Christopher Ashley Cline
INLS 723: Databases III
Arcot Rajasekar
2012-11-05

Current Technology Investigation: GIS Databases

Background

GIS stands for geographic information system. According to Wikipedia, at its most reduced level, GIS implies the combination of "cartography, statistical analysis, and database technology" (Geographic information system, 2012). At present, we are concerned with the database technology component of this combination, though most GIS systems rely upon a technology stack in order to function.

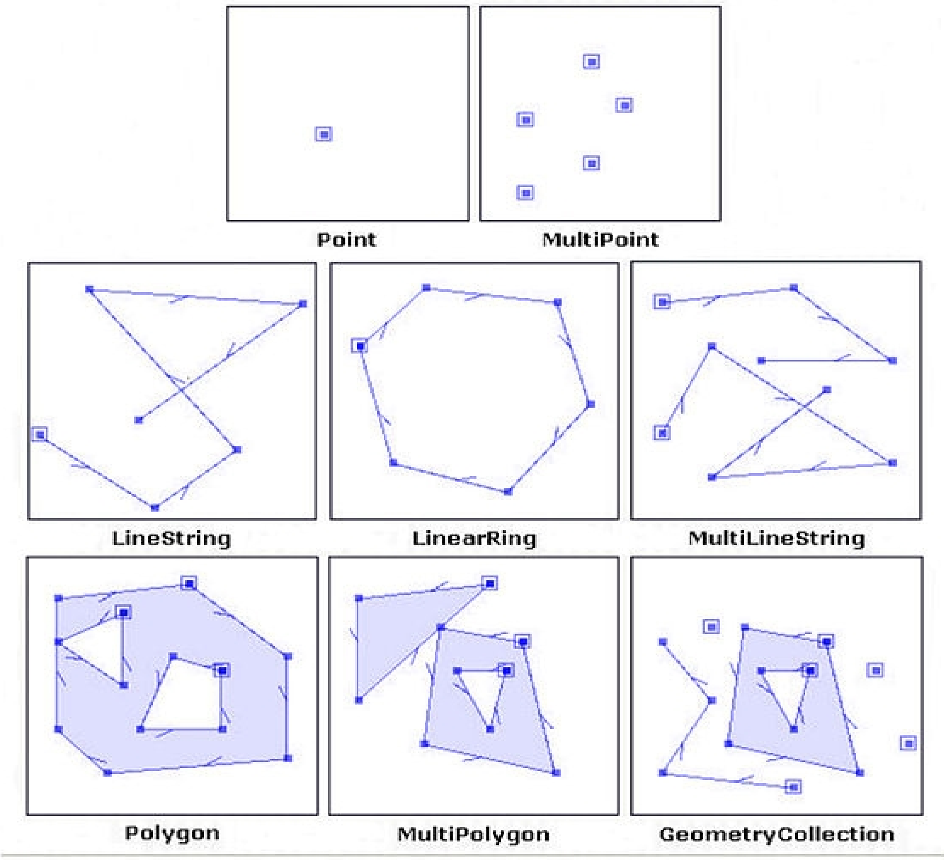
Because of this reliance upon an interconnected arrangement of technology components, much energy has been devoted to developing standards for the data types and operations performed using GIS data. While many standards organizations exist, the North Carolina Department of Transportation has its own GIS standards panel, for instance, the powerhouse in the world of GIS standards seems to be the Open Geospatial Consortium (NCDOT GIS Standards and Practices, n.d.; Open Geospatial Consortium, n.d.). The history of the OGC goes back to 1994, when the development of GIS standards first became a priority. The organization based its early decisions on design work with the Geographic Resource Analysis Support System (G.R.A.S.S.) and received interested buy-in from the Army Corps of Engineers, Department of Agriculture, and Sun Microsystems.

The most widely adopted standard produced by the OGC is the organization's specification for Simple Feature Access (Simple feature access, 2012, October 22; Herring, 2010). This standard lays out a baseline list of geometric object types and operations to be performed on those objects for geographic systems. It is a recognized standard both to the OGC and the ISO. Many database systems implement this standard, and the implementation requires a database system to support three new features: new data types, new operations, and new indices.

Data Types

The Simple Feature Access standard defines a storage model for geographical data according to 18 basic geometry types:

- Geometry
- Point
- MultiPoint
- LineString
- MultiLineString
- Polygon
- MultiPolygon
- Triangle
- CircularString
- Curve
- MultiCurve
- CompoundCurve
- CurvePolygon
- Surface
- MultiSurface
- PolyhedralSurface
- TIN (Triagnulated Irregular Network)
- GeometryCollection



Simple Feature Access Geometry Type Examples

These are stored or transferred in either Well-known text format, or Well-known binary format, with the latter being the common choice for database storage. Both are ISO-established standards. The Well-known text format is aptly named, providing an easily human-readable markup language for geometric objects. For instance a geometry point can be represented by the following: *POINT (25 50)*. Well-known binary format merely converts this markup into a numeric, machine-readable form. Many well-known database systems incorporate these standards, either natively or with the aid of additional plugins:

- Postgresql
- Oracle
- MySQL
- IBM DB2
- Microsoft SQL Server

Operations

Besides requiring data types specific to geospatial needs, GIS standards also require backend databases to support operations beyond those traditionally provided by SQL. Common queries relating to GIS data cannot be answered with SQL alone and require operators which perform mathematical calculations based upon stored geographic data. The PostGIS addon for Postgresql, for instance, provides the following methods for working with GIS data:

- Distance(geometry, geometry) : number
- Equals(geometry, geometry) : boolean
- Disjoint(geometry, geometry) : boolean
- Intersects(geometry, geometry) : boolean
- Touches(geometry, geometry) : boolean
- Crosses(geometry, geometry) : boolean
- Overlaps(geometry, geometry) : boolean
- Contains(geometry, geometry) : boolean
- Length(geometry) : number
- Area(geometry) : number
- Centroid(geometry) : geometry

The values in the parentheses represents the object-type parameter provided to the method, while the values after the colon describes the object type returned by the method call. These methods allow a query to easily ask, for example, whether or not a given two-dimensional space overlaps another two-dimensional, which is not something one could easily do with a basic SQL query.

Indices

Effectively supporting the sorts of queries which are typically asked of a GIS database requires different indices from those employed by text-based databases. Traditionally, B+ trees and hash indices are the sorts of index used by text or numeric databases. In order to support efficient queries; however, GIS objects must be indexed differently. For instance, for some queries, it might be helpful if stored geospatial objects were indexed by their relative geospatial proximity. One method that achieves this sort of indexing is the use of an R-tree which employs minimum bounding rectangles to index groups of geospatial objects. Other types of geospatial index include:

- Grid (spatial index)
- Z-order (curve)
- Quadtree
- Octree
- UB-tree
- R+ tree
- R* tree
- Hilbert R-tree
- X-tree
- kd-tree
- m-tree

Administration

Challenges

There are many potential challenges to administering a GIS database system. Interoperability problems are potentially significant. While GIS data standards, such as those put forth by the OGC, exist, they are not universally adopted, and conflicting standards are an issue. Further, GIS don't typically involve traditional spreadsheet-like reporting as seen with other systems supported by database backends. GIS are almost instrinsically graphical, and have significant application and presentation layer components with which the backend database must effectively interact. If there is incongruence between the understandings of data types or operations between theses three layers, it can result in bad data or misrepresentations of good data. GIS data also tends to be large. Achieving appropriate indexing and efficient data retrieval are therefore paramount.

Bibliography

File:GeoServer GeoNetwork with web app.png. (n.d.).Wikipedia, the free encyclopedia. Retrieved from http://en.wikipedia.org/wiki/File:GeoServer_GeoNetwork_with_web_app.png

Geographic information system. (2012, November 3).Wikipedia, the free encyclopedia. Retrieved from http://en.wikipedia.org/w/index.php?

title=Geographic_information_system&oldid=521136246

Simple feature access. (2012, October 22).Wikipedia, the free encyclopedia. Retrieved from http://en.wikipedia.org/w/index.php?title=Simple_feature_access&oldid=515909953

Spatial database. (2012, October 12).Wikipedia, the free encyclopedia. Retrieved from http://en.wikipedia.org/w/index.php?title=Spatial_database&oldid=517379046

Spatial query. (2012, October 16).Wikipedia, the free encyclopedia. Retrieved from http://en.wikipedia.org/w/index.php?title=Spatial_query&oldid=430843056

Well-known text. (2012, October 11).Wikipedia, the free encyclopedia. Retrieved from http://en.wikipedia.org/w/index.php?title=Well-known_text&oldid=515910084

NCDOT GIS Standards and Practices. (n.d.).*NCDOT: GIS*. Official Information Site. Retrieved November 5, 2012, from http://www.ncdot.gov/it/gis/GISStandardsAndPractices/default.html

Open Geospatial Consortium. (n.d.).*Open Geospatial Consortium | OGC(R)*. Official Information Site. Retrieved November 5, 2012, from http://www.opengeospatial.org/

Herring, J. R. (Ed.). (2010, August 4). OpenGIS® Implementation Standard for Geographic information - Simple feature access - Part 2: SQL option. Open Geospatial Consortium, Inc.