

A Relational Algebra Approach to Solve Problems of Spatial Databases Representation

Dr. Neeraj Bhargava¹, Dr. Ritu Bhargava², Prakash Singh Tanwar³

¹Associate Professor, Dept. of Computer Science, School of Engineering & Systems Sciences, MDS University, Ajmer, India

²Lecturer, Department of MCA, Govt. Women's Engineering College, Ajmer, India

³Research Scholar, Department of Computer Science, MJRP University, Jaipur, India

Abstract

This paper describes an approach to represent the operations of geographical data. Relational algebra as the basis for database has been greatly used in GIS business applications. Geographical data are divided into two main categories i.e. geo-objects and geo-fields. This paper is in the form of theoretical discussion with real illustrations and practical testing. Finally, geo-relational database systems and their implementation and analysis strategy has been workout using SQL server 2008 with fundamental set of spatial operations. Spatial analysis is the important part of GIS. It can be done in two ways. One is the vector-based and the other is raster-based analysis.

Keywords— Data Models, Relational Algebra, Spatial Databases, Spatial Query, Spatial Operation, Geographical Information System;

1. Introduction

As the theoretical basis for development of database, relational algebra has been enormously significant and successful in GIS applications.

Geometric, geographic, or spatial data are needed to be managed in various fields. A spatial database system is a database system with added capabilities for handling spatial or geo-referenced data. The GIS represent data in the form of raster and vector data. Other data models are derived from these two basic models [5].

In GIS and SDBMS there are three basic data types i.e. point, line, and polygon. They provide basic abstractions for modeling the structure of geometric entities, properties, and operations. They also provide relationships between geometric entities[9].

Basic relational operations like select, project, and rename has been implemented with suitable examples and implementation has been done using SQL Server 2008. Relational algebra is closely related with first-order logic (and of algebra of sets) and which deals with a set of finite relations that is closed under certain operators. These operators operate on one or more relations to generate a relation [1].

2. Literature Survey

In a paper entitled, “Issues in Spatial Databases and Geographical Information Systems (GIS)”, explored the challenges in the development of relational GeoSpatial databases [11]. Samet H[12] explains design and analysis of spatial data structures.

Batty P. [2][3][4], the author explained the importance of spatial industry in GIS. He presented fundamental challenges of spatial databases.

Herring J.R.[7] proposed TIGRIS for Object oriented GIS.

Longlaey P. et. al.[8], Gutting R.H., et. al.[6] and Sekhar S. et al.[12] gives an introduction to spatial databases.

Markus S.[7] proposed ROSA, the author explains rough spatial objects in it.

Openshaw S. [10] shows how to develop appropriate spatial analysis methods in GIS.

3. Spatial Data Modeling

GIS data represents real objects like roads, elevation, land parcels, trees, rivers, lakes , etc. with digital data. They can determine the mix real objects which can be divided into two abstraction levels, i). discrete objects (e.g., a house) and ii). continuous fields (such as rainfall amount, or elevations). There are two broad methods used to store both type of objects in GIS. These methods are raster data (images) and vector data. Vector data includes points, lines, and polygons which are mapped location attribute references. A raster data type is any type of digital image [5].

There are basically two types of data models in spatial databases i.e. raster data model and vector data model.

3.1. Raster Data Model

In raster data model, data is stored in the form of rectangular cells. Raster data type is made up of rows and columns of cells. Each cell requires at least a single value to store the data. Raster data can be images with each pixel (or cell) containing a

color value in it. A raster cell stores a single value. This value can be extended by using raster bands, color mappings or extended by attribute table. In raster bands representation of each cell is given by RGB color having red green and blue triplets. In color mapping, a mapping between a thematic code and RGB value occurred. In extended attribute table each row represents a unique cell value. Cell width in ground units is called resolution of the raster data.

3.2. Vector Data Model

In a GIS, mostly geographical features are expressed in the form of vectors. Vectors are geometrical shapes. Different geographical features are expressed by different types of geometry like point, line or polylines and polygons. A simple vector map can have a well as a point theme, rivers as a polyline theme, and lakes as a polygon theme.

3.2.1. Points. Simple location or geographical features having a single point reference can be expressed as Point theme. Points have the least amount of information on the map. Wells, schools, temples, ATMs, are some example of point themes. The same entity can be represented as point in one scale but in another scale it may be a polygon. For example cities in world map can be represented as a point but on city map it can be represented as polygon. Point feature can't have any length or area.

3.2.2. Lines or Polylines. Lines or polylines are used for linear features such as rivers, railway path, roads, and topographic lines, contours. Again, like point features, line or polyline features displayed at a small scale will be represented as line or linear features but at large scale it will be represented as a polygon. Linear features can measure distance but they don't have area.

3.2.3. Polygons. Polygons are used to represent an area of earth's surface in 2D. Such features may include countries, states, districts, lakes, park boundaries, buildings, or land uses. Polygons convey the most amount of information. It can measure perimeter and area.

Point clouds combine 3D points with RGB information at each point. And it returns a 3D color image. This type of method of storing data is known as hybrid method. It shows more realistic visual description of a map.

The Table 1 shows various data models and their application area [7].

Table 1. Data model and its example applications

Data Model	Example Application
Computer-aided design (CAD)	Automated engineering design and drafting
Graphical (non-topological)	Simple mapping
Image	Image processing and simple grid analysis
Raster/grid	Spatial analysis and modeling, especially in environmental and natural resource applications
Vector/Geo-relational topological	Many operations on vector geometric features in cartography, socio-economic and resource analysis, and modelling
Network	Network analysis in transportation, hydrology and utilities
Triangulated irregular network (TIN)	Surface/terrain visualization
Object	Many operations on all types of entities (raster/vector/TIN etc.) in all types of application

4. The Relational Algebra

According to Codd's algebra, there are five primitive operators i.e. the selection, the projection, the cartesian product (also called the cross product or cross join), the set union, and the set difference.

5. Spatial Relational Algebra

5.1. Basic Operations

For implementation of problem LandParcel is considered as a relation which is having attributes like ID, Shape_Leng, Shape_Area, Type, Floor, Height, Owner, Address, Symbol, Base and Geom. These attributes formed the database such as :

LandParcels (ID, Shape_Leng, Shape_Area, Type, Floor, Height, Owner, Address, Symbol, Base, Geom)

5.1.1. Select. A selection is a process in which the tuples from the entity set satisfying a given predicate are returned. In spatial select, it additionally describes a selection criteria based on spatial location based predicate the rest part is same as relational select. Spatial select is a unary operator. It operates on a single operand

If a Relation R1 is given then spatial select operator can be written as:

$$\sigma_{\text{predicate}}(R1)$$

Where the *Predicate* is a condition. It contains relation attributes and/or constants with some operators such as Logical Operator i.e 'and', 'or' 'not' Comparision Operators i.e. '<', '<=' , '>',

'>=' etc. This selection selects all those tuples in R1 for which *Predicate* is true.

Predicate is a condition for selection having various numbers of clauses.

<predicate>: Number of clauses:

- <attribute> <comp_op> <const>
or
- <attribute> <comp_op> <attribute>
<attribute> : Attribute from R1
<Comp_op> : One of the comparison operator
<Const> : Constant value from the attribute domain

Example:

Find the Landparcels of Owner 'X'

$\sigma_{\text{Owner} = "X"}(\text{LandParcels})$

The cardinality of the Spatial Select Operation can be expressed as:

$$0 \leq |\sigma_{\text{predicate}}(R1)| \leq |R1|$$

If the relation instance $r(R1)$ is given then we can express select operation as :

$$\sigma_{\text{predicate}}(R1) = \{t_i \mid t_i \in r(R1)\}$$

5.1.2. Project. The spatial project operation is also a unary operation. It returns its argument relation, with certain given attributes. The result of such projection is defined as the set of attributes that is obtained when all tuples in R are restricted to this set of attributes.

Let R1 be a relation and *variable_list* is their respective field representation in the form of a symbolic relation

$$\Pi_{\text{variable_list}}(R1)$$

variable_list: <attribute list> - Desired attributes from the relation R1

The cardinality of the Spatial Project Operation can be expressed as:

$$0 \leq |\Pi_{\text{A1, A2, ..., An}}(R1)| \leq |R1|$$

Given $R2 = \Pi_{\text{A1, A2, ..., An}}(R1)$ the relation instances $r(R1)$ and $r(R2)$ can be expressed in set notation as:

$$\Pi_{\text{A1, A2, ..., An}}(R1) = \{t_i \mid t_j \in r(R1) \wedge t_i = t_j (r(R2))\}$$

Example

Show ID, Type, Owner and Geom attributes of LandParcels Relation

$$\Pi_{\text{ID, Type, Owner, Geom}}(\text{LandParcels})$$

5.1.3. Rename. The spatial rename is a unary operator used to change the name of a relation. If a relation R1 is given then the spatial rename operator is denoted as:

$$\rho_{\text{newname}}(\text{oldRelName})$$

A rename is a unary operation written as $\rho_x(R1)$ where the result is identical to R1 except that the attribute in all tuples, is renamed to a new attribute name. This is simply used to rename the attributes of a relation or the relation itself.

Ex: Renaming the relation LandParcels to LP is expressed as

$$\rho_{\text{LP}}(\text{LandParcels})$$

6. Methodology

Spatial operations typically follow a two step algorithm filter and refinement as shown in figure 1

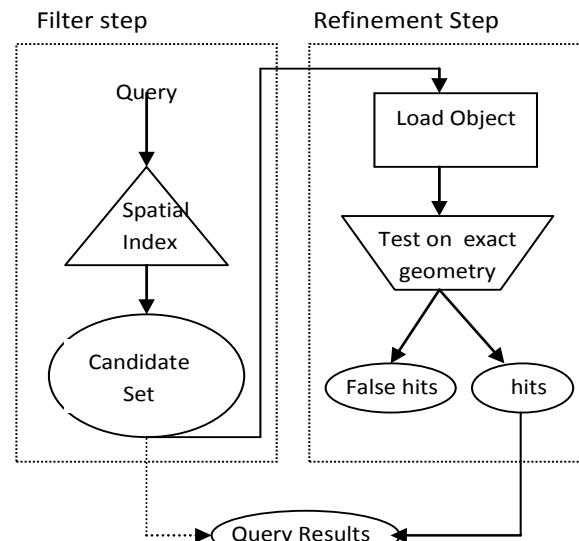


Figure 1. Two step processing

In first step the bounding rectangle of an extended spatial object quickly filter out the irrelevant objects. Then in second step i.e. refinement, exact geometry is used for the remaining spatial objects to complete the processing [14].

Selection Operation: the Selection operation can be performed in either indexed or non indexed spatial data. The major differences between these two methods are to evaluate the predicate and the type of indexes.

As in a two-step approach, the geometry of a spatial object is approximated with the help of rectangle. This geometry is used to evaluate the predicate of spatial relational algebra.

For Spatial data many indexing techniques are used like R-tree, B+ Tree, Quad Tree, Octree and space-filling curves [14].

7. Spatial Queries in Relational Database

The relational database software SQL Server 2008 supports two different spatial data types i.e. GEOMETRY and GEOGRAPHY. Table 2 explains spatial datatype with their usage.

GEOMETRY - This data type stores data in projected planar surfaces.

GEOGRAPHY - This data type stores data in an ellipsoidal model.

Table 2. Spatial Datatype and its Usage

Datatype	Usage
Geography	Geodetic Vector Spatial Data
Geometry	Planar Vector Spatial Data

Comparison between GEOGRAPHY and GEOMETRY data types are as shown in Table 3.

The structure of AdventureWorks_IPs and AdventureWork.LandParcels Table is as shown in given figure 2 and figure 3. The following structures show important places and land parcel tables.

Column Name	Data Type	Allow Nulls
ID	int	<input type="checkbox"/>
SHAPE_Leng	real	<input checked="" type="checkbox"/>
SHAPE_Area	real	<input checked="" type="checkbox"/>
Type	nvarchar(255)	<input checked="" type="checkbox"/>
Floor	int	<input checked="" type="checkbox"/>
Height	real	<input checked="" type="checkbox"/>
Owner	nvarchar(255)	<input checked="" type="checkbox"/>
Address	nvarchar(255)	<input checked="" type="checkbox"/>
Symbol	nvarchar(255)	<input checked="" type="checkbox"/>
base	real	<input checked="" type="checkbox"/>
geom	geography	<input checked="" type="checkbox"/>

Figure 2. Structure of AdventureWorks.dbo.IPs Table.**Table 3.** Comparison between Geography and Geometry Data type

Property	Geometry Datatype	Geography Datatype
Shape of the Earth	Flat	Round
Coordinate System	Cartesian (x,y)	Latitude and Longitude
Unit of measurement	Same as coordinate values	Defined in sys.spatial_reference_system
Spatial Reference Identifier	Not enforced	Enforced
Default SRID	0	4326(WGS 1984)
Size Limitations	None	No object may occupy more than one hemisphere
Ring Orientation	Not Significant	Significant

Column Name	Data Type	Allow Nulls
ID	int	<input type="checkbox"/>
Location	nvarchar(255)	<input checked="" type="checkbox"/>
Type	nvarchar(255)	<input checked="" type="checkbox"/>
Angle	real	<input checked="" type="checkbox"/>
Length	real	<input checked="" type="checkbox"/>
Width	real	<input checked="" type="checkbox"/>
geom	geography	<input checked="" type="checkbox"/>
geometry	geometry	<input checked="" type="checkbox"/>

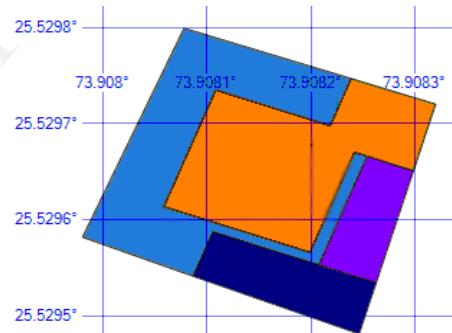
Figure 3. Structure of AdventureWorks. dbo. LandParcels Table

7.1. Select

Let us consider the query “Select LandParcels of ‘Society House’ type”

```
SELECT *
FROM AdventureWorks.dbo.LandParcels
WHERE [Type] = 'Society House';
```

The result of the above query in spatial form is shown in figure 4.

**Figure 4.** Spatial Output of select query for type = ‘Society House’.

7.2. Project

Let us consider the query defined as “projection of attributes id, type, owner and Geom attributes from relation LandParcels”.

```
SELECT [ID],[type],[owner],geom
FROM ADventureWorks.dbo.LandParcels;
```

The result of the query in spatial form is shown in figure 5, and in relational table form is as shown in figure 6.

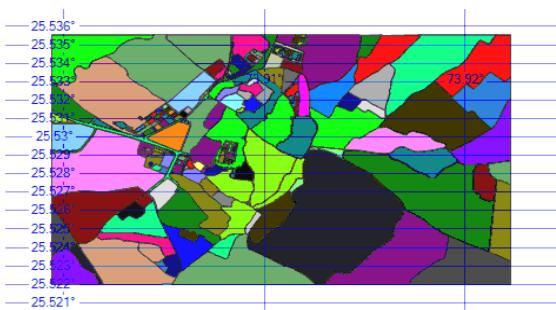


Figure 5. Spatial Output of query to project id, type, owner and Geom.

Results				Spatial results	Messages
ID	type	owner	geom		
1	Colony		0xE6100000010416000000000B8511BCA85394080AFBEFBEC...		
2	Colony		0xE610000001040600000000F6E48CFC85394000C36184917...		
3	Colony		0xE610000001041E000000000C60A2AC85394000387FE8F77...		
4	Field		0xE610000001042C0000000004A3456A18539408009644E7B67...		
5	Field		0xE610000001042E00000000A08B55A185394080025938BA7...		
6	Colony		0xE610000001041C000000000AD9D702863940003A0901ED...		
7	Field		0xE610000001043500000000467D3437863940000A1740D77...		
8	Field		0xE610000001041D000000000E76401486394080C79478F47...		
9	Field		0xE610000001043200000000020A5254863940001CA89F2E7...		

Figure 6. Relational Output of query to project id, type, owner and Geom.

7.3. Rename

Let us consider the rename operation on Deo_IP relation to rename it to Imp_Places.

```
USE AdventureWorks;
GO
sp_RENAME 'Deo_IP','Imp_Places';
```

This outputs as Imp_Places as a new name for Deo_IP relation.

8. Conclusion

GIS is considered as a decision making tool in problem solving environment. Spatial analysis is a vital part of GIS and can be used for many applications like site suitability, disaster management, natural resource monitoring, health care system and many more. Vector and raster based analysis functions like arithmetic, logical and conditional operations are used based on the recovered derivations.

This research presented a relational database primitives for geospatial data sets with which system developers can use to develop geospatial databases. The advantages of relational spatial database are elimination of redundant spatial data, spatial data sharing in various application area

Data sharing is facilitated and a corporate view of data can be provided to all managers and users. This makes it easy to integrate geospatial data with other related external system

The main contribution of this paper is the development of a set of fundamental spatial operations by using relational functions and

manipulating a relational database. The main intent of this research was to develop queries based on relational database that accommodates geospatial data sets. It provides the fundamental elements for relational algebra in the GIS and also presents a conceptual framework and demonstrates its use with an example prototype.

9. References

- [1] Alastair A., "Beginning spatial with SQL Server 2008", APress, New York, 2009
- [2] Batty Peter, "Future Trends & the Spatial Industry: Part On. Geospatial Solutions", E-Newsletter, geospatialonline,2004.
- [3] Batty Peter. "GIS Databases are Different. Geospatial Solutions", Geospatial E-Newsletter, geospatialonline,2005.
- [4] Batty, P., "Exploiting relational database technology in GIS.", Computers and Geosciences, vol. 18(4) pp. 453-462.,1992
- [5] Dr. Neeraj Bhargava, Dr. Ritu Bhargava, Prakash Singh Tanwar, and Ankit Sharma, "Rainfall Spatial Analysis using GIS", IJARCCE, Volume 2, issue 5, p.p. 2197-2200, May 2013.
- [6] Güting, R. H., et. al., "An Introduction to Spatial Database Systems". VLDB Journal, Vol. 3, pp 357-399, 1994.
- [7] Herring, J. R., "TIGRIS: A data model for an object-oriented geographic information system.", Computers and Geosciences, vol.18(4), pp. 443-452, 1992
- [8] Longley, P., et. al. , "Geographical Information systems and Sciences", Wiley Publication, Chechester, 2001
- [9] Markus S., "ROSA: An Algebra for Rough Spatial Objects in Databases", Lecture Notes in Computer Science, Vol. 4481/2007, pp. 411-418, 2007.
- [10] Openshaw S., "Developing appropriate spatial analysis methods for GIS. Geographical Information Systems" , Vol. 1, Principles. Longman Scientific and Technical, D. J.Maguire, M. F. Goodchild and D. W. Rhind. UK, pp. 389-402, 1991.
- [11] Samet H., "Issues in Spatial Databases and Geographic Information Systems (GIS)", Institute of Advanced Computer Studies, 2003
- [12] Samet H., "Design and Analysis of Spatial Data Structures", Addison-Wesley,1990

Shekhar, S., et. al., “*Spatial Databases: A Tour*”, Prentice Hall, 2003



Dr. Neeraj Bhargava: Presently working as Associate Professor & Head, Department of Computer Science, School of Engineering & System Sciences, MDS University, Ajmer. He has more than 23 yr of experience for teaching in the University. His areas of interest are Spatial Database, Image Processing and Ad hoc Network.



Dr. Ritu Bhargava: Presently working as lecturer, Department of MCA, Govt. Women's Engineering College, Ajmer. She has more than 9 yr of experience for teaching MCA. Her areas of research are Web GIS, Wireless Network.



Prakash Singh Tanwar: Pursing Ph.D in Computer Science from MJRP University, Jaipur, Rajasthan. His research area is 3D GIS and Web GIS. He has 7 years PG teaching experience and 3 years industrial experience. He is M.Phil(CS), MCA and B.Sc. (Maths).