



computer science
illuminated

Information Systems

Nell Dale & John Lewis

(adaptation by Michael Goldwasser)



Managing Information

- An **information system** can be generally defined as software that helps us organize and analyze data
 - Flexible application software tools that allow the user to dictate and manage the organization of data
 - Two of the most popular general application information systems are
 - electronic spreadsheets* (generally simple)
 - database management systems* (more industrial)



Spreadsheets

	A	B	C	D	
1					
2					
3					
4					
5					

Figure 12.1 A spreadsheet, made up of a grid of labeled cells

- A **spreadsheet** is a software application that allows the user to organize and analyze data using a grid of labeled **cells**
 - A cell can contain data or a formula that is used to calculate a value
 - Data stored in a cell can be text, numbers, or “special” data such as dates
 - Spreadsheet cells are referenced by their row and column designation



Spreadsheets

- Suppose we have collected data on the number of students that came to get help from a set of tutors over a period of several weeks

	A	B	C	D	E	F	G	H
1								
2				Tutor				
3			Hal	Amy	Frank	Total	Avg	
4		1	12	10	13	35	11.67	
5		2	14	16	16	46	15.33	
6	Week	3	10	18	13	41	13.67	
7		4	8	21	18	47	15.67	
8		5	15	18	12	45	15.00	
9		Total	59	83	72	214	71.33	
10		Avg	11.80	16.60	14.40	42.80	14.27	
11								
12								

Figure 12.1

A spreadsheet containing data and computations



Spreadsheet Formulas

- The power of spreadsheets comes from the formulas that we can create and store in cells
 - When a formula is stored in a cell, the result of the formula is displayed in the cell
 - If we've set up the spreadsheet correctly, we could add or remove tutors, add additional weeks of data, or change any of the data we have already stored and the corresponding calculations would automatically be updated



Spreadsheet Formulas

	A	B	C	D	E	F	G	H
1								
2				Tutor				
3			Hal	Amy	Frank	Total	Avg	
4		1	12	10	13	35	11.67	
5		2	14	16	16	46	15.33	
6	Week	3	10	18	13	41	13.67	
7		4	8	21	18	47	15.67	=SUM(C4..E4)
8		5	15	18	12	45	15.00	=F7/COUNT(C7..E7)
9		Total	59	83	72	214	71.33	
10		Avg	11.80	16.60	14.40	42.80	14.27	=F9/COUNT(C4..E8)
11								
12								

=SUM(C4..C8)

=E9/COUNT(E4..E8)

Figure 12.1 The formulas behind some of the cells



Spreadsheet Analysis

- The dynamic nature of spreadsheets provides the powerful ability to do **what-if analysis**
 - What if the number of attendees decreased by 10%?
 - What if we increase the ticket price by \$5?
 - What if we could reduce the cost of materials by half?



Database Management Systems

- A **database** can simply be defined as a structured set of data
- A **database management system** (DBMS) is a combination of software and data made up of:
 - Physical database—a collection of files that contain the data
 - Database engine—software that supports access to and modification of the database contents
 - Database schema—a specification of the logical structure of the data stored in the database



Database Management Systems

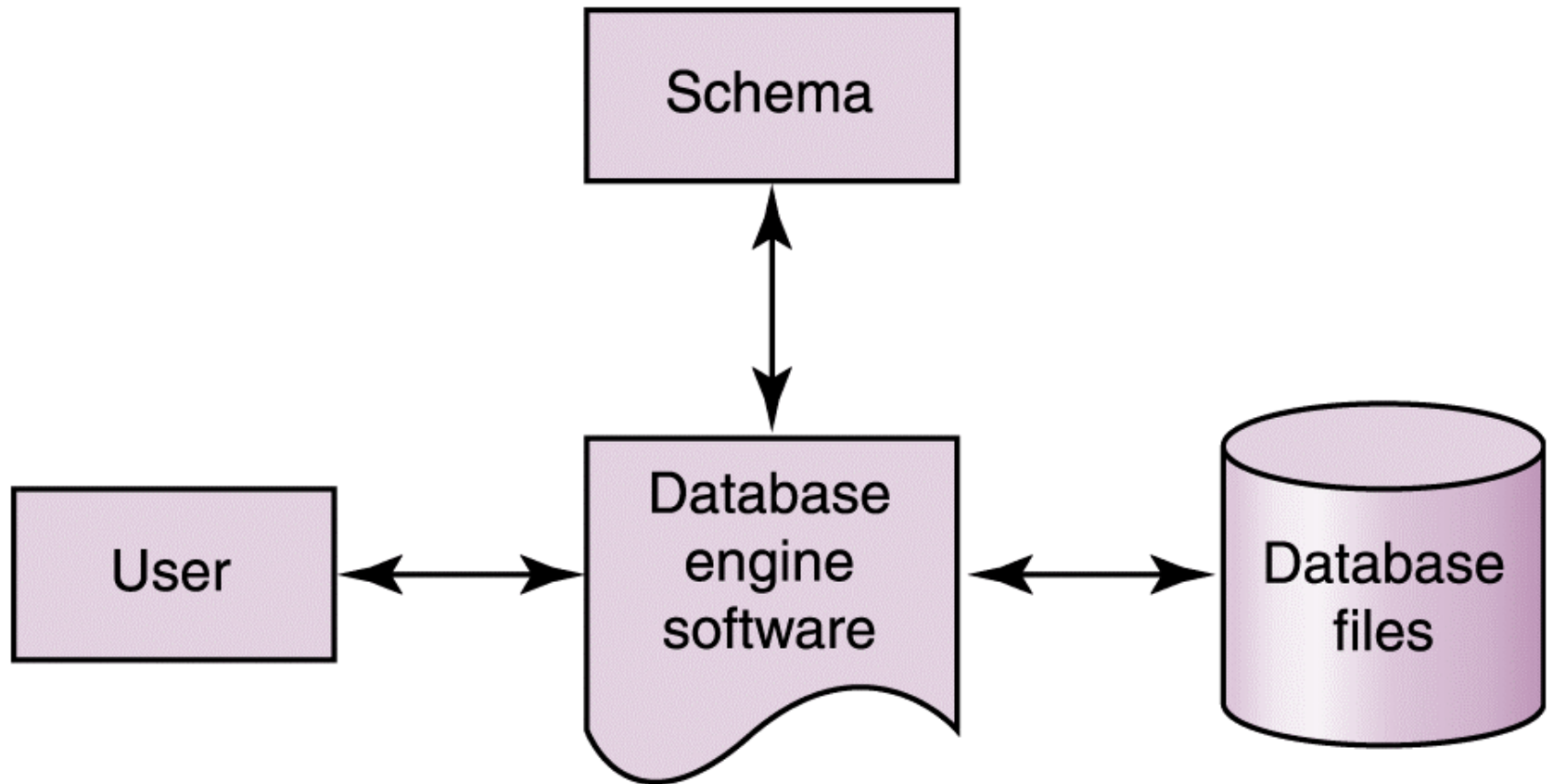


Figure 12.6 The elements of a database management system



Database Management Systems

- Goals:
 - Search on a single field (e.g., Name)
(might binary search work?)
 - Search on multiple fields (find all employees with 10+ years seniority but salary < \$50K)
 - Privacy (some users should only have access to certain information)



Duplication/Redundancy is bad!!

(except in a title)

- Whenever possible you want to avoid having the same underlying information stored in more than one place in the underlying database.
 - Wastes memory!
 - Updating database while ensuring consistency becomes challenging.
 - Deletions present several pitfalls.



An Example

- An example (for the Hollywood fan clubs)

Name	Birthday	Zodiac	Birthstone
Art Alexakis	April 12	Aries	diamond
Hank Azaria	April 25	Taurus	diamond
Antonio Banderas	August 10	Leo	peridot
Lucas Black	November 29	Sagittarius	citrine
Matthew Broderick	March 21	Pisces	aquamarine
Sandra Bullock	July 26	Leo	ruby
Steve Buscemi	December 13	Sagittarius	turquoise
Nicholas Cage	January 7	Capricorn	garnet
Jim Carrey	January 17	Capricorn	garnet
George Clooney	May 6	Taurus	emerald
Courtney Cox	June 15	Gemini	pearl
John Cusack	June 28	Cancer	pearl
Joan Cusack	October 11	Libra	opal
Matt Damon	October 8	Libra	opal



The Relational Model

- In a relational DBMS, the data items and the relationships among them are organized into **tables**
 - A table is a collection of **records**
 - A record is a collection of related **fields**
 - Each field of a database table contains a single data value
 - Each record in a table contains the same fields



The Relational Model

- Essentially all commercial databases today are based on the relational model.
 - Oracle
 - Microsoft Access
 - Informix



A (Relational) Example

Actor

Name	Birthday
Art Alexakis	April 12
Hank Azaria	April 25
Antonio Banderas	August 10
Lucas Black	November 2
Matthew Broderick	March 21
Sandra Bullock	July 26
Steve Buscemi	December 1
Nicholas Cage	January 7
Jim Carrey	January 17
George Clooney	May 6
Courtney Cox	June 15
John Cusack	June 28
Joan Cusack	October 11
Matt Damon	October 8

Birthstone

Start Date	End Date	Birthstone
January 1	January 31	garnet
February 1	February 29	amethyst
March 1	March 31	aquamarine
April 1	April 30	diamond
May 1	May 31	emerald
June 1	June 30	pearl
July 1	July 31	ruby
August 1	August 31	peridot
September 1	September 30	sapphire
October 1	October 31	opal
November 1	November 30	citrine
December 1	December 31	turquoise

Zodiac

Start Date	End Date	Zodiac
February 19	March 20	Pisces
March 21	April 19	Aries
April 20	May 20	Taurus
May 21	June 20	Gemini
June 21	July 22	Cancer
July 23	August 22	Leo
August 23	September 22	Virgo
September 23	October 22	Libra
October 23	November 21	Scorpio
November 22	December 21	Sagittarius
December 22	January 19	Capricorn
January 20	February 18	Aquarius



Relationship “algebra”

- Three basic operations
 - Can choose to display only certain fields (columns) from a given table
 - Can choose to include only certain records (rows) from a given table
 - Can “join” two or more tables by taking the Cartesian product (see next slide for example)



The Join Operation

- A record is created for every pair of records in the original two tables.

Name	Birthday
Art Alexakis	April 12
Hank Azaria	April 25

join

Start Date	End Date	Zodiac
February 19	March 20	Pisces
March 21	April 19	Aries
April 20	May 20	Taurus

equals

Name	Birthday	Start Date	End Date	Zodiac
Art Alexakis	April 12	February 19	March 20	Pisces
Art Alexakis	April 12	March 21	April 19	Aries
Art Alexakis	April 12	April 20	May 20	Taurus
Hank Azaria	April 25	February 19	March 20	Pisces
Hank Azaria	April 25	March 21	April 19	Aries
Hank Azaria	April 25	April 20	May 20	Taurus



The Join Operation

- Rarely do we want all such records. A join is usually combined with other operations.
- E.g., choose only those records from join with $\text{Start Date} \leq \text{Birthday} \leq \text{End Date}$

Name	Birthday	Start Date	End Date	Zodiac
Art Alexakis	April 12	February 19	March 20	Pisces
Art Alexakis	April 12	March 21	April 19	Aries
Art Alexakis	April 12	April 20	May 20	Taurus
Hank Azaria	April 25	February 19	March 20	Pisces
Hank Azaria	April 25	March 21	April 19	Aries
Hank Azaria	April 25	April 20	May 20	Taurus



Another (Blockbuster?) Example

Movie

MovieId	Title	Genre	Rating
101	Sixth Sense, The	thriller horror	PG-13
102	Back to the Future	comedy adventure	PG
103	Monsters, Inc.	animation comedy	G
104	Field of Dreams	fantasy drama	PG
105	Alien	sci-fi horror	R
106	Unbreakable	thriller	PG-13
107	X-Men	action sci-fi	PG-13
5022	Elizabeth	drama period	R
5793	Independence Day	action sci-fi	PG-13
7442	Platoon	action drama war	R

Figure 12.7 A database table, made up of records and fields



Relationships

Customer

CustomerId	Name	Genre	CreditCardNumber
101	Dennis Cook	123 Main Street	2736 2371 2344 0382
102	Doug Nickle	456 Second Ave	7362 7486 5957 3638
103	Randy Wolf	789 Elm Street	4253 4773 6252 4436
104	Amy Stevens	321 Yellow Brick Road	9876 5432 1234 5678
105	Robert Person	654 Lois Lane	1122 3344 5566 7788
106	David Coggin	987 Broadway	8473 9687 4847 3784
107	Susan Klaton	345 Easy Street	2435 4332 1567 3232

Figure 12.8 A database table containing customer data



Relationships

- We can use a table to represent a collection of relationships between objects

Rents

CustomerId	Movielid	DateRented	DateDue
103	104	3-12-2002	3-13-2002
103	5022	3-12-2002	3-13-2002
105	107	3-12-2002	3-15-2002

Figure 12.9 A database table storing current movie rentals



Structured Query Language

- The **Structured Query Language (SQL)** is a comprehensive database language for managing relational databases
- Originally created by IBM in early 70s.
Standardized by ANSI in 1986.



Queries in SQL

select *attribute-list*
from *table-list*
where *condition-list*

- If more than one table in *table-list*, the join of the tables will be computed.
- Of the many possible fields, only those given in *attribute-list* are displayed.
- The *condition-list* can be an arbitrary Boolean Expression used to select records



attribute-list

- “*” means to include all attributes
- New attributes can be created as a combination of existing attributes.



condition-list

- Based on Attributes of potential records
- Arbitrary Boolean Expressions (AND, OR, NOT)
- Can use operators ($>$, \geq , $=$, $<$, \leq)
- Can do partial matches for text such as:
 - Name LIKE 'Mich%'
- Can use set theory, such as:
 - Direction IN ('North', 'East')



Examples with One Table

```
select Title from Movie where Rating = 'PG'
```

```
select Name, Address from Customer
```

```
select * from Movie where Genre like '%action%'
```

```
select * from Movie where Rating = 'R' order by Title
```



Examples with Two Tables

If an identical attribute name is used in several of the underlying tables, then you reference a particular attribute by using both the Table name and the Attribute name.

```
select Title  
from Movie,Rents  
where Movie.MoviId = Rents.MoviId
```



A Gentle Introduction to SQL

- To get hands-on practice, we will use a wonderful website developed by Andrew Cumming of the School of Computing of Napier University in the UK.

<http://www.dcs.napier.ac.uk/~andrew/gisq/>

- We will use two databases from that site.



CIA World Factbook

- Information on all countries, according to the 1995 version of CIA World Factbook (www.cia.gov/cia/publications/factbook/)

table 'cia'

name	region	area	population	gdp
Afghanistan	Asia	652000	25838797	21000000000
Albania	Europe	28748	3490435	5600000000
Algeria	Africa	2381740	31193917	147600000000
...



Internet Movie Database

- Information on movies and their stars according to the 1997 version of the Internet Movie Database (www.imdb.com/)

table 'movie'

id	title	yr	score	votes
1	Star Wars	1977	8.8	53567
2	Shawshank Redemption, The	1994	9.0	44974
3	Pulp Fiction	1994	8.6	43993
4	Titanic	1997	7.2	43371
...

table 'actor'

id	name
1	Woody Allen
2	Clint Eastwood
3	Robert DeNiro
4	Sean Connery
...	...

table 'casting'

movieid	actorid	ord
972	588	1
849	588	2
1575	588	3
47	590	4
...



Modifying Database Content

insert into Customer values (9876, 'John Smith', '602 Greenbriar Court', '2938 3212 3402 0299')

update Movie set Genre = 'thriller drama'
where title = 'Unbreakable'

delete from Movie where Rating = 'R'