Business Intelligence

Case Study

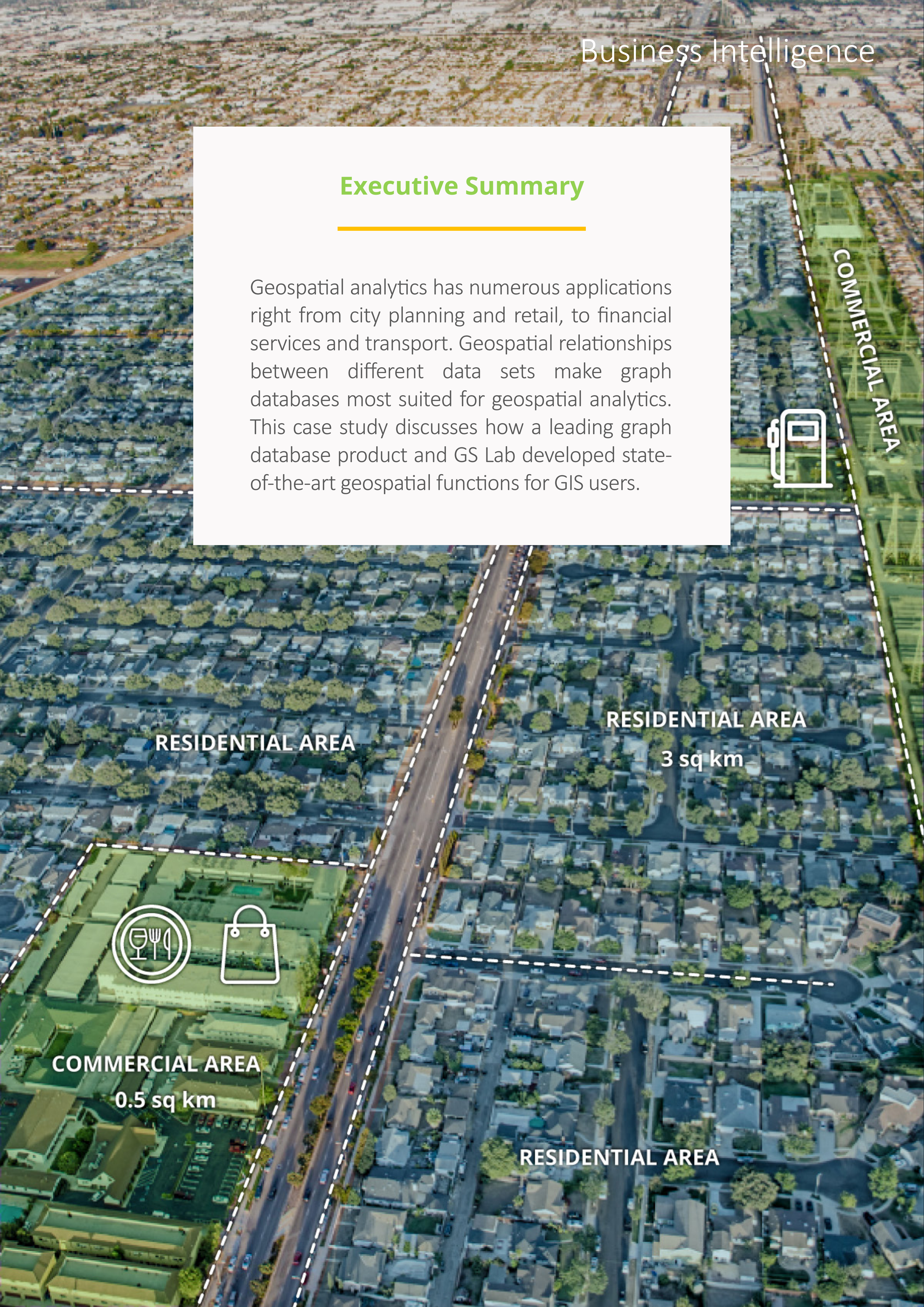
gslab

Developing state-of-the-art geospatial functions for graph database users

A graph database gives enterprises the ability to leverage massive geographic data for improved & accurate decision making

Executive Summary

Geospatial analytics has numerous applications right from city planning and retail, to financial services and transport. Geospatial relationships between different data sets make graph databases most suited for geospatial analytics. This case study discusses how a leading graph database product and GS Lab developed state-of-the-art geospatial functions for GIS users.



Overview

Many analytics projects need information about people and events in relation to where they occur. Solutions need to determine a location and its relationship to borders, regions, zones and other geographic indicators. Graph databases provide this much-needed relational information.

Our customer is a market leader in enterprise big data management and exploratory analytics. Competing graph database solutions offer grid based geospatial analytics capabilities. This creates limitations for users as real terrain often exhibits asymmetric irregular shapes. Some products have a more evolved implementation aimed at deployment on a single server. This causes severe scalability issues when users want to process large volume of data.

Challenge

Our customer decided to implement geospatial analytics with full support for Open Geospatial Consortium's GeoSPARQL standard with which most users are familiar. Their solution had to satisfy the following requirements:

1. Different geometric shapes had be to be supported.
2. These geometries had to support several operations.
3. These geometries needed to have abilities to transform and translate.
4. The product had to support multiple file formats to read & display different geometric shapes, data and additional properties present.
5. The solution had to easily scale on multiple servers.

Solution

GS Lab studied readily available libraries and frameworks. We decided to use ESRI libraries and framework.

However, this framework had the following limitations:

1. It did not have all the features required by the GeoSPARQL standard.
2. Functions like *Area/Distance/Length* were not available in different earth shapes like *Sphere* and *Ellipsoid.*
3. The Geospatial Coordinate System was not well supported.
4. Not all geospatial formats were supported.

We overcame these limitations using various approaches:

1. GS Lab did extensive research to solve these challenges by going through a number of geospatial research papers, algorithms use cases and accuracy comparisons.
2. In some cases we introduced better-fitting open source frameworks like Apache SIS and LocationTech JTS.
3. Geospatial functions in RDBMS databases were used as benchmarks for accuracy and compliance.
4. We implemented custom specifications in many cases instead of external frameworks for better parallel performance.

Leveraging graph database UDXs

Our customer's graph database product exposes a number of extension points through which developers can customize and extend the system. The extension point interfaces and the user code that implements them are called User Defined eXtensions (UDXs). The APIs for these UDXs are available in C++ and JVM. The UDXs as available as User Defined Functions (UDFs), User Defined Aggregates (UDAs) and User Defined Services (UDSs). GS Lab developed more than 160 UDX geospatial functions.

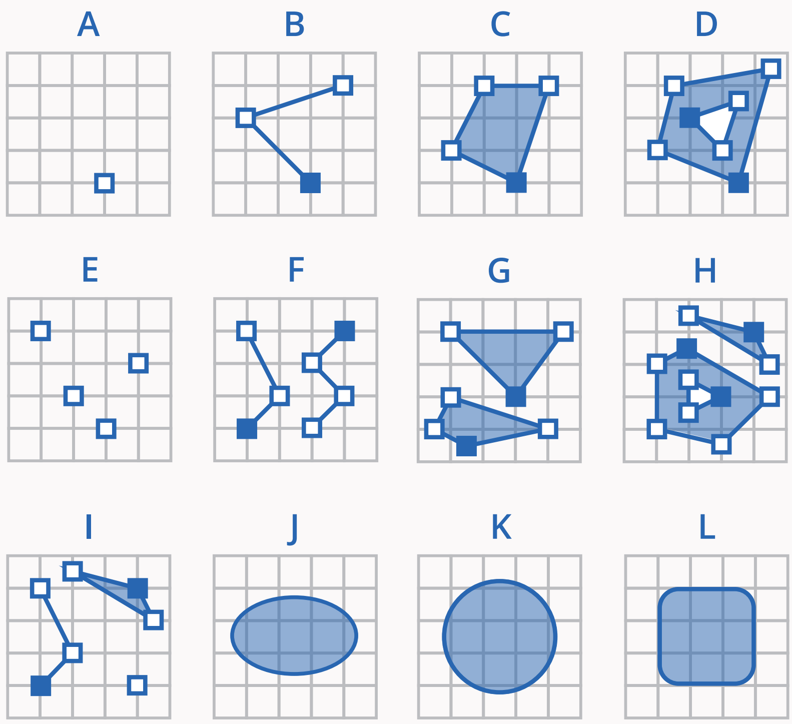
We developed:

1. UDFs and standard operations for multiple geometric shapes.

2. UDAs to construct new geometric shapes using the existing ones.

3. UDSs to extract geometric information from various source files like SHP, KML and JSON.

Key geospatial geometries supported by the product:



A. Point
B. Line
C. Polygon
D. Polygon with Rings
E. MultiPoint
F. MultiLine
G. MultiPolygon
H. MultiPolygon with Rings
I. Geometry Collection
J. Ellipse
K. Circle
L. Squircle

Key geospatial operations supported by the product:

1. Union
2. Intersection
3. Contains
4. Overlaps
5. Equals
6. Disjoint
7. Crosses
8. Within
9. Distnace
10. Touches
11. Difference
12. SymDifference
13. Relate
14. EnvIntersects
15. Centroid
16. Buffer
17. Convexhull
18. GeomtryN
19. PointN
20. Scale
21. Transform/Translate
22. Shear
23. Rotate
24. Type
25. Boundry
26. IsRing
27. InteriorRingN
28. ExteriorRing
29. IsSimple
30. Is3D
31. Min[x/y/z/m]
32. Max[x/y/z/m]
33. SRID
34. SetSRID
35. Dimension
36. Length
37. Area
38. Envelope
39. NumPoints
40. NumGeometries

The new solution gave additional capabilities to graph database users

- Users could define regions using shapes.
- They could use existing geographical information.
- Users could change coordinate systems.
- They could calculate relationships such as zones, overlaps and distances.

Impact

New capabilities

Improved flexibility

High scalability

- The graph database product now offers a first-of-its-kind MPP based geospatial analytics solution.
- The solution easily scales to billions of triples.
- It does not require complicated schemas and relationship tables.
- High performance was achieved without costly JOINs .
- The solution is highly flexible and allows integration of sparsely populated data at any point.

Great Software Laboratory (GS Lab) has been the technology partner of choice to 100+ organizations across North America, Europe and Asia-Pacific for over 17 years. Leveraging our expertise in 130+ tools & technologies, we have created 300+ ‘first-of-its-kind’ solutions to real-world problems. Our ‘Beyond code’ philosophy ensures that we not only push boundaries of existing technologies but also try out newer problem solving approaches to keep our customers one step ahead of their competitors. Our global team of 1200+ employees is adept at creating ‘real value’ at each stage of the customer growth journey, right from proof-of-concepts to completely scaled up products. For more information about our solutions & offerings, please visit www.gslab.com

gslab™

www.gslab.com