Introduction

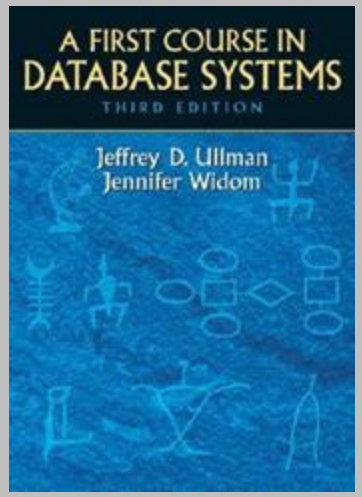
Reading:
“The world of database systems” chapter of the textbook.

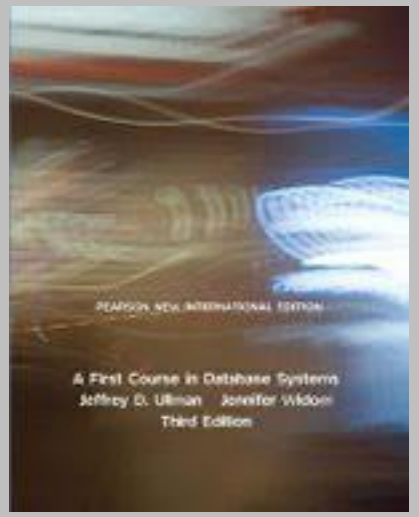
Weekly schedule

- Two hours of lecture recording.
  - Watch beforehand.
- One hour of lectorial
  - Participate discussions.
- Two hours of practical class.
- Read the main text and references (details on the next slide).

Main text

- J. Ullman and J. Widom. *A First Course in Database Systems*. 3rd. Ed. Pearson Education.
- E-book and hard copies available via the Library.
  - http://www1.rmit.edu.au/library → search for “A first course in database systems”





Communication

- Course Homepage:
  - http://canvas.rmit.edu.au → Database Concepts ISYS1057.
- First point of communication. Visit regularly!
  - Announcements
  - Discussion forum – Ask questions here! Read discussions here!
- Emails:
  - xiuzhen.zhang@rmit.edu.au (Xiuzhen (Jenny) Zhang, lecturer)

Communication …

What essential information is missing in the following email from a student?

Good Afternoon,

I would like to know if there is any inconvenience in changing my group session. I am currently after lecture session and I would like to assist to session on Thursday from 12:30 to 2:30.

George

Communication …

Essential Information to include in an email:

- Only use your RMIT student email
- Subject line: course code + description
- Content: your student No, your name, course code, and tute group (if needed)

Assessments

- Assessment 1: Assignment. 10 marks. Due in Week 5.
- Assessment 2: In-class exercise. 30 marks. Due in Week 8.
- Assessment 3: Assignment. 20 marks. Due in Week 11.
- Assessment 4: Project (milestone 1 due in Week 5). 40 marks. Due in Week 13.

The world of data

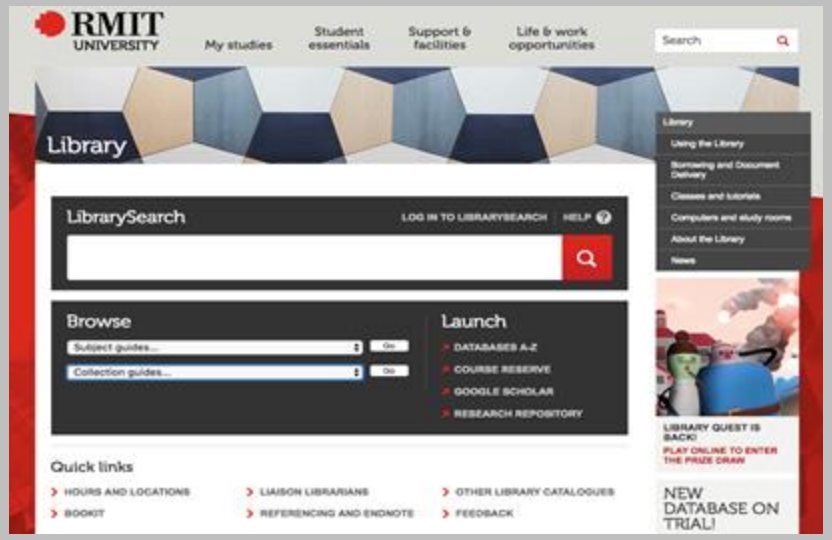
Data is generated every second, all over the world.

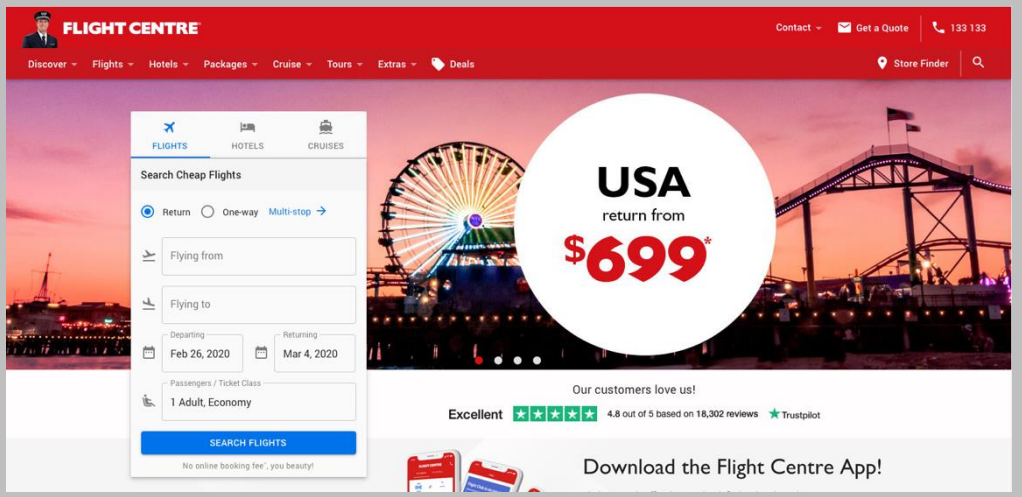
- The Web: homepages, Facebook Twitter,
- Mobiles and Sensors,
- Scientific and medical data: satellites,
- Enterprise information systems.

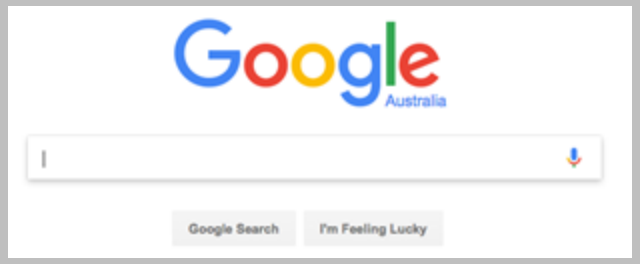
Data management

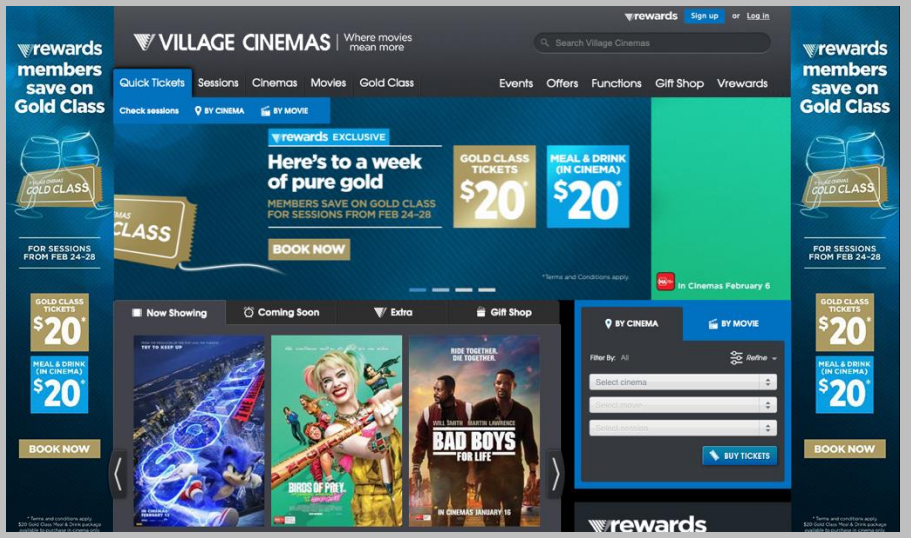
- Management of data, of different types, is behind almost everything you do on the Web.
  - Google: Web pages, images, maps
  - Amazon: books and other products
  - Facebook: homepage, images, friendship, posts
  - RMIT: library, student administration, course management, timetable.
  - Digital libraries: Lynda.com

Data Management …



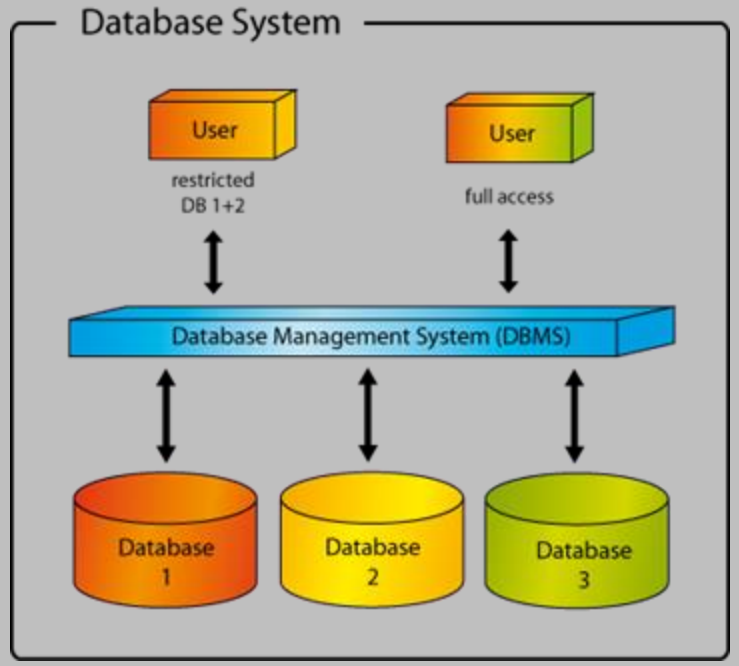






The database system

- A database system is for efficiently and effectively managing and using data.



Database system dimensions

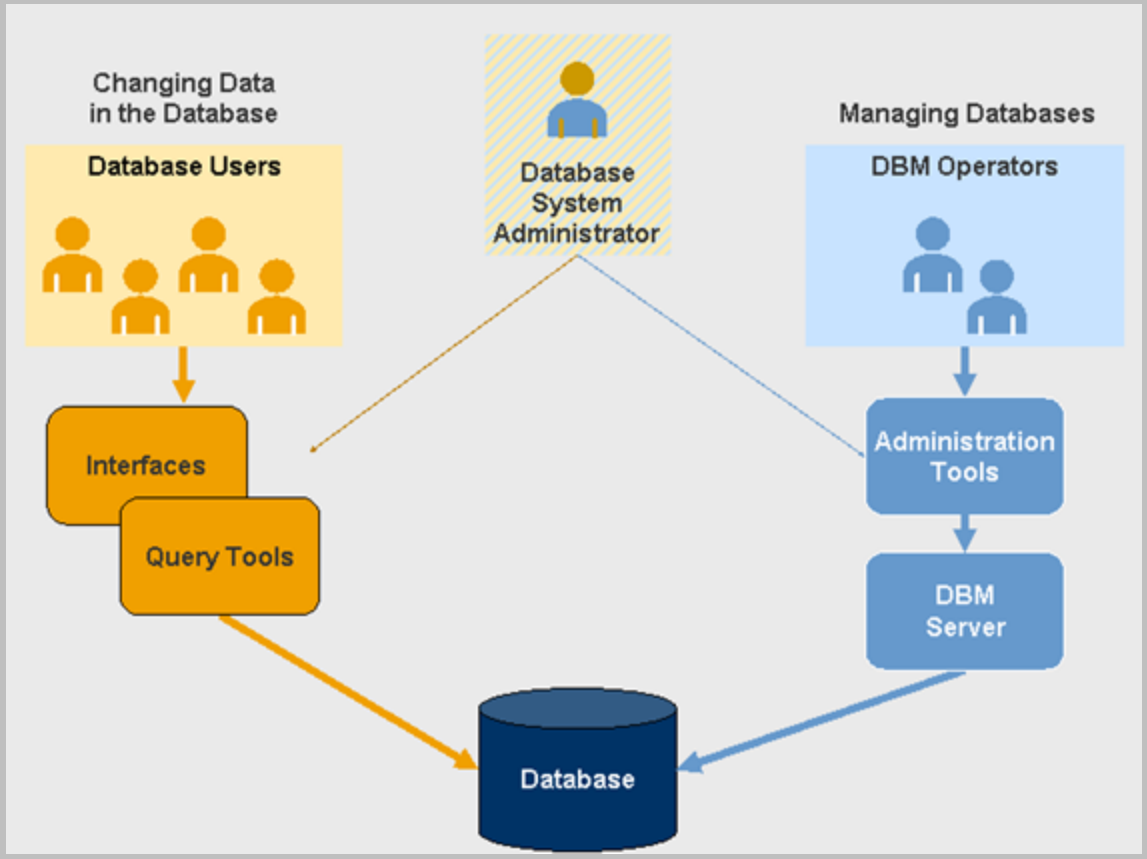
- Users, roles and security
- Databases
- A database programming language
- Transactions and concurrency control

Let us look at each topic in details.

Users

- There are multiple users for a database and they have different roles and privileges:
  - Database Administrator (DBA): full control of the database --- create and define database schemas, grant privileges to other users.
  - Database developers: manage/control his/her own database.
  - End users/data operators: only retrieve or update information in some parts of the database.

Users …


Changing Data
in the Database
Database Users
Database
System
Administrator
Managing Databases
DBM Operators
Interfaces
Query Tools
Administration
Tools
DBM
Server
Database

Databases

A database is a collection of data managed by a Database Management System (DBMS).

- Schema + Data files.

Databases: data models

- A data model is a language for describing data or information. The description generally consists of three parts:
  - Structure of the data, or data structure.
  - Operations on the data.
  - Constraints on the data
- Popular data models:
  - The relational model for structured data, including object-relational extensions.
  - The semi-structured data model, including XML and related standards.

Databases: Schema

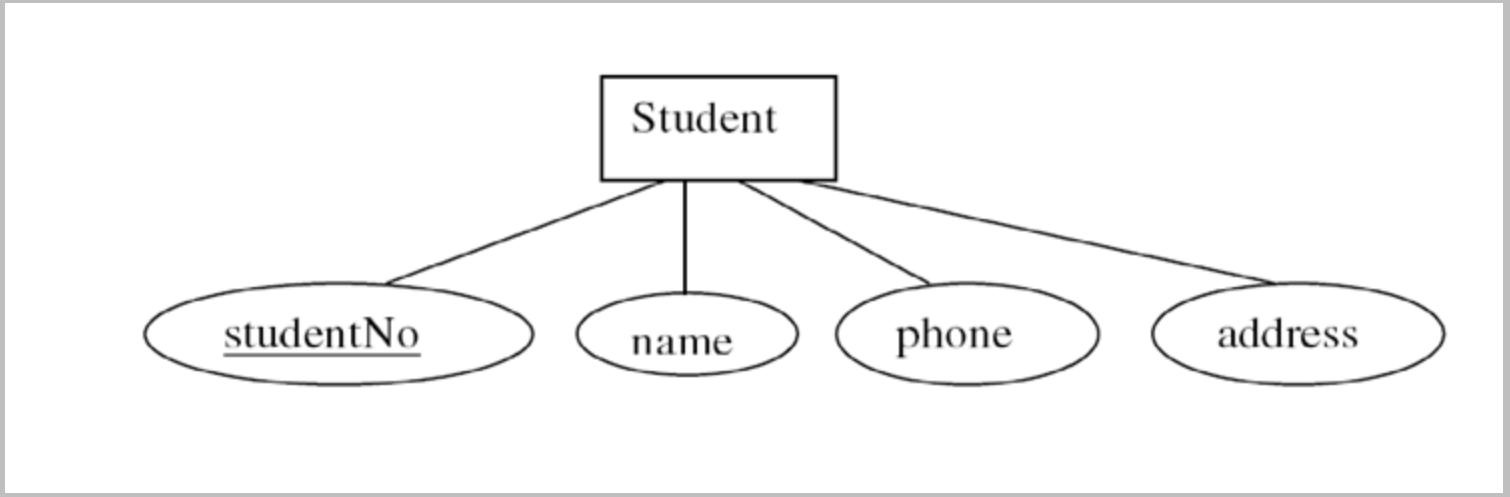
- A description of data in the database. The description can use data models at different levels.
  - The Entity Relationship (ER) model.
  - The Relational model.
- Constraints for data integrity.

The Schema -- example

- Real world:
  - Information: Student records should have student ID, name, contact phone number, and address.
  - Constraints:
    - Each student should be identified by Student No – Each student record must have the student No information.
    - No two students can have the same student ID --- Each student should have a unique student ID.

The Schema --- example

- ER model:



- Relation schema:

  Student(studentNo, name, phone, address)

Databases: data

- Data files on hard disks of computers.
  - The files are managed by a DBMS (on top of some file system).
- How to access the data?
  - Via the programming language (SQL) provided by the DBMS.
  - Data (and the meta-data) can be read, retrieved and updated.

Databases: Metadata

- Metadata (sometimes called data dictionary) is data describing the database, which may include:
  - Schema definition
  - Index – to speed up query processing
  - Data types
  - Constraints on data
  - User information and privileges

Database programming Language --- SQL

- SQL -- Structured Query Language --- is not only a query language:
  - Data Definition Language (DDL)
    - Define database schema
  - Data Query Language (DQL)
    - Query the database to extract information
  - Data Manipulation Language (DML)
    - Update the database – insertion/deletion/update.

SQL ...

- Database programming language SQL is declarative --- you tell the system what problem to solve.
  - Short programs.
  - System optimization.
- Most other programming languages are procedural --- you tell the system how to solve the problem.
  - Long programs.
  - User optimization.

SQL ...

Given a list of records for students (student ID, name, address...) enrolled in ISYS1055/1057, calculate how many students are there in total?

- In SQL:
  - select count(studentID) from Student;
- In C: read files, data structure, loop structure, output.
  - Example program on the next slide.

```c
#include <stdio.h>
#include <stdlib.h>

int main(void)
{
   FILE *fp;
   int i, isquared;

   /* open the file */
   fp = fopen("results.dat", "r");
   if (fp == NULL) {
      printf("I couldn't open results.dat for reading.\n");
      exit(0);
   }

   while (fscanf(fp, "%d,%d\n", &i, &isquared) == 2)
      printf("i: %d,  isquared: %d\n", i, isquared);

   /* close the file */
   fclose(fp);

   return 0;
}
```

Transactions and Concurrency Control

Many users operate the same database *concurrently*:

- Many students are booking tute classes on a timetable database.
- Many people book flight tickets online.

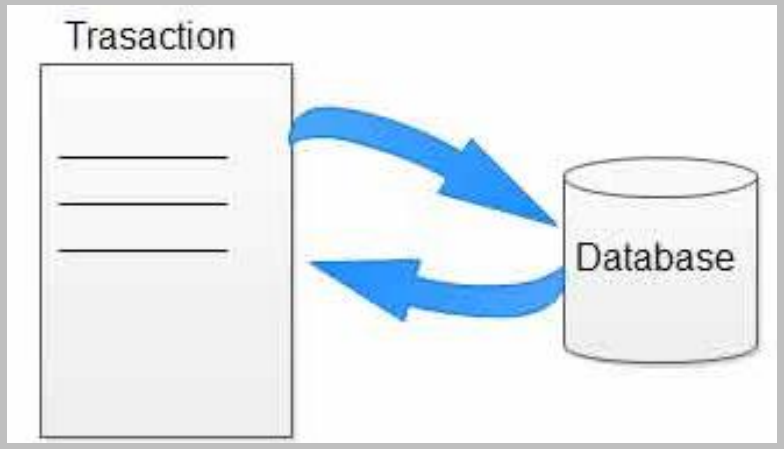
Transactions

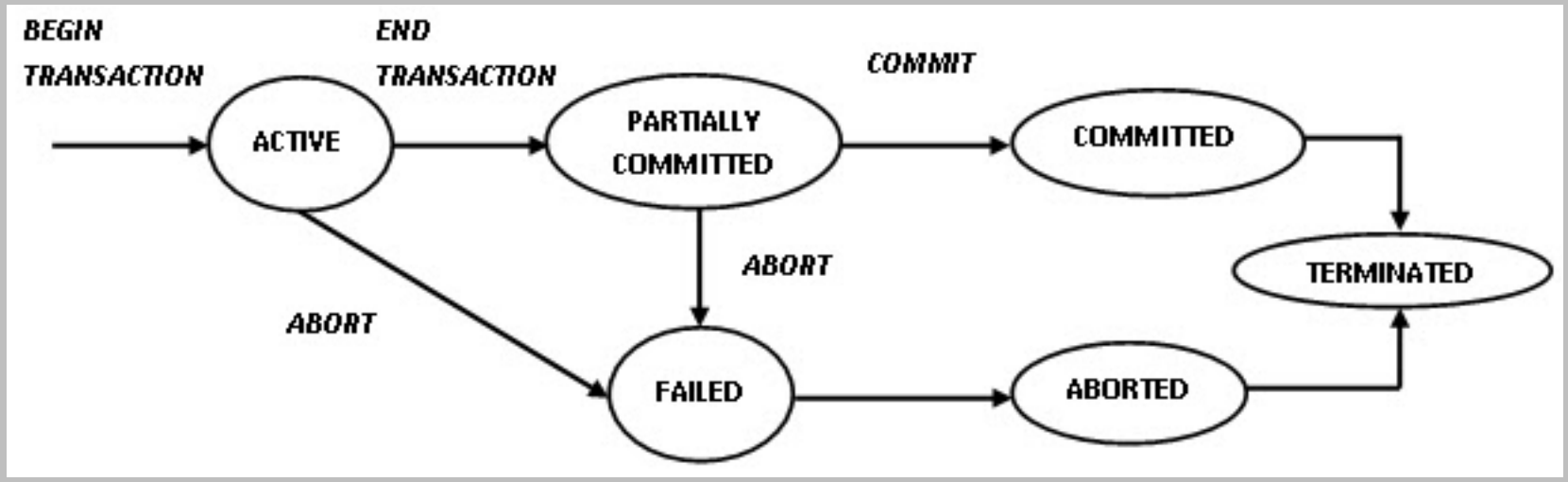
- Operations on databases (DDL/DML/DQL) are organized into transactions – an *atomic* unit that must finish in whole or nothing happens at all. No partial effect on the database.
  - A transaction comprises several operations of READ and WRITE on databases.
- The execution of transactions should be durable --- the effect of any completed transaction is permanent, even if system failure happens.
  - By way of logging.

The ACID properties of transactions

- Atomicity – the all-or-nothing execution of transactions
- Consistency – a transaction brings a database from one consistent state to another. In other words, meeting consistency constraints like balance can not be negative after a transaction.
- Isolation -- each transaction must appear to execute as if no other transactions are executing at the same time.
- Durability – effect of a transaction is must never be lost, once transaction complete.

Transaction processing





Concurrency control

- Concurrency control allows multiple users --- transactions from each user are executed in “isolation”.

Transactions: Commit and Rollback

- The “Commit” and “Rollback” command in SQL.
- Example:

```
select * from movies;                    // show content of the Movie table.
delete from movies where mvID =2;        // updates to the Movie table.
select * from movies;                    // show content of the Movie table.
rollback;                                // rollback the changes.
```

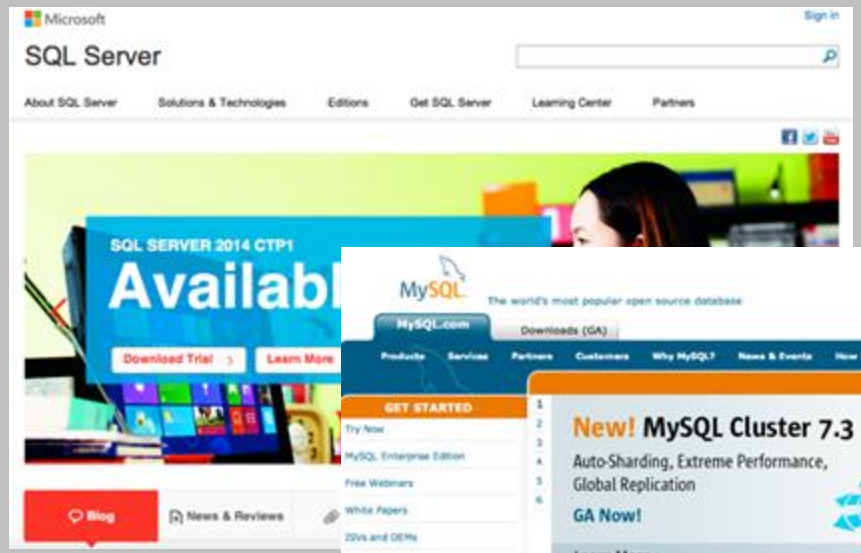
- Starting the SQL Developer application starts a new session. When the application is killed (not when you close a connection), if there are updates to the database, you are asked whether to “commit the changes” or “rollback changes”.

The DBMS

A DBMS typically has three parts:

- Storage management: how secondary storage (hard disks) is used effectively to store and access data.
- Query processing: how to execute (SQL) queries
- Transaction management: how to support transactions with the ACID properties.

DBMS Examples



What will you learn in this course?

- Basic database concepts
- Design of databases
  - Design of the ER diagram
  - Design of the relational database schema
- Database programming
  - SQL programming

Software

- SQLite studio. Download and install on your machine.

  https://sqlitestudio.pl

- Tutors will help you set up these software in your first practical class in Week 2.

Summary

- Databases and database schema
- DBMS vs. file systems
  - Components of a DBMS
- Database systems and relational database systems
  - Sometimes databases and database systems are used interchangeably.
- The database programming language
  - DDL, DQL, DML
- Transactions and the ACID properties