



Data Management

Leveraging Spatial Capabilities in Your Database

May 27, 2010

ibm.com/db2/labchats

> Executive's Message



Sal Vella

Vice President, Development,
Distributed Data Servers and Data Warehousing

IBM



> Featured Speaker



David Adler

**DB2 Spatial Extender Development,
DB2 for Linux, UNIX, and Windows**

IBM



Agenda

- **Introduction to Spatial**
- **DB2 Spatial Extender Overview**
- **Q & A**



What is Spatial Data?

- **Information about anything that can be located on the earth's surface (geo-spatial)**
 - Natural: rivers, lakes, mountains, earthquake fault lines
 - Manmade: buildings, roads, utility facilities, railroad tracks
 - Cadastral: property boundaries, voting districts, state lines, country borders



What is Spatial Data? – cont.

- **Existing business data has a spatial component (e.g. address)**
 - Customers
 - Stores, branches, ATMs, etc
- **Location based information and map data**
 - Address: '555 Bailey Ave, San Jose, CA 95141'
 - Coordinates: latitude and longitude values, or a pair of X,Y values in a specific unit of measure (e.g. feet)
 - Name: 'Statue of Liberty', 'Lake Tahoe', 'LAX'

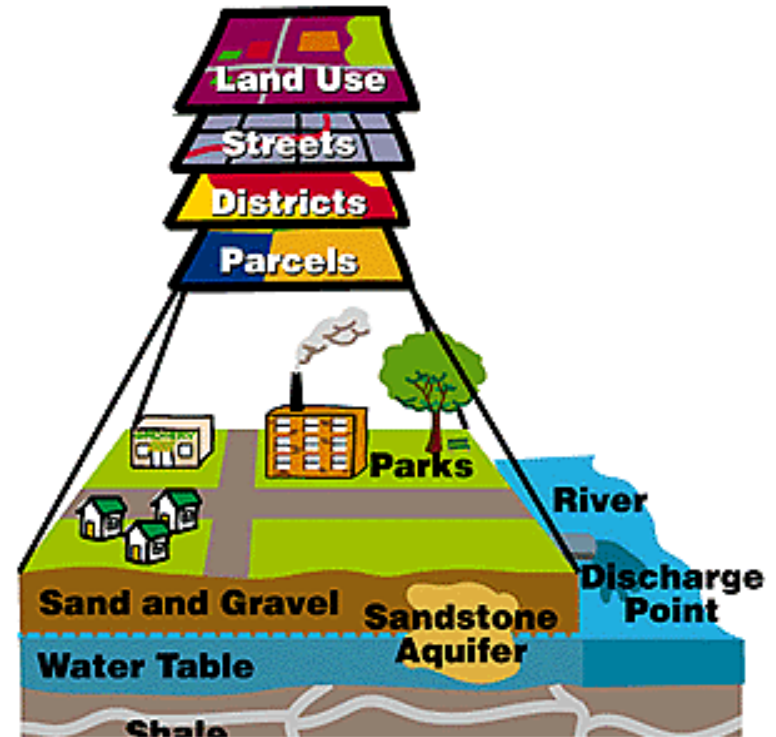
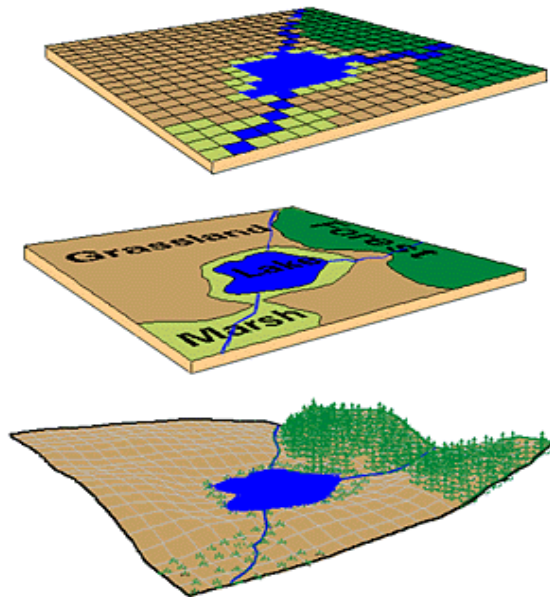


How is Spatial Data Represented?

- **Data can be represented by a single geometry or a collection of geometries**
 - Point: LAX airport (-121.736658, 37.201095)
 - Linestring: road, earthquake fault line
 - Polygon: property boundary, lake
 - Multipoint: seismic data for multiple sensor locations
 - Multilinestring: all the public bus routes of a county
 - Multipolygon: all the lakes in county

What is Spatial Data? – cont.

- Spatial data is represented as raster or vector, and ...
- ... organized as collections of thematic layers



Where do we get spatial data?

- **Geocoding**

- “Look up” address in reference data to determine location coordinates, most often latitude & longitude
- Many vendors of geocoding software

- **GPS (Global Positioning System) devices**

- Many GPS devices allow you to save locations or tracks and upload to a computer

- **Create with GIS (Geographic Information System)**

- User defines points, lines or polygons, generally on top of a reference base map.



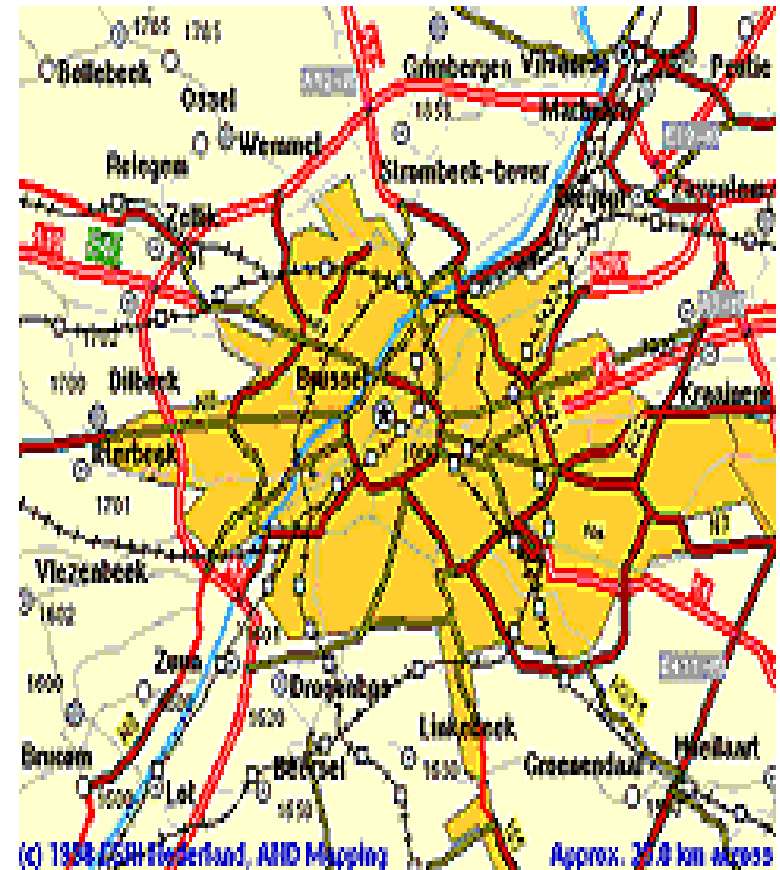
Where do we get spatial data?

- **Import from spatial data file**
 - “Shapefile” most common de-facto standard
 - Types of data:
 - Roads, census blocks, demographic data
 - Fire risk, flood polygons, earthquake locations
 - Data sources:
 - Sample datasets provided with Spatial Extender
 - Datasets available at no charge via Internet
 - Datasets available from private sources



So how do we answer these questions?

- **What is the closest retail outlet for each of the customers who have spent > \$3K during 2009 within the Chicago area?**
- **Identify customers with a home insurance policy living within 1000 meters of a river who DO NOT have flood insurance.**
- **What are the patterns of malignant cells in an MRI brain scan?**



We can answer those questions by ...

- **Paper map and pencil**
 - Results can be inaccurate and time consuming
- **Geographic Information System (GIS) Software**
 - Excellent for land and space management
 - Can answer spatial related questions effectively but it does require trained GIS personnel
- **Spatially-enabled DBMS**
 - Common SQL access to spatial data
 - Little or no GIS knowledge required
 - Visualization may or may not be needed



Agenda

- **Introduction to Spatial**
- **DB2 Spatial Extender Overview**
- **Q & A**

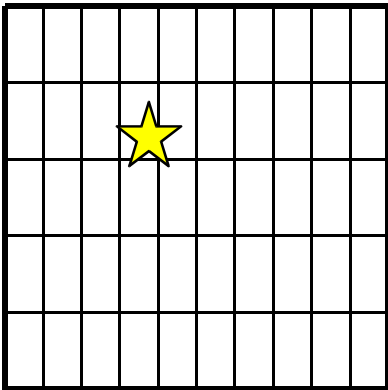


DB2 Spatial Extender - Overview

Spatial Data Analysis

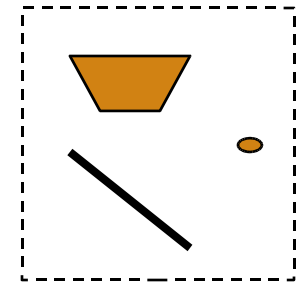


Spatial Grid Index

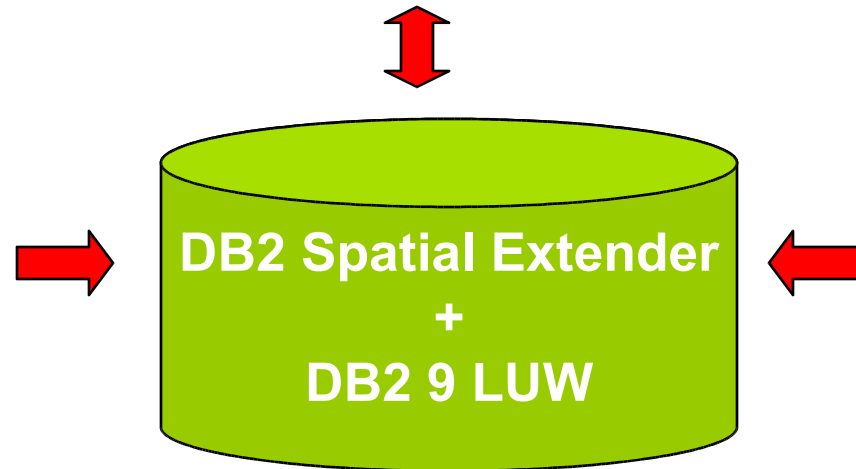


Allow Fast Access to Spatial Data

Spatial Data Types



All Spatial Data Stored in the Database



Spatial Access Through SQL



DB2 Spatial Extender

- **Spatial support based on DB2 object-relational technology**
 - User-defined structured types
 - User-defined functions
 - Index extension
- **Spatial index**
 - Greatly improves spatial query performance
 - Exploited by SQL query optimizer
- **Administration tools – command line & GUI**
 - Spatially-enable database
 - Shapefile import/export utility
 - Setup geocoding



DB2 Spatial Extender – InfoSphere Warehouse & Mining

- **Partitioned database support (DPF)**
 - Spatial data distributed across nodes
 - Spatial index distributed across nodes
 - Parallel spatial function / predicate processing
- **Design Studio**
 - Spatial SQL supported in data flows



DB2 Spatial Extender – Standards Based

- **Implements:**

- Open Geospatial Consortium (OGC) Simple Features for SQL using Types and Functions
- ISO SQL/MM part 3: spatial
- Geography Markup Language (GML) for geometries
- Well-known Text/Binary (WKT/WKB) for geometries
- Shapefile import / export (de-facto)
- WKT for coordinate systems (>3000 predefined)



DB2 Spatial Extender – Spatial Types

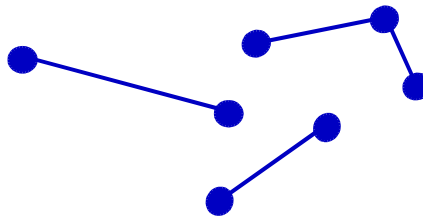
Point



LineString



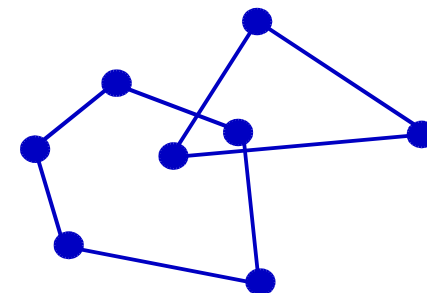
MultiLineString



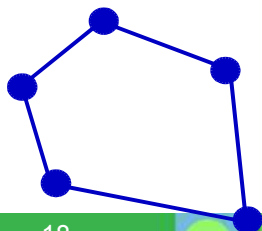
Multipoint



MultiPolygon



Polygon

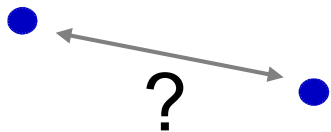


```
CREATE TABLE CUSTOMER (
  POLICY_NUMBER INT,
  NAME          VARCHAR(32),
  ADDRESS       VARCHAR(250),
  LOCATION      DB2GSE.ST_POINT);
```

```
INSERT INTO CUSTOMER VALUES(
  123, 'David Adler', '26 Mill Hill Rd, Woodstock, NY',
  DB2GSE.ST_POINT(-76.512, 42.357, 1));
```

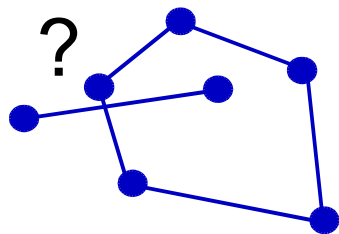
DB2 Spatial Extender – Spatial Functions and Predicates

ST_Distance(g1,g2)



```
SELECT c.name, s.name,  
       ST_Distance(c.loc, s.loc, 'METER') as distance  
FROM customers c, stores s  
WHERE ST_Distance(c.loc, s.loc, 'METER') < 10000  
      AND s.state = 'NY'
```

ST_Intersects(g1,g2)



```
SELECT a.name, a.type  
FROM highways a, floodzones b  
WHERE ST_Intersects(a.location, b.location) = 1  
      AND b.last_flood > 1950
```

And Many More ...

ST_Area	ST_GeometryType	ST_OrderingEquals
ST_AsBinary	ST_InteriorRingN	ST_Overlaps
ST_AsText	ST_Intersection	ST_Perimeter
ST_Boundary	ST_Intersects	ST_Point
ST_Buffer	ST_IsClosed	ST_PointFromText
ST_Centroid	ST_IsEmpty	ST_PointFromWKB
ST_Contains	ST_IsRing	ST_PointN
ST_ConvexHull	ST_IsSimple	ST_PointOnSurface
ST_CoordDim	ST_IsValid	ST_PolyFromText
ST_Crosses	ST_Length	ST_PolyFromWKB
ST_Difference	ST_LineFromText	ST_Polygon
ST_Dimension	ST_LineFromWKB	ST_Relate
ST_Disjoint	ST_MLineFromText	ST_SRID
ST_Distance	ST_MLineFromWKB	ST_StartPoint
ST_Endpoint	ST_MPointFromText	ST_SymmetricDiff
ST_Envelope	ST_MPointFromWKB	ST_Touches
ST_Equals	ST_MPolyFromText	ST_Transform
ST_ExteriorRing	ST_MPolyFromWKB	ST_Union
ST_GeomFromWKB	ST_NumGeometries	ST_WKBToSQL
ST_GeometryFromText	ST_NumInteriorRing	ST_WKTTToSQL
ST_GeometryN	ST_NumPoints	ST_Within
		ST_X
		ST_Y

And more...

Simplified Constructors
from:

x,y

WKT

WKB

GML

shape

Linear referencing

Spatial aggregation

ST_AsGML

ST_AsShape



Spatial constructor functions

- **ST_Point(x, y, srs_id) – create point at this location**
- **ST_Point('POINT (-121.5, 37.2)', 1)**
- **ST_Linestring('LINESTRING (-121.5 37.2,-121.7 37.1)',1)**
- **ST_Polygon(CAST (? AS CLOB(1M)),1)**
 - For host variable containing well-known text, well-known binary or shape representation

Spatial predicates – WHERE clause

- **ST_Distance(geom1, geom2) < distance_constant or var**
- **ST_Contains(geom1, geom2) = 1**
- **ST_Within(geom1,geom2) = 1**
- **EnvelopesIntersect(geom1, geom2) = 1**
- **EnvelopesIntersect(geom1, x1, y1, x2, y2, srs_id) = 1**
- **ST_Area(geom) < some_value**



Spatial functions that create new spatial values

- **ST_Buffer(geom, distance)**
- **ST_Centroid(geom)**
- **ST_Intersection(geom1, geom2)**
- **ST_Union(geom1, geom2)**



Functions that return information about a spatial value

- **ST_Area(geom), ST_Length(geom)**
- **ST_MinX(geom), ST_MinY(geom), ST_MaxX(geom), ST_MaxY(geom)**
- **ST_Is3d(geom), ST_IsMeasured(geom)**
- **ST_X(geom), ST_Y(geom), ST_Z(geom), ST_M(geom)**
- **ST_AsText(geom)**



Harness the full power of SQL

- Outer join
- Common table expressions
- Recursive queries, sub-queries
- Aggregate functions
- Order by, group by, having clauses
- OLAP, XML and more ...

Example problem: Determine the average household income for the sales zone of each store in the San Diego area.

```
WITH sdStores AS (SELECT * FROM stores
                   WHERE st_within(location, :sandiego) = 1)
SELECT s.id, s.name, AVG(h.income) FROM houseHolds h,
sdStores s
WHERE st_intersects(s.zone, h.location) = 1
GROUP BY s.id, s.name
ORDER BY s.name
```



Harness the full power of SQL

Example problem: Identify customers with a home insurance policy living within 1000 meters of a river who DO NOT have flood insurance.

```
SELECT
  c.name
  ,MIN( st_distance(c.location, r.geom, 'METER') ) as distance
FROM
  customers c
  ,rivers r
WHERE  c.policy_type = 'Home'
AND   c.coverage <> 'Flood'
AND   st_distance(c.location, r.geom, 'METER') < 1000
AND   st_distance(c.location, r.geom) < 0.01
GROUP BY c.name
```

Harness the full power of SQL

Example problem: What is the closest retail outlet for each of the customers who have spent > \$3K during 2009 within the Chicago area?

```
SELECT cust_name, outlet_name FROM (  
  SELECT  
    c.name  
    ,o.name  
    ,RANK() OVER  
    (PARTITION BY c.name ORDER BY st_distance(c.location, o.location, 'STATUTE  
MILE') ASC)  
  FROM outlets o  
    , customers c  
  WHERE c.spending > 3000  
    AND c.year = 2009  
    AND c.city = 'Chicago'  
    AND o.city = 'Chicago'  
  ) subq (cust_name, outlet_name, rank)  
WHERE rank = 1
```



Data - Spatial Project/Customer-Distance.sql - IBM Data Studio

File Edit Navigate Search Project Data Script Window Help

Optim Solutions

Projec Data P *Customer-Distance.sql

No Connection

```

set current function path = db2gse;
select
  b.name
  ,c.name
  ,st_distance(b.location, c.location, 'STATUTE MILE')
  as distance
from se_demo.customers as c, se_demo.branches as b
where b.name='San Carlos'
order by distance;
    
```

Property Capture SQL Cat Workload Monitor L SQL Res

Type query expression here Status Result1

Status	Operation	NAME	NAME	DISTANCE	
✓	Succ Customer-Distance.sql	1	San Carlos	Margie Simpson	0.2916467482...
✓	Run SQL	2	San Carlos	Idek Go	0.3801575946...
✓	Run SQL	3	San Carlos	Max Power	0.9232827210...
		4	San Carlos	Knut Stolze	0.9589921584...
		5	San Carlos	Jean Ho	0.9667169964...
		6	San Carlos	Bart Simpson	0.9682752316...
		7	San Carlos	Bryan Patterson	0.9802307267...
		8	San Carlos	Rafael Coss	0.9825555894...
		9	San Carlos	Kuai Chang Kaine	1.0139389929...
		10	San Carlos	Lisa Simpson	1.0825212747...
		11	San Carlos	Homer Simson	1.0975067296...

Total 99 records shown

SE_BANK (Connected)

Data Source Explorer

Tables

- ACCOUNTS
- BRANCH_BUFFERS
- BRANCHES
- CITY_LIMITS
- CUSTOMERS
- SALES_REGIONS
- SJ_CENSUS_BLOCKS

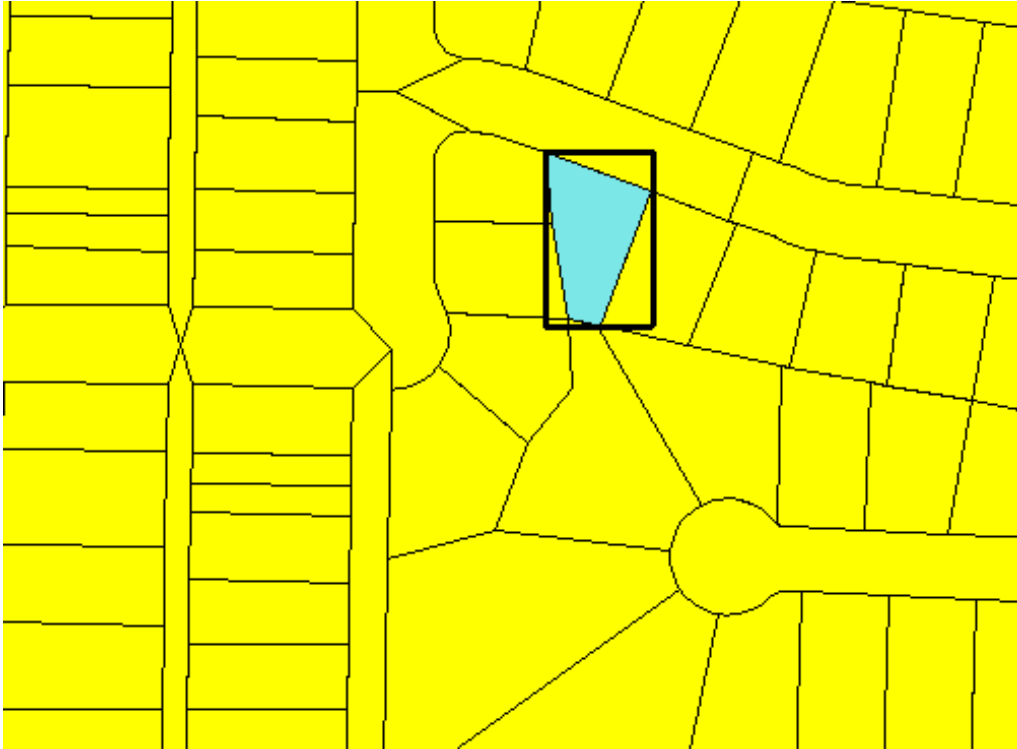


Spatial index operation

- **When a geometry is inserted**
 - One or more index entries created using 1 of 4 possible grid levels
 - 3 user-specified grid sizes, System-defined “overflow” level
- **When a query is performed**
 - Query envelope computed from “non-index” column value
 - Grid cells intersecting query envelope are identified
 - Index entries corresponding to grid cells are scanned
 - Index extension rejects index entries outside query window
 - (Predicate UDF may do tertiary filtering)
- **Optimization factors:**
 - Minimize # of index entries generated per geometry
 - **Larger grid size is better**
 - Minimize # of geometries analyzed at query time
 - **Smaller grid size is better**

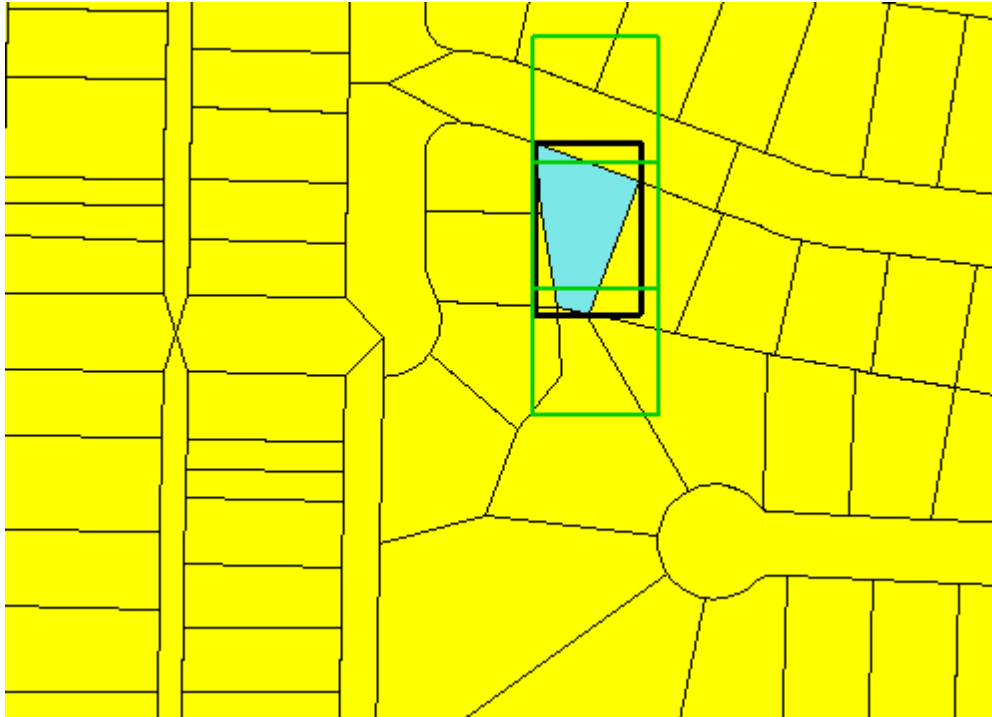


Grid index generation



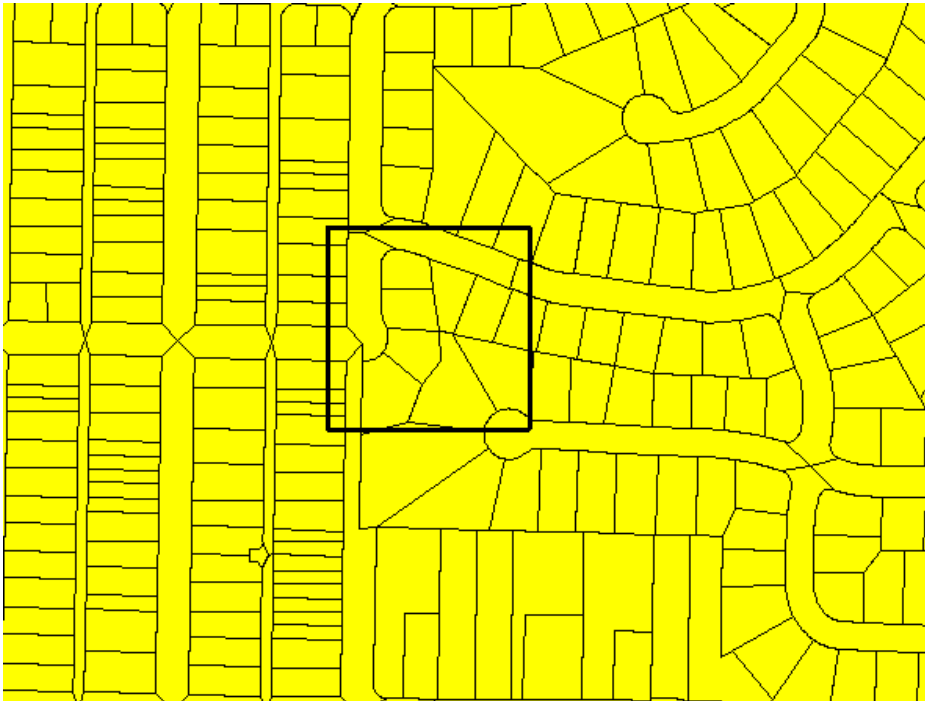
- **Geometry MBR computed**

Grid index generation - continued



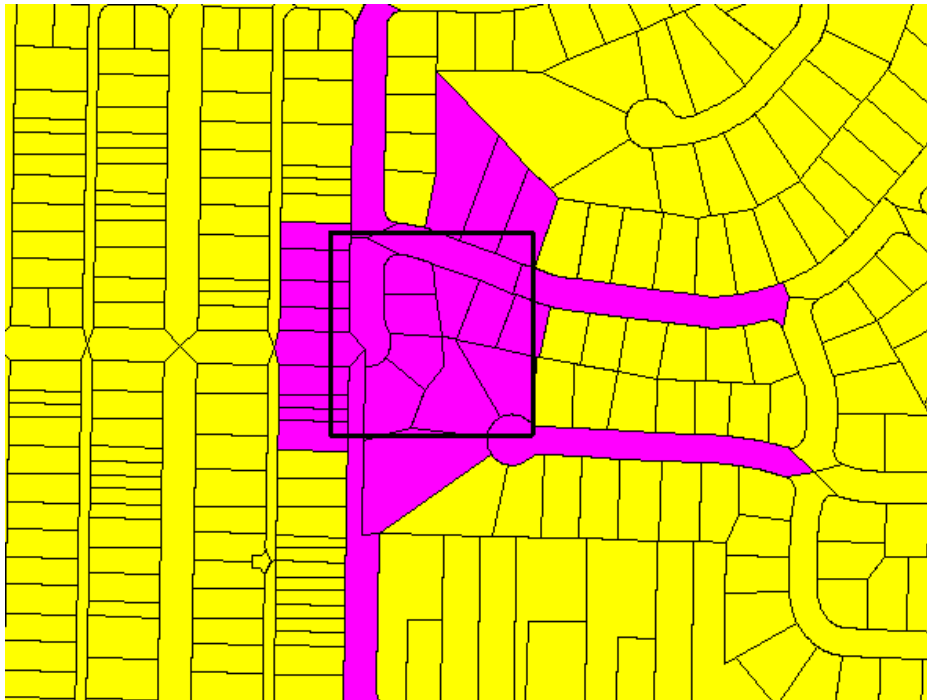
- In this example 3 grid cells intersect at grid size 100

Index query



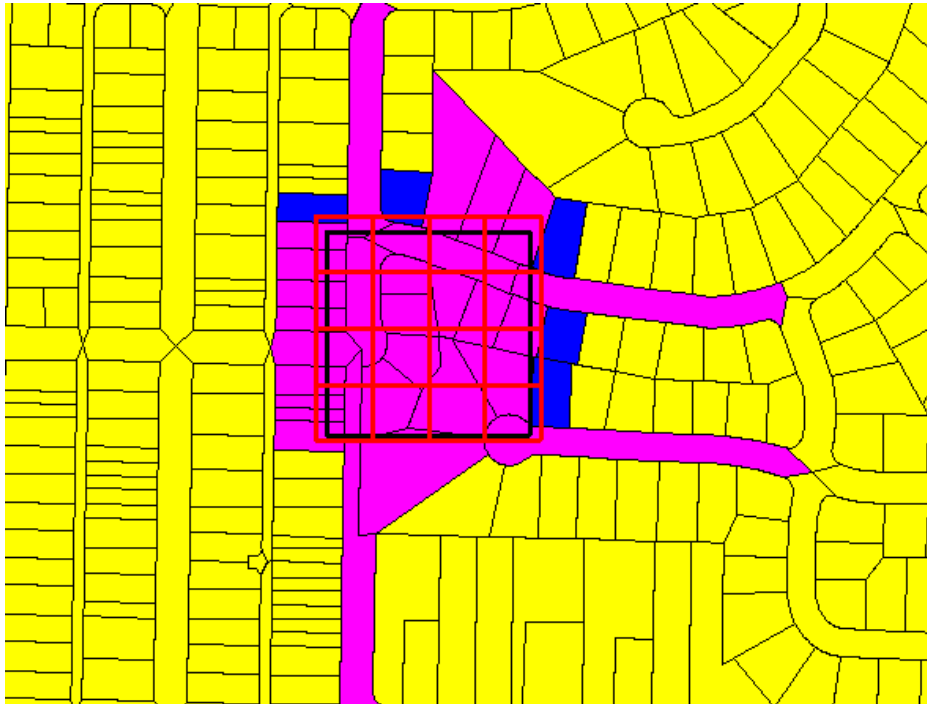
- **We want to find all geometries that intersect the query window represented by the black rectangle**

Index query - continued



- **The geometries whose envelope actually intersect the query window are highlighted in magenta.**
- **There are 28 such geometries**

Index query - continued



- 33 geometries intersect the grid cells that intersect the query window
- 91 index entries would be scanned
- Only the 28 geometries in magenta will be returned, but the 5 additional geometries need to be scanned and discarded.

How is a spatial index created?

- **CREATE INDEX myindex ON mytable(mycolumn)
EXTEND USING db2gse.spatial_index(grid1, grid2, grid3)**
- **Spatial index advisor to analyze spatial column and recommend appropriate grid sizes**
 - 'gseidx' index advisor shipped with Spatial Extender
 - Java index advisor downloadable via the Spatial Extender web site

Spatial Application / Tool Support

- **Commercial**

- IBM ILOG JViews Maps
- IBM InfoSphere Warehouse
- ESRI ArcGIS
- (Pitney Bowes) MapInfo MapXtreme for Java
- Safe FME

- **Open Source**

- GeoTools – Java toolkit, DB2 plug-in contributed by IBM
- GeoServer – Web Map/Feature Server
- uDig – rich GIS application / framework



IBM DB2 and Informix Sample for JViews Maps

File

100% (0 tasks)

DB2 Layer

- BANKDEMO.BRANCH.FEATURE.S
- BANKDEMO.CUSTOMER.FEATUR
- BANKDEMO.MAINROADS.FEATU
- BANKDEMO.V10.FEATURE.SRID-

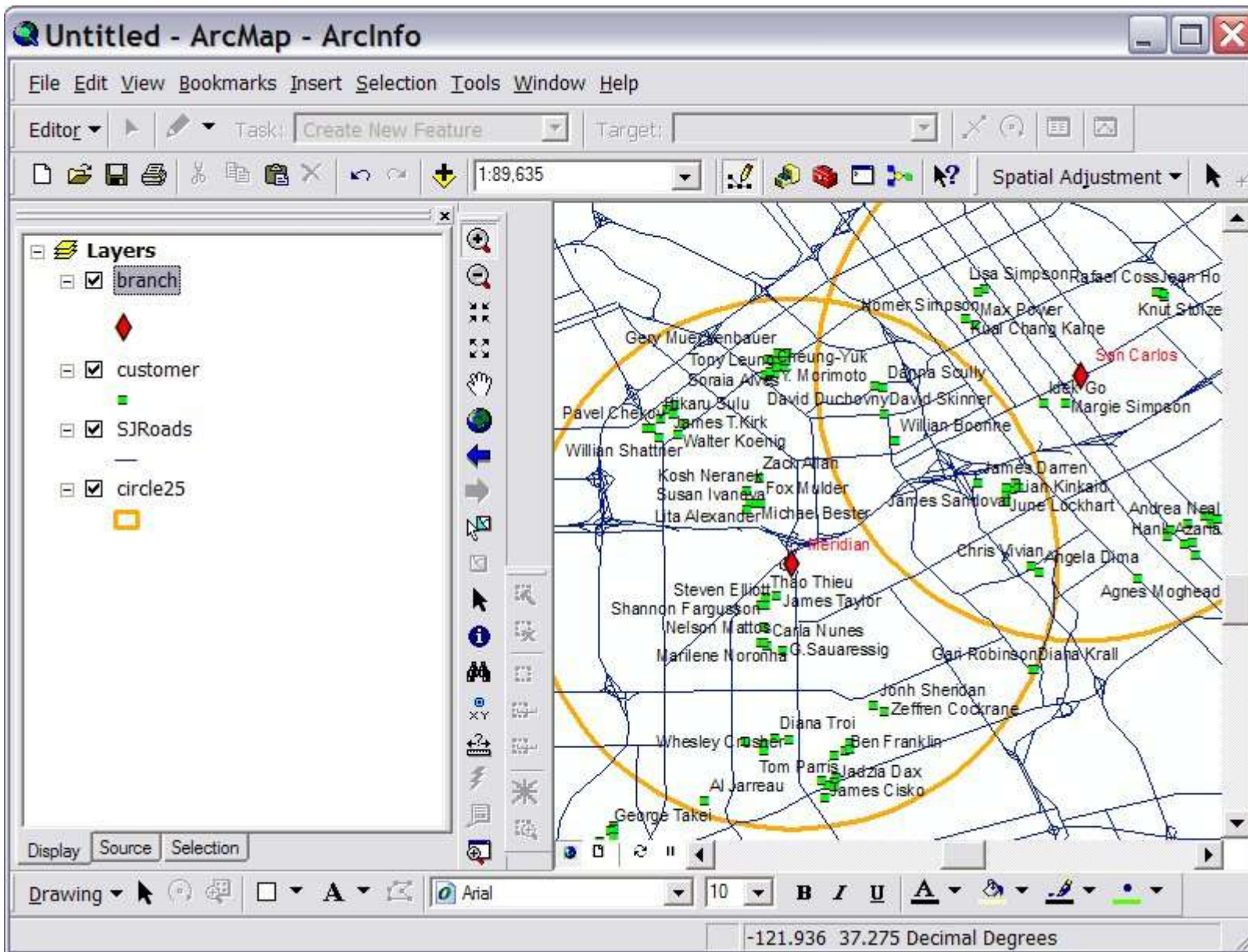
Dynamic Style Setting

Default Style

Transparency	0%
Decoration	none
Decoration Only	false
Decoration Color	
Outlined	true
Filled	true
Line Color	Yellow
Labeling	No Label
Line Style	
Line Width	2
Paint	
Visible in Overview	false

*IBM ILOG
JViews
Maps
Example*





ESRI
ArcMap
Example



The screenshot shows the uDig application window with the following components:

- Menu Bar:** File, Edit, Navigation, Layer, Map, Data, Window, Help
- Toolbar:** Standard GIS navigation and editing tools.
- Projects Panel:** Shows a project named 'V10' with layers 'CUSTO' and 'MAINR'.
- Legend:**
 - CUSTOMER (Green square)
 - MAINROADS (Blue line)
 - BRANCH (Red star)
 - V10 (Orange circle)
- Map View:** A street map with green squares representing customer locations and orange circles representing V10 service areas. Labeled customers include: Homer Simpson, Jean Ho, Max Power, San Carlos, Margie Simpson, Laura Stewart, David Duchovny, David Skinner, Michelle Nichols, Walter Koenig, William Boone, Zack Allan, Fox Mulder, James Sandoval, Paul Buchman, James Oconor, Angela Dima, Gari Robinson, Meridian, James Taylor, Marilene Noronha, Jonh Sheridan, Whesley Crusher, Katryn Janeway, Al Jarreau, and George Takei.
- Status Bar:** Scale 1:76,884, Coordinate System WGS 84, and Coordinates -121.910, 37.361.
- Bottom Panel:** Includes buttons for Catalog, Web, Search, Table, and Information. A text box below reads: "To display information, select the info tool and click on a Map."

uDig Example



EZMapping with DB2 Spatial Extender - Mozilla Firefox

File Edit View History Delicious Bookmarks Tools Help

http://dadler.dynalias.org/BankdemoPHP/main.php

leopalace japari

Most Visited Google - W3 wamc.asx (video/x-ms... AT&T Yahoo! http://clc.dynalias.org:... All Public Files

firefox:toolbar AT&T Yahoo! ArcIMS ciwtest ArcIMS Viewer using t... Map Intelligence Tools IBM DB2 9.5 Info

EZMapping with DB2 Spatial Ext...

EZMapping with DB2 Spatial Extender

Load Customers

Map Satellite Hybrid

William Boone
150 Autumn St
San Jose, CA 95110

Legend
Customers

Google
Maps
Example





Commander Dashboard

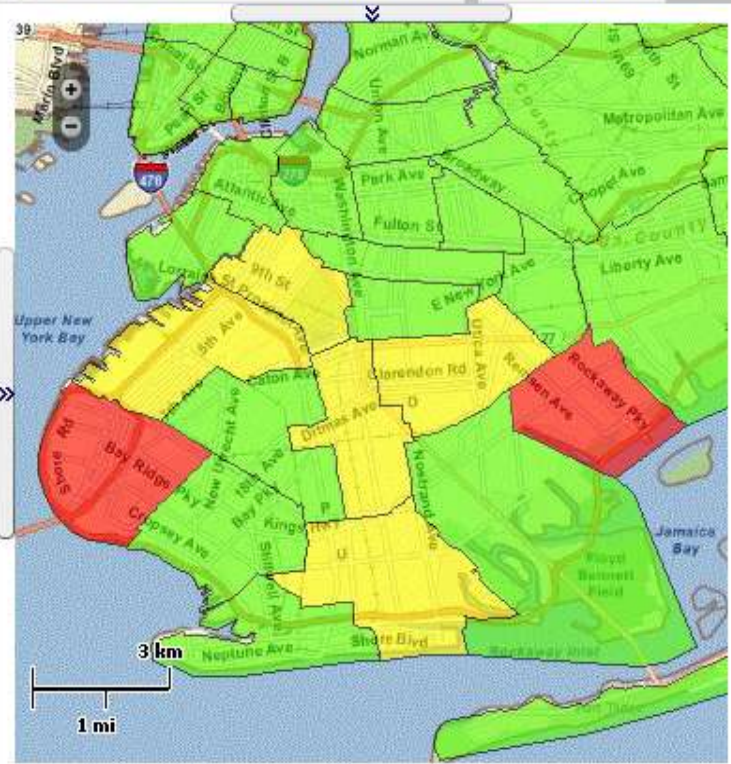
Real Time View (Past 24 hours)

Show/Hide Real Time Params Submit



Predicted Levels (Next 4 Hours)

Show/Hide Predictive Params Submit



**IBM
Crime
Prevention
And
Prediction
Example**

Harley-Davidson - Ride Planner - Windows Internet Explorer

http://rideplanner.harley-davidson.com/rideplanner/rideP

File Edit View Favorites Tools Help

Harley-Davidson - Ride Planner

HARLEY-DAVIDSON RIDE PLANN
PLAN. RIDE. SHARE.

SHOW ON MAP
Click icons to display info below

ROADS DEALERS HOTELS EVENTS

Welcome, Guest
[Sign In](#) or [Create Personal Profile](#)
[Need Help?](#)

My Ride

Start Location
1 NY, 12401 GO
Enter location or right-click on the map

Roads

Harley-Davidson Dealers

1 to 1 of 1

Woodstock Harley-Davidson, Inc.
Dealer's web site
949 Route 28
Kingston, NY, 12401, USA
Phone: (845) 338-2800
[More Information](#)
[Add location to My Ride](#)

FOR RIDERS, BY RIDERS
Learn how to share your route with other riders in the new Ride Planner.

Road Aerial Mixed

GPS Synchronization

Catskill State Park

West Hurley

Morgan Hill

Kingston

2.5 miles

bing

© 2009 Microsoft Corporation © 2009 HARLEY-DAVIDSON

Harley-Davidson RidePlanner Example



IBM Spatial Offerings

- **DB2 Spatial Extender for DB2 Linux, Unix & Windows (LUW)**
 - No charge on all supported DB2 platforms
- **Spatial Support for DB2 for z/OS**
 - Similar to Spatial Extender on LUW
- **DB2 Geodetic Data Management Feature**
 - Seamless, whole-earth geometry representation and query
 - Additional license charge, available only on Enterprise Ed.
 - Integrated with DB2 Spatial Extender
- **Informix Spatial Datablade**
 - No charge, similar to DB2 Spatial Extender
- **Informix Geodetic Datablade**
 - Additional license charge; includes time dimension



Resources

- DB2 InfoCenter:
<http://publib.boulder.ibm.com/infocenter/db2luw/v9r7/index.jsp>
 - Main section “Spatial and geodetic data”
- DB2 Spatial Offerings web site:
<http://www.ibm.com/software/data/spatial/>
- DeveloperWorks DB2 Spatial Extender forum:
<http://www.ibm.com/developerworks/forums/forum.jspa?forumID=>
- DeveloperWorks articles: <http://www.ibm.com/developerworks/>
 - Search for “DB2 Spatial”
- Contact the author: David Adler – dadler@us.ibm.com



> Questions



Thank You!

ibm.com/db2/labchats



Thank you for attending!

