Blogging is:

A. Someone's online journal.
B. A Celtic dance with wooden shoes.
C. How the Celtics keep the Knicks away from the ball.
Blogging is:

A. Someone's online journal.
B. A Celtic dance with wooden shoes.
C. How the Celtics keep the Knicks away from the ball.
Announcements

- Project 2B due tonight at 10pm
  - I will join the 1:30 drop-in lab in MGH 430 right after lecture
  - CLUE Tutoring tonight at 7pm in MGH 058
Data Storage and Transfer with XML and Databases

D.A. Clements
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
• 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
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 Markup Language

XML
An Example from Tahiti

- Area in km² for Tahiti & neighboring islands

```xml
<?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
  <?xml version="1.0" encoding="ISO-8859-1" ?>
  ```
- File should be ASCII text
- File extension should be `.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 version="1.0"
    encoding="ISO-8859-1" ?>
<geo_feature>
    <archipelago>
        <a_name>Galapagos Islands</a_name>
    </archipelago>
    <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>
</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:
    
    ```xml
    <archipelago name="Galapagos">
    </archipelago>
    ```

  - Better to give alternate form of the data
    
    ```xml
    <a_name
    accents="Gal&aacute;pagos">Galapagos</a_name>
    ```
Identification Rule: Label Data with Tags Consistently

You can choose whatever tag names you wish 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.
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
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)
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
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:

<table>
<thead>
<tr>
<th>Records (rows)</th>
<th>Fields (columns)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anderson Thomas A 123 Marine View Dr. 237-1234</td>
<td></td>
</tr>
<tr>
<td>Benson Karen C 1300 California Ave 237-1098</td>
<td></td>
</tr>
<tr>
<td>Casserly Rick W 12492 Rd 19 342-0502</td>
<td></td>
</tr>
<tr>
<td>Drummond Lynn M 12059 30th Ave W 931-1105</td>
<td></td>
</tr>
</tbody>
</table>

#### 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
Ultimate Purpose of a Database Management System (DBMS)

To transform

Data → Information → Knowledge → Action
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
- Can answer very complex questions
### Flat-File Example

#### Staff Telephone List

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

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


Objects
- Tables
  - Create table in Design view
  - Create table by using wizard
  - Create table by entering data
  - Course List
  - Room List
  - Schedule
  - Staff List
  - Staff List (Small)
  - Staff Salary Data
  - Trainers
  - Training Register
  - xtblCourseManuals
  - xtblDeptCodes

Staff List (Small) : Table
<table>
<thead>
<tr>
<th>Emp ID</th>
<th>Fname</th>
<th>Lname</th>
<th>Hiredate</th>
<th>Dept</th>
<th>Grade</th>
</tr>
</thead>
<tbody>
<tr>
<td>19588</td>
<td>Wes</td>
<td>Adams</td>
<td>0/23/1997</td>
<td>PROD</td>
<td>22</td>
</tr>
<tr>
<td>21533</td>
<td>George</td>
<td>Alberts</td>
<td>2/18/2003</td>
<td>PROD</td>
<td>4</td>
</tr>
<tr>
<td>20256</td>
<td>Susan</td>
<td>Allen</td>
<td>8/21/1999</td>
<td>PROD</td>
<td>14</td>
</tr>
<tr>
<td>10544</td>
<td>Maria</td>
<td>Allert</td>
<td>1/17/1973</td>
<td>PROD</td>
<td>2</td>
</tr>
<tr>
<td>22184</td>
<td>Kim</td>
<td>Ally</td>
<td>1/30/2004</td>
<td>PROD</td>
<td>3</td>
</tr>
<tr>
<td>22113</td>
<td>Mike</td>
<td>Andrews</td>
<td>9/20/2004</td>
<td>PROD</td>
<td>13</td>
</tr>
<tr>
<td>12244</td>
<td>Ward</td>
<td>Apperly</td>
<td>9/13/1977</td>
<td>PROD</td>
<td>15</td>
</tr>
<tr>
<td>12370</td>
<td>Diane</td>
<td>Arthur</td>
<td>1/17/1978</td>
<td>MKTG</td>
<td>20</td>
</tr>
<tr>
<td>11222</td>
<td>Jane</td>
<td>Asher</td>
<td>1/26/1974</td>
<td>ACCT</td>
<td>18</td>
</tr>
<tr>
<td>20286</td>
<td>Lawrence</td>
<td>Astor</td>
<td>9/20/1999</td>
<td>PROD</td>
<td>5</td>
</tr>
<tr>
<td>22263</td>
<td>William</td>
<td>Ayres</td>
<td>2/17/2005</td>
<td>PROD</td>
<td>1</td>
</tr>
<tr>
<td>19042</td>
<td>Gerald</td>
<td>Baker</td>
<td>4/24/1996</td>
<td>ACCT</td>
<td>6</td>
</tr>
<tr>
<td>17996</td>
<td>William</td>
<td>Barker</td>
<td>6/13/1993</td>
<td>PROD</td>
<td>9</td>
</tr>
</tbody>
</table>

Record: 1 of 321
Query from Two Tables
Forms

XYZ Company
Employee Salary Data

Emp ID: 19589  Dept: PROD
Last Name: Adams  First Name: Wes
Hire Date: 10/23/1997  Years: 9.2
Grade: 22  Monthly Salary: $3,812.55
Annual Salary: $45,750.60  Hourly Rate: $21.91

XYZ Department Summary
Dept: Accounting & Finance

Dept Code: ACCT  Building: A

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

Valerie Smith
Vice President
# Training Records by Employee

<table>
<thead>
<tr>
<th>Emp ID</th>
<th>Last Name</th>
<th>First Name</th>
<th>Course Name</th>
<th>Start Date</th>
<th>Grade</th>
<th>Room</th>
<th>Instructor</th>
<th>Dept</th>
<th>Hours</th>
<th>Cost</th>
</tr>
</thead>
<tbody>
<tr>
<td>95589</td>
<td>Adams</td>
<td>Wes</td>
<td>Microsoft Access, Level 1</td>
<td>2/12/2005</td>
<td>3.4</td>
<td>G210</td>
<td>Tamir Urias</td>
<td>HRIO</td>
<td>7</td>
<td>$147.00</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Microsoft Access, Level 2</td>
<td>2/13/2005</td>
<td>3.4</td>
<td>G210</td>
<td>Tamir Urias</td>
<td>HRIO</td>
<td>7</td>
<td>$147.00</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Microsoft Access, Level 3</td>
<td>2/14/2005</td>
<td>3.4</td>
<td>G210</td>
<td>Tamir Urias</td>
<td>HRIO</td>
<td>7</td>
<td>$147.00</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Microsoft Access, Level 4</td>
<td>4/16/2005</td>
<td>3.6</td>
<td>G210</td>
<td>Tamir Urias</td>
<td>HRIO</td>
<td>7</td>
<td>$147.00</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Microsoft Access, Level 5</td>
<td>4/17/2005</td>
<td>3.6</td>
<td>G210</td>
<td>Tamir Urias</td>
<td>HRIO</td>
<td>7</td>
<td>$147.00</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Microsoft Windows 95, Level 1</td>
<td>9/20/2005</td>
<td>5.1</td>
<td>R121</td>
<td>Bob Larson</td>
<td>HRIO</td>
<td>7</td>
<td>$147.00</td>
</tr>
<tr>
<td>21533</td>
<td>Altel</td>
<td>George</td>
<td>Microsoft Excel, Level 1</td>
<td>4/25/2005</td>
<td>2.7</td>
<td>G210</td>
<td>Amia Fander</td>
<td>HRIO</td>
<td>7</td>
<td>$147.00</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Microsoft Excel, Level 2</td>
<td>7/20/2005</td>
<td>2.6</td>
<td>G210</td>
<td>Bob Larson</td>
<td>HRIO</td>
<td>7</td>
<td>$147.00</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Microsoft Excel, Level 3</td>
<td>7/24/2005</td>
<td>2.6</td>
<td>G210</td>
<td>Bob Larson</td>
<td>HRIO</td>
<td>7</td>
<td>$147.00</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Microsoft Excel, Level 4</td>
<td>6/19/2005</td>
<td>3.4</td>
<td>G210</td>
<td>Bob Larson</td>
<td>HRIO</td>
<td>7</td>
<td>$147.00</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Microsoft Excel, Level 5</td>
<td>6/20/2005</td>
<td>3.4</td>
<td>G210</td>
<td>Bob Larson</td>
<td>HRIO</td>
<td>7</td>
<td>$147.00</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Microsoft Windows 95, Level 1</td>
<td>9/20/2005</td>
<td>5.1</td>
<td>R121</td>
<td>Doug Russell</td>
<td>HRIO</td>
<td>7</td>
<td>$147.00</td>
</tr>
<tr>
<td>20246</td>
<td>Allen</td>
<td>Susan</td>
<td>Microsoft Word, Level 1</td>
<td>6/20/2005</td>
<td>4.0</td>
<td>R123</td>
<td>Sally Larson</td>
<td>HRIO</td>
<td>7</td>
<td>$147.00</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Microsoft Word, Level 2</td>
<td>6/25/2005</td>
<td>4.0</td>
<td>R123</td>
<td>Sally Larson</td>
<td>HRIO</td>
<td>7</td>
<td>$147.00</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Microsoft Word, Level 3</td>
<td>6/26/2005</td>
<td>4.0</td>
<td>R123</td>
<td>Sally Larson</td>
<td>HRIO</td>
<td>7</td>
<td>$147.00</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Microsoft PowerPoint, Level 1</td>
<td>7/13/2005</td>
<td>3.9</td>
<td>G107</td>
<td>Bob Larson</td>
<td>HRIO</td>
<td>7</td>
<td>$147.00</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Microsoft Excel, Level 1</td>
<td>3/30/2007</td>
<td></td>
<td></td>
<td></td>
<td>HRIO</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Microsoft Excel, Level 2</td>
<td>3/30/2007</td>
<td></td>
<td></td>
<td></td>
<td>HRIO</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Microsoft Excel, Level 3</td>
<td>3/30/2007</td>
<td></td>
<td></td>
<td></td>
<td>HRIO</td>
<td></td>
<td></td>
</tr>
<tr>
<td>92844</td>
<td>Allert</td>
<td>Malta</td>
<td>Microsoft Windows 95, Level 1</td>
<td>6/25/2005</td>
<td>3.0</td>
<td>R121</td>
<td>Doug Kuhlman</td>
<td>HRIO</td>
<td>7</td>
<td>$144.00</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Microsoft Windows 95, Level 2</td>
<td>12/15/2005</td>
<td>5.2</td>
<td>R121</td>
<td>Doug Kuhlman</td>
<td>HRIO</td>
<td>7</td>
<td>$144.00</td>
</tr>
</tbody>
</table>

*Thursday, January 04, 2007*
RELATIONAL DATABASES
• Relational databases and tables