A Table with a View

Data Storage and Transfer with
XML and Databases

Differences Between Tables and Databases

- When we think of databases, we often think of tables of information

- Comparing Tables
  - Database tables
    - Metadata tag identifying each of the data fields
  - Spreadsheet tables
    - Rely on position to keep the integrity of their data
  - HTML tables
    - Data as table entries with no unique identity at all
    - Concerned only with how to display the data, not with its meaning

The Database Advantage

- Metadata is key advantage of databases over other systems recording data as tables
- Two of the most important roles in defining metadata
  * Identify the type of data with a unique tag
  * Define the relationships of the data

16-3

XML: A Language for Metadata Tags

- **Extensible Markup Language**
  - ∗ Tagging scheme similar to XHTML
  - ∗ No standard tags to learn
    - Self-describing, think up the tags you need
  - ∗ Works well with browsers and Web-based applications
  - ∗ Use a simple text editor
  - ∗ XML tag names cannot contain spaces

Extensible **M**arkup **L**anguage

XML

An Example from Tahiti

- Area in km^2 for Tahiti & neighboring islands

```
<?xml version = "1.0" encoding="ISO-8859-1" ?>
<archipelago>
<island><iName>Tahiti</iName>     <area>1048</area></island>
<island><iName>Moorea</iName>     <area>130</area></island>
<island><iName>Maiao</iName>      <area>9.5</area></island>
<island><iName>Mehetia</iName>    <area>2.3</area></island>
<island><iName>Tetiaroa</iName>  <area>12.8</area></island>
</archipelago>
```

Figure 16.1 XML file encoding data for the Windward Islands database. The first line states that the file contains XML tags.

An Example from Tahiti (cont'd)

- ## First line

    `<?xml version="1.0" encoding="ISO-8859-1" ?>`

- ## File should be ASCII text
- ## File extension should be `.xml`

Table 16.1 Rules for writing XML.

Required first line	`<?xml version="1.0" encoding="ISO-8859-1"?>` must appear on the first line, starting in the first position.
First tag	The first tag encountered is the *root* element, and it must enclose all of the file's content; it appears on the second or possibly third line.
Closing tags	All tags must be closed.
Element naming	Observe these rules: • Names can contain letters, numbers, and underscore characters. • Names must not start with a number or punctuation character. • Names must not start with the letters xml (or XML, or Xml, etc.). • Names cannot contain spaces.
Case sensitivity	Tags and attributes are case sensitive.
Proper nesting	All tags must be well-nested.
Attribute quoting	All attribute values must be quoted; paired single quotes (apostrophes) or paired double quotes are okay; use "dumb" quotes only; choose 'opposite' quotes to enclose quoted values.
White space	White space is preserved and converted to a single space.
Comments	XML comments have the form `<!-- This is a comment. -->`.

Expanding Use of XML

- Combine encodings of two archipelagos – the Windward and the Galapagos Islands
- Root element is the tag that encloses all of the content of the XML file
  * <archipelago> in Fig. 16.1
  * <geo_feature> in Fig. 16.2
- Indenting for readability and structure

```xml
<?xml version = "1.0"
    encoding="ISO-8859-1" ?>
<geo_feature>
    <archipelago>
        <a_name>Windward Islands
        </a_name>
        <island>
            <iName>Tahiti</iName>
            <area>1048</area>
        </island>
        <island>
            <iName>Moorea</iName>
            <area>130</area>
        </island>
        <island>
            <iName>Maiao</iName>
            <area>9.5</area>
        </island>
        <island>
            <iName>Mehetia</iName>
            <area>2.3</area>
        </island>
        <island>
            <iName>Tetiaroa</iName>
            <area>12.8</area>
        </island>
    </archipelago>
    <archipelago>
        <a_name>Galapagos Islands
        </a_name>
        <island>
            <iName>Isabella</iName>
            <area>4588</area>
            <elevation>1707</elevation>
        </island>
        <island>
            <iName>Fernandina</iName>
            <area>642</area>
            <elevation>1494</elevation>
        </island>
        <island>
            <iName>Tower</iName>
            <area>14</area>
            <elevation>76</elevation>
        </island>
        <island>
            <iName>Santa Cruz</iName>
            <area>986</area>
            <elevation>846</elevation>
        </island>
    </archipelago>
</geo_feature>
```

Figure 16.2 XML file for the Geographic Features database. XML ignores white space, so the text in the file has been indented for easier reading.

Attributes in XML

- Use attributes for additional metadata, not for additional content
    - ∗ Not good, name is content:

        `<archipelago name="Galapagos">`

    - ∗ Better to give alternate form of the data

        `<a_name`
        `accents="Galápagos">Galapagos</a_name>`

Effective Design with XML Tags

- Identification Rule: Label Data with Tags Consistently
  - ∗ You can choose whatever tag names you with to name data, but once you've decided on a tag for a particular kind of data, you must always surround it with that tag.

Effective Design with XML Tags (cont'd)

- **Affinity Rule: Group Related Data**
    - ∗ Enclose in a pair of tags all tagged data referring to the same entity. Grouping it keeps it all together, but the idea is much more fundamental: Grouping makes an association of the tagged data items as being related to each other, properties of the same thing.
    - ∗ Groups together data for a single thing – an island
   - Association is among properties of an object

Effective Design with XML Tags (cont'd)

- ## Collection Rule: Group Related Instances
    - ∗ When you have several instances of the same kind of data, enclose them in tags; again, it keeps them together and implies that they are related by being instances of the same type.
    - ∗ Groups together data of several instance of the same thing – islands
        - • Association is among the objects themselves (entities)

The XML Tree

- XML encodings of information produce hierarchical descriptions that can be thought of as trees
  - ∗ Hierarchy a consequence of how tags enclose one another and the data
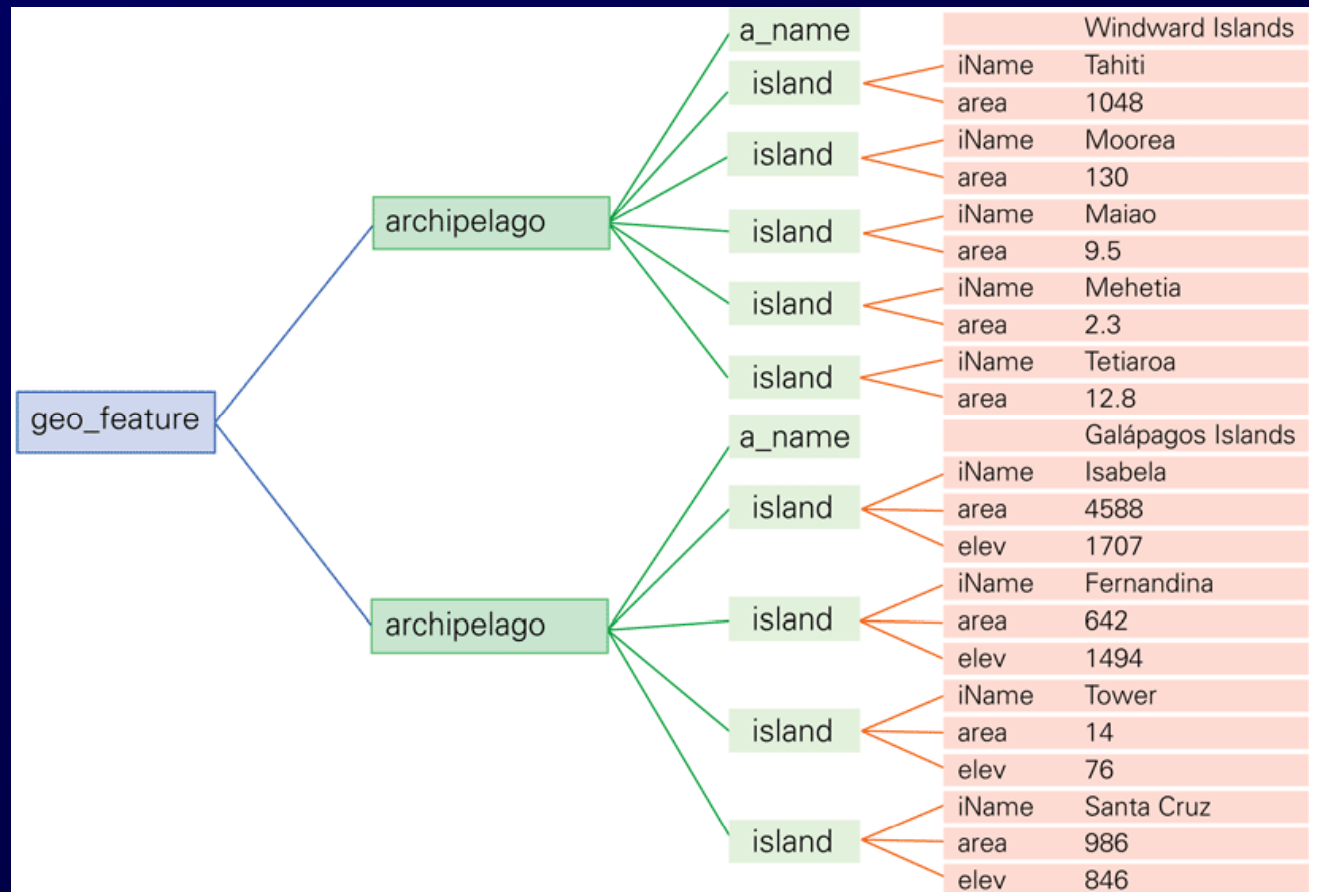
Figure 16.3 The XML displayed as a tree. The encoding from Figure 16.2 is shown with the root element (geo_feature) to the left and the leaves (content) shown to the right.

DATABASES

What is a Database

- Any organized collection of data
- A collection of similar data
- Examples of databases:
  - Telephone book white pages
  - T.V. Guide
  - Airline reservation system
  - Motor vehicle registration records
  - Papers in your filing cabinet
  - Files on your computer hard drive

Data | Information | Knowledge

- Data
  - ∗ Can be defined in many ways
  - ∗ IS defines data as unprocessed information
- Information
  - ∗ Data that have been organized and communicated in a coherent and meaningful manner
- Knowledge
  - ∗ Knowledge—information evaluated and organized so that it can be used purposefully

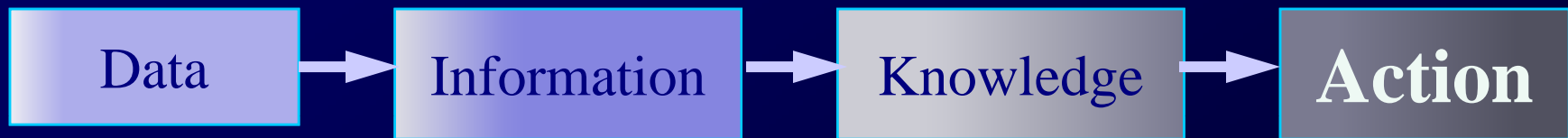Data is converted into information, and information is converted into knowledge

Data vs. Information

- We collect data
- Information is harvested from data

- Many companies are good at collecting data
- Fewer are good at harvesting information

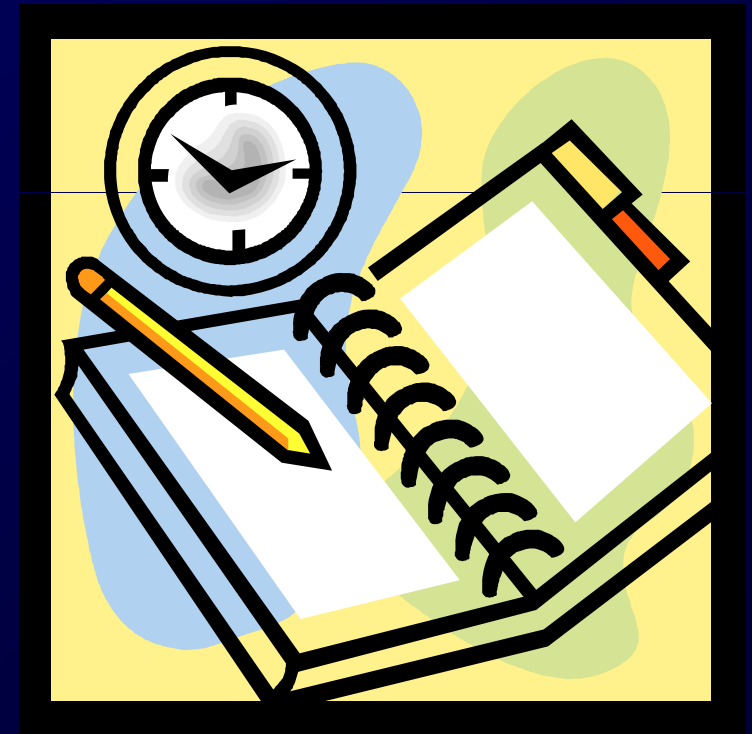# Ultimate Purpose of a Database Management System (DBMS)

To transform

Data → Information → Knowledge → **Action**

Why do we need a database?

- Keep records of our:
  - Clients
  - Staff
  - Volunteers
- To keep a record of activities and interventions
- Keep sales records
- Develop reports
- Perform research

Database Terminology

Phone book:

Records (rows)

Fields (columns)

Anderson	Thomas	A	123 Marine View Dr.	237-1234
Benson	Karen	C	1300 California Ave	237-1098
Casserly	Rick	W	12492 Rd 19	342-0502
Drummond	Lynn	M	12059 30th Ave W	931-1105

Table

Field (the columns in a table)	• Smallest unit of information in a table • Sometime called "attributes"	• First name • Last name • Middle initial • Street address • Phone number(s)
Record (the rows in a table)	• All related fields are collectively called a record	• All fields for one person are a record
Table	• A collection of records is a data table	• Collection of everyone's records
Database Management System (DBMS)	• All the related tables, queries, data entry and edit forms, reports, macros and VBA modules constitute a database	

Database Management System (DBMS)

- Software tools for working with data
- Designed to:
  - ∗ Store (tables)
  - ∗ Organize (sort)
  - ∗ Add, modify or delete
  - ∗ Ask questions (queries)
  - ∗ Produce forms and reports
    - Summarizing
    - Displaying details
- Toolbox is a good analogy

DBMS Examples

- Microsoft Access
- Structured Query Language (SQL)
  * Microsoft SQL Server
  * Oracle
  * MySQL
- FileMaker Pro
- Lotus Notes
- Open Office Base

Flat-File vs. Relational Database

- **Flat-File Database**
  - All relevant data in a single table, or series of unrelated tables
  - Work best for small quantities of data; where viewing and sorting the data in a single list does not create a time-consuming task
  - Typically a person's first databases
  - Example: Excel spreadsheet or Word data list file
- **Relational Database**
  - Provide a solution to data entry redundancy problems
  - Linked through common fields (columns) with exactly the same data
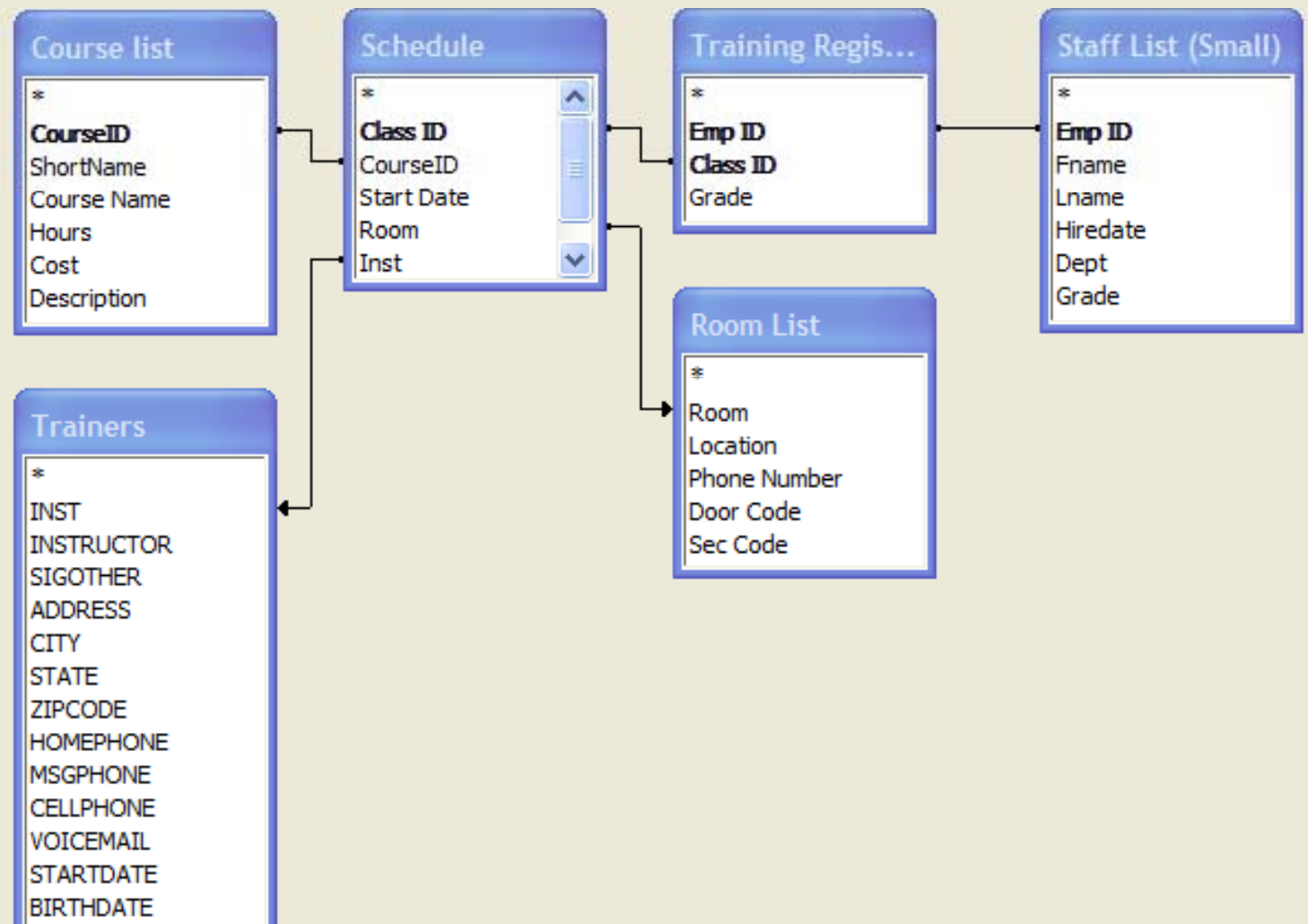 - Tables linked together can be queried as if one table
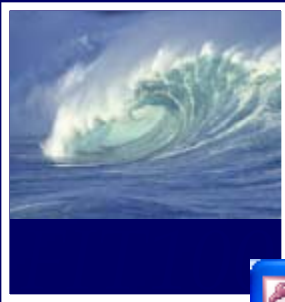
Flat-File Example

Staff Telephone List

Last Name	First Name	Emp ID	Dept	Location	Work Phone	M/S	Supervisor Name	Supr Phone
Adams	Wes	19589	PROD	Seattle	(206) 221-1958	QR-07	Susan Buckle	(206) 221-2241
Alberts	George	21533	PROD	Seattle	(206) 221-2153	QR-35	Marsha Mosley	(206) 221-1975
Allen	Susan	20256	PROD	Renton	(206) 393-2025	PB-18	Frank Sullivan	(206) 393-1000
Allert	Maria	10544	PROD	Seattle	(206) 221-1054	QR-27	Lynn Jarret	(206) 221-1366
Andrews	Mike	22113	PROD	Seattle	(206) 221-2211	QR-12	Harry Hillis	(206) 221-2179
Apperly	Ward	12244	PROD	Renton	(206) 393-1224	PB-14	Molly Goldberg	(206) 393-1513
Arthur	Diane	12370	MKT G	Bellevue	(206) 882-1237	RL-27	Wes Adams	(206) 221-1958
Asher	Jane	11222	ACCT	Seattle	(206) 221-1122	BX-45	Val Johnson	(206) 221-1958
Astor	Lawrence	20286	PROD	Seattle	(206) 221-2028	QR-10	Peggy Kramer	(206) 221-2083
Ayres	William	22263	PROD	Seattle	(206) 221-2226	QR-10	P. Kramer	(206) 221-2083
Baker	Gerald	19042	ACCT	Seattle	(206) 221-1904	BX-45	Valerie Johnson	(206) 221-1958

- Weaknesses common to flat-file systems
  * Duplicate information in the table
  * Inconsistencies in the way Supervisor Names are entered

Relational Database Example

Course list
- *
- **CourseID**
- ShortName
- Course Name
- Hours
- Cost
- Description

Schedule
- *
- **Class ID**
- CourseID
- Start Date
- Room
- Inst

Training Regis...
- *
- **Emp ID**
- **Class ID**
- Grade

Staff List (Small)
- *
- **Emp ID**
- Fname
- Lname
- Hiredate
- Dept
- Grade

Room List
- *
- Room
- Location
- Phone Number
- Door Code
- Sec Code

Trainers
- *
- INST
- INSTRUCTOR
- SIGOTHER
- ADDRESS
- CITY
- STATE
- ZIPCODE
- HOMEPHONE
- MSGPHONE
- CELLPHONE
- VOICEMAIL
- STARTDATE
- BIRTHDATE

Database Tables

Query from Two Tables

qryStaffListSmallSortedByName : Select Query

Emp ID	Lname	Fname	Hiredate	Dept	Grade	Salary
19589	Adams	Wes	0/23/1997	PROD	22	$3,812.55
21533	Alberts	George	2/18/2003	PROD	4	$1,611.05
20256	Allen	Susan	8/21/1999	PROD	14	$2,761.80
10544	Allert	Maria	1/17/1973	PROD	2	$1,380.90
22184	Ally	Kim	1/30/2004	PROD	3	$1,495.98
22113	Andrews	Mike	9/20/2004	PROD	13	$2,646.72
12244	Apperly	Ward	9/13/1977	PROD	15	$2,876.88
12370	Arthur	Diane	1/17/1978	MKTG	20	$3,452.25
11222	Asher	Jane	1/26/1974	ACCT	18	$3,222.10
20286	Astor	Lawrence	9/20/1999	PROD	5	$1,726.12
22263	Ayres	William	2/17/2005	PROD	1	$1,265.83
19042	Baker	Gerald	4/24/1996	ACCT	6	$1,841.20
17996	Barker	William	6/13/1993	PROD	9	$2,186.42

Record: 1 of 321

qryStaffListSmallSortedByName : Select Query

Staff List (Small)
- *
- **Emp ID**
- Fname
- Lname
- Hiredate
- Dept
- Grade

Staff Salary Data
- *
- **Grade**
- Salary

Field:	Emp ID	Lname	Fname	Hiredate	Dept	Grade	Salary
Table:	Staff List (Small)	Staff List (Small)	Staff List (Small)	Staff List (Small)	Staff List (Small)	Staff List (Small)	Staff Salary Data
Sort:		Ascending	Ascending				
Show:	☑	☑	☑	☑	☑	☑	☑
Criteria:							
or:							

Forms

Reports

Training Records by Employee

Emp ID	Last Name	First Name	Course Name	Start Date	Grade	Room	Instructor	Dept	Hours	Cost
19589	Adams	Wes	Microsoft Access, Level 1	2/12/2006	3.4	G218	Terrie Urbas	PROD	7	$145.00
			Microsoft Access, Level 2	2/13/2006	3.4	G218	Terrie Urbas	PROD	7	$145.00
			Microsoft Access, Level 3	2/14/2006	3.4	G218	Terrie Urbas	PROD	7	$145.00
			Microsoft Access, Level 4	4/9/2006	3.6	G218	Terrie Urbas	PROD	7	$145.00
			Microsoft Access, Level 5	4/10/2006	3.6	G218	Terrie Urbas	PROD	7	$145.00
			Microsoft Windows 95, Level 1	9/2/2006	3.1	R121	Bob Larson	PROD	7	$145.00
21533	Alberts	George	Microsoft Excel, Level 4	4/29/2006	2.7	G218	Andrea Forster	PROD	7	$145.00
			Introduction to Computers	6/5/2005	3.7	G219	Dan McIalwain	PROD	7	$145.00
			Microsoft Excel, Level 1	7/2/2005	2.6	G218	Bob Larson	PROD	7	$145.00
			Microsoft Excel, Level 2	7/3/2005	2.6	G218	Bob Larson	PROD	7	$145.00
			Microsoft Excel, Level 3	7/4/2005	2.6	G218	Bob Larson	PROD	7	$145.00
			Microsoft Access, Level 1	6/18/2006	3.4	G218	Bob Larson	PROD	7	$145.00
			Microsoft Access, Level 2	6/19/2006	3.4	G218	Bob Larson	PROD	7	$145.00
			Microsoft Access, Level 3	6/20/2006	3.4	G218	Bob Larson	PROD	7	$145.00
			Microsoft Access, Level 4	8/13/2006	3.6	G218	Bob Larson	PROD	7	$145.00
			Microsoft Access, Level 5	8/14/2006	3.6	G218	Bob Larson	PROD	7	$145.00
			Microsoft Windows 95, Level 1	11/2/2006	3.1	R121	Doug Hitchman	PROD	7	$145.00
20256	Allen	Susan	Microsoft Word, Level 1	6/2/2006	4.0	R123	Sally Larson	PROD	7	$145.00
			Microsoft Word, Level 2	6/3/2006	4.0	R123	Sally Larson	PROD	7	$145.00
			Microsoft Word, Level 3	6/4/2006	4.0	R123	Sally Larson	PROD	7	$145.00
			Microsoft PowerPoint, Level 1	7/31/2006	3.9	G107	Bob Larson	PROD	7	$145.00
			Microsoft Excel, Level 1	3/29/2007				PROD	7	$145.00
			Microsoft Excel, Level 2	3/30/2007				PROD	7	$145.00
			Microsoft Excel, Level 3	3/31/2007				PROD	7	$145.00
10544	Allert	Maria	Microsoft Windows 95, Level 1	8/26/2005	3.0	R121	Doug Hitchman	PROD	7	$145.00
			Microsoft Windows 95, Level 2	12/16/2005	3.2	R121	Doug Hitchman	PROD	7	$145.00

Dept: Accounting & Finance   **Vice President:** Valerie Smith

Number of Employees:	28
Total Monthly Salaries:	$59,048.53
Average Monthly Salary:	$2,108.88
Largest Monthly Salary:	$3,812.55
Lowest Monthly Salary:	$1,265.83
Average Pay Grade:	8

Employee	Emp ID	Location	Work Phone	Mail Stop	Budget
Asher, Jane	11222	Seattle	(206) 221-1122	BX-45	A-1834
Baker, Gerald	19042	Seattle	(206) 221-1904	BX-45	A-1834
Carrera, Barbara	20002	Renton	(425) 393-2000	BB-27	A-1834
Ewing, Robert	12672	Seattle	(206) 221-1267	BX-45	A-1834
Fairchild, Earl	16332	Renton	(425) 393-1633	BB-27	A-1834
Farmer, Lou	14082	Seattle	(206) 221-1408	BX-45	A-1834
Giles, Peter	12752	Seattle	(206) 221-1275	BX-45	A-1834
Graham, Margaret	21902	Seattle	(206) 221-2190	BX-45	A-1834
Graves, Bert	10702	Seattle	(206) 221-1070	BX-45	A-1834
Henderson, Peter	18572	Seattle	(206) 221-1857	BX-45	A-1834
Hickok, Joe	21752	Renton	(425) 393-2175	BB-27	A-1834
Hoover, Toni	22062	Seattle	(206) 221-2206	BX-45	A-1834
Isaacs, Rick	18412	Renton	(425) 393-1841	BB-27	A-1834
Lautenbach, Duane	21932	Seattle	(206) 221-2193	BX-45	A-1834
McGinnis, Gerald	21592	Renton	(425) 393-2159	BB-27	A-1834
Murray, Beverly	19932	Renton	(425) 393-1993	BB-27	A-1834
Perkins, Leslie	15602	Seattle	(206) 221-1560	BX-45	A-1834
Randall, Maxine	20742	Renton	(425) 393-2074	BB-27	A-1834
Roberts, Dick	19622	Seattle	(206) 221-1962	BX-45	A-1834
Robertson, Georgia	11582	Seattle	(206) 221-1158	BX-45	A-1834
Sample, Roger	21942	Renton	(425) 393-2194	BB-27	A-1834
Selleck, Shirley	21962	Renton	(425) 393-2196	BB-27	A-1834
Smith, Steve	12002	Renton	(425) 393-1200	BB-27	A-1834
Smith, Valerie	15002	Seattle	(206) 221-1500	BX-45	A-1834
Strassberger, Tom	22503	Seattle	(206) 221-2279	BX-45	A-1834
Thompson, Tom	15992	Seattle	(206) 221-1599	BX-45	A-1834
Wentworth, Cathy	11972	Renton	(425) 393-1197	BB-27	A-1834
Zoom, Zelda	22517	Renton	(425) 393-2326	BB-27	A-1834

RELATIONAL DATABASES

Video

- <u>Relational databases and tables</u>

Relational Database Model

- Introduced by E. F. Codd in 1970
- A Logical View of Data
  - Enables developer to view data logically rather than physically
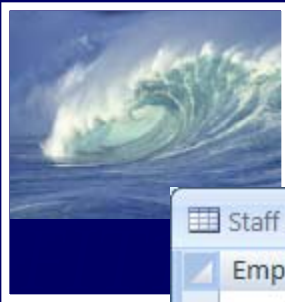 - Greater logical simplicity tends to yield simpler and more effective database design methodologies

Tables

Emp ID	Last Name	First Name	Address	City	State	Zip	Telephone
19589	Adams	Wes	3132 C N. E.	Auburn	WA	98002	(253) 833-1958
21533	Alberts	George	1819 Westlake Ave. N.	Seattle	WA	98109	(206) 452-2153
20256	Allen	Susan	17314 140th Ave S. E.	Renton	WA	98058	(425) 226-2025
10544	Allert	Maria	865 Lind S. W.	Renton	WA	98055	(425) 227-1054
22184	Ally	Kim	2904 A St. S. E.	Auburn	WA	98002	(253) 833-2218
22113	Andrews	Mike	23605 - 156th S.E.	Kent	WA	98042	(253) 872-2211

Record: 14 of 321 — Unfiltered — Search

- Cornerstone of Relational DBMS
- Advantages – structural and data independence
- Conceptually Resembles a file
  - Note a file is actually a physical structure
- Easier to understand than its hierarchical and network database predecessors

Table Characteristics

	Staff List								
	Emp ID	Last Name	First Name	Address	City	State	Zip	Telephone	
	19589	Adams	Wes	3132 C N. E.	Auburn	WA	98002	(253) 833-1958	
	21533	Alberts	George	1819 Westlake Ave. N.	Seattle	WA	98109	(206) 452-2153	
	20256	Allen	Susan	17314 140th Ave S. E.	Renton	WA	98058	(425) 226-2025	
	10544	Allert	Maria	865 Lind S. W.	Renton	WA	98055	(425) 227-1054	
	22184	Ally	Kim	2904 A St. S. E.	Auburn	WA	98002	(253) 833-2218	
	22113	Andrews	Mike	23605 - 156th S.E.	Kent	WA	98042	(253) 872-2211	

Record: 14 of 321 — Unfiltered — Search

1	A table is perceived as a two-dimensional structure composed of rows and columns.
2	Each table row (**tuple**) represents a single entity occurrence within the entity set.
3	Each table column represents an attribute, and each column has a distinct name.
4	Each row/column intersection represents a single data value.
5	All values in a column must conform to the same data format. For example, if the attribute is assigned an integer data format, all values in the column representing that attribute must be integers.
6	Each column has a specific range of values known as the **attribute domain**.
7	The order of the rows and columns is immaterial to the DBMS.
8	Each table must have an attribute or a combination of attributes that uniquely identifies each row.