

CS 245: Database System Principles

Notes 01: Introduction

Steven Whang

Isn't Implementing a Database System Simple?

Relations \Rightarrow Statements \Rightarrow 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!!

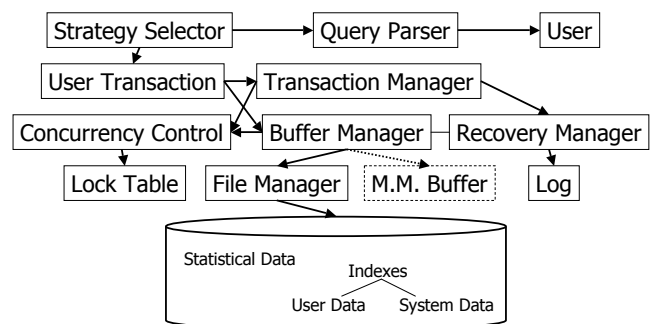
Course Overview

- **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,...

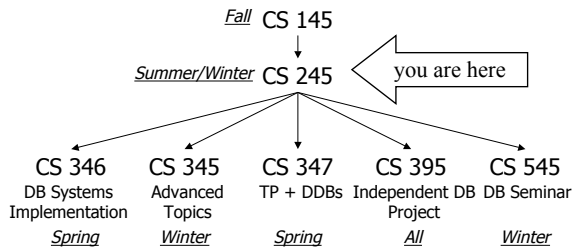
Course Overview

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

System Structure



Stanford Database Courses



Some Terms

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

Mechanics

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

Staff

- INSTRUCTOR: Steven Whang
 - Office: Gates 432
 - Email: swhang@cs.stanford.edu
 - Office Hours: Mondays, Wednesdays 2pm to 3pm
 - Phone: 650-796-6006
- TEACHING ASSISTANT: Shipra Agrawal
 - Office: Terman 329
 - Staff Email: cs245-sum0809-staff@lists.stanford.edu
 - Office Hours: Tuesdays, Thursdays 2pm to 3pm
 - Phone: 650-704-3751
- SECRETARY: Marianne Siroker
 - Office: Gates 436
 - Email: siroker@cs.stanford.edu
 - Phone: 650-723-0872

Details

- LECTURES: Monday, Wednesday 11:00am to 12:50pm, Skilling Aud
- TEXTBOOK: Garcia-Molina, Ullman, Widom "DATABASE SYSTEMS, THE COMPLETE BOOK" [First or Second edition]
- ASSIGNMENTS: Five homework assignments through Gradiance. Two written homeworks. No programming. Also readings in Textbook.
- GRADING: Gradiance Homeworks: 20%, Additional Written Homeworks: 10%, Midterm: 30%, Final: 40%.
- 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.

Gradiance System

- Go to <http://www.gradiance.com/pearson> and create a new account or use your previous CS145 account
- Use the following class token to subscribe to the class: E5E12A4B

Tentative Syllabus 2009

DATE	CHAPTER [2nd Ed]	TOPIC
• Wednesday June 24	Ch. 11 [13]	Introduction / Hardware
• Monday June 29	Ch. 12 [13]	File and System Structure
• Wednesday July 1	Ch. 12 [13]	File and System Structure
• Monday July 6	Ch. 13 [14]	Indexing and Hashing
• Wednesday July 8	Ch. 13 [14]	Indexing and Hashing
• Monday July 13	Ch. 15 [15]	Query Processing
• Wednesday July 15	Ch. 16 [16]	Query Processing
• Monday July 20		MIDTERM
• Wednesday July 22	Ch. 17 [17]	Crash Recovery
• Monday July 27	Ch. 17 [17]	Crash Recovery
• Wednesday July 29	Ch. 18 [18]	Concurrency Control
• Monday August 3	Ch. 18 [18]	Concurrency Control
• Wednesday August 5	Ch. 19 [19]	Transaction Processing
• Monday August 10	Ch. 20 [21,22]	Information Integration
• Wednesday August 12		Review
• Friday August 14, 8:30-11:30am		FINAL EXAM

CS 245

Notes 1

37

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]

CS 245

Notes 1

38