

CS 525: Advanced Database Organization



01: Introduction

Boris Glavic

Slides: adapted from a [course](#) taught by
[Hector Garcia-Molina](#), Stanford InfoLab

CS 525



Notes 1 - Introduction

1

IIT College of
Science and Letters
ILLINOIS INSTITUTE OF TECHNOLOGY

Advanced Database Organization?

- =Database Implementation
- =How to implement a database system
- ... and have fun doing it ;-)

CS 525



Notes 1 - Introduction

2

IIT College of
Science and Letters
ILLINOIS INSTITUTE OF TECHNOLOGY

Isn't Implementing a Database System Simple?

Relations \implies Statements \implies Results

CS 525



Notes 1 - Introduction

3

IIT College of
Science and Letters
ILLINOIS INSTITUTE OF TECHNOLOGY

Introducing the MEGATRON 3000 Database Management System

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

CS 525



Notes 1 - Introduction

4

IIT College of
Science and Letters
ILLINOIS INSTITUTE OF TECHNOLOGY

Megatron 3000 Implementation Details

! First sign non-disclosure agreement !

CS 525



Notes 1 - Introduction

5

IIT College of
Science and Letters
ILLINOIS INSTITUTE OF TECHNOLOGY

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 525



Notes 1 - Introduction

6

IIT College of
Science and Letters
ILLINOIS INSTITUTE OF TECHNOLOGY

Megatron 3000 Implementation Details

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

```

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

CS 525
Notes 1 - Introduction
7
IIT College of
Science and Letters
ILLINOIS INSTITUTE OF TECHNOLOGY

Megatron 3000 Sample Sessions

```

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

CS 525
Notes 1 - Introduction
8
IIT College of
Science and Letters
ILLINOIS INSTITUTE OF TECHNOLOGY

Megatron 3000 Sample Sessions

```

& select *
from R #

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

CS 525
Notes 1 - Introduction
9
IIT College of
Science and Letters
ILLINOIS INSTITUTE OF TECHNOLOGY

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 525
Notes 1 - Introduction
10
IIT College of
Science and Letters
ILLINOIS INSTITUTE OF TECHNOLOGY

Megatron 3000 Sample Sessions

```

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

Result sent to LPR (printer).

CS 525
Notes 1 - Introduction
11
IIT College of
Science and Letters
ILLINOIS INSTITUTE OF TECHNOLOGY

Megatron 3000 Sample Sessions

```

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

New relation T created.

CS 525
Notes 1 - Introduction
12
IIT College of
Science and Letters
ILLINOIS INSTITUTE OF TECHNOLOGY

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 525



Notes 1 - Introduction

13

IIT College of
Science and Letters
ILLINOIS INSTITUTE OF TECHNOLOGY

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 525



Notes 1 - Introduction

14

IIT College of
Science and Letters
ILLINOIS INSTITUTE OF TECHNOLOGY

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 525



Notes 1 - Introduction

15

IIT College of
Science and Letters
ILLINOIS INSTITUTE OF TECHNOLOGY

What's wrong with the Megatron 3000 DBMS?

CS 525



Notes 1 - Introduction

16

IIT College of
Science and Letters
ILLINOIS INSTITUTE OF TECHNOLOGY

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 525



Notes 1 - Introduction

17

IIT College of
Science and Letters
ILLINOIS INSTITUTE OF TECHNOLOGY

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 525



Notes 1 - Introduction

18

IIT College of
Science and Letters
ILLINOIS INSTITUTE OF TECHNOLOGY

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 525



Notes 1 - Introduction

19

IIT College of
Science and Letters
ILLINOIS INSTITUTE OF TECHNOLOGY

What's wrong with the Megatron 3000 DBMS?

- No buffer manager

e.g., Need caching

CS 525



Notes 1 - Introduction

20

IIT College of
Science and Letters
ILLINOIS INSTITUTE OF TECHNOLOGY

What's wrong with the Megatron 3000 DBMS?

- No concurrency control

CS 525



Notes 1 - Introduction

21

IIT College of
Science and Letters
ILLINOIS INSTITUTE OF TECHNOLOGY

What's wrong with the Megatron 3000 DBMS?

- No reliability

e.g., - Can lose data

- Can leave operations half done

CS 525



Notes 1 - Introduction

22

IIT College of
Science and Letters
ILLINOIS INSTITUTE OF TECHNOLOGY

What's wrong with the Megatron 3000 DBMS?

- No security

e.g., - File system insecure

- File system security is coarse

CS 525



Notes 1 - Introduction

23

IIT College of
Science and Letters
ILLINOIS INSTITUTE OF TECHNOLOGY

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 525



Notes 1 - Introduction

24

IIT College of
Science and Letters
ILLINOIS INSTITUTE OF TECHNOLOGY

What's wrong with the Megatron 3000 DBMS?

- Cannot interact with other DBMSs.

CS 525



Notes 1 - Introduction

25

IIT College of
Science and Letters
ILLINOIS INSTITUTE OF TECHNOLOGY

What's wrong with the Megatron 3000 DBMS?

- Poor dictionary facilities

CS 525



Notes 1 - Introduction

26

IIT College of
Science and Letters
ILLINOIS INSTITUTE OF TECHNOLOGY

What's wrong with the Megatron 3000 DBMS?

- No GUI

CS 525



Notes 1 - Introduction

27

IIT College of
Science and Letters
ILLINOIS INSTITUTE OF TECHNOLOGY

What's wrong with the Megatron 3000 DBMS?

- Lousy salesman!!

CS 525



Notes 1 - Introduction

28

IIT College of
Science and Letters
ILLINOIS INSTITUTE OF TECHNOLOGY

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 525



Notes 1 - Introduction

29

IIT College of
Science and Letters
ILLINOIS INSTITUTE OF TECHNOLOGY

Course Overview

- **Concurrency Control**
Correctness, locks,...
- **Transaction Processing**
Logs, deadlocks,...
- **Security & Integrity**
Authorization, encryption,...
- **Advanced Topics**
Distribution, More Fancy Optimizations, ...

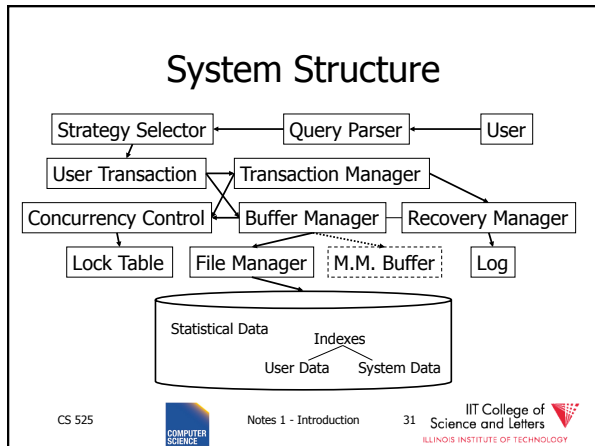
CS 525





Notes 1 - Introduction


30


IIT College of
Science and Letters
ILLINOIS INSTITUTE OF TECHNOLOGY




- ## Some Terms
- Database system
 - Transaction processing system
 - File access system
 - Information retrieval system
- CS 525  Notes 1 - Introduction 32 IIT College of Science and Letters ILLINOIS INSTITUTE OF TECHNOLOGY

- ## Course Information
- **Webpage:** <http://www.cs.iit.edu/~cs525/>
 - **Instructor:** Boris Glavic
 - <http://www.cs.iit.edu/~glavic/>
 - **DBGroup:** <http://www.cs.iit.edu/~dbgroup/>
 - **Office Hours: Thursdays, 1pm-2pm**
 - **Office:** Stuart Building, Room 226 C
 - **TA: Xi Zhang** (xzhang22@hawk.iit.edu)
 - **Time:** Mon + Wed 3:15pm – 4:30pm
- CS 525  Notes 1 - Introduction 33 IIT College of Science and Letters ILLINOIS INSTITUTE OF TECHNOLOGY

- ## Google Group
- <https://groups.google.com/forum/#!forum/cs525-2014-spring-group>
 - Mailing-list for announcements
 - Discussion forum
 - Student - Instructor/TA
 - Student - Student
 - ->please join the group to keep up to date
- CS 525  Notes 1 - Introduction 34 IIT College of Science and Letters ILLINOIS INSTITUTE OF TECHNOLOGY

- ## Workload and Grading
- Schedule and Important Dates
 - On webpage & updated there
 - Programming Assignments (50%)
 - 4 Assignments
 - Groups of 3 students
 - Plagiarism -> 0 points and administrative action
 - Quizzes (10%)
 - Mid Term (20%) and Final Exam (20%)
- CS 525  Notes 1 - Introduction 35 IIT College of Science and Letters ILLINOIS INSTITUTE OF TECHNOLOGY

- ## Textbooks
- Elmasri and Navathe , **Fundamentals of Database Systems**, 6th Edition , Addison-Wesley , 2003
 - Garcia-Molina, Ullman, and Widom, **Database Systems: The Complete Book**, 2nd Edition, Prentice Hall, 2008
 - Ramakrishnan and Gehrke , **Database Management Systems**, 3rd Edition , McGraw-Hill , 2002
 - Silberschatz, Korth, and Sudarshan , **Database System Concepts**, 6th Edition , McGraw Hill , 2010
- CS 525  Notes 1 - Introduction 36 IIT College of Science and Letters ILLINOIS INSTITUTE OF TECHNOLOGY

Programming Assignments

- 4 assignments one on-top of the other
- Optional 5th assignment for extra credit
- Code has to compile & run on server account
 - Email-ID@fourier.cs.iit.edu
 - Linux machine
 - SSH with X-forwarding
- Source code managed in **git** repository on Bitbucket.org
 - Handing in assignments = submit (push) to repository
 - One repository per student
 - You should have gotten an invitation (if not, contact me/TA)
 - Git tutorials linked on course webpage!

CS 525



Notes 1 - Introduction

37

IIT College of
Science and Letters
ILLINOIS INSTITUTE OF TECHNOLOGY

Next:

- Hardware

CS 525



Notes 1 - Introduction

38

IIT College of
Science and Letters
ILLINOIS INSTITUTE OF TECHNOLOGY