ibm.com/db2/labchats

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.

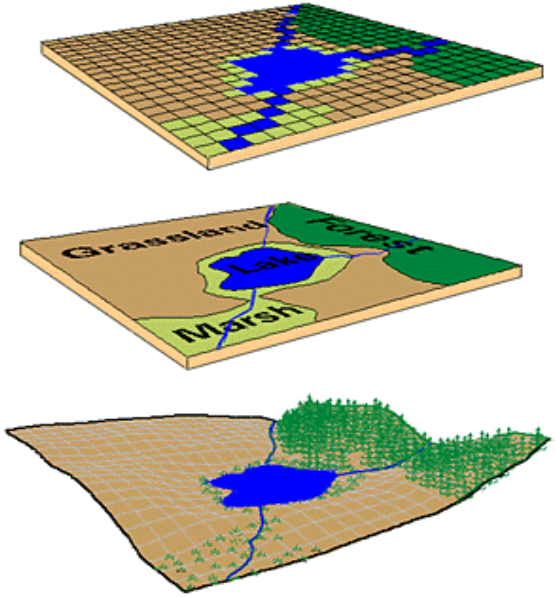
- **Existing business data has a spatial component (e.g. address)**
  - Customers
  - Stores, branches, ATMs, etc
- **Location based information and map data**
  - Address: '555 Bailey Ave, San Jose, CA 95141'
  - Coordinates: latitude and longitude values, or a pair of X,Y values in a specific unit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

How is Spatial Data Represented?

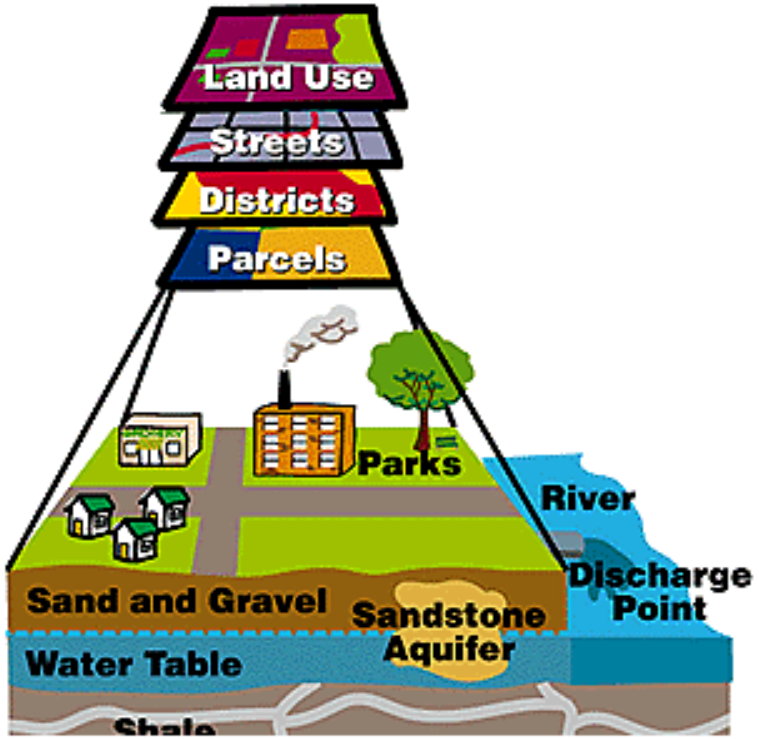
- **Data can be represented by a single geometry or a collection of geometries**
  - Point: LAX airport (-121.736658, 37.201095)
  - Linestring: road, earthquake fault line
  - Polygon: property boundary, lake
  - Multipoint: seismic data for multiple sensor locations
  - Multilinestring: all the public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

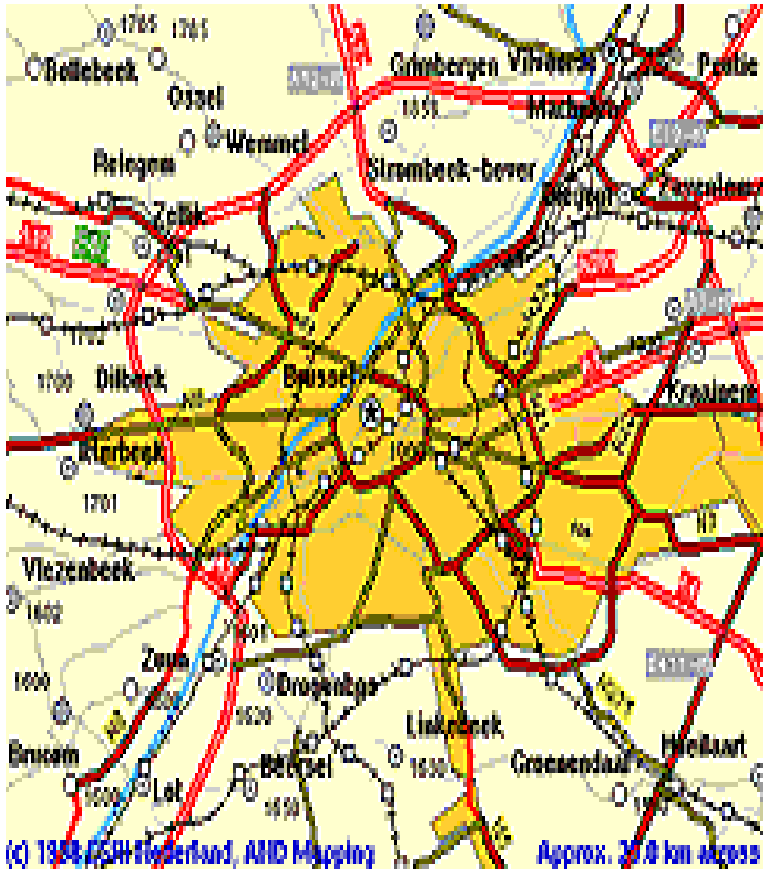
- **Geocoding**
  - "Look up" address in reference data to determine location coordinates, most often latitude & longitude
  - Many vendors of geocoding software
- **GPS (Global Positioning System) devices**
  - Many GPS devices allow you to save locations or tracks and upload to a computer
- **Create 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

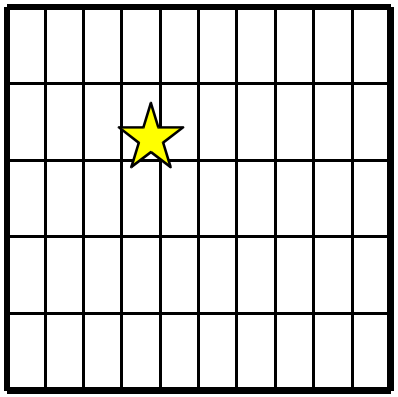
Agenda

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

DB2 Spatial Extender - Overview

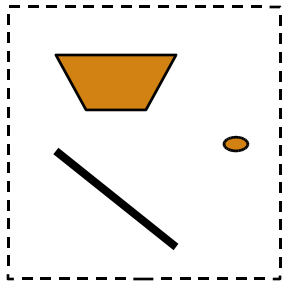
Spatial Data Analysis

Spatial Grid Index



Allow Fast Access to Spatial Data

Spatial Data Types



All Spatial Data Stored in the Database

DB2 Spatial Extender
+
DB2 9 LUW

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

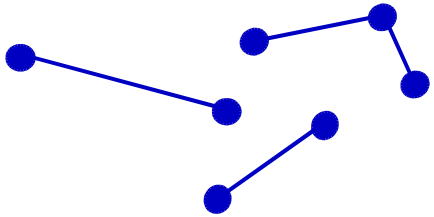
- **Implements:**
  - Open Geospatial Consortium (OGC) Simple Features for SQL using Types and Functions
  - ISO SQL/MM part 3: spatial
  - Geography Markup Language (GML) for geometries
  - Well-known Text/Binary (WKT/WKB) for geometries
  - Shapefile import / export (de-facto)
  - WKT for coordinate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

DB2 Spatial Extender – Spatial Types

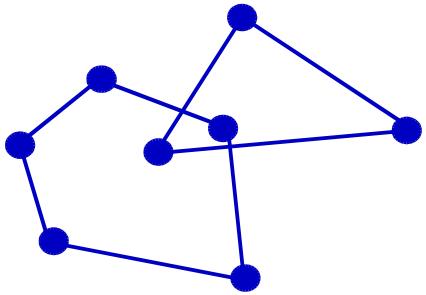
Point

LineString

MultiLineString

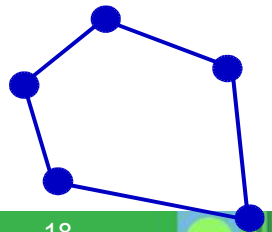


MultiPolygon



Multipoint

Polygon

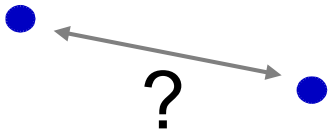


```
CREATE TABLE CUSTOMER (
        POLICY_NUMBER     INT,
        NAME              VARCHAR(32),
        ADDRESS           VARCHAR(250),
        LOCATION          DB2GSE.ST_POINT);

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

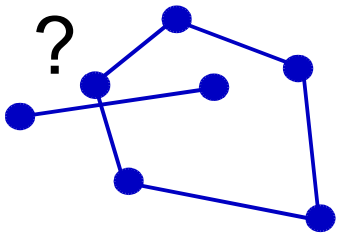
DB2 Spatial Extender – Spatial Functions and Predicates

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_AsBinary
ST_AsText
ST_Boundary
ST_Buffer
ST_Centroid
ST_Contains
ST_ConvexHull
ST_CoordDim
ST_Crosses
ST_Difference
ST_Dimension
ST_Disjoint
ST_Distance
ST_Endpoint
ST_Envelope
ST_Equals
ST_ExteriorRing
ST_GeomFromWKB
ST_GeometryFromText
ST_GeometryN

ST_GeometryType
ST_InteriorRingN
ST_Intersection
ST_Intersects
ST_IsClosed
ST_IsEmpty
ST_IsRing
ST_IsSimple
ST_IsValid
ST_Length
ST_LineFromText
ST_LineFromWKB
ST_MLineFromText
ST_MLineFromWKB
ST_MPointFromText
ST_MPointFromWKB
ST_MPolyFromText
ST_MPolyFromWKB
ST_NumGeometries
ST_NumInteriorRing
ST_NumPoints

ST_OrderingEquals
ST_Overlaps
ST_Perimeter
ST_Point
ST_PointFromText
ST_PointFromWKB
ST_PointN
ST_PointOnSurface
ST_PolyFromText
ST_PolyFromWKB
ST_Polygon
ST_Relate
ST_SRID
ST_StartPoint
ST_SymmetricDiff
ST_Touches
ST_Transform
ST_Union
ST_WKBToSQL
ST_WKTToSQL
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

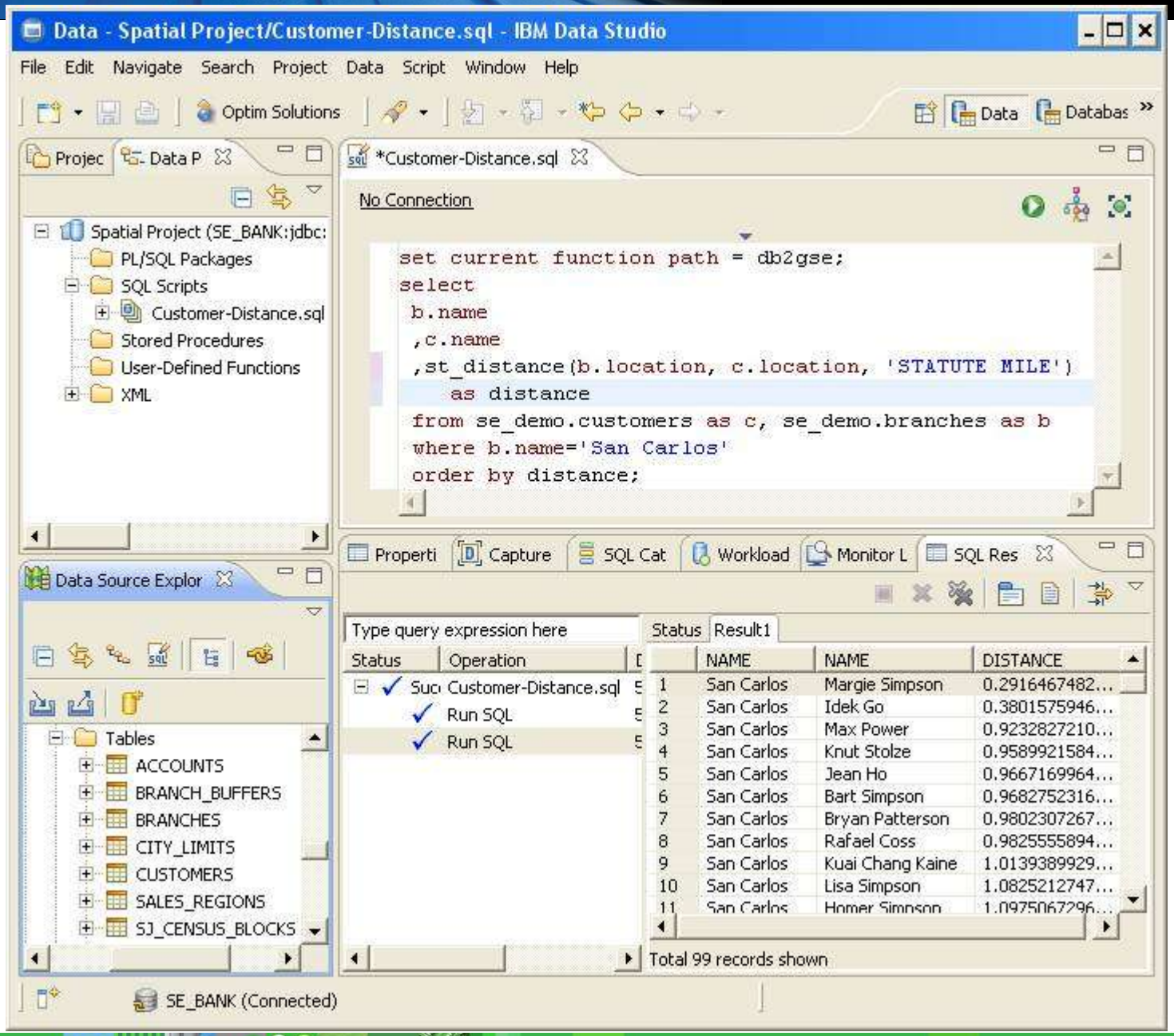
Example problem: Identify customers with a home insurance policy living within 1000 meters of a river who DO NOT have flood insurance.

```
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?

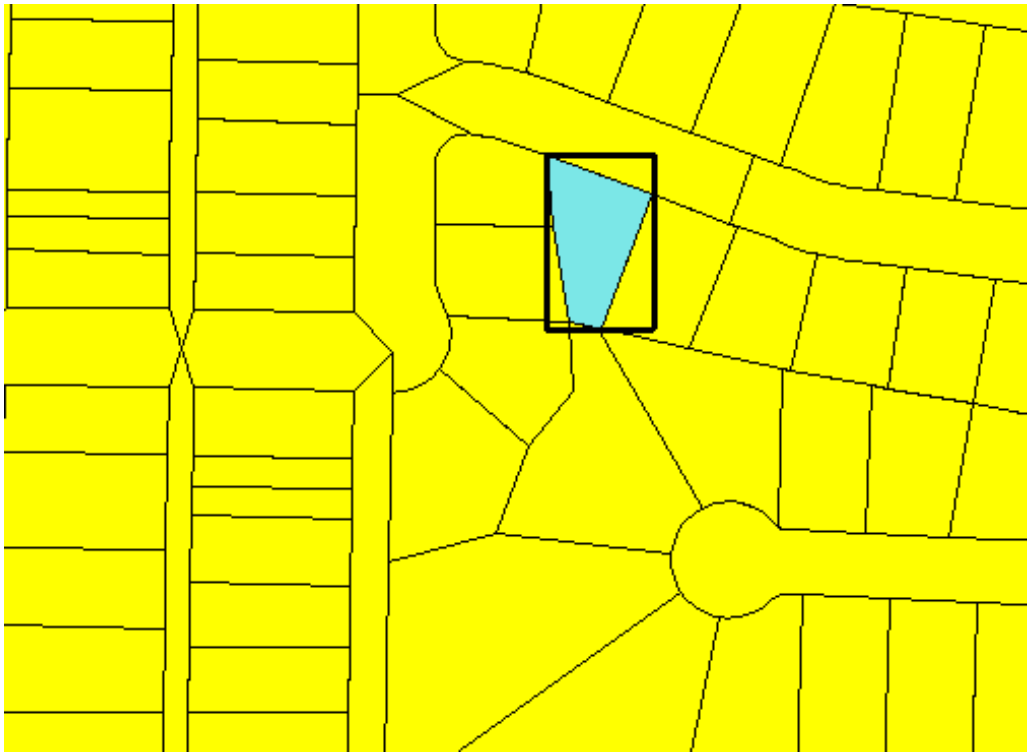
```
SELECT cust_name, outlet_name FROM (
SELECT
   c.name
   ,o.name
   ,RANK() OVER
   (PARTITION BY c.name ORDER BY st_distance(c.location, o.location, 'STATUTE
MILE') ASC)
FROM outlets o
   , customers c
WHERE c.spending > 3000
AND c.year = 2009
AND c.city = 'Chicago'
AND o.city = 'Chicago'
) subq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



Spatial index operation

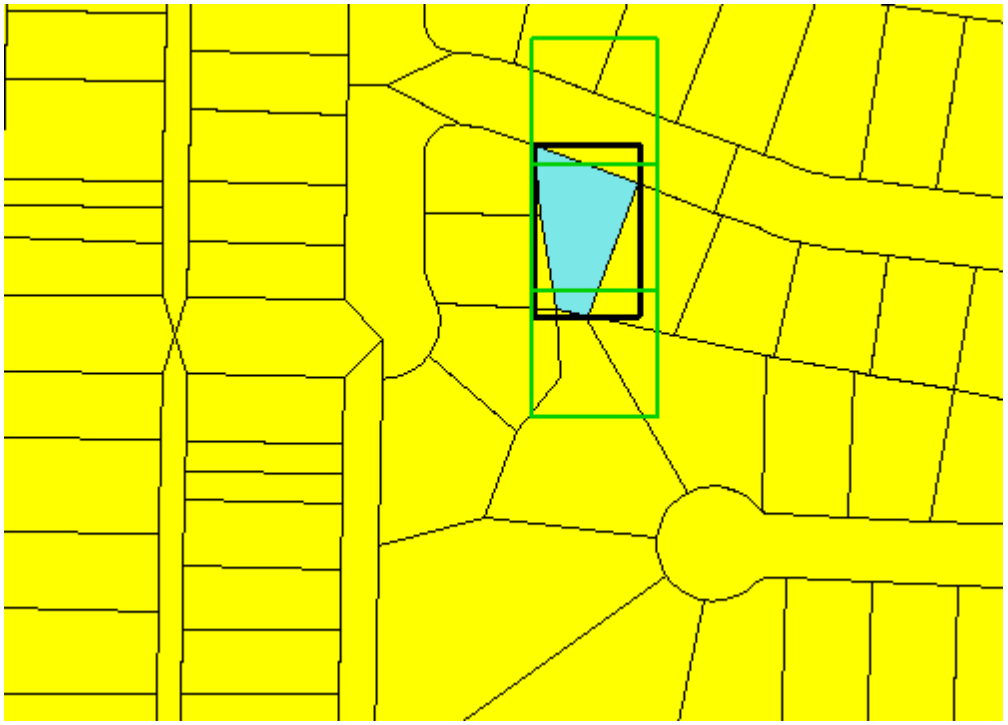
- **When a geometry is inserted**
  - One or more index entries created using 1 of 4 possible grid levels
  - 3 user-specified grid sizes, System-defined “overflow” level
- **When a query is performed**
  - Query envelope computed from “non-index” column value
  - Grid cells intersecting query envelope 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



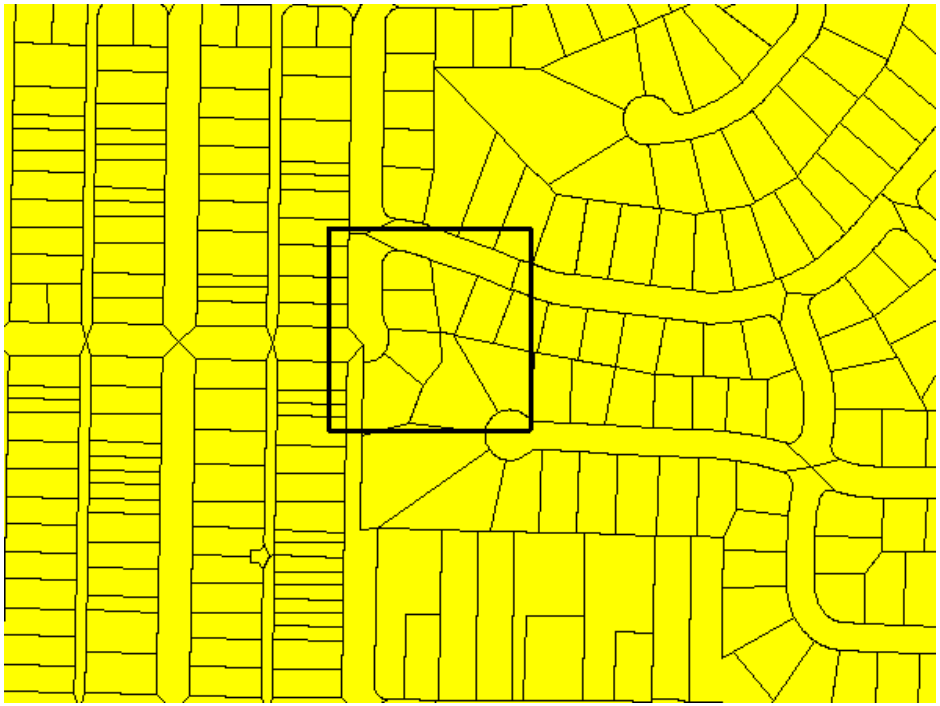
- **Geometry MBR computed**

Grid index generation - continued



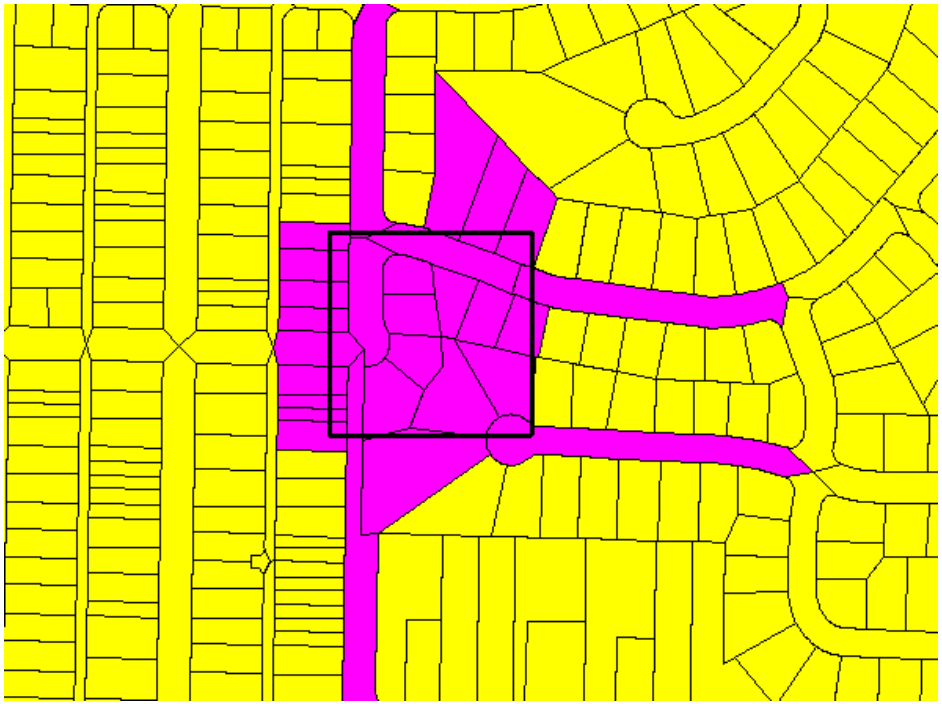
- **In this example 3 grid cells intersect at grid size 100**

Index query



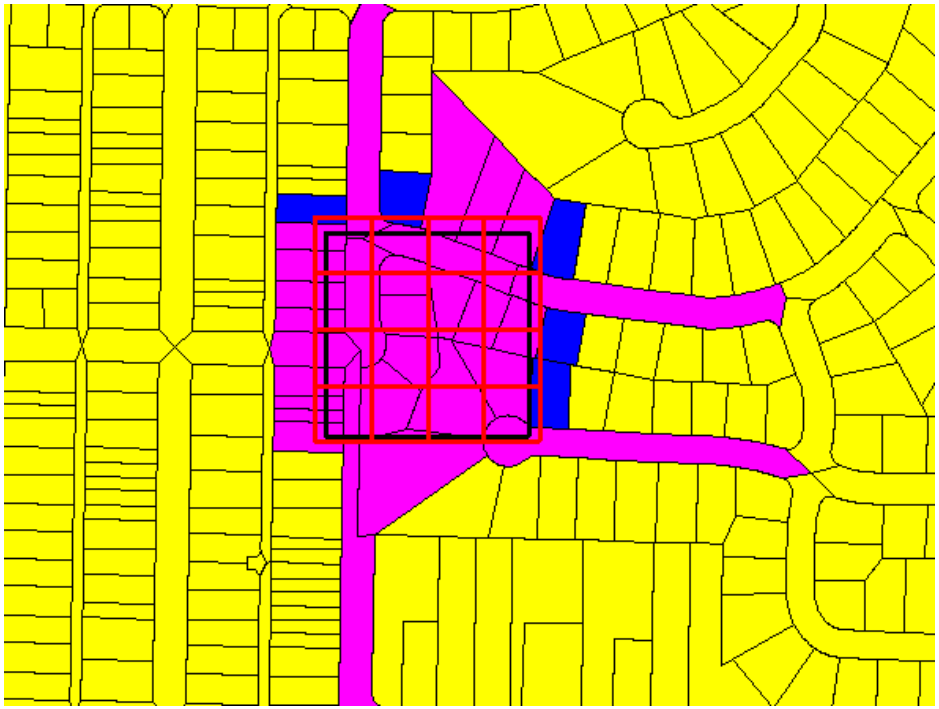
- **We want to find all geometries that intersect the query window represented by the black rectangle**

Index query - continued



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

Index query - continued



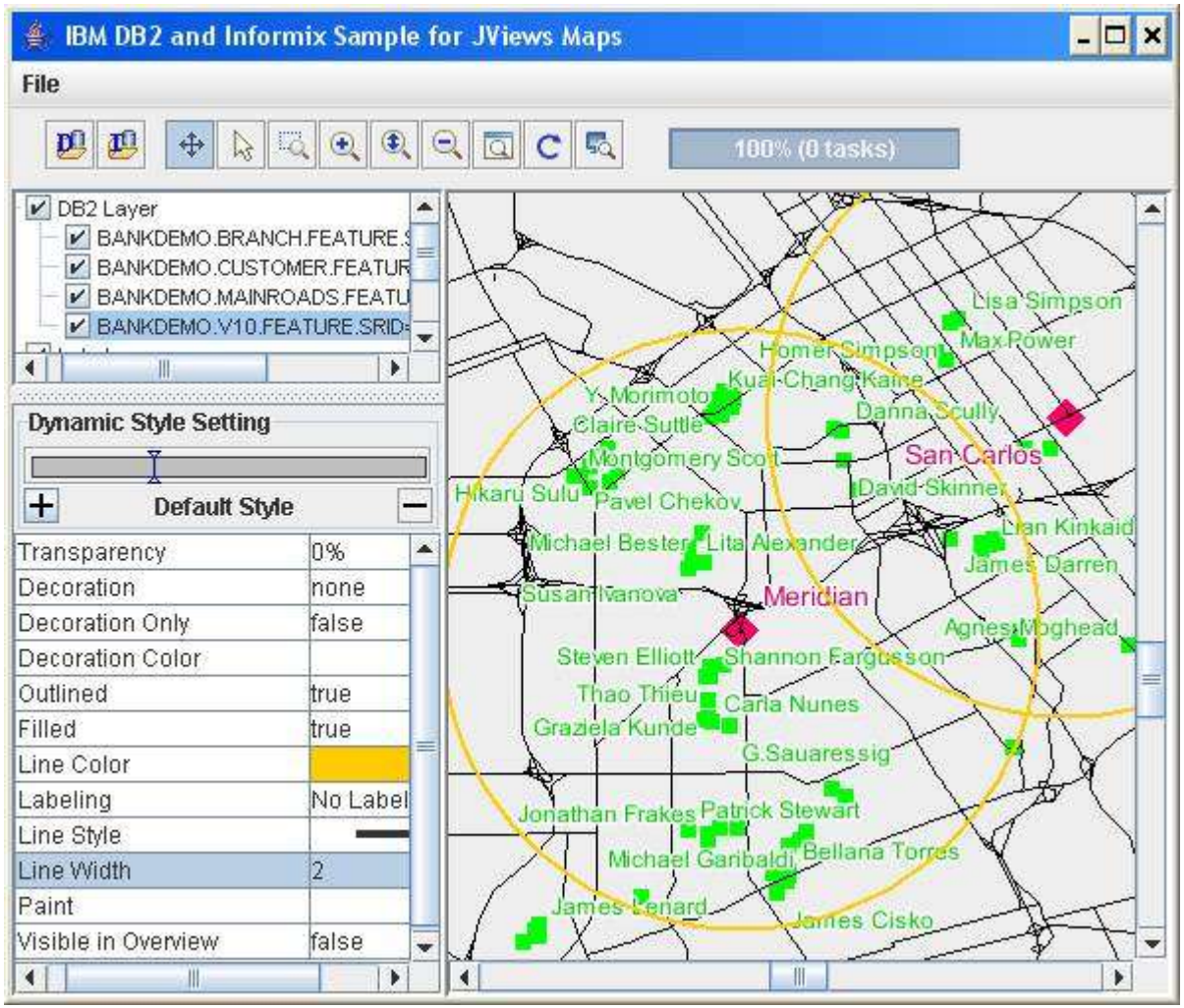
- 33 geometries intersect the grid cells that intersect the query window
- 91 index entries would be scanned
- Only the 28 geometries in magenta will be returned, but the 5 additional geometries need to be scanned and discarded.

How is a spatial index created?

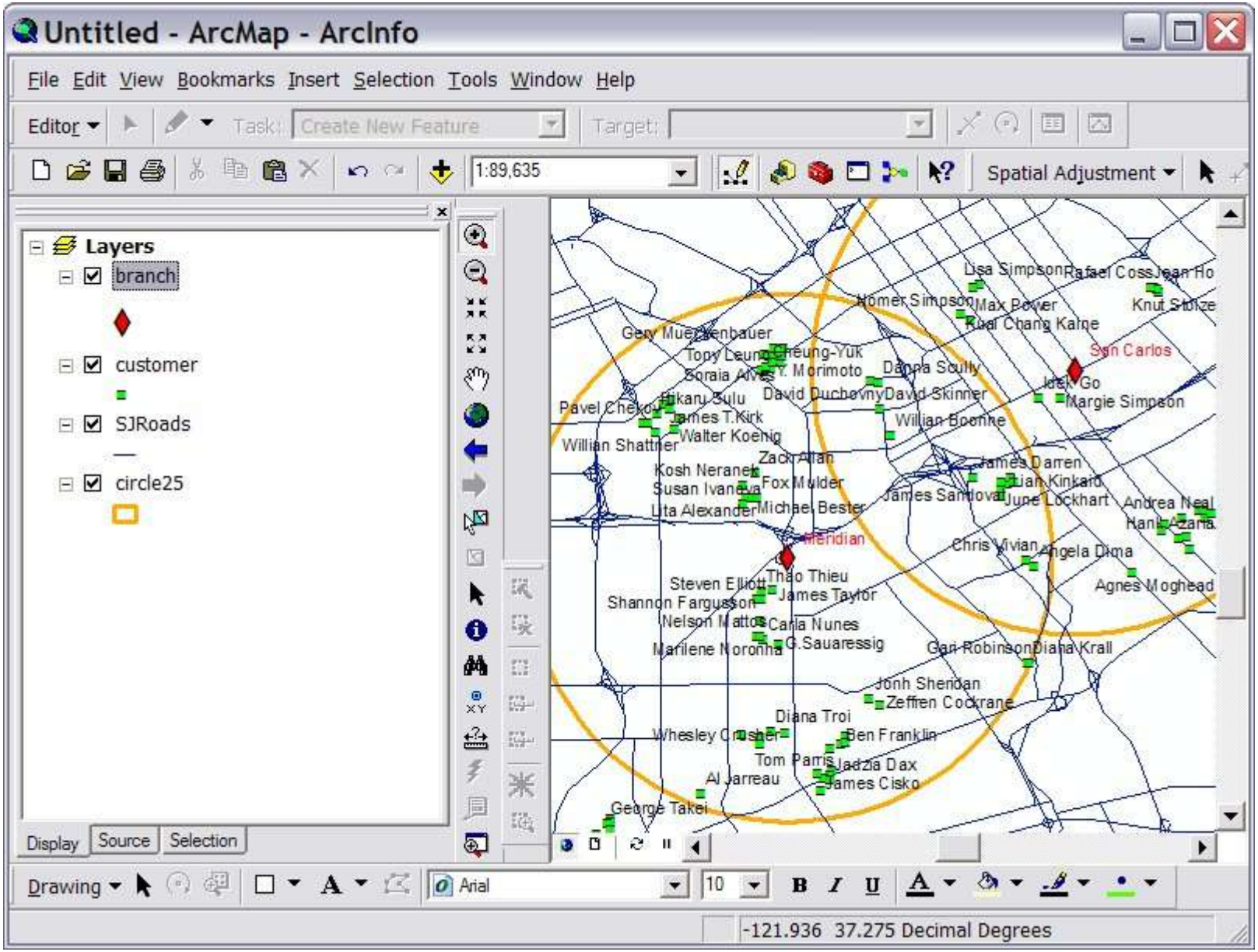
- **CREATE INDEX myindex ON mytable(mycolumn) EXTEND USING db2gse.spatial_index(grid1, grid2, grid3)**
- **Spatial index advisor to analyze spatial column and recommend appropriate grid sizes**
  - 'gseidx' index advisor shipped with Spatial Extender
  - Java index advisor downloadable via the Spatial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

Spatial Application / Tool Support

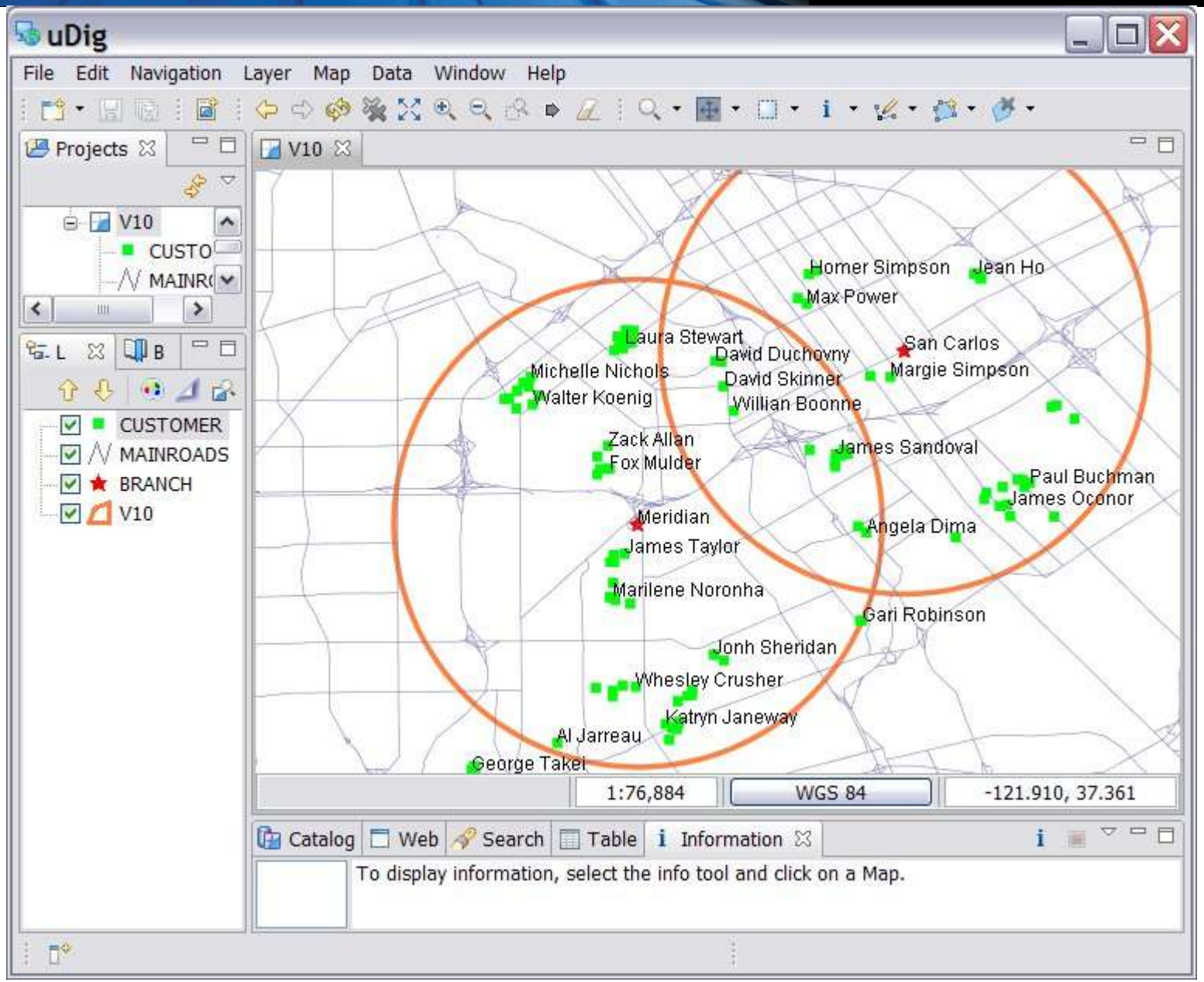
- **Commercial**
  - IBM ILOG JViews Maps
  - IBM InfoSphere Warehouse
  - ESRI ArcGIS
  - (Pitney Bowes) MapInfo MapXtreme for Java
  - Safe FME
- **Open Source**
  - GeoTools – Java toolkit, DB2 plug-in contributed by IBM
  - GeoServer – Web Map/Feature Server
  - uDig – rich GIS application / framework



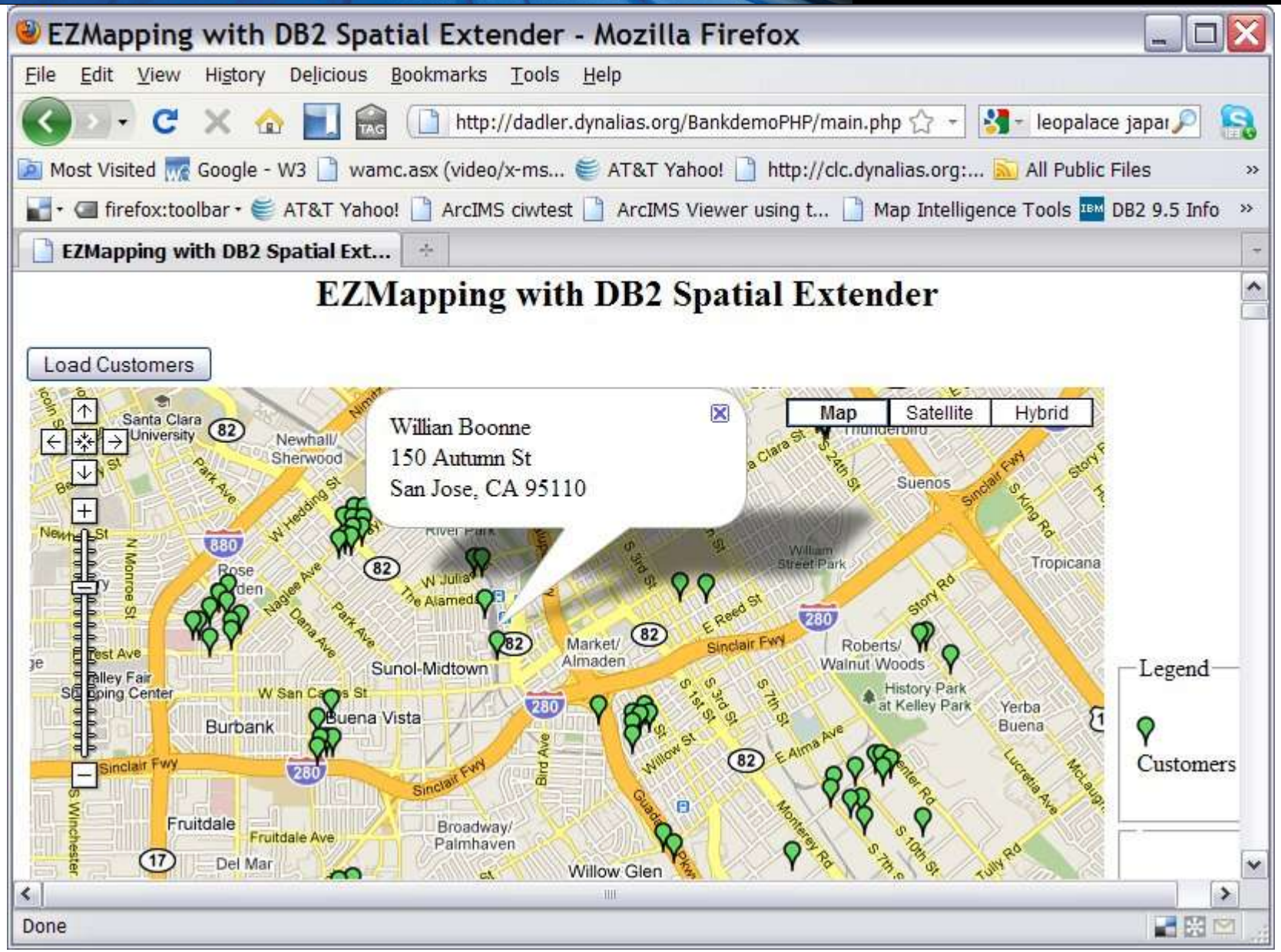
IBM ILOG JViews Maps Example



ESRI ArcMap Example



uDig Example



Google Maps Example

Commander Dashboard - Predictive - Cognos Viewer - Windows Internet Explorer

http://cognosdemo/cognos8/cgi-bin/cognos.cgi?b_action=cognosViewer&ui.action=run&ui.object=%2fcontent%2fpackage

Commander Dashboard - Predictive - Cognos Viewer

Cognos Viewer - Commander Dashboard - Predictive

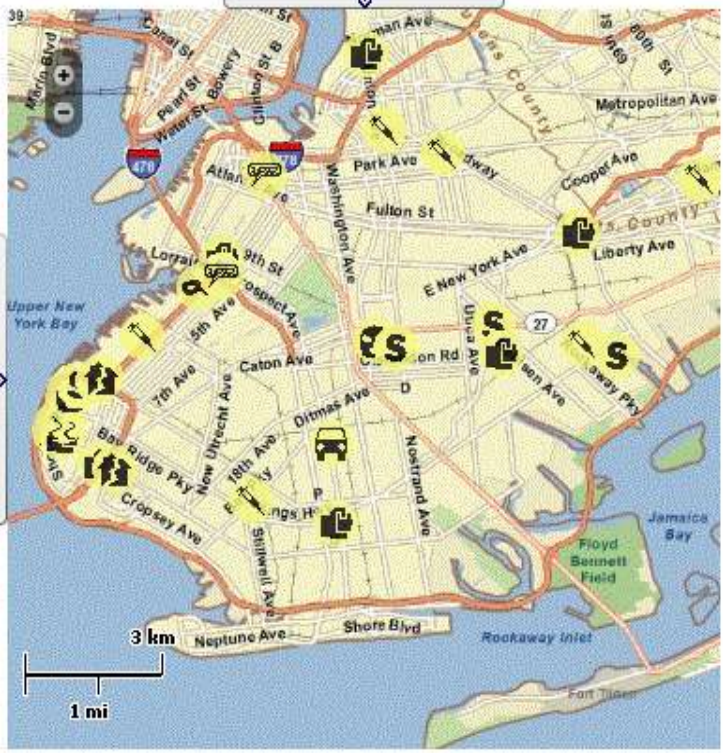
Commander | Log Off | About

Keep this version | Add this report

Commander Dashboard

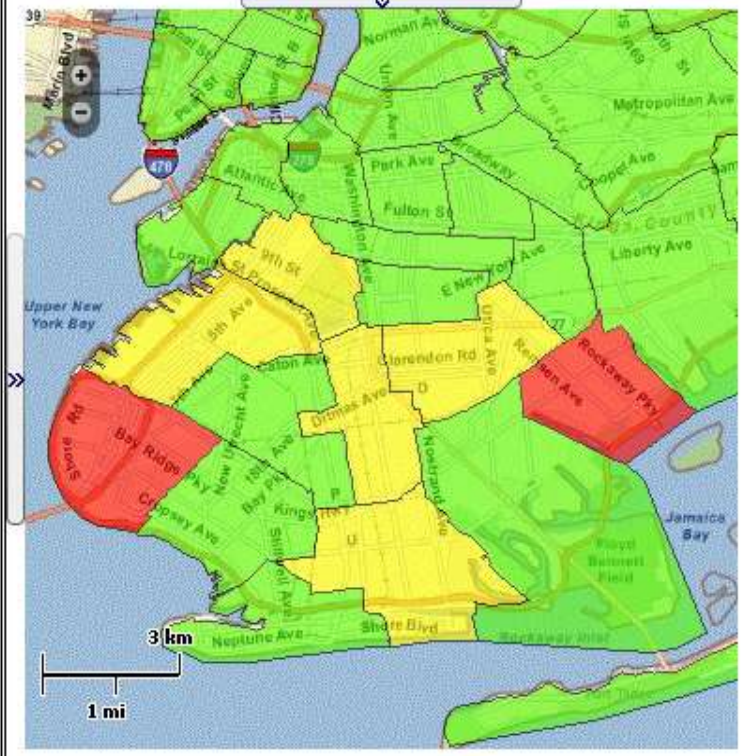
Real Time View (Past 24 hours)

Show/Hide Real Time Params | Submit



Predicted Levels (Next 4 Hours)

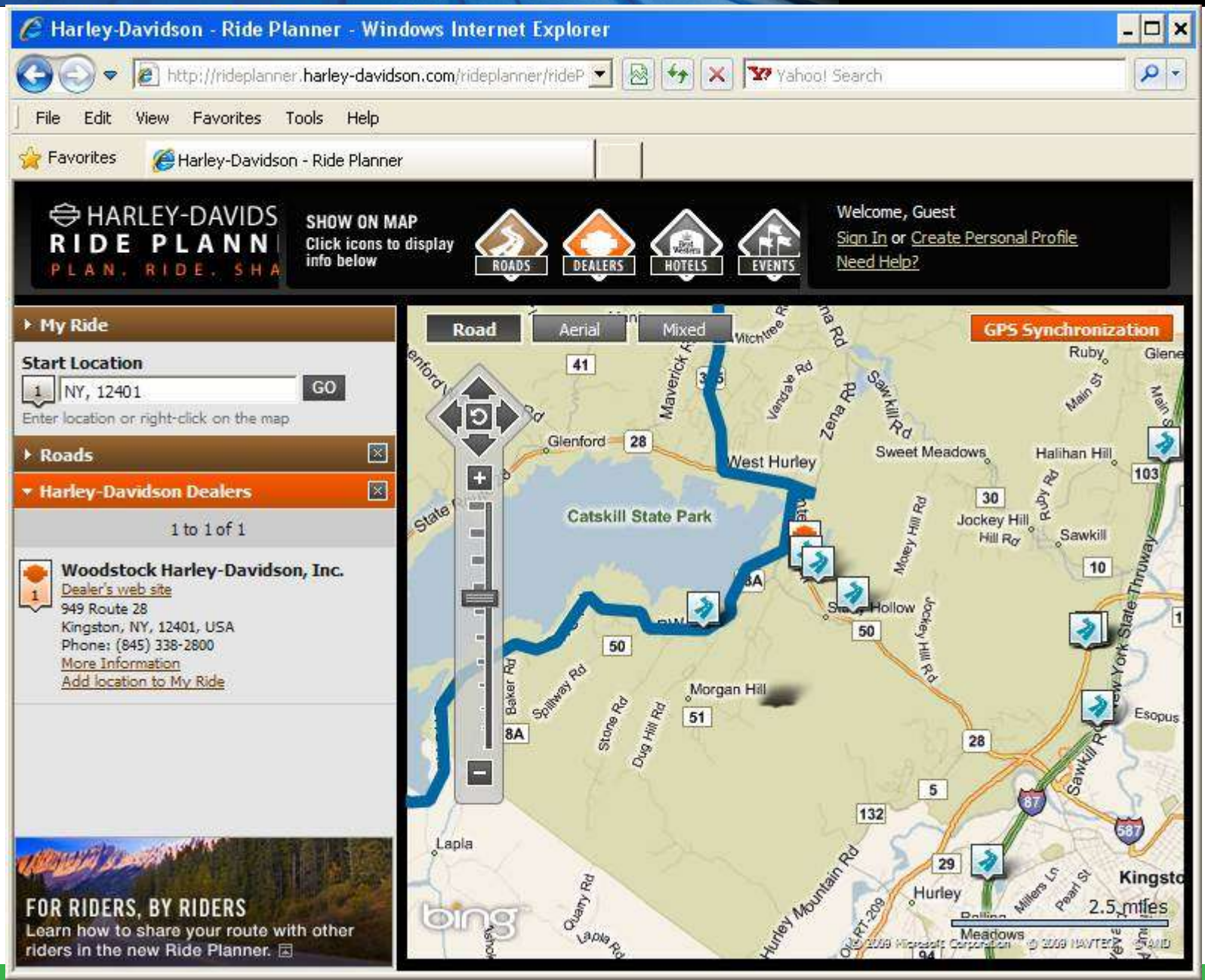
Show/Hide Predictive Params | Submit



IBM Crime Prevention And Prediction Example

Top | Page up | Page down | Bottom

Done | Internet | 100%



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

45

Thank You!

ibm.com/db2/labchats

Thank you for attending!

46