CS 245: Database System Principles

Notes 01: Introduction

Hector Garcia-Molina

Isn't Implementing a Database System Simple?

Relations ⇨ Statements ⇨ Results

Introducing the

MEGATRON 3000

Database Management System

- The latest from Megatron Labs
- Incorporates latest relational technology
- UNIX compatible

Megatron 3000 Implementation Details

! First sign non-disclosure agreement !

Megatron 3000 Implementation Details

- Relations stored in files (ASCII)
  e.g., relation R is in /usr/db/R

```
Smith # 123 # CS
Jones # 522 # EE
  .
  .
```


Megatron 3000 Implementation Details

- Directory file (ASCII) in /usr/db/directory

```
R1 # A # INT # B # STR …
R2 # C # STR # A # INT …
  .
  .
```

Megatron 3000
Sample Sessions

```
% MEGATRON3000
   Welcome to MEGATRON 3000!
&
  .
  .
& quit
%
```


Megatron 3000
Sample Sessions

```
& select *
  from R #

    Relation R
  A       B     C
  SMITH   123   CS

&
```


Megatron 3000
Sample Sessions

```
& select A,B
  from R,S
  where R.A = S.A and S.C > 100 #

   A    B
  123  CAR
  522  CAT

&
```


Megatron 3000
Sample Sessions

```
& select *
  from R | LPR #
&
```

Result sent to LPR (printer).

Megatron 3000
Sample Sessions

```
& select *
  from R
  where R.A < 100 | T #
&
```

New relation T created.

Megatron 3000

- To execute "**`select * from R where`** ***`condition`***":
  - (1) Read dictionary to get R attributes
  - (2) Read R file, for each line:
    - (a) Check condition
    - (b) If OK, display

Megatron 3000

- To execute "**`select * from R where condition | T`**":
  - (1) Process select as before
  - (2) Write results to new file T
  - (3) Append new line to dictionary

Megatron 3000

- To execute "**`select A,B from R,S where condition`**":
  - (1) Read dictionary to get R,S attributes
  - (2) Read R file, for each line:
    - (a) Read S file, for each line:
      - (i) Create join tuple
      - (ii) Check condition
      - (iii) Display if OK

What's wrong with the Megatron 3000 DBMS?

What's wrong with the Megatron 3000 DBMS?

- Tuple layout on disk

e.g., - Change string from 'Cat' to 'Cats' and we have to rewrite file
- ASCII storage is expensive
- Deletions are expensive

What's wrong with the Megatron 3000 DBMS?

- Search expensive; no indexes

e.g., - Cannot find tuple with given key quickly
- Always have to read full relation

What's wrong with the Megatron 3000 DBMS?

- Brute force query processing

e.g.,
```
select *
from R,S
where R.A = S.A and S.B > 1000
```
- Do select first?
- More efficient join?

What's wrong with the Megatron 3000 DBMS?

- No buffer manager

e.g., Need caching

What's wrong with the Megatron 3000 DBMS?

- No concurrency control

What's wrong with the Megatron 3000 DBMS?

- No reliability

e.g., - Can lose data
- Can leave operations half done

What's wrong with the Megatron 3000 DBMS?

- No security

e.g., - File system insecure
- File system security is coarse

What's wrong with the Megatron 3000 DBMS?

- No application program interface (API)

e.g., How can a payroll program get at the data?

What's wrong with the Megatron 3000 DBMS?

- Cannot interact with other DBMSs.

What's wrong with the Megatron 3000 DBMS?

- Poor dictionary facilities

What's wrong with the Megatron 3000 DBMS?

- No GUI

What's wrong with the Megatron 3000 DBMS?

- Lousy salesman!!

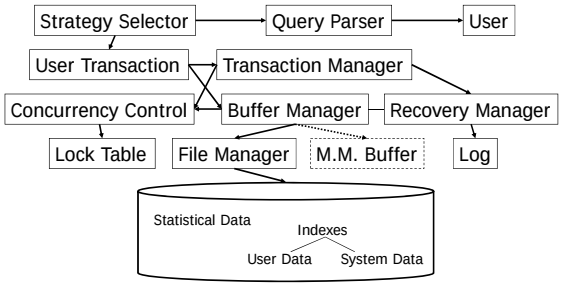
Course Overview

- File & System Structure
  Records in blocks, dictionary, buffer management,...
- Indexing & Hashing
  B-Trees, hashing,...
- Query Processing
  Query costs, join strategies,...
- Crash Recovery
  Failures, stable storage,...

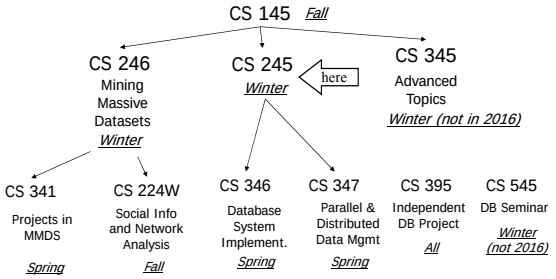
Course Overview

- Concurrency Control
  Correctness, locks,...
- Transaction Processing
  Logs, deadlocks,...
- Security & Integrity
  Authorization, encryption,...
- Distributed Databases
  Interoperation, distributed recovery,...

System Structure



Stanford Data Management Courses



If you did not take CS145:

- You can still take this class
- Read in textbook:
  - Chapter 2 (Relational Model) through Section 2.4
  - Chapter 6 (SQL) through Section 6.2

Some Terms

- Database system
- Transaction processing system
- File access system
- Information retrieval system

Mechanics

- http://www.stanford.edu/class/cs245/

Staff

- INSTRUCTOR: Hector Garcia-Molina Office: Gates 434 Email: hector@cs.stanford.edu
- Office Hours: Tuesdays, Thursdays 11am to 11:50am
- (Try to make an appointment with Marianne Siroker to ensure I can see you.)
- TEACHING ASSISTANTS (Tentative):
  - Aditya Bhandari adityasb@stanford.edu
  - Seo Jin Park seojin@cs.stanford.edu
  - Sudarshan Srinivasan sudarsh2@stanford.edu
  - Stephanie Tsai stsai612@stanford.edu
  - Stephen Yang syang0@stanford.edu
- SECRETARY: Marianne Siroker Office: Gates 436; Email: siroker@cs.stanford.edu
- Phone: 650-723-0872

Details

- LECTURES: Tuesday, Thursday 12:00pm to 1:20pm, NVidia Auditorium
- TEXTBOOK: Garcia-Molina, Ullman, Widom
  "DATABASE SYSTEMS, THE COMPLETE BOOK"
  [Second edition]
- ASSIGNMENTS: Six written homework assignments.
  Two (or three) MySQL "code analysis" homeworks.
  Also readings in Textbook.
-
- GRADING: Homeworks: 20%, Midterm: 30%, Final: 50%.
- WEB SITE: All handouts & assignments will be posted on our Web site at http://www.stanford.edu/class/cs245
- Please check it periodically for last minute announcements.

Tentative Syllabus 2016

DATE	CHAPTER [2nd Ed]	TOPIC
Tuesday January 5		Introduction
Thursday January 7	Ch. 11 [13]	Hardware
Tuesday January 12	Ch. 12 [13]	File and System Structure
Thursday January 14	Ch. 12 [13]	File and System Structure
Tuesday January 19	Ch. 13 [14]	Indexing and Hashing
Thursday January 21	Ch. 13 [14]	Indexing and Hashing
Tuesday January 26	Ch. 14 [14]	Indexing and Hashing
Thursday January 28	Ch. 15 [15]	Query Processing
Tuesday February 2	Ch. 15 [16]	Query Processing
Thursday February 4	Ch. 16 [16]	Query Processing
Tuesday February 9		**MIDTERM (in class)**
Thursday February 11	Ch. 17 [17]	Crash Recovery
Tuesday February 16	Ch. 17 [17]	Crash Recovery
Thursday February 18	Ch. 18 [18]	Concurrency Control
Tuesday February 23	Ch. 18 [18]	Concurrency Control
Thursday February 25	Ch. 18 [18]	Concurrency Control
Tuesday March 1	Ch. 19 [19]	Transaction Processing
Thursday March 3	Ch. 19 [19]	Transaction Processing
Tuesday March 8	Ch. 20 [21,22]	Information Integration
Thursday March 10		Review
Wednesday March 16,	12:15-3:00pm	**FINAL EXAM**

Read: Chapters 11-20 [13-22 in Second Edition]

- Except following optional material
  [brackets for Second Edition Complete Book]:
  - Sections 11.7.4, 11.7.5 [13.4.8, 13.4.9]
  - Sections 14.3.6, 14.3.7, 14.3.8 [14.6.6, 14.6.7, 14.6.8]
  - Sections 14.4.2, 14.4.3, 14.4.4 [14.7.2, 14.7.3, 14.7.4]
  - Sections 15.7, 15.8, 15.9 [15.7, 15.8]
  - Sections 16.6, 16.7 [16.6, 16.7]
  - In Chapters 15, 16 [15, 16]: material on duplicate elimination operator, grouping, aggregation operators
  - Section 18.8 [18.8]
  - Sections 19.2 19.4, 19.5, 19.6 [none, i.e., read all Ch 19]
  - [In the Second Edition, skip all of Chapter 20, and Sections 21.5, 21.6, 21.7, 22.2 through 22.7]

Next time:

- Hardware
- Read chapter 11 [13.1 through 13.4]