

Intro DB 2, ERD

27 October 2009
Lecture 2

Topics for Today

- Schemas: Physical, Conceptual, External
- Queries
- Transaction Management
- Structure of a DBMS
- Sources:
 - Ramakrishnan and Gehrke 1.1-1.9

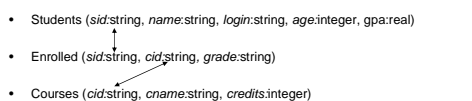
Levels of Abstraction

- Relational model introduces levels of abstraction
 - Lets us ignore details of lower levels
 - Each level has its own schema, written in terms of the level below
 - Three important levels:
 - Physical
 - Conceptual
 - External
- } Write their schemas with the *Data Definition Language (SQL)*

Physical Schema

- How data is organized on disk
 - Relations usually stored as unordered collections of records
 - Indexes are used to make sorting and searching faster
 - We'll talk a bit about them next semester
- The physical schema is normally invisible to the SQL coder
 - We won't get to it in this course

Conceptual Schema

- Describes data as *entities* and *relationships* or links between relations
 - A student is enrolled in a course
 - Students (*sid:string, name:string, login:string, age:integer, gpa:real*)
 - Enrolled (*sid:string, cid:string, grade:string*)
 - Courses (*cid:string, cname:string, credits:integer*)
- 

External Schema

- How the database is seen from the outside
 - Written in terms of the conceptual schema
- What interface or relations are shown to the outside world?
 - May just be the original relations
 - Or may be composite relations built from other relations in the schema - this is called a *View*

Views and External Schema

- A (Dynamic) View is a virtual relation calculated solely based on the contents of other relations. Each time it is accessed, it is recalculated before use.
- Instead of Students – Enrolled – Courses, we can define summaries of the data
- Enrollment (*cid:string, enrollment:integer*)
 - Shows only how many students per course
- GpaPerCourse(*cid:string, avggpa:float*)
 - Shows the average grade for students in the course

October 27, 2009

ISE 322: Database Systems

7

Why Views?

- Data abstraction
 - The logistics people don't need student names, just the count
- Data hiding
 - The lecturer shouldn't need to know each student's grade average, just the average for the class
- Data protection
 - Logistics people can't update the enrollment figures, just view them
 - How else could we enforce this?
- Data updates
 - Just update the base relations

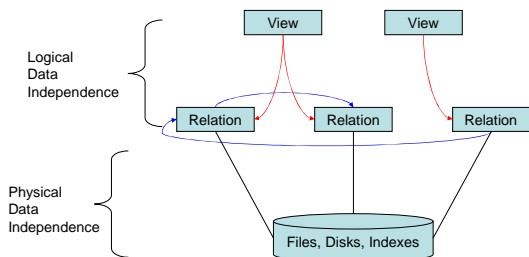
October 27, 2009

ISE 322: Database Systems

8

Data Independence

- One nice outcome of the levels of abstraction – data independence



October 27, 2009

ISE 322: Database Systems

9

Why Logical/Physical Independence?

- Physical Flexibility
 - What if we upgrade the hard disks or reformat them?
 - Change from a single server to a network for storage?
 - Defragment the disk?
- Logical Flexibility
 - Add/delete a column from a base relation?
 - Change a relation's type?
 - Merge/split two relations
 - Need to hide a column from some users

October 27, 2009

ISE 322: Database Systems

10

So Far

- Schemas: Physical, Conceptual, External
- Queries
- Transaction Management
- Structure of a DBMS

October 27, 2009

ISE 322: Database Systems

11

Queries

- The power of database is how it lets the user **ask questions** about its contents
- Queries are questions we ask the DBMS – it answers them by looking at the database
 - SQL is the most popular **Query Language**
- It's important to understand the meaning of queries - their *semantics*
 - To help, abstract query languages have been devised – Relational Algebra underlies SQL

October 27, 2009

ISE 322: Database Systems

12

Two Types of Queries

- Read Queries
 - Ask the DBMS to answer a question by returning a single value or a table of results
 - Using the Query Language
- Write Queries
 - Ask the DBMS to add/delete/modify a record
 - Using the Data Manipulation Language
- SQL is both

October 27, 2009

ISE 322: Database Systems

13

Three faces of SQL

- SQL is multipurpose:
 - We can use it to define, modify, and delete relations (Data Definition Language)
 - We can use it add, modify, and delete records (Data Manipulation Language)
 - We can use it to ask questions about the data (Query Language)



October 27, 2009

ISE 322: Database Systems

14

How to Query

- Could log into the DBMS with a command line or GUI interface and ask queries directly
- Could use a remote DB tool which builds a connection to a DB and has a GUI to receive user queries and display results
- Could write application code which builds a database connection, sends queries, and does things with the results
- Could build a web site which has a backend application which runs queries implicitly or explicitly and shows the results on the page

October 27, 2009

ISE 322: Database Systems

15

Transaction Management

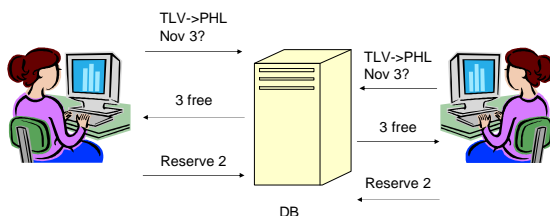
- Transactions essential for two reasons:
 - Concurrency (multiple users)
 - Failure recovery

October 27, 2009

ISE 322: Database Systems

16

Concurrency

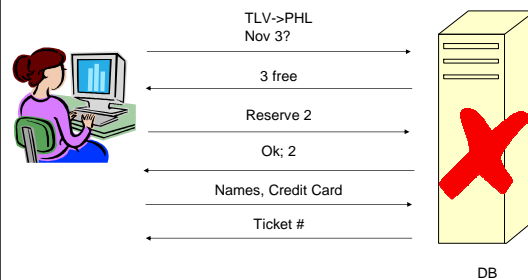


October 27, 2009

ISE 322: Database Systems

17

Failure Recovery



October 27, 2009

ISE 322: Database Systems

18

Transactions in DBs

- A transaction is one execution of a user program on a DBMS
 - Similar to the Operating System and Distributed Systems definitions
 - Unlike the definition we'll see in the other course
- DBMS performs actions at the transaction level
 - Atomic
 - All or nothing

Tools for Concurrency Control

- DBMS must manage multiple transactions
 - Enforce non-conflict
- Most use **locks**
 - There are other ways to do it, though

Locks

- An **exclusive lock** – only the one transaction can read or write the object
- A **shared lock** – one or more transactions can read the object
- A simple, but well studied locking protocol:
 - To read – acquire a shared lock first
 - To write – acquire an exclusive lock first
- We'll talk more about this later in the semester and next semester

Locks and Interleavings

- Two or more concurrently executing transactions are an interleaving
 - Many are possible for a given set of transactions and locking protocol
 - Two interleavings of the same transactions may yield different results
- The goal: To make each transaction "think" it is operating in isolation

Interleaving 1

- T1: Read(A); Read(B)
- T2: Read(A)

T1	T2
	shared-acquire(A)
shared-acquire(A)	
shared-acquire(B)	
shared-release(A)	
	shared-release(A)
shared-release(B)	

T1	T2
shared-acquire(A)	
	shared-acquire(A)
shared-acquire(B)	
	shared-release(A)
shared-release(A)	
shared-release(B)	

Interleaving 2

- T3: Write(A)
- T4: Read(A)
- T5: Read(A), Write(B)

T3	T4	T5
exclusive-acquire(A)		
exclusive-release(A)		
	shared-acquire(A)	
		shared-acquire(A)
		exclusive-acquire(B)
	shared-release(A)	
		shared-release(A)
		exclusive-release(B)

T3	T4	T5
	shared-acquire(A)	
		shared-acquire(A)
		exclusive-acquire(B)
	shared-release(A)	
exclusive-acquire(A)		shared-release(A)
exclusive-release(A)		exclusive-release(B)

Dealing with Crashes

- The main tool to deal with crashes: the **log**
 - Write ahead – each operation is entered in the log before it is executed
 - Log is flushed every so often, always when a transaction completes
- Also:
 - Checkpoints
 - Ability to selectively force pages to disk

October 27, 2009

ISE 322: Database Systems

25

Dealing with Crashes

- When a crash occurs:
 - Reload the database to where it was before
 - Go over the log and remove actions from transactions which didn't commit
- More in the book on this – we may get to it this semester

October 27, 2009

ISE 322: Database Systems

26

So Far

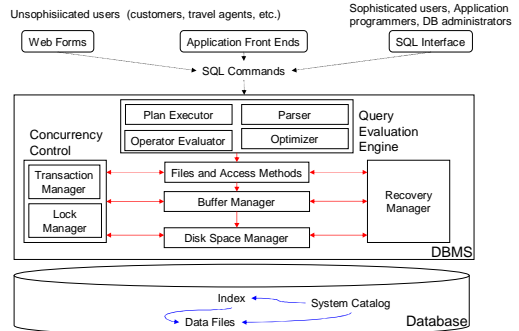
- Schemas: Physical, Conceptual, External
- Queries
- Transaction Management
- Structure of a DBMS

October 27, 2009

ISE 322: Database Systems

27

Structure of a DBMS



October 27, 2009

ISE 322: Database Systems

28

Who uses databases?

- Database Administrators (DBAs)
- Application programmers
- Information Systems Architects/Programmers
- Anybody who uses:
 - Amazon
 - Google
 - Youtube
 - ...
- Oracle is the biggest commercial DB vendor
 - They are second biggest application development company

October 27, 2009

ISE 322: Database Systems

29

Summary

- Schemas: Physical, Conceptual, External
- Queries
- Transaction Management
- Structure of a DBMS

October 27, 2009

ISE 322: Database Systems

30

Word of the day

- “in vogue” - in the current fashion or style