11. Creating and Maintaining Geographic Databases

Geographical Information Systems and Science
Longley P A, Goodchild M F, Maguire D J, Rhind D W
(2001) John Wiley and Sons Ltd

Outline

- Definitions
- Characteristics of DBMS
- Types of database
- Relational model
- SQL
- Spatial databases
- Indexing methods

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
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 - OODBMS
- Object-relational - ORDBMS
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
Role of DBMS System

- Geographic Information System
- Database Management System
- Data

Task
- Data load
- Editing
- Visualization
- Mapping
- Analysis
- Storage
- Indexing
- Security
- Query

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)

Table
- Column = property
- Row = object

Object Classes with Geometry called Feature Classes
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

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

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

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

Grid Index (multi-level)
Summary

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