Location, location, location

Geospatial capability gives a new direction to business intelligence. *by Alan Greenspan*

Is it important to your business to understand how the distance between a customer's home and your store affects that person's shopping patterns? Should you know how your risk is affected by the concentration of insured properties or company facilities in a particular area? Does it matter to your bottom line which route a supply delivery truck takes?

Where a customer lives, the location of at-risk property and the roads traveled all require understanding the geospatial information contained within an organization's data. The Teradata Database now processes location data with agility and scalability to offer a new dimension in analytics and understanding.

What is geospatial?

Geospatial refers to storing and processing location information in a database and using it in business intelligence (BI) and other applications. "Location" can be either a single point on the earth or a series of points that together represent a line or polygon. This information exists in almost every type of business and is contained in common data such as customer and store addresses; origin and destination points for trips, shipments or routes; or hazard zones where natural phenomena such as a storm, disease or fire occur.

Within the database, points are generally translated into longitudes and latitudes in a process called geocoding. This step enables location relationships to be calculated in later analyses. Just as the database can subtract columns such as "price" and "cost"
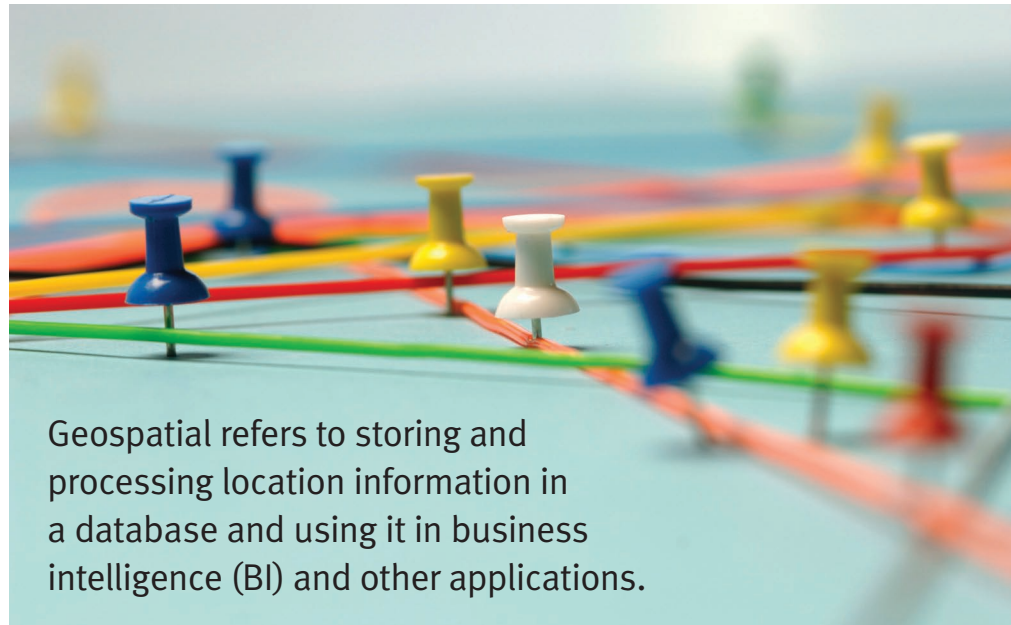
Geospatial refers to storing and processing location information in a database and using it in business intelligence (BI) and other applications.

to determine "net profit," so too can the distance between two location points be calculated. When such details are combined with other business data—from a marketing strategy, for example—the results produce actionable information.

While the home address of each customer and the address of each store are just data, this information becomes actionable when it is added to information gleaned from a customer's response to ad fliers. A marketing department might find that customer response is low for those who live more than five miles from a store.

Geospatial enhances BI

Most people are familiar with the mapping tools from Google, Mapquest and

others. These geospatial applications show data on a map and can help users visualize locations of customers, facilities, etc. The Teradata Database can provide the data needed for this type of visualization. By themselves, however, these applications lack true analytical capabilities. BI applications using geospatial functions from the Teradata Database can produce tremendous analytical results, often without ever rendering a map.

For example, insurance companies do not want to expose themselves to too much risk in the same area. An operational application used during the underwriting process could evaluate the added risk when a new location is introduced into the current portfolio of insured properties. While the application

may not actually map the properties, calculations based on the location data would indicate the concentration of properties in the area and, in turn, influence the scoring of an insurance application. This same location data could also be used to reveal the financial exposure of current insured properties that exist in the path of a forecast storm.

Teradata geospatial implementation

The Teradata geospatial capability comprises a new data type and database, along with import and export facilities for common spatial file formats. The new data type, which is a user-defined type (UDT), and its associated methods implement the SQL Multimedia (SQL/MM) spatial standard. The UDT, which can be downloaded and installed into Teradata Database V2R6.2 or Teradata 12, is enhanced with Teradata 13 and implemented as a system type for higher performance.

The new column type ST_GEOMETRY represents all of the geospatial types specified in the SQL/MM standard, including Points, LineStrings, Polygons, MultiPoint,

> **Well-known binary (WKB)** representation of geometry objects
> **A set of vertices** (2,500 in Teradata Database V2R6.2 and Teradata 12, and upwards of 1 million in Teradata 13)

Well-known text (WKT), a text markup language, or WKB, a more compact binary equivalent, is used when ST_ GEOMETRY is represented externally.

Associated with the ST_GEOMETRY type are a large number of methods that

value for functions such as area, distance, perimeter or length.

The new database within the geospatial capability contains spatial aggregates, transformations between various representations and spatial reference systems, tessellation capabilities for performance, and spatial metadata.

Spatial data loading

Besides the data points (e.g., addresses) that are geocoded and stored within the Teradata Database, existing business data loading processes can be augmented to load spatial data (e.g., territory boundary definitions, roads) received from a data supplier or from another system. If data is received in the commonly used WKT format, it can be loaded using SQL or Teradata TPump. In addition, the new TDGeoImport utility, a Windows program, converts data from other common file formats to WKB binary files and loads them into the Teradata Database. Common file formats include ESRI Shapefiles, MapInfo data sets in native TAB format, U.S. Census Bureau TIGER/Line files and comma-separated value.

Prior to Teradata 13, TDGeoImport used MultiLoad to load the data. With Teradata 13, spatial objects can hold more detail about the object shape and are, therefore, large data objects (LOBs). To accommodate this change,

The Teradata Database now processes location data with agility and scalability to offer a new dimension in analytics and understanding.

MultiLineStrings, MultiPolygons and GeometryCollections. Geosequence type, developed and used by Teradata, stores a line sequence to efficiently track movement along a route. Within ST_GEOMETRY, the following attributes describe the geospatial data:

> **Spatial reference identifier,** indicating the system used to specify coordinates for a point
> **Minimum bounding rectangle,** a set of numbers indicating the upper and lower bounds of the smallest rectangle around the spatial object

are used within SQL to perform calculations on the spatial data. With the exception of those that take or return arrays, the methods defined in the SQL/MM specification are provided and can be grouped into four categories: attribute, spatial, topology interactions and measurement. Spatial methods return a new geometry object and calculate intersection, boundary or the distance between locations, etc.

Topology interaction methods return 0 or 1 to indicate relationships such as overlaps, intersects, touches, within or crosses. Measurement methods return a numeric

TDGeoImport uses the JDBC FastLoad interface to load spatial objects smaller than 64KB and the JDBC parameter array interface to load objects larger than 64KB.

Using spatial data

Spatial data can be analyzed in conjunction with other business data in BI applications by using the spatial methods in SQL. Even when the application does not automatically generate this SQL, handwritten SQL can often be used to include location information in reports, dashboards, etc., and to produce answers to complicated questions in "location aware" BI.

Many products developed by Teradata partners allow users or developers within IT departments to provide any valid SQL for a query. Teradata partner Tableau Software published a blog that illustrates the use of handwritten SQL in more detail. The blog includes an example with distances and other spatial relationships used in a dashboard. As customer interest grows, more tools are expected to automatically generate SQL to analyze spatial data.

Many spatial visualization and analysis tools use popular file formats defined by ESRI, MapInfo and Google. These Shapefile, TAB and KML formats can be produced from Teradata by the open-source ogr2ogr utility.

The value of integration

In the past, standard business data with location implications was stored in the data warehouse but then had to be extracted and put into another system for efficient location-aware processing. Often, specialized systems that are good at mapping cannot handle complex BI queries on large databases. Not only did it limit the usefulness of the data but it also incurred added costs associated with data movement, and duplicate data and systems. Therefore, asking important business questions based on customer segmentation using variables such as location or distance from a store was impossible.

With geospatial capabilities inside the Teradata Database, all of this analysis can be done using standard SQL without ever extracting the data. This gives organizations a better opportunity to use location data as a new dimension to answering their business questions. **T**

Alan Greenspan joined Teradata in 1988 and is product marketing manger for the Teradata Database and other core software products.