

CS245: Database System Principles

Summer 2001

Instructor: Chris Olston

TA: Masood Siddiqi

Mondays, Wednesdays 1:00-2:50
McCullough 115

CS 245

Notes 1

1

CS 245: Database System Principles

Notes 01: Introduction

Slides by: Hector Garcia-Molina
(Some modifications by Chris Olston)

CS 245

Notes 1

2

Isn't Implementing a Database System Simple?

Relations \Rightarrow Statements \Rightarrow Results

CS 245

Notes 1

3

Introducing the MEGATRON 3000 Database Management System

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

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

...Done

Now we can all go home, right?

CS 245

Notes 1

15

What's wrong with the Megatron 3000 DBMS?

CS 245

Notes 1

16

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

17

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

18

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

19

What's wrong with the Megatron 3000 DBMS?

- No buffer manager

e.g., Need caching

CS 245

Notes 1

20

What's wrong with the Megatron 3000 DBMS?

- No concurrency control

CS 245

Notes 1

21

What's wrong with the Megatron 3000 DBMS?

- No reliability

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

CS 245

Notes 1

22

What's wrong with the Megatron 3000 DBMS?

- No security

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

CS 245

Notes 1

23

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

24

What's wrong with the Megatron 3000 DBMS?

- Cannot interact with other DBMSs.

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 (1): Efficient Data Management

- **Efficient Storage**
 - File & System Structure
Records in blocks, buffer management,...
- **Efficient Access**
 - Indexing & Hashing
B-Trees, hashing,...
- **Efficient Processing**
 - Query Optimization and Execution
Query costs, selection ordering, join strategies,...

CS 245

Notes 1

28

Course Overview (2): Transaction Management

- **ACID Properties:**
 - Atomicity, Consistency, Isolation, Durability
- **Logging**
 - Commit / abort,...
- **Crash Recovery**
 - Failures, stable storage,...
- **Concurrency Control**
 - Correctness, locks,...

CS 245

Notes 1

29

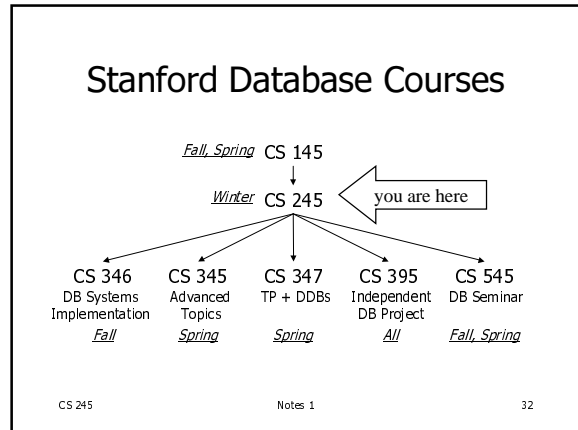
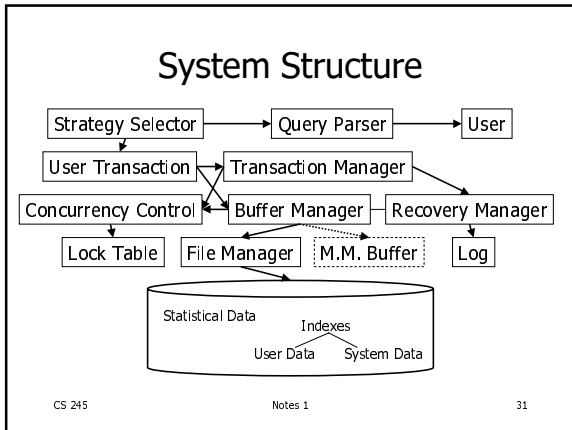
Course Overview (3)

- **Security & Integrity**
 - Authorization, encryption,...
- **Information Integration**
 - Mediators, data mining,...

CS 245

Notes 1

30



- ### What you (should have) learned in CS145
- Data modeling
 - SQL
 - Database application programming
 - **In short: how to use a database system**
- CS 245 Notes 1 33

- ### What you will learn in this course (CS245)
- What lies "under the hood" of a database system
- CS 245 Notes 1 34

- ### Who cares?
- Database technology is prevalent
 - Business, scientific, e-commerce, etc...
 - Implement and maintain database systems
 - Database administration (DBA)
 - Database technology useful in many settings
- CS 245 Notes 1 35

- ### Some Terms
- Database management system (DBMS)
 - Transaction processing (TP) system
 - File access system
 - Information retrieval system
- CS 245 Notes 1 36

Mechanics

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

CS 245

Notes 1

37

Staff

- INSTRUCTOR: Chris Olston Office: Gates 412 Email: olston@cs.stanford.edu
• Office Hours: Mondays and Wednesdays 4-5pm
- TEACHING ASSISTANT: Masood Siddiqi, masood@cs.stanford.edu
• Office Hours: Tuesdays and Thursdays 4:15-6:15pm
- 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

38

Details

- LECTURES: Monday, Wednesday 1:00-2:50, McCullough 115
- TEXTBOOK: Garcia-Molina, Ullman, Widom;
"DATABASE SYSTEM IMPLEMENTATION"
- 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

39

Mailing Lists & Newsgroup

- Subscribe to: cs245@lists.stanford.edu
– Instructions on Web page
- To mail the class staff (myself & TA)
– Mail to: cs245-staff@lists.stanford.edu
- Class newsgroup
– su.class.cs245
– For discussion among students

CS 245

Notes 1

40

Tentative Syllabus

DATE	CHAPTER	TOPIC
Wednesday June 27		Introduction
Monday July 2	Ch. 2	Hardware
Wednesday July 4		NO CLASS
Monday July 9	Ch. 3	File and System Structure
Wednesday July 11	Ch. 4	Indexing and Hashing
Monday July 16	Ch. 6	Query Processing
Wednesday July 18	Ch. 6	Query Processing
Monday July 23	Ch. 7	Query Optimization
Wednesday July 25		Midterm
Monday July 30	Ch. 8	Crash Recovery
Wednesday August 1	Ch. 9	Concurrency Control
Monday August 6	Ch. 9	Concurrency Control
Wednesday August 8	Ch. 10	Transaction Processing
Monday August 13	Ch. 11	Information Integration
Wednesday August 15		Review

CS 245

Notes 1

41

Read: All Chapters

- Except the following optional material:
 - Sections 2.6.4, 2.6.5
 - Sections 5.3.3, 5.3.4, 5.3.5, 5.3.6
 - Sections 5.4.2, 5.4.3, 5.4.4
 - Sections 6.8.3, 6.9, 6.10.2, 6.10.3, 6.10.4
 - Sections 7.6, 7.7.2, 7.7.3, 7.7.4, 7.7.5, 7.7.6
 - In Chapters 6, 7: material on duplicate elimination operator, grouping, aggregation operators
 - Section 8.5
 - Section 9.5, 9.6, 9.7, 9.8
 - Sections 10.4, 10.5, 10.6

CS 245

Notes 1

42

Next time:

- Hardware
- Read chapters 1 and 2