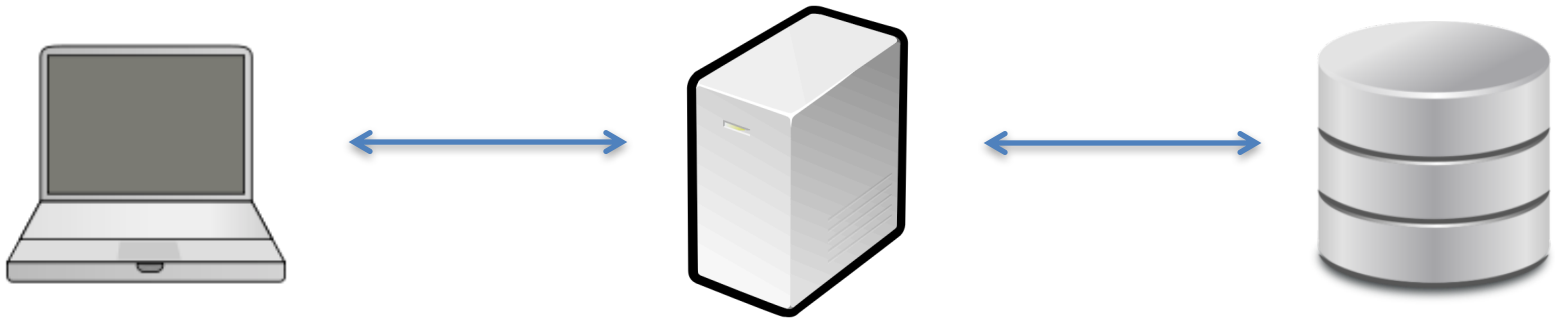
Regulating the internet giants

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

Databases Drives the Web



User Web service Database

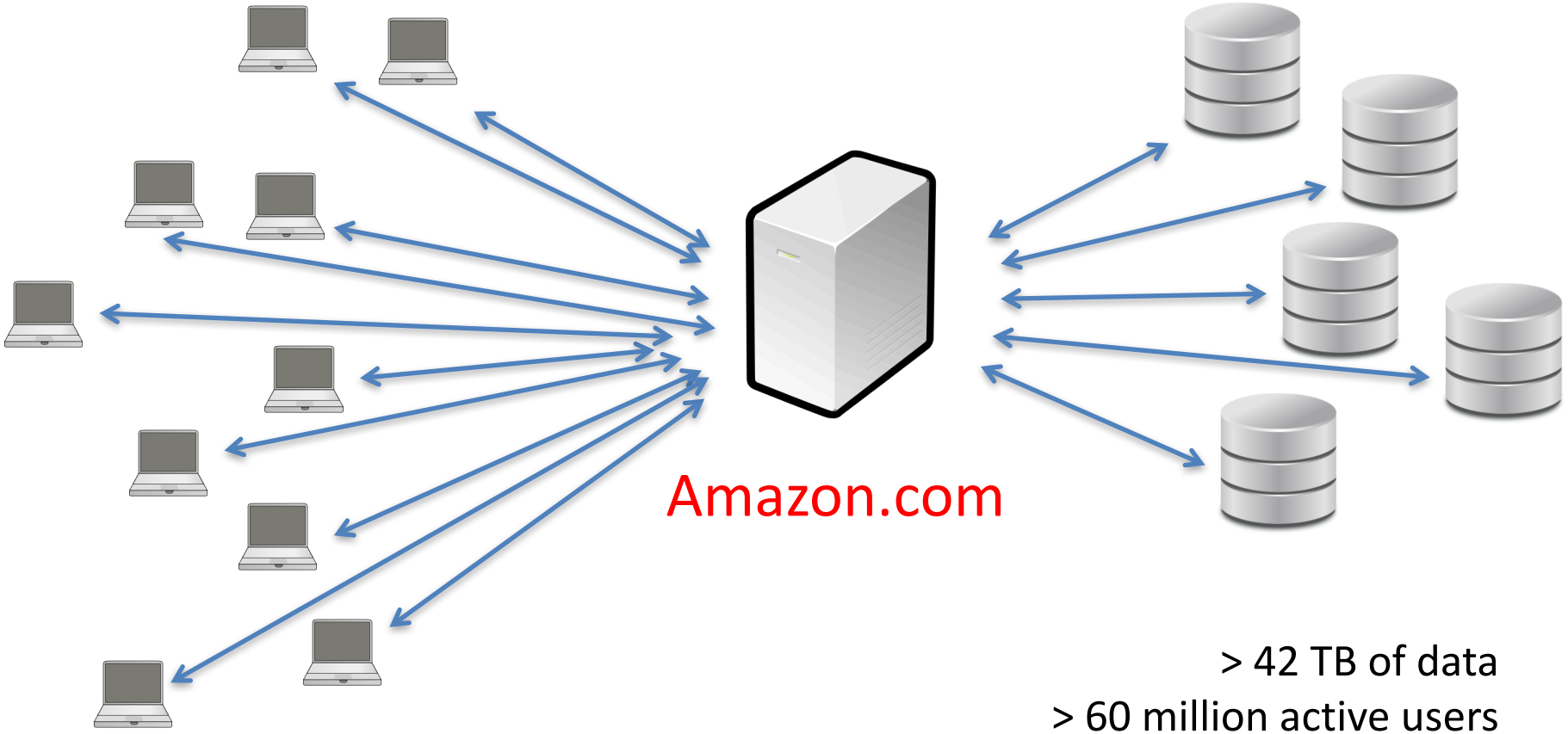
Online store

Social media site

Airline/train/hotel reservation

– many more –

Databases Drives the Web

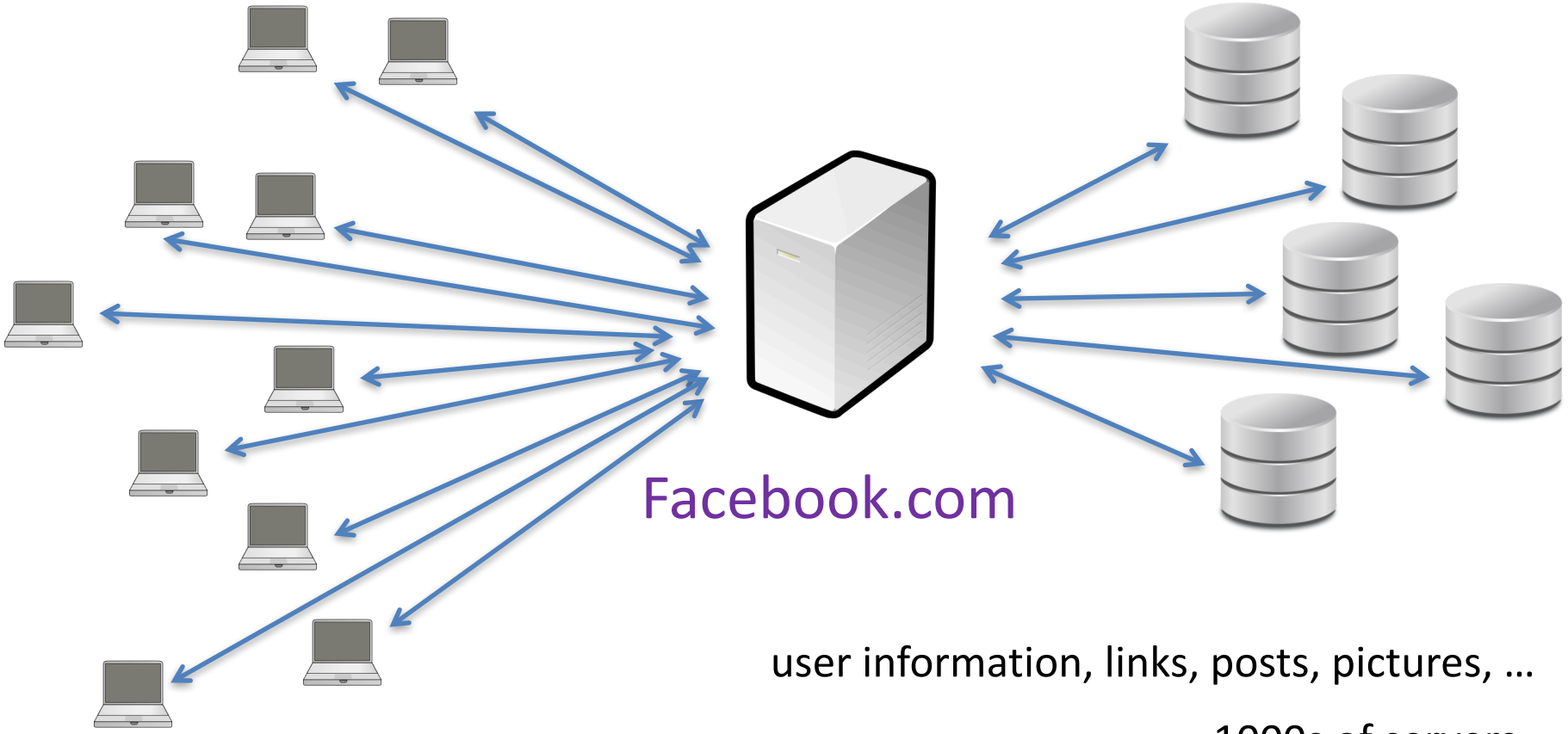


millions of users per day
many at the same time

> 42 TB of data
> 60 million active users
millions of products

100s of servers

Databases Drives the Web



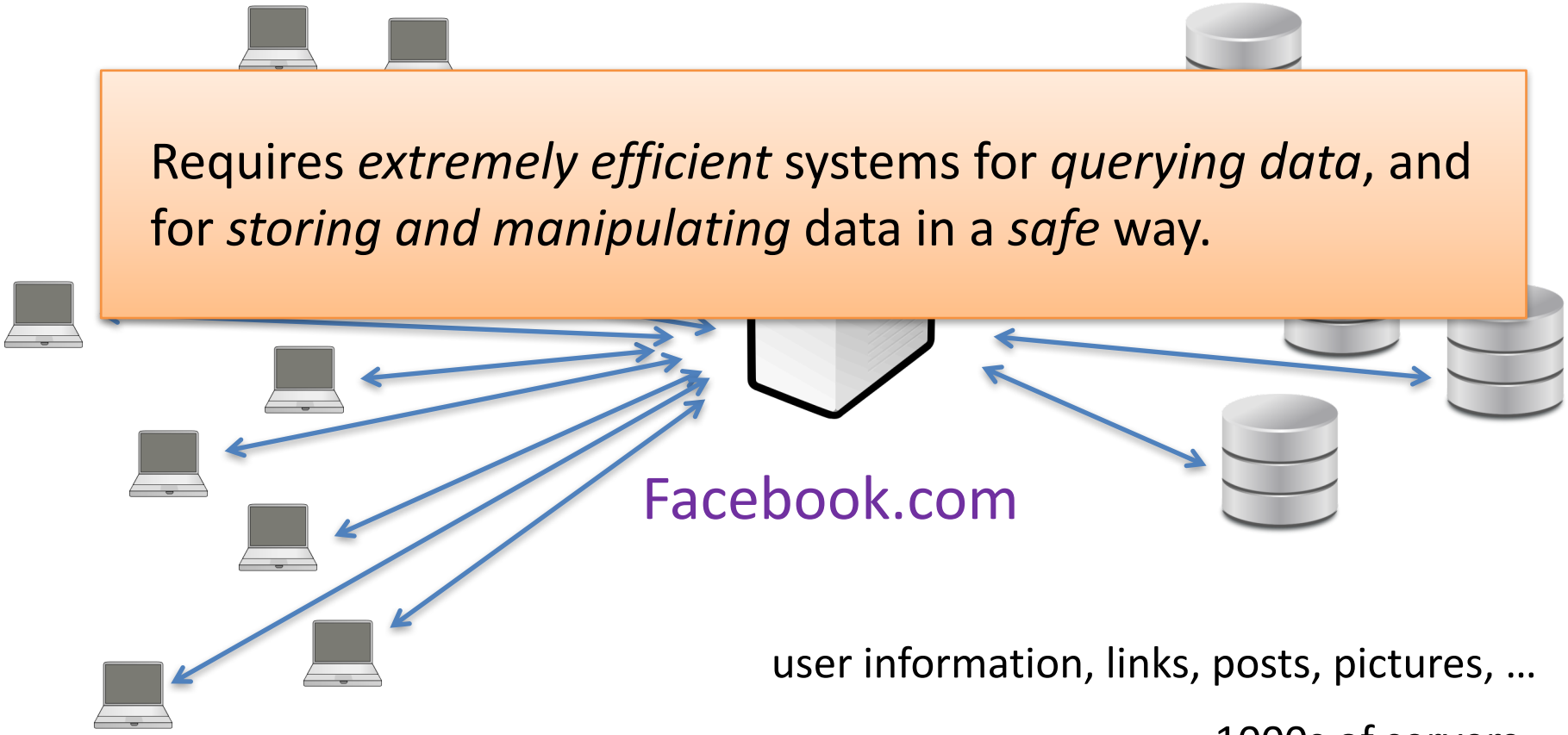
user information, links, posts, pictures, …

1000s of servers

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

> 1.13 billion users daily

Databases Drives the Web



user information, links, posts, pictures, …

1000s of servers

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

> 1.13 billion users daily

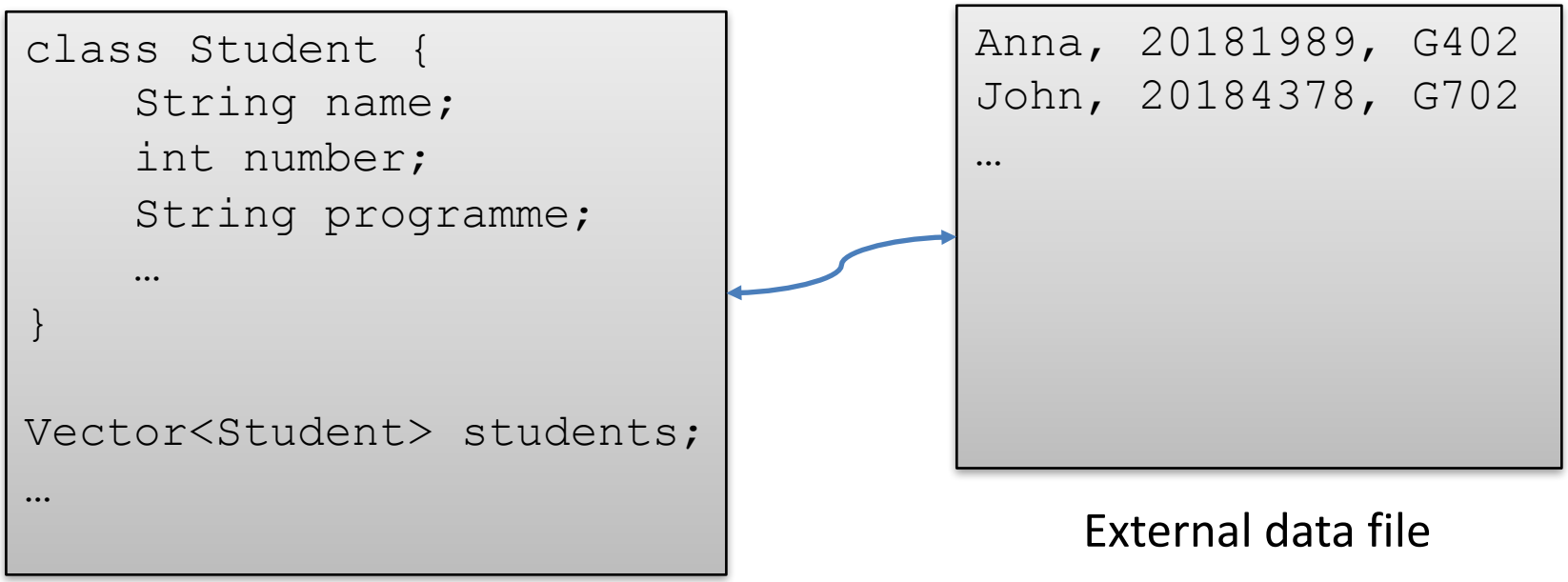
Databases Drives Businesses

- 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

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

Relational Databases to the Rescue…

- Modern database management systems (DBMSs) are based on 40+ 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®

PostgreSQL

Relational Model (Reminder)

- Data organised as relations (“tables”)

Student

name	number	…
Anna	20181989	
John	20184378	
…	…	…

Module

code	name	lecturer
COMP207	Database Development	R. Ibsen-Jensen
COMP219	Artificial Intelligence	F. Oliehoek
…	…	…

- 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

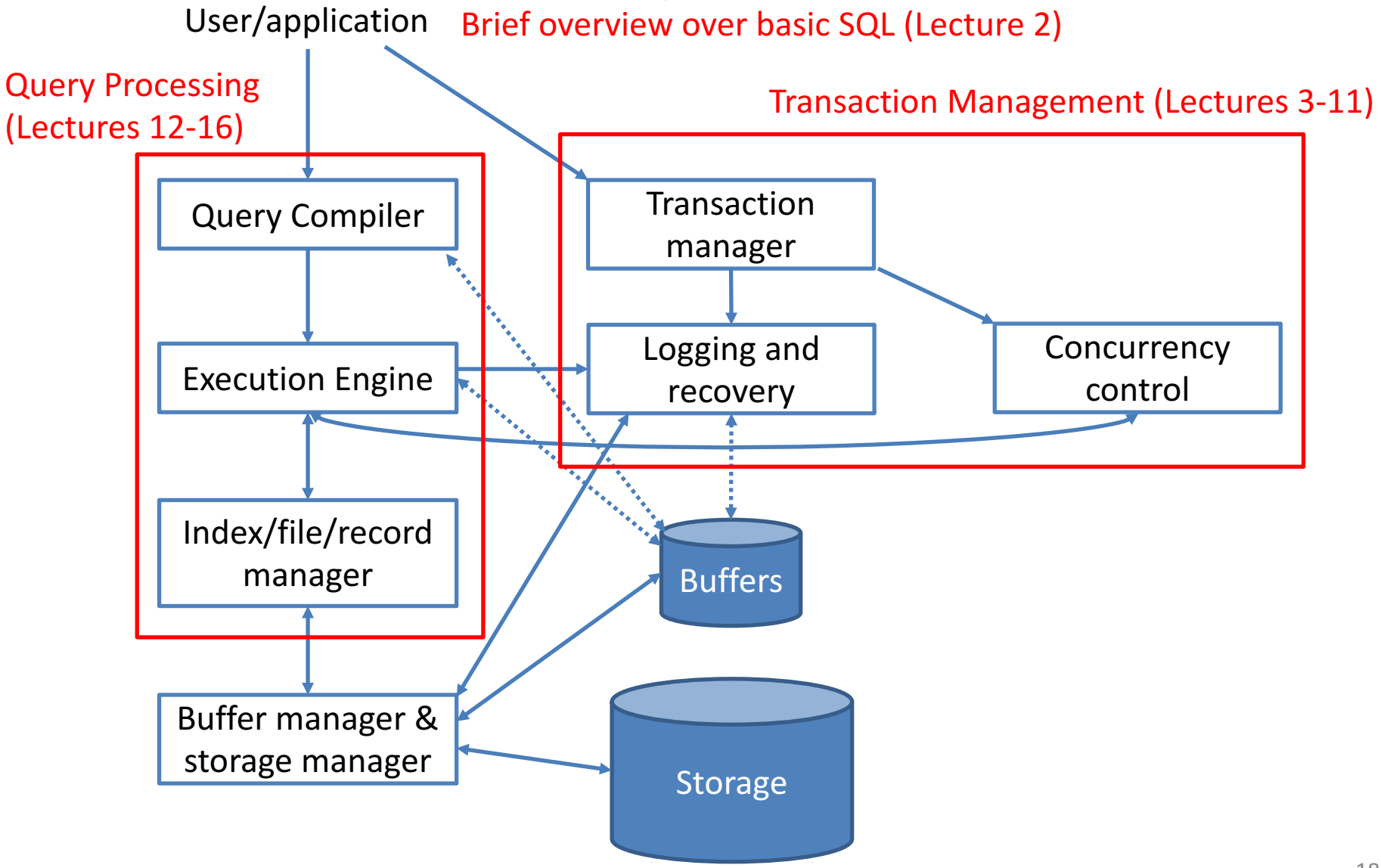
“Relational DBMS are one of the biggest success stories in computer science.”

– a colleague, 2012

Also: one of the most complex pieces of software…

Relational DBMS Components

(Simplified)



Beyond Plain Relational DBMS

- Lectures 17-18:
  - Distributed DBMSs
- Lecture 19: MapReduce
- Lectures 20-26: Beyond relational data
  - Object-relational databases
  - Semi-structured databases: XML
  - NoSQL databases
- Lectures 27-30:
  - Data warehouses
  - Data analysis/mining

What you should know at the end (Learning Outcomes)

- Transaction management:
  - Identification & application of the principles underpinning transaction management within DBMS
- Advanced SQL:
  - SQL from COMP107 extended with indexes, transactions, query optimisation
  - Application in problem solving
- Object-relational models:
  - Identification of principles
- Web technologies:
  - Illustrate issues related to web technologies as a semi-structured data representation formalism
- Data warehouses and data mining:
  - Interpret the main concepts and security aspects in data warehousing
  - Interpret the main concepts of data mining

If you care, what can you become?

- Data analysts:
  - Writing high-performant queries requires understanding of how these systems work
  - Same as programming in Java
- Database administrators:
  - Even more so
  - Need to design and tune system so that typical SQL statements are executed fast and deciding who is allowed to access the database
- System analysts/software developers:
  - Should be familiar with the full range of capabilities of DBMS
  - Writes reuseable queries for end users
- Research:
  - DBMS are still evolving, especially now: “NoSQL”, “big data”, …

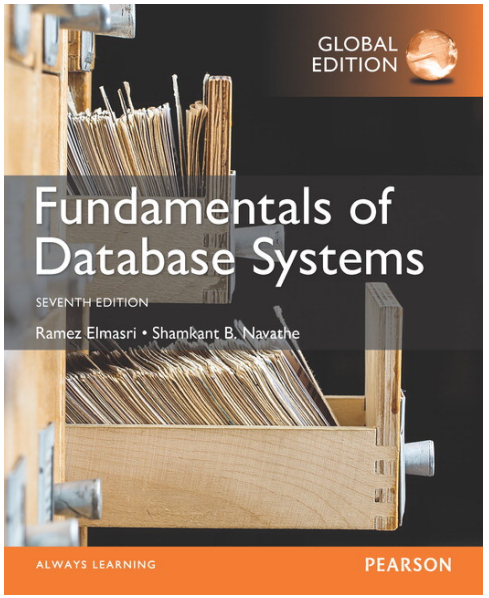
Topics We Will Not Cover

- No modelling of databases, but see…
  - COMP107/CSE103
  - 2nd year group project and COMP283
- No administration of DBMS
- No implementation of database system components

Some words on logistics…

Resources

- Course page on Vital:
  - Announcements, slides, stream lectures, course work, 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 because we participate in a trial
  - 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



Lectures & Labs/Tutorials

- 30 lectures in total (week 1-10; might be longer)
- 10 weeks of labs (starts in week 3)
  - You should be assigned to one of the lab sessions
  - *If not: let me know as soon as possible.*

Interaction

- Use lectures, labs, discussion board, and office hours
- Discussion board on Vital:
  - For all questions, comments, etc. regarding COMP207
  - I will try not to give complete answers to questions
    - I might provide hints that you can use to answer questions, either yourself or with the help of others
    - Try to answer others' questions (also benefits yourself)
  - Anonymous posts are activated, but try to not use this as the default mode
- For messages that should not be shared with others: r.ibsen-jensen@liverpool.ac.uk

Assessment

- 20% continuous assessment:
  - 2 assignments
  - Each 10% of the final mark
- 80% final exam:
  - Multiple-choice questions
  - 120 minutes
- There’ll be a mock exam towards the end

Private Study

- Invest ~ 6 hours of private study per week
- Study actively, not passively
  - Work through (don't just read) lecture notes, book chapters, online resources, etc.
  - Work through examples and experiment with these (e.g., modify them and try to see what happens)
  - Engage with other students, if only on the discussion board
  - Discuss/explain material to others (extremely good to see if you understood it)
- Start now & study regularly – don't start the night before the exam (this won't work)

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 Vital course page
  - Review relevant material from COMP107/CSE103:
    - Relational model, theory of decentralised DBs, DBMS
    - SQL
  - Keep checking timetables for first few weeks