CS 525:
Advanced Database Organization

01: Introduction

Boris Glavic

Slides: adapted from a course taught by
Hector Garcia-Molina, Stanford InfoLab

IIT College of
Science and Letters
ILLINOIS INSTITUTE OF TECHNOLOGY

COMPUTER SCIENCE

Advanced Database Organization?

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

IIT College of
Science and Letters
ILLINOIS INSTITUTE OF TECHNOLOGY

COMPUTER
SCIENCE

Isn't Implementing a Database System Simple?

Relations ⟹ Statements ⟹ Results

COMPUTER SCIENCE

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

COMPUTER SCIENCE

IIT College of Science and Letters
ILLINOIS INSTITUTE OF TECHNOLOGY

Megatron 3000 Implementation Details

! First sign non-disclosure agreement !

COMPUTER SCIENCE

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
   :
   :
```

IIT College of
Science and Letters
ILLINOIS INSTITUTE OF TECHNOLOGY

COMPUTER
SCIENCE

Megatron 3000 Implementation Details

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

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

IIT College of
Science and Letters
ILLINOIS INSTITUTE OF TECHNOLOGY

COMPUTER SCIENCE

Megatron 3000 Sample Sessions

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

IIT College of
Science and Letters
ILLINOIS INSTITUTE OF TECHNOLOGY

COMPUTER SCIENCE

Megatron 3000
Sample Sessions

```
& select *
  from R #

    Relation R
  A         B      C
  SMITH    123    CS


&
```

COMPUTER SCIENCE

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

&
```

IIT College of
Science and Letters
ILLINOIS INSTITUTE OF TECHNOLOGY

COMPUTER
SCIENCE

Megatron 3000
Sample Sessions

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

Result sent to LPR (printer).

IIT College of
Science and Letters
ILLINOIS INSTITUTE OF TECHNOLOGY

COMPUTER SCIENCE

Megatron 3000
Sample Sessions

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

New relation T created.

IIT College of
Science and Letters
ILLINOIS INSTITUTE OF TECHNOLOGY

COMPUTER
SCIENCE

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

COMPUTER SCIENCE

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

COMPUTER SCIENCE

IIT College of Science and Letters
ILLINOIS INSTITUTE OF TECHNOLOGY

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

IIT College of
Science and Letters
ILLINOIS INSTITUTE OF TECHNOLOGY

COMPUTER SCIENCE

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

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?

COMPUTER SCIENCE

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

IIT College of
Science and Letters
ILLINOIS INSTITUTE OF TECHNOLOGY

COMPUTER SCIENCE

What's wrong with the Megatron 3000 DBMS?

- No concurrency control

COMPUTER SCIENCE

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

IIT College of
Science and Letters
ILLINOIS INSTITUTE OF TECHNOLOGY

COMPUTER SCIENCE

What's wrong with the Megatron 3000 DBMS?

- ## No security

e.g.,   - File system insecure

- File system security is coarse

IIT College of
Science and Letters
ILLINOIS INSTITUTE OF TECHNOLOGY

COMPUTER
SCIENCE

What's wrong with the Megatron 3000 DBMS?

- No application program interface (API)

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

COMPUTER SCIENCE

IIT College of
Science and Letters
ILLINOIS INSTITUTE OF TECHNOLOGY

What's wrong with the Megatron 3000 DBMS?

- Cannot interact with other DBMSs.

COMPUTER SCIENCE

IIT College of Science and Letters
ILLINOIS INSTITUTE OF TECHNOLOGY

What's wrong with the Megatron 3000 DBMS?

- Poor dictionary facilities

Notes 1 - Introduction

What's wrong with the Megatron 3000 DBMS?

- No GUI

COMPUTER SCIENCE

IIT College of
Science and Letters
ILLINOIS INSTITUTE OF TECHNOLOGY

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,…

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, …

IIT College of
Science and Letters

ILLINOIS INSTITUTE OF TECHNOLOGY

COMPUTER
SCIENCE

System Structure

Notes 1 - Introduction
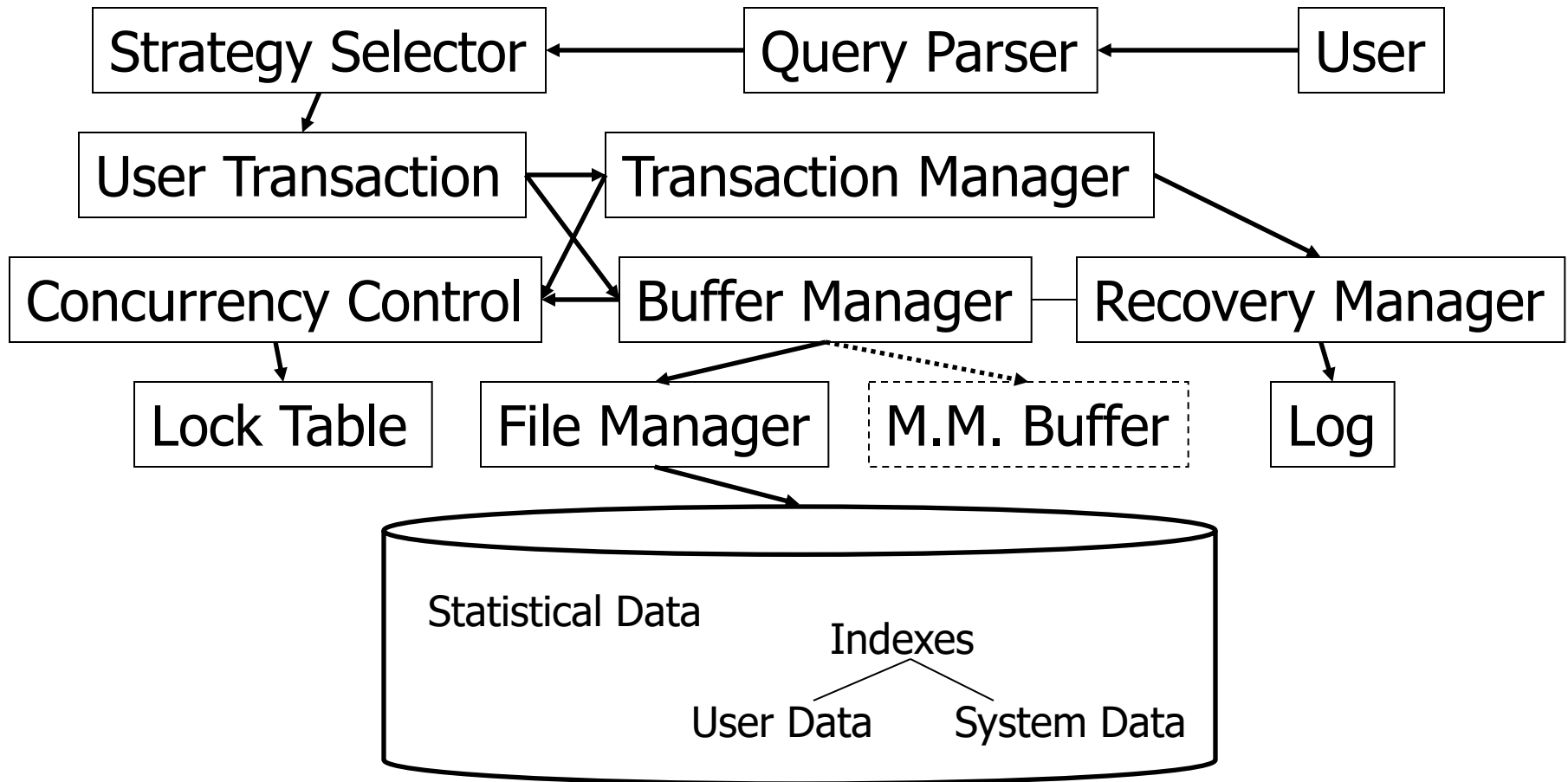
Some Terms

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

COMPUTER SCIENCE

IIT College of
Science and Letters

ILLINOIS INSTITUTE OF TECHNOLOGY

Instructor Info

- **Webpage**: http://www.cs.iit.edu/~glavic/cs525/

- **Instructor**: Boris Glavic
  - **http://www.cs.iit.edu/~glavic/**
  - **DBGroup: http://www.cs.iit.edu/~dbgroup/**
  - **Office Hours: Wednesdays, 1pm-2pm**
    - **https://meet.google.com/pqy-qnyd-kqf**

- **TA: TBA**

IIT College of
Science and Letters
ILLINOIS INSTITUTE OF TECHNOLOGY

Course Information

- **Time:** Mon + Wed 5:10pm – 8:30pm

- **Live Video Call**:

  - **https://meet.google.com/siw-youu-kme**

- **Live Lectures will be recorded and uploaded to blackboard and to**

  - **https://drive.google.com/drive/folders/1CxZy2sXe8 v1dOb4WrNufJ0uMkwbjjKRd?usp=sharing**

IIT College of
Science and Letters
ILLINOIS INSTITUTE OF TECHNOLOGY

COMPUTER SCIENCE

Piazza

- https://piazza.com/class/kdapgqw7az2vq

- Announcements

- Discussion forum
  - Student - Instructor/TA
  - Student – Student

- ->please join piazza to keep up to date

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%)

COMPUTER SCIENCE

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**, 3nd Edition , McGraw-Hill , 2002

- Silberschatz, Korth, and Sudarshan , **Database System Concepts**, 6th Edition , McGraw Hill , 2010

COMPUTER SCIENCE

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!

IIT College of
Science and Letters
ILLINOIS INSTITUTE OF TECHNOLOGY

COMPUTER SCIENCE

Next:

- Hardware

COMPUTER SCIENCE

IIT College of Science and Letters

ILLINOIS INSTITUTE OF TECHNOLOGY