

CSC 343

Introduction to Databases

Diane Horton
University of Toronto
Winter 2011

Our first hour or so

- introduction to databases
- a few admin details
- introduction to the relational model

Databases and DBMSs

- Databases are everywhere, often behind the scenes.
- Database: large, integrated collection of data managed by a DBMS.
- DBMS: Database Management System.
- Interesting theory, highly complex software.

3

Doing it all with files

- You can manage a large collection of data with files.
- In fact, the first commercial databases evolved in this way.
- How would this work? What problems would arise?

4

What do we need from a DBMS?

- Ability to specify the logical structure of the data.
- Ability to query or modify the data.
- Separation of the logical structure from the implementation details.

5

... more needs

- Speed. Even when the database is huge, eg
 - Facebook 2010: 700 terabytes (TB = 10^{12})
 - Yahoo: 2 petabytes, (PB = 10^{15})
 - National Energy Research Scientific Computing Center 2009: 3.9 petabytes
- And even when there are many operations on the data, eg
 - Twitter: 1 billion queries per day
 - Yahoo: 24 billion events per day

6

... more needs

- Isolation: Many users at once, with no interference.
- Atomicity: Actions on the data happen completely or not at all.
- Durability: Data persists even if there are failures (e.g., power failure, system crash), or intentional attacks.

7

Instances and schemas

- Schema: definition of structure in the database
- Instance: particular data in the database at a point in time.
- Instances change constantly; schemas rarely, if ever.
- Note: conventional databases only store the current version of the data. Databases that record the history are called *temporal* databases.

8

Data models

- Data model: a notation for describing data, including
 - the structure of the data
 - constraints on the content of the data
 - operations on the data
- Some specific data models:
 - old: network data model & hierarchical data model
 - relational data model -- the most widely used
 - semistructured model (later in the term)

9

The relational data model

- Main concept is a “relation.”
- Based on relations in math. Most easily thought of as tables of rows and columns.

Name	Home Field	Coach	Teams
Rangers	Runnymede CI	Tarvo Sinervo	
Ducks	Humber Public	Maeve Mahar	
Choppers	High Park	Tom Cole	

Games	Home team	Away team	Home goals	Away goals
	Rangers	Ducks	3	0
	Ducks	Choppers	1	1
	Rangers	Choppers	4	2
	Choppers	Ducks	0	5

10

Overall architecture

- The DBMS sits between the database and the users/application program
- Within the DBMS there are layers for, eg:
 - planning and optimizing complex queries
 - implementing the fundamental operations
 - accessing the files that store the data and indices
 - management of buffers
 - management of disk space

11

Why study DBs?

- Some very cool concepts and techniques are used to satisfy those needs we identified.
- They span computer science, including OS, languages, theory, AI, multimedia, logic.
- Databases are increasingly important, as
 - we shift from computation to information
 - data is increasing in volume and diversity.

12

And for your career

- If you are heading for industry:
 - Database professionals are in demand and well paid
- If you want to do research:
 - There are many problems still open
- Further studies in databases at DCS:
 - csc443: under the hood
 - csc2508 (grad course, taken with permission): Advanced DBMSs

13

Admin Stuff

Important: Read the course info sheet

- Contact:
 - website and discussion board are required reading
 - your questions: to the discussion board
 - personal matters: email or visit me
- Office hours:
 - Please come: I love talking with students

14

... more

- Textbook plus Gradiance
- Course prerequisites:
 - (1) CSC165/240 or MAT137/157 and
 - (2) CSC207.See me immediately if you don't have them.
(Engineers are already covered.)
- Assignments:
 - You may work with a partner
 - You have 3 grace day points; no other lates!
- Academic offences: don't!
(Read info sheet for specifics.)

15

Course Marking Scheme

Work	Weight	Comment
3 Assignments	30%	10% each
2 Tests	20%	10% each
5 Gradiance exercises	10%	2% each
Final exam	40%	You must get $\geq 40\%$ to pass the course

16