Definitions

- Database – an integrated set of data on a particular subject
- Geographic (=spatial) database - database containing geographic data of a particular subject for a particular area
- Database Management System (DBMS) - software to create, maintain and access databases
A DBMS is a software application designed to organize the efficient and effective storage and access of data. It separates the physical form of the data from any applications, allowing applications to be developed without changing the data structure.

Capabilities of DBMS:
- Include a data model, data load capability, indexes, a query language, security, controlled update, backup and recovery, database administration tools, applications and APIs.
- This list of DBMS capabilities is very attractive to GIS users and so, not surprisingly, virtually all large GIS databases are based on DBMS technology.

Advantages of Databases over Files:
- Avoids redundancy and duplication
- Reduces data maintenance costs
- Applications are separated from the data
- Applications persist over time
- Support multiple concurrent applications
- Better data sharing
- Security and standards can be defined and enforced

Disadvantages of Databases over Files:
- Expense
- Complexity
- Performance – especially complex data types
- Integration with other systems can be difficult

Types of DBMS Model:
- Hierarchical
- Network
- Relational - RDBMS
- Object-oriented - ODBMS
- Object-relational - ORDBMS

Hybrid object-relational DBMS
- ODBMS have not proven to be as commercially successful as some predicted because of the massive installed base of RDBMS. Thus appeared...
- Hybrid object-relational DBMS (ORDBMS). These can be thought of as an RDBMS engine with an extensibility framework for handling objects.
- The ideal geographic ORDBMS is one that has been extended to support geographic object types and functions through the addition of a geographic query parser, a geographic query optimizer, a geographic query language, multidimensional indexing services, storage management for large files, long transaction services, and replication services.
Two of the commercial DBMS vendors have released spatial database extensions to their standard ORDBMS products:

- IBM – DB2 Spatial Extender and Informix Spatial Datablade
- ORACLE - Oracle Spatial

Both are extensions to the DBMS software kernel to manage geographic data and to build support for spatial data types and functions into a middle-tier (or middleware) application server.

GIS middleware extensions

An alternative to extending the DBMS software kernel to manage geographic data is to build support for spatial data types and functions into a middle-tier (or middleware) application server.

- ESRI’s Spatial Data Engine (SDE)

Characteristics of DBMS (1)

- Data model support for multiple data types
  - e.g., MS Access: Text, Memo, Number, Date/Time, Currency, AutoNumber, Yes/No, OLE Object, Hyperlink, Lookup Wizard
- Load data from files, databases and other applications
- Index for rapid retrieval

Characteristics of DBMS (2)

- Query language – SQL
- Security – controlled access to data
  - Multi-level groups
- Controlled update using a transaction manager
- Backup and recovery
- DBA tools
  - Configuration, tuning

Characteristics of DBMS (3)

- Applications
  - CASE tools
  - Forms builder
  - Reportwriter
  - Internet Application Server
  - Programmable API

Relational DBMS (1)

- Data stored as tuples (tup-el), conceptualized as tables
- Table – data about a class of objects
  - Two-dimensional list (array)
  - Rows = objects
  - Columns = object states (properties, attributes)
Relational DBMS (2)

- Most popular type of DBMS
- Over 95% of data in DBMS is in RDBMS

Commercial systems
- IBM DB2
- Informix
- Microsoft Access
- Microsoft SQL Server
- Oracle
- Sybase

Table

<table>
<thead>
<tr>
<th>Column = property</th>
</tr>
</thead>
<tbody>
<tr>
<td>Table = Object Class</td>
</tr>
<tr>
<td>Row = object</td>
</tr>
<tr>
<td>Object Classses with Geometry called Feature Classes</td>
</tr>
</tbody>
</table>

Figure 10.2b

<table>
<thead>
<tr>
<th>Date/Assessed</th>
<th>Value</th>
<th>Zoning</th>
<th>Type</th>
<th>Code</th>
<th>Postal</th>
<th>Owner Address</th>
<th>Owner Name</th>
<th>Parcel Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>2005 350000</td>
<td>Residential</td>
<td>211439</td>
<td>110034</td>
<td>110034</td>
<td>110034</td>
<td>Sheila Sullivan</td>
<td>674-113</td>
<td></td>
</tr>
<tr>
<td>2004 450575</td>
<td>Residential</td>
<td>211439</td>
<td>119171</td>
<td>119171</td>
<td>119171</td>
<td>Sam Camarata</td>
<td>670-231</td>
<td></td>
</tr>
<tr>
<td>2001 290000</td>
<td>Commercial</td>
<td>311439</td>
<td>1452145</td>
<td>1452145</td>
<td>1452145</td>
<td>D Widseler</td>
<td>674-100</td>
<td></td>
</tr>
<tr>
<td>2000 275500</td>
<td>Commercial</td>
<td>311439</td>
<td>1452145</td>
<td>1452145</td>
<td>1452145</td>
<td>D Widseler</td>
<td>674-100</td>
<td></td>
</tr>
<tr>
<td>1999 249000</td>
<td>Commercial</td>
<td>311439</td>
<td>1452145</td>
<td>1452145</td>
<td>1452145</td>
<td>D Widseler</td>
<td>674-100</td>
<td></td>
</tr>
<tr>
<td>2003 545500</td>
<td>Residential</td>
<td>211439</td>
<td>1151115</td>
<td>1151115</td>
<td>1151115</td>
<td>Joel Campbell</td>
<td>673-101</td>
<td></td>
</tr>
<tr>
<td>2002 220000</td>
<td>Residential</td>
<td>211439</td>
<td>1010110</td>
<td>1010110</td>
<td>1010110</td>
<td>Jeff Peters</td>
<td>673/100</td>
<td></td>
</tr>
<tr>
<td>2001 220000</td>
<td>Residential</td>
<td>211439</td>
<td>1010110</td>
<td>1010110</td>
<td>1010110</td>
<td>Jeff Peters</td>
<td>673/100</td>
<td></td>
</tr>
<tr>
<td>1999 249000</td>
<td>Commercial</td>
<td>311439</td>
<td>1010110</td>
<td>1010110</td>
<td>1010110</td>
<td>D Widseler</td>
<td>674-100</td>
<td></td>
</tr>
</tbody>
</table>

Figure 10.3A

<table>
<thead>
<tr>
<th>Parcel Number</th>
<th>Description</th>
<th>Description Address</th>
<th>Parcel Code</th>
<th>Zoning Code</th>
<th>Zoning Type</th>
<th>Date/Assessed</th>
</tr>
</thead>
<tbody>
<tr>
<td>02 1005425</td>
<td>Residential</td>
<td>211439 110034 110034</td>
<td>12345</td>
<td>12345</td>
<td>12345</td>
<td>2005 350000</td>
</tr>
<tr>
<td>02 1005425</td>
<td>Residential</td>
<td>211439 110034 110034</td>
<td>12345</td>
<td>12345</td>
<td>12345</td>
<td>2004 450575</td>
</tr>
<tr>
<td>02 1005425</td>
<td>Residential</td>
<td>211439 110034 110034</td>
<td>12345</td>
<td>12345</td>
<td>12345</td>
<td>2001 290000</td>
</tr>
<tr>
<td>02 1005425</td>
<td>Commercial</td>
<td>311439 1452145 1452145</td>
<td>12345</td>
<td>12345</td>
<td>12345</td>
<td>2000 275500</td>
</tr>
<tr>
<td>02 1005425</td>
<td>Commercial</td>
<td>311439 1452145 1452145</td>
<td>12345</td>
<td>12345</td>
<td>12345</td>
<td>1999 249000</td>
</tr>
<tr>
<td>02 1005425</td>
<td>Residential</td>
<td>211439 1151115 1151115</td>
<td>12345</td>
<td>12345</td>
<td>12345</td>
<td>2003 545500</td>
</tr>
<tr>
<td>02 1005425</td>
<td>Residential</td>
<td>211439 1010110 1010110</td>
<td>12345</td>
<td>12345</td>
<td>12345</td>
<td>2002 220000</td>
</tr>
<tr>
<td>02 1005425</td>
<td>Residential</td>
<td>211439 1010110 1010110</td>
<td>12345</td>
<td>12345</td>
<td>12345</td>
<td>2001 220000</td>
</tr>
<tr>
<td>02 1005425</td>
<td>Commercial</td>
<td>311439 1010110 1010110</td>
<td>12345</td>
<td>12345</td>
<td>12345</td>
<td>1999 249000</td>
</tr>
</tbody>
</table>

Figure 10.3B
Figure 10.3C

Figure 10.3D Joined table

Fig 10.4

Relation Rules (Codd, 1970)
- Only one value in each cell (intersection of row and column)
- All values in a column are about the same subject
- Each row is unique
- No significance in column sequence
- No significance in row sequence

Normalization
- Process of converting tables to conform to Codd’s relational rules
- Split tables into new tables that can be joined at query time
  - The relational join
  - Several levels of normalization
    - Forms: 1NF, 2NF, 3NF, etc.
- Normalization creates many expensive joins
- De-normalization is OK for performance optimization

Relational Join
- Fundamental query operation
- Occurs because
  - Normalization
  - Data created/maintained by different users, but integration needed for queries
- Table joins use common keys (column values)
- Table (attribute) join concept has been extended to geographic case
SQL
- Structured (Standard) Query Language - (pronounced SEQUEL)
- Developed by IBM in 1970s
- Now de facto and de jure standard for accessing relational databases
- Three types of usage
  - Stand alone queries
  - High level programming
  - Embedded in other applications

Types of SQL Statements
- Data Definition Language (DDL)
  - Create, alter and delete data
  - CREATE TABLE, CREATE INDEX
- Data Manipulation Language (DML)
  - Retrieve and manipulate data
  - SELECT, UPDATE, DELETE, INSERT
- Data Control Languages (DCL)
  - Control security of data
  - GRANT, CREATE USER, DROP USER

Geometry Class Hierarchy

Spatial Relations
- Equals - same geometries
- Disjoint - geometries share common point
- Intersects - geometries intersect
- Touches - geometries intersect at common boundary
- Crosses - geometries overlap
- Within - geometry within
- Contains - geometry completely contains
- Overlaps - geometries of same dimension overlap
- Relate - intersection between interior, boundary or exterior

Spatial Methods
- Distance - shortest distance
- Buffer - geometric buffer
- ConvexHull - smallest convex polygon geometry
- Intersection - points common to two geometries
- Union - all points in geometries
- Difference - points different between two geometries
- SymDifference - points in either, but not both of input geometries
Examples of spatial analysis methods on geometries

Four levels of data model available for use in GIS projects.

Stages in Database Design
- Reality
- Conceptual
- Logical
- Physical

Topology
- Two main database approaches
  - Normalized
    - Arc-node primitives
  - Physical
    - Simple features + rules

Normalized Database Topology Model

Physical Database Topology Model
Indexing

- Used to locate rows quickly
- RDBMS use simple 1-d indexing (R-tree, B-tree, etc.)
- Spatial DBMS need 2-d, hierarchical indexing
  - Grid
  - Quadtree
  - R-tree
  - Others
- Multi-level queries often used for performance (MBR)
Summary

- Database – an integrated set of data on a particular subject
- Databases offer many advantages over files
- Relational databases dominate
- Some limitations for GIS