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

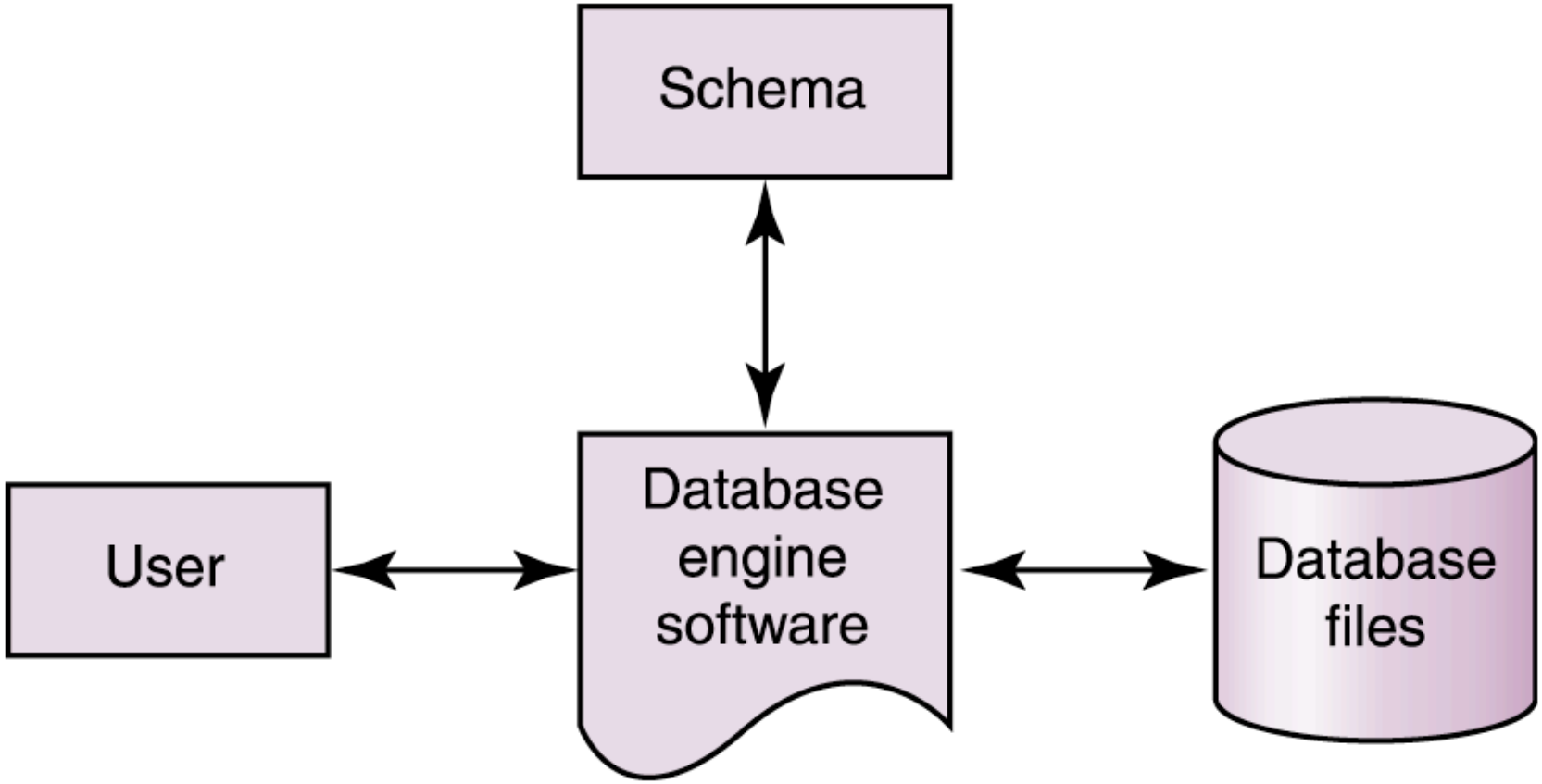

Figure 12.6 The elements of a database management system

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	topaz
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
Courtnet 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 record contains the same set of fields)

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
Courtnet 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	aquamari
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	topaz
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 Azeria	April 25	February 19	March 20	Pisces
Hank Azeria	April 25	March 21	April 19	Aries
Hank Azeria	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 Start Date <= Birthday <= 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 Azeria	April 25	February 19	March 20	Pisces
Hank Azeria	April 25	March 21	April 19	Aries
Hank Azeria	April 25	April 20	May 20	Taurus

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 (>, >=, =, <, <=)
- Can do partial matches for text such as:
  - Name LIKE ‘Mich%’
- Can use set theory, such as:
  - Direction IN (‘North’, ‘East’)

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://sqlzoo.net/

- 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
…		…