

CS 245: Database System Principles

Notes 01: Introduction

Hector Garcia-Molina

CS 245

Notes 1

1

Isn't Implementing a Database System Simple?

Relations \Rightarrow Statements \Rightarrow Results

CS 245

Notes 1

2

Introducing the MEGATRON 3000 Database Management System

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

CS 245

Notes 1

3

Megatron 3000 Implementation Details

! First sign non-disclosure agreement !

CS 245

Notes 1

4

Megatron 3000 Implementation Details

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

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

CS 245

Notes 1

5

Megatron 3000 Implementation Details

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

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

CS 245

Notes 1

6

Megatron 3000 Sample Sessions

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

CS 245

Notes 1

7

Megatron 3000 Sample Sessions

```
& select *
from R #

Relation R
A      B      C
SMITH  123    CS
&
```

CS 245

Notes 1

8

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
&
```

CS 245

Notes 1

9

Megatron 3000 Sample Sessions

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

Result sent to LPR (printer).

CS 245

Notes 1

10

Megatron 3000 Sample Sessions

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

New relation T created.

CS 245

Notes 1

11

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

CS 245

Notes 1

12

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

CS 245

Notes 1

13

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

CS 245

Notes 1

14

What's wrong with the Megatron 3000 DBMS?

CS 245

Notes 1

15

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

CS 245

Notes 1

16

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

CS 245

Notes 1

17

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?

CS 245

Notes 1

18

What's wrong with the Megatron 3000 DBMS?

- No buffer manager
- e.g., Need caching

CS 245

Notes 1

19

What's wrong with the Megatron 3000 DBMS?

- No concurrency control

CS 245

Notes 1

20

What's wrong with the Megatron 3000 DBMS?

- No reliability
- e.g., - Can lose data
- Can leave operations half done

CS 245

Notes 1

21

What's wrong with the Megatron 3000 DBMS?

- No security
- e.g., - File system insecure
- File system security is coarse

CS 245

Notes 1

22

What's wrong with the Megatron 3000 DBMS?

- No application program interface (API)
- e.g., How can a payroll program get at the data?

CS 245

Notes 1

23

What's wrong with the Megatron 3000 DBMS?

- Cannot interact with other DBMSs.

CS 245

Notes 1

24

What's wrong with the Megatron 3000 DBMS?

- Poor dictionary facilities

CS 245

Notes 1

25

What's wrong with the Megatron 3000 DBMS?

- No GUI

CS 245

Notes 1

26

What's wrong with the Megatron 3000 DBMS?

- Lousy salesman!!

CS 245

Notes 1

27

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,...

CS 245

Notes 1

28

Course Overview

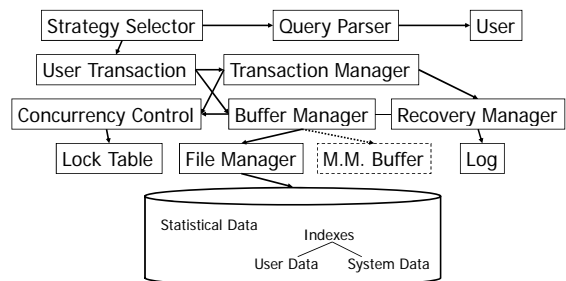
- Concurrency Control
Correctness, locks,...
- Transaction Processing
Logs, deadlocks,...
- Security & Integrity
Authorization, encryption,...
- Distributed Databases
Interoperation, distributed recovery,...

CS 245

Notes 1

29

System Structure

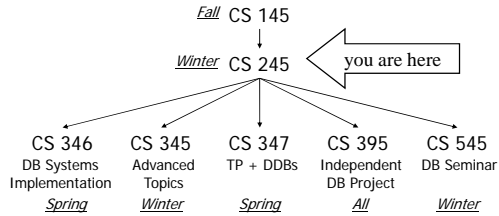


CS 245

Notes 1

30

Stanford Database Courses



CS 245

Notes 1

31

Some Terms

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

CS 245

Notes 1

32

Mechanics

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

CS 245

Notes 1

33

Staff

- INSTRUCTOR: Hector Garcia-Molina Office: Gates 434 Email: hector@cs.stanford.edu
- Office Hours: Tuesdays, Thursdays 11am to 11:55am
- (Try to make an appointment with Marianne Siroker to ensure I can see you.)
- TEACHING ASSISTANTS:
 - Steven Whang
 - Petros Venetis
 - Eldar Sadikov
- SECRETARY: Marianne Siroker Office: Gates 436; Email: siroker@cs.stanford.edu
- Phone: 650-723-0872

See course web page for office location & hours.

CS 245

Notes 1

34

Details

- LECTURES: Tuesday, Thursday 12:50pm to 2:05pm, Gates B03
- TEXTBOOK: Garcia-Molina, Ullman, Widom "DATABASE SYSTEMS, THE COMPLETE BOOK" [First or Second edition]
- ASSIGNMENTS: Seven written homework assignments. No programming. 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.

CS 245

Notes 1

35

Tentative Syllabus 2008

DATE	CHAPTER (2nd Ed)	TOPIC
• Tuesday January 6		Introduction
• Thursday January 8	Ch. 11 [13]	Hardware
• Tuesday January 13	Ch. 12 [13]	File and System Structure
• Thursday January 15	Ch. 12 [13]	File and System Structure
• Tuesday January 20	Ch. 13 [14]	Indexing and Hashing
• Thursday January 22	Ch. 13 [14]	Indexing and Hashing
• Tuesday January 27	Ch. 14 [14]	Indexing and Hashing
• Thursday January 29	Ch. 15 [15]	Query Processing
• Tuesday February 3	Ch. 15 [16]	Query Processing
• Thursday February 5	Ch. 16 [16]	Query Processing
• Tuesday February 10		MIDTERM
• Thursday February 12	Ch. 17 [17]	Crash Recovery
• Tuesday February 17	Ch. 17 [17]	Crash Recovery
• Thursday February 19	Ch. 18 [18]	Concurrency Control
• Tuesday February 24	Ch. 18 [18]	Concurrency Control
• Thursday February 26	Ch. 18 [18]	Concurrency Control
• Tuesday March 3	Ch. 19 [19]	Transaction Processing
• Thursday March 5	Ch. 19 [19]	Transaction Processing
• Tuesday March 10	Ch. 20 [21,22]	Information Integration
• Thursday March 12		Review
• Thursday March 19, 7:00-10:00pm		FINAL EXAM

CS 245

Notes 1

36

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

37

Next time:

- Hardware
- Read chapter 11 [13.1 through 13.4]

CS 245

Notes 1

38