INTRODUCTION TO DATABASES

Corresponding Reading: Chapter 1

Databases in our everyday lives

- We interact with databases throughout the day:

At the Bank	(Transactions)
At the Airport	(Reservations)
At the Library	(Card Catalog)
At the Grocery Store	(Inventory)
At UNH	(Blackboard)

Databases

- a tool for handling ginormous amounts of data.
  - Data/Information being stored and processed is either
    - Textual (abcdef…)
    - Numeric (0123…)
    - enumerated
    - pictorial
- data resides on disks - permanent.
- programs that act on the data may change.

What exactly is a database?

- A database (DB) is a collection of related data
  - What do we mean by data?
    - Information that can be recorded and that has meaning
      - Example: Product names, Prices (wholesale/retail), Inventory
    - Typically describing the activities of an organization
      - Grocery Store, On-line Retailor
- Databases are created for a specific purpose.
  - Not randomly developed
  - Intended group of users/applications
    - Airline Company:
      - Reservations
      - Pilots
      - Mechanics

Database Sizes

- Databases can be of any size and complexity.

Address Book:	100 entries (KB)
Library:	Author list ~500,000 entries (MB)
IRS Tax Returns:	800 GB each year
Amazon.com:	2TB over 200 Computers
CERN LHC:	2-4PB data creation annually

Databases are BIG MONEY

- Ebay:
  - Relies on databases for all aspects of their sites.
  - All data contained in multiple databases.
  - 2006 Stats:
    - 250,000,000 users
    - 1Billion page views per day
    - Trading $1600 of goods per second
    - 105,000,000 listings at any given time
    - 2 PB of data being stored
    - 26 Billion database queries/instructions per day
    - Roll out 100,000+ lines of new code every 2 weeks.

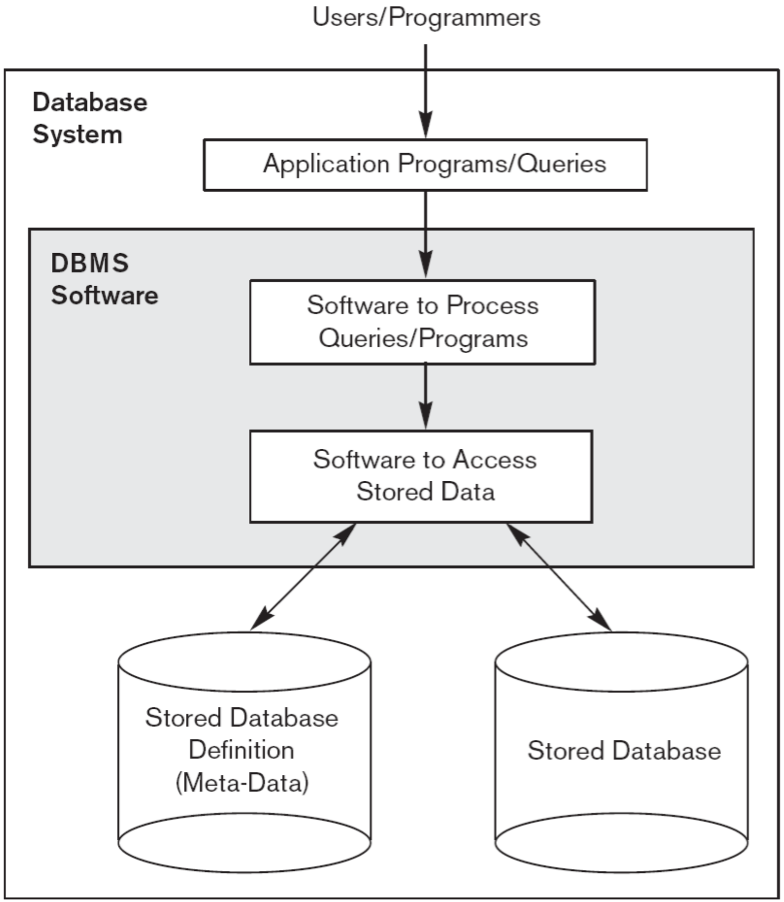
Database Management System (DBMS)

- Collection of programs that enable users to create and maintain a database
  - Software System that facilitates:
    - Defining a DB
      - Specify data types, structures, constraints on the data
    - Constructing a DB
      - Storing the data on some storage medium controlled by DBMS
    - Manipulating a DB
      - Asking the database to retrieve specific data (**Queries**)
      - Updating/modifying the database (**Transactions**)

Database Management System (DBMS)

- Collection of programs that enable users to create and maintain a database
  - Software System that facilitates:
    - Sharing a DB
      - Allow multiple users/programs to access the DB simultaneously
    - Protecting a DB
      - System protection against hardware/software malfunction
      - Security protection against unauthorized access
      - availability 99.99999% up time.
    - Maintaining a DB
      - Allow the system to evolve as requirements change over time
- In general, DBMS are very complex (expensive!) software systems.
  - Oracle's FY11 Revenue: $36 billion

High Level View



Users/Programmers
Database System
Application Programs/Queries
DBMS Software
Software to Process Queries/Programs
Software to Access Stored Data
Stored Database Definition (Meta-Data)
Stored Database

Data Models

- Relational model
  - data and database is viewed as a set of records.

- XML documents
  - hierarchical (tree) structure of labeled values.

Schema versus Data

- Defining the DB: **Schema**
  - Specify the structure of the records of each file
    - What are the different kinds of **data elements** to be stored in each record?
      - Student name, Class Year,…
    - What is the **data type** fore each data element within a record?
      - Integer, string, …
- Constructing the DB: **Data**
  - Store data to represent each student, course, section, grade report, and prerequisite as a record in the appropriate file

Schema versus Data

- Schema is set up using **DDL** (Data Definition Language).
- Data is then loaded into the schema.

- Schema set up at start; does not change. Data changes rapidly.

- Query and modify the database using **DML** (Data Manipulation Language).

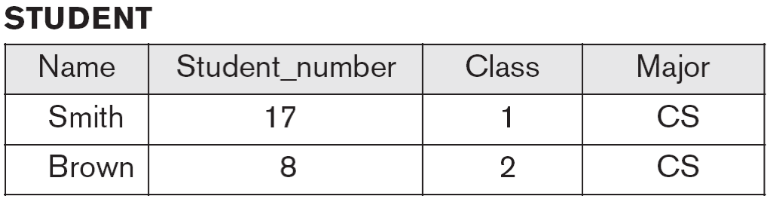
Database Example

- The University Database
  - Common example used in many texts and tutorials
- Purpose of the DB:
  - Maintain information regarding students, courses and grades
- Database Structure:
  - Organized as 5 files (or collections) of data records

STUDENT	Student data
COURSE	Course data
SECTION	Data about each section of a course
GRADE_REPORT	Grades that students receive in a course
PREREQUISITE	Prerequisites for courses

University Database

- STUDENT



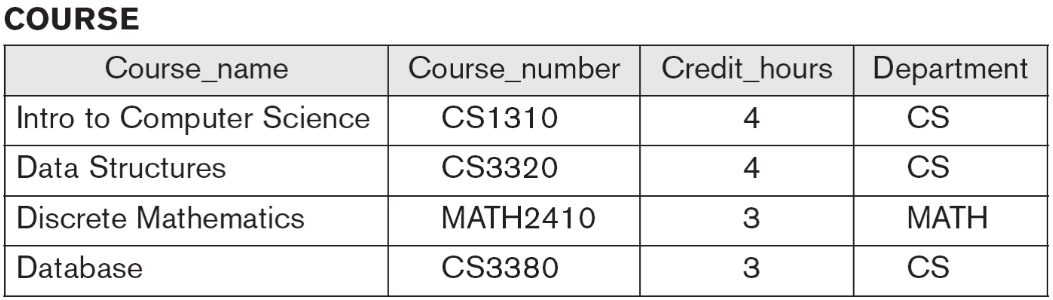
STUDENT

Name	Student_number	Class	Major
Smith	17	1	CS
Brown	8	2	CS

- Each student record is assigned:
  - Name (string – alphabetical characters)
  - Student_number (integer)
  - Class (integer) – Freshman (1), Sophmore (2)
  - Major (string)

University Database

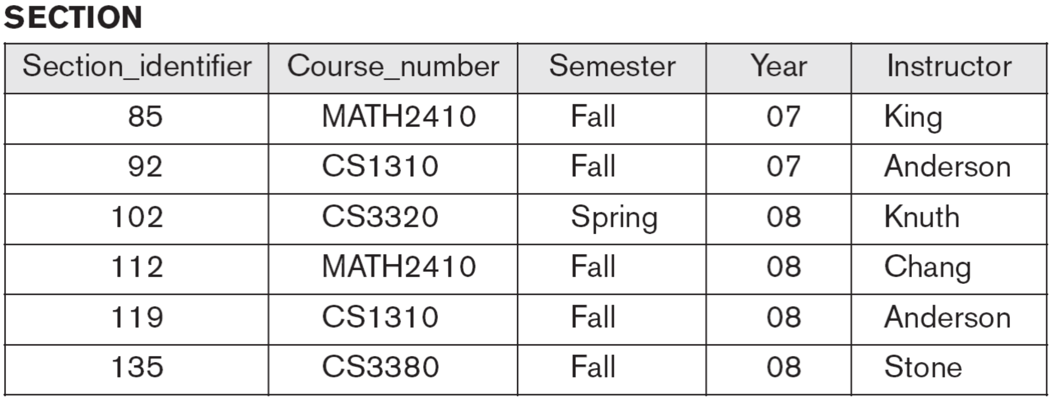
- COURSE



COURSE

Course_name	Course_number	Credit_hours	Department
Intro to Computer Science	CS1310	4	CS
Data Structures	CS3320	4	CS
Discrete Mathematics	MATH2410	3	MATH
Database	CS3380	3	CS

- SECTION

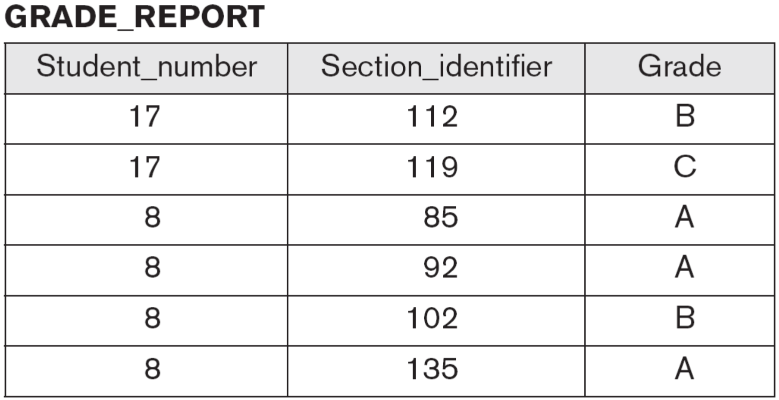


SECTION

Section_identifier	Course_number	Semester	Year	Instructor
85	MATH2410	Fall	07	King
92	CS1310	Fall	07	Anderson
102	CS3320	Spring	08	Knuth
112	MATH2410	Fall	08	Chang
119	CS1310	Fall	08	Anderson
135	CS3380	Fall	08	Stone

University Database

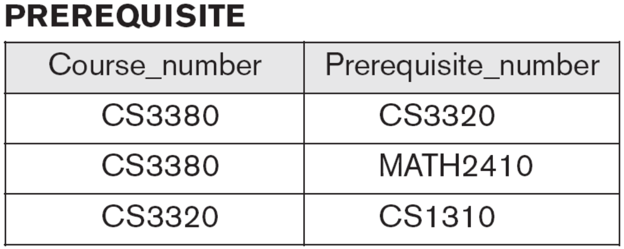
- GRADE_REPORT



GRADE_REPORT

Student_number	Section_identifier	Grade
17	112	B
17	119	C
8	85	A
8	92	A
8	102	B
8	135	A

- PREREQUISITE



PREREQUISITE

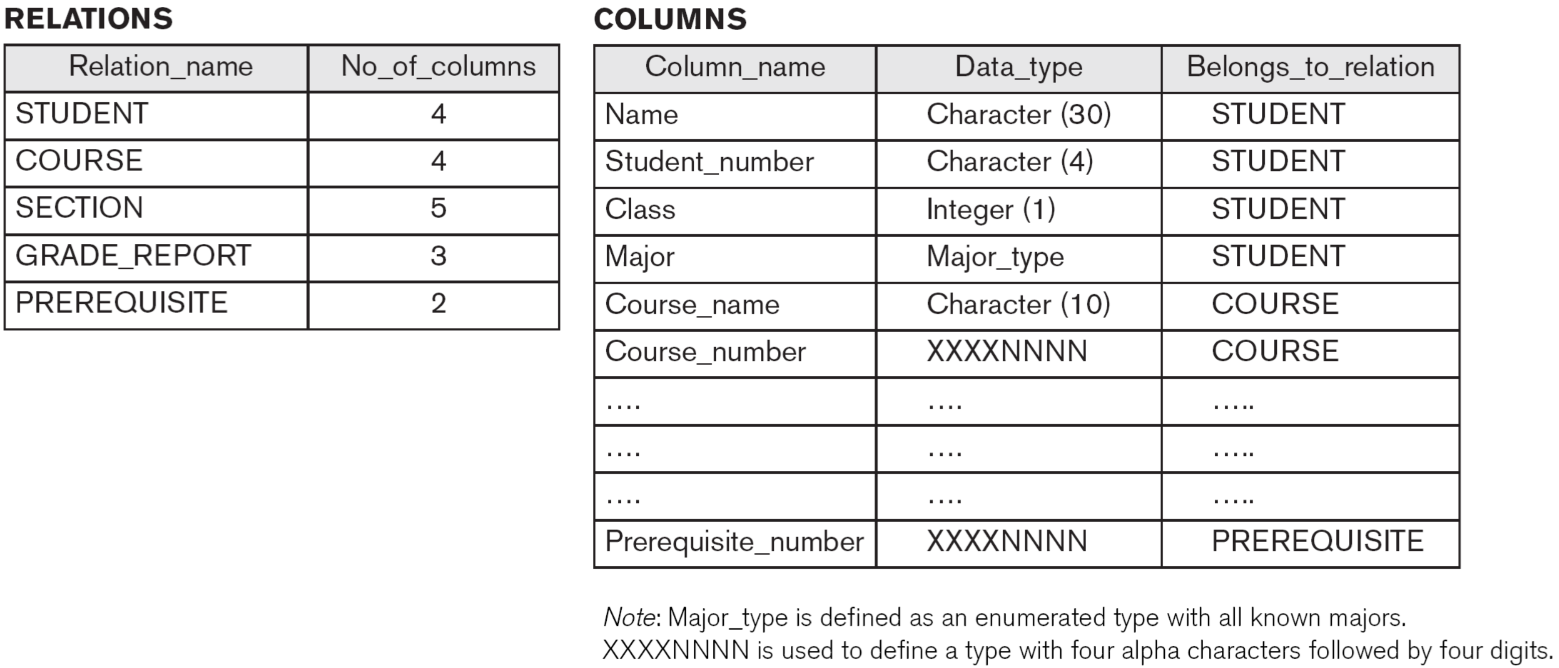
Course_number	Prerequisite_number
CS3380	CS3320
CS3380	MATH2410
CS3320	CS1310

University Database

- Data Relations in a DB:
  - *Notice any columns that are related in the files?*
- Manipulating the DB:
  - Queries to the DB:
    - Retrieve the transcript for ‘Brown’ (Requires which files?)
    - List the prerequisites for the ‘Database’ course
  - Updates to the DB:
    - Change the class of “Smith” to sophomore
    - Enter a grade of ‘A’ for ‘Smith’ in the ‘Database’ section of last semester

University Database

- Describing the DB:
  - DBMS Catalog contains a complete definition of the DB structure and constraints (**meta-data**)



RELATIONS

Relation_name	No_of_columns
STUDENT	4
COURSE	4
SECTION	5
GRADE_REPORT	3
PREREQUISITE	2

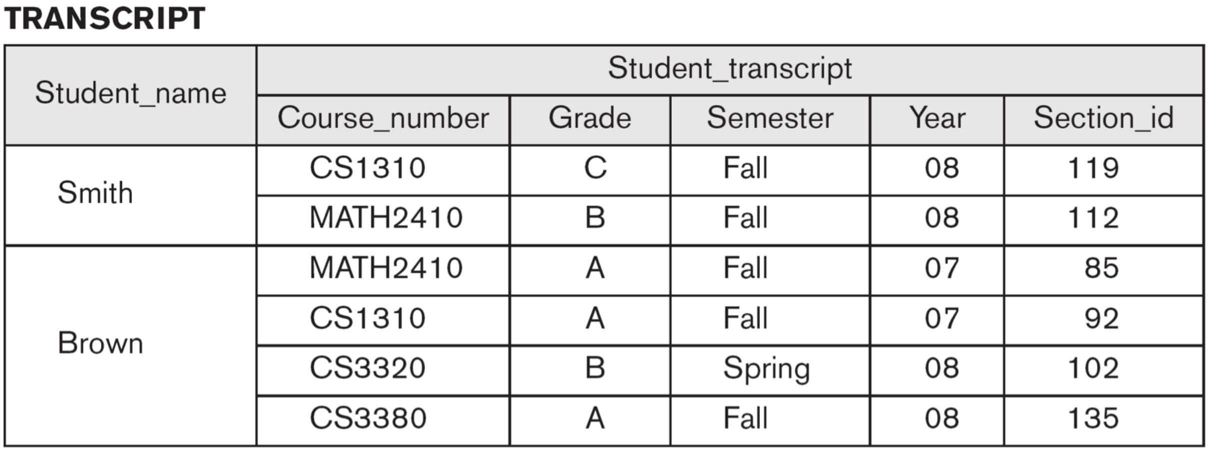
COLUMNS

Column_name	Data_type	Belongs_to_relation
Name	Character (30)	STUDENT
Student_number	Character (4)	STUDENT
Class	Integer (1)	STUDENT
Major	Major_type	STUDENT
Course_name	Character (10)	COURSE
Course_number	XXXXNNNN	COURSE
....	....	.....
....	....	.....
....	....	.....
Prerequisite_number	XXXXNNNN	PREREQUISITE

Note: Major_type is defined as an enumerated type with all known majors.
XXXXNNNN is used to define a type with four alpha characters followed by four digits.

University Database

- Viewing the DB:
  - Through the DBMS, a **view** of the database can be created.
  - Subset of the data or combined data from multiple files
  - Different views for different types of users



TRANSCRIPT

Student_name	Student_transcript				
	Course_number	Grade	Semester	Year	Section_id
Smith	CS1310	C	Fall	08	119
	MATH2410	B	Fall	08	112
Brown	MATH2410	A	Fall	07	85
	CS1310	A	Fall	07	92
	CS3320	B	Spring	08	102
	CS3380	A	Fall	08	135

University Database

- Sharing the DB:
  - Multiuser DBMS allows multiple users access the DB at the same time
  - What happens when two users want to update the same record?
    - **Concurrency Control** at the **data item level** *(not at the level of files/databases - too slow)*.

University Database

- Physical data independence.
  - programs independent of data storage.
  - **high level query languages:** what you want, not how. The algorithms are part of DBMS.
- User Interfaces for the DB:
  - Command-line [boring]
  - Graphical user interface (GUI) [better]
  - Web GUIs [awesome!]
- Performance.
  - speed
  - availability

People

- Database implementer.
- Database designer.
- Application developer and user.
- Database administrator.

Very Brief History of DB Applications

- Early DB apps maintained records for companies
  - Large number of records of similar structure
- Very little data abstraction and program-data independence
  - Specify physical storage and placement on hard disk
  - Reorganizing DB was very time consuming and complex
- DB systems were implemented on large (expensive) mainframe computers (1960-1980)

Very Brief History of DB Applications

- **Relational databases** separate physical storage from conceptual representation.
  - Introduced high-level query languages
- Commercial **RDBMS** started in 80’s (very slow) and became the dominant DB systems (hey there Oracle!)
- Object orientated databases (**OODB**s) mainly used for specialized applications. [Less than 5% in the market]
- **XML Databases** evolved with the popularity of the WWW and provide flexible storage (more on this later.)

Summary

- Introduction to Databases
  - Database – Collection of related data (facts, info.)
  - DBMS – software for implementing/maintaining DB
  - Discussed requirements/capabilities of the DBMS
    - Defining, constructing, describing, manipulating, sharing, and performance tuning the DB
  - History of DB – Relational DBMS are the most popular and widely used DBMS