Databases and Database Management System

Goals

- comprehensive introduction to
  - the design of databases
  - the use of database management systems for applications
  - data organization and query processing

2

Topics

- the entity-relationship (E/R) and relational data models
- approaches to database design
- abstract query languages (relational algebra)
- SQL
- physical data organization
- query processing and optimization
- transaction processing
- advanced topics (data warehouses, data mining, temporal databases, and XML) *[time permit – usually in CS582]*

3

After this course

- You should be able to
  - design a database (preferably optimal)
  - use/write SQL and relational algebra
  - implement database interface with the web
  - discuss the properties of a design of a database
  - write efficient SQL queries
  - know what should and shouldn't in database design and SQL queries

4

Book

1. Textbook: Databases and Transaction Processing – An Application-Oriented Approach by Kifer , Bernstein, and Lewis (Addison Wesley)
2. Additional reading:
   1. A first course in Database Systems by *Ullman & Widom*
   2. Database systems the complete book by *Garcia-Molina, Ullman, & Widom*

5

Organizational issues

- Class web site: hosted by WebCT
- 7 weeks: First mid-term (October 1)
- 12 weeks: Second mid-term (November 7)
- December 14: Final
- Office hours: 11 am -12 pm Tues & Thus.
- Use of emails (WebCT): questions, notifications
- Homework submission – web based

6

Goal of today's class

- Get you excited in this course ☺
- Let begin with some questions …

7

Background

- Do you know what is set union, set difference, complement of a set, Cartesian product? If you don't, I'll have a hard time!
- Do you know how to prove a theorem? Fixpoint (or fixed point)?
  If you don't, it will be even harder for me!
- Do you know SQL? C? JAVA? PHP?

My assumption: you are very good programmers in many languages! (at least in C and JAVA)

8

Today

- Examples of a few well-known databases
- Definition of database and its properties
- DBMS (what is and requirements)
- Short comparison between DB now & then
- Issues related to database system studies
- Applications of DB

9

Registrar Office (NMSU)

Work with tables
- add rows to tables
- delete rows from tables
- combine tables
- answer questions
  - What is Bo's GPA?
  - How many students in CS482?
  - …

StudentID	Name	Address
1	J. Bo	123 Main
2	J. Carr	23 SW 5th

Class	StudentID	Grade
CS482	1	A
CS482	2	B

10

HMR Office (NMSU)

Work with tables
- add rows to tables
- delete rows from tables
- combine tables
- answer questions
  - what is A's salary?
  - what is the average salary of CS?
  - …

Emp-ID	Name	Address
1	S. West	456 Mountain
2	E. Corolly	23 NW 1th

Emp-ID	Salary	Department
1	500K	CS
2	700K	EE

11

Airline Reservation

Work with tables
- add rows to tables
- delete rows from tables
- combine tables
- answer questions
  - Is flight X full?
  - Who are on flight X?
  - Is there flight form X to Y?

FlightID	From	To
1	ELP	DFW
2	ELP	LAX

FlightID	Name	Seat
1	S. Tran	1A
2	M. Fox	25C
1	J. O'Hara	1B

12

Observations

- They all
  - work with tables
  - execute similar actions
  - are not so interested in other's tables
- They should
  - maintain their tables with *most updated information*
  - maintain their tables so that their answers to questions posed to them are *correct*

NOW – let get serious about database …

13

What is a Database?

- American Heritage Dictionary: a *collection of data* organized for *easy* and *speed* of search and retrieval

14

What is a Database?

- Related to data, perhaps lot of them
  - same characteristics
  - over long period of time (months, years, …)
- Pertinent to people who use the data
- Might or might not be of interest to other

15

Definition and Properties

- **Definition**: A database is a collection of data central to some enterprise that is managed by a Database Management System (**DBMS**)
- **Properties**:
  - Essential to operation of enterprise (contains the only record of enterprise activity)
  - Valuable to the enterprise (Historical data can guide enterprise strategy, might be of interest to other enterprises)
  - Reflection of the state of the enterprise (database is persistent)

16

DBMS

- Specialized software manages databases
  - create new databases
  - modify existing databases (update data, create reports for different purposes)
- Supports
  - high-level access language (e.g. SQL)
  - application describes database accesses using that language.

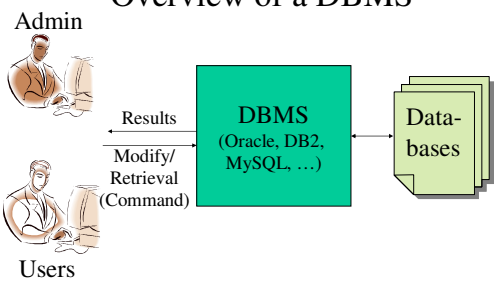
17

DBMS

- Provides users with
  - Persistent storage: like file system but much more flexible
  - Programming interface: accessing and modifying data through a query language
  - Transaction management: concurrent access to data

18

Overview of a DBMS



19

Databases (Now vs. Then)

- Relational model using SQL - high-level view of data
  - Older systems presented low-level view
- Might contain multimedia data
  - Older systems restricted to alphanumeric data
- On-line: database accessed at time of event
  - Older systems were off-line, batch

20

Databases (Now vs. Then)

- Concurrent - multiple transactions execute simultaneously
  - Older systems processed transactions sequentially
- Distributed computation - different parts of the application execute on different computers
  - Older systems were centralized

21

Databases (Now vs. Then)

- Distributed data - different parts of the data are stored in different databases on different computers
  - Older systems were centralized
- Heterogeneous - involves HW and SW modules from different manufacturers
  - Older systems were homogeneous
- Accessed by everyone (e.g., e-commerce)
  - Older systems restricted to trained personnel

22

Database (System) Requirements

Database (System) Requirements

- **High Availability**: on-line => must be operational while enterprise is functioning
- **High Reliability**: correctly tracks state, does not lose data, controlled concurrency
- **High Throughput**: many users => many transactions/sec
- **Low Response Time**: on-line => users are waiting

24

Requirements (cont.)

- **Long Lifetime**: complex systems are not easily replaced
  - Must be designed so they can be easily extended as the needs of the enterprise change
- **Security**: sensitive information must be carefully protected since system is accessible to many users
  - Authentication, authorization, encryption

25

People in Design, Implementation, and Maintenance of a Database

- **System Analyst** - specifies system using input from customer; provides complete description of functionality from customer's and user's point of view
- **Database Designer** - specifies structure of data that will be stored in database
- **Application Programmer** - implements application programs (transactions) that access data and support enterprise rules

26

People (cont.)

- **Database Administrator** - maintains database once system is operational: space allocation, performance optimization, database security
- **System Administrator** - maintains transaction processing system: monitors interconnection of HW and SW modules, deals with failures and congestion

27

Database System Studies

Design of databases

- **how** to design a database
- **what** should be stored
- **which** structure for the data
- **what** assumptions should be made
- **how** is the connection between data

29

Database programming

- how to write queries on the database
- how to use other capabilities of a DBMS in an application
- how is database programming combined with conventional programming

30

Database System Implementation

- how to build a DBMS (query processing, transaction processing, storage manager etc.)

This will not be discussed in this course.

31

Application of Database

Decision Support System (OLTP vs. OLAP)

- **On-line Transaction Processing** (OLTP)
  - Day-to-day handling of transactions that result from enterprise operation
  - Maintains correspondence between database state and enterprise state
- **On-line Analytic Processing** (OLAP)
  - Analysis of information in a database for the purpose of making management decisions

33

On-Line Analytical Processing

- Analyzes historical data (terabytes) using complex queries
- Due to volume of data and complexity of queries, OLAP often uses a data warehouse
- **Data Warehouse -** (offline) repository of historical data generated from OLTP or other sources
- **Data Mining** - use of warehouse data to *discover* relationships that might influence enterprise strategy

34

Exp – Airline reservation system

- OLTP
  - Event: customer A books tickets from ELP to NY; update database to reflect that event
- OLAP
  - During the last holiday season, how many customers fly from ELP to Dallas and NY?
- Data Mining
  - Are there any airports in which more than 50% of travelers from ELP need to change their flight?

35