10 Creating and Maintaining Geographic Databases

OVERVIEW

- After people, the database is arguably the most important part of a GIS because of the costs of collection and maintenance, and because the database forms the basis of all queries, analysis, and decision making.

- Today, virtually all large GIS implementations store data in a database management system (DBMS), a specialist piece of software designed to handle multi-user access to an integrated set of data.

- Databases need to be designed with great care, and to be structured and indexed to provide efficient query and transaction performance.

- A comprehensive security and transactional access model is necessary to ensure that multiple users can access the database at the same time.

- On-going maintenance is also an essential, but very resource-intensive, activity.

LEARNING OBJECTIVES

By the end of this chapter students should:

- **Understand the role of database management systems in GIS;**

- **Recognize structured query language (SQL) statements;**

- **Understand the key geographic database data types and functions;**

- **Be familiar with the stages of geographic database design;**

- **Understand the key techniques for structuring geographic information, specifically creating topology and indexing;**

- **Understand the issues associated with multi-user editing and versioning.**

DBMS, (RDBMS, ODBMS, ORDBMS), parsers, middleware, object classes, database tables, keys, normal forms, SQL, SQL/MM, database design, indexes, B-tree indexes, grid indexes, quadtree indexes, R-tree indexes, database editing and update, transactions, long transactions, versioning

OVERVIEW

10.1 Introduction
10.2 Database management systems
10.3 Storing data in DBMS tables
10.4 SQL
10.5 Geographic database types and functions
10.6 Geographic database design
10.7 Structuring geographic information
10.8 Editing and data maintenance
10.9 Multi-user editing of continuous databases
10.10 Conclusion

CHAPTER SUMMARY

10.1 Introduction

- A database can be thought of as an integrated set of data on a particular subject.
- Geographic databases are simply databases containing geographic data for a particular area and subject.
- Lists the advantages of the database approach to storing geographic data over traditional file-based datasets including reducing redundancy, decreasing costs, allowing multiple applications, transfer of knowledge, data sharing, security and standards, concurrent users
- Disadvantages include cost, complexity, single user performance decreased
- Describes how to create and maintain geographic databases, and the concepts, tools, and techniques that are available to manage geographic data in databases.

10.2 Database management systems

- A DBMS is a software application designed to organize the efficient and effective storage and access of data.

- Briefly outlines the capabilities of DBMS which include a data model, a 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.

10.2.1 Types of DBMS

- Three main types of DBMS are available to GIS users today: relational (RDBMS), object (ODBMS), and object-relational (ORDBMS).
- A relational database comprises a set of tables, each a two-dimensional list (or array) of records containing attributes about the objects under study.
- Object database management systems (ODBMS) were initially designed to address weaknesses of RDBMS, including the inability to store complete objects directly in the database (both object state and behavior), poor performance for many types of geographic query
- 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) 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, replication services

10.2.2 Geographic DBMS extensions

- 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 Spatial
  - spatial capabilities in the core of Microsoft SQLServer
  - Opensource DBMS PostgreSQL has also been extended with spatial types and functions (PostGIS).
- None is a complete GIS software system
- Focus is on data storage retrieval and management
- Technical box 10.1 details Oracle Spatial

10.3 Storing data in DBMS tables

- The lowest level of user interaction with a geographic database is usually the *object class* (also called a layer or feature class), which is an organized collection of data on a particular theme
- Object classes are stored in a standard database *table*, a two-dimensional array of rows and columns.
    - Rows contain objects (*instances* of object classes)
    - Columns contain object properties or attributes
- The data stored at individual row, column intersections are usually referred to as values.
- Geographic database tables are distinguished from non-geographic tables by the presence of a geometry column (often called the shape column).
- To save space and improve performance, the actual coordinate values may be stored in a highly compressed binary form.
- Tables are joined together using common row/column values or *keys*.
- Following joins, all tables can be treated as a single table
- Lists Codd's five principles for the efficient and effective design of tables and introduces the concept of *normal forms*
- Normal forms improve the simplicity and stability of a database and reduce redundancy of tables by splitting them into sub-tables that are re-joined at query time
- Notes that large tables common in geographic applications leads to tendency for non-normalized table designs in GIS
- Includes a worked example of normalization of a simple land parcel tax assessment table

10.4 SQL

- The standard database query language adopted by virtually all mainstream databases is SQL (Structured or Standard Query Language: ISO Standard ISO/IEC 9075).
- May be used directly via command line, compiled in a general purpose programming language or via a GUI
- The third major revision of SQL (SQL 3) which came out in 2004 defines spatial types and functions as part of a multi-media extension called *SQL/MM*.
- There are three key types of SQL statements:
- DDL (data definition language) used to create, alter and delete relational database structures