Big Data Spatial Performance with Oracle Database 12c
A Practitioner’s Panel

Siva Ravada
Senior Director, Software Development
Oracle Spatial Technologies
Big Data Spatial Performance with Oracle Database 12c
A Practitioners Panel

Siva Ravada
Senior Director, Software Development
Oracle Spatial Technologies
Safe Harbor Statement

The following is intended to outline our general product direction. It is intended for information purposes only, and may not be incorporated into any contract. It is not a commitment to deliver any material, code, or functionality, and should not be relied upon in making purchasing decisions. The development, release, and timing of any features or functionality described for Oracle’s products remains at the sole discretion of Oracle.
Today’s Panelists

• Nick Salem  
  Distinguished Engineer  
  Neustar, Inc.

• Dan Geringer  
  Solutions Architect  
  Oracle Spatial Development

• Steve Pierce  
  Chief Executive Officer  
  Think Huddle

• Siva Ravada  
  Senior Director  
  Oracle Spatial Development
Oracle Spatial and Graph
Complete, Open, Integrated, Most Widely Used

- "Points"
- "Lines"
- "Polygons"

Web Services (OGC)
SPARQL End Point
Rasters
Network Graphs
Topologies
RDF Semantic Graphs

Geocoding
Routing
Inferencing

3D
Oracle’s Spatial Stack

- Location-Enabled Applications
- Fusion Middleware
- MapViewer
- Oracle Spatial and Graph
- Designed for Exadata
- Bundled and Cloud-based Map Content
Integrated Geospatial Middleware Features

- Map Visualization
- HTML5 and Javascript APIs
- Web Map Server, Web Tile Server

- Geocoding and Routing Engine deployment
- Geospatial-enabled Event Processing
- Geospatial-enabled Business Intelligence
Advances with Oracle Spatial and Graph 12c

Dramatic Performance

Simplified Application Development
Oracle Spatial and Graph 12c
Dramatic Performance for most common operations

Up to 300 times faster performance

Join
Touch
AnyInteract
Aggregate
Contains
Overlaps

Complex Masks
Distance
Within Distance
Inside
Validate Geometry
Many other...
Benchmark

Simon Greener (SpatialDBAdvisor.com)

• Hardware and OS
  – Linux 6 OEL
  – Database 12cR1
  – 16 CPU (x4 cores)
  – Memory 48GB

Benchmark Dataset

<table>
<thead>
<tr>
<th>Table Name</th>
<th>Number of Rows</th>
<th>Geometry Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>PLACE</td>
<td>546652</td>
<td>POINT</td>
</tr>
<tr>
<td>ROAD</td>
<td>15624811</td>
<td>LINE</td>
</tr>
<tr>
<td>BUILDING</td>
<td>3091262</td>
<td>Very Small POLYGON</td>
</tr>
<tr>
<td>BLOCKGROUP</td>
<td>215881</td>
<td>Small POLYGON</td>
</tr>
<tr>
<td>US_STATES</td>
<td>50</td>
<td>Large POLYGON</td>
</tr>
</tbody>
</table>
Database Query Tests

**AGGREGATE UNION:**
BlockGroups aggregated by States.

**TOUCH:**
Find BlockGroups Touching a given blockgroup.

**ANYINTERACT:**
Find Roads that intersect a given BlockGroup.

Total time for 1000 queries

Source: SpatialDB Advisor

Copyright © 2014, Oracle and/or its affiliates. All rights reserved.
Database Query Tests

**INSIDE:**
Find Roads that are inside a given BlockGroup
Total time for 1000 queries

**INSIDE:**
Find Roads that are inside a given State
Total time for 1000 queries

Source: SpatialDB Advisor
Panelist

Nick Salem
Distinguished Engineer
Neustar, Inc.
Neustar Overview

- Publicly held $902M+ company serving internet, telco, information services, retail, media and advertising industries

- ElementOne Platform
  - Geospatial analytics in the cloud
  - Feature rich, secure and highly scalable
  - 1.9TB database size and growing

- Oracle Database 12c Enterprise Edition
  - Spatial and Graph Option with GeoRaster, Network Data Model
  - Partitioning
  - Real Application Clusters
  - Advanced Security

- Oracle Fusion Middleware 11.1.1.7
  - MapViewer
Geospatial Analytics in the Cloud

- Highly flexible, detailed and high performance mapping
- Road network / drive time analysis (NDM)
- GeoRaster image overlays
- Heat Maps (SDO_TIN)
- Aerials / terrain map overlays (SDO_CS)
- SDO_UTIL and SDO_GEOM for various geography creation and manipulation
- Shape intersections, XOR, union, difference, point at bearing, radian, distance, convex hull, polygon smoothing and cleansing
Oracle Spatial - 12c Benchmarks

- Oracle VM Server 3.0 Test Benchmark Environment
  - 2 identical VMs with Oracle Linux 6.5
  - 8G of SGA / 2GA of PGA
  - 4 x 8 Core CPUs (Intel Xeon 2.70GHz)
  - Comparison: 11g vs. 12c (12.1.0.1)

- 300x faster using SDO_AGGR_UNION aggregating CA ZIPs

- 8x performance gain using SDO_TOUCH (California ZIPs with block groups)

- 6x faster using SDO_WITHIN_DISTANCE to retrieve all US biz within a 10 mile radius of 544 shopping centers

- 10x faster using SDO_CS.TRANSFORM to convert all US block groups to world mercator projection
Oracle Spatial - 12c Benchmarks

- **3x** faster using SDO_GEOM.RELATE to determine topological relationships for California ZIPs and surrounding block groups
- **6.5x** faster using SDO_JOIN to perform any interact operation on the complete ZIP and block group layer
- **2.6x** faster using SDO_GEOM.DISTANCE to calculate distances between a site and surrounding 112k sites
- **2.1x** faster using SDO_INSIDE to retrieve all businesses inside NY block groups
- **1.8x** faster using VALIDATE_GEOM to validate all US county geometries
Panelist

Steve Pierce
Chief Executive Officer
Think Huddle
Our Approach / A Join Use Case

• About Think Huddle
  – Oracle Spatial and MapViewer
  – Big Data and BI for Spatial
  – Fortune 500 clients

• Test Configuration
  – 1 server / 2 Oracle databases
  – Oracle Linux 6.5
  – 8GB of SGA / 2GB of PGA
  – 2 x E5 Xeon 2.9Ghz


Elapsed Time (s)
Geometry Aggregation Use Case

- Merge Adjoining Census Tracts in Virginia
  - SDO_ANY_INTERACT + SDO_AGGR_UNION
  - ~2,000 aggregations
  - Tracts have 7 neighbors = 14,000 tracts
  - 280 tracts processed per second

```sql
SELECT SDO_AGGR_UNION(sdoaggrtype(a.geom, .05)) INTO geom
FROM tracts a
WHERE SDO_ANYINTERACT(a.geom, rec.geom) = 'TRUE';
```
Point in Polygon Use Case

**In 11g:**

```
SELECT lon, lat
FROM   (SELECT sdo_geometry(2001, 8307, sdo_point_type(lon, lat, NULL), NULL, NULL) geom
        FROM   large) c,
       states s
WHERE SDO_ANYINTERACT(s.geom, c.geom) = 'TRUE'
    AND s.state_abrv = 'CA'
```

**In 12c:**

```
SELECT /*+ PARALLEL(#) */ lon, lat
FROM   TABLE(
    SDO_POINTINPOLYGON ( CURSOR(
        SELECT lon, lat
        FROM   large),
        v_state_geom)
    , .05)
```

- Sdo_PointInPolygon
- Near Linear Performance Gain with Parallel
Query comparison – 11g vs 12c:

Video

https://www.youtube.com/watch?v=fzUktlY9ewM&feature=youtu.be
Panelist

Dan Geringer
Solutions Architect
Oracle Spatial Development
Oracle Spatial and Graph
LIDAR Data Model
Optimized For Oracle Engineered Systems
What is LIDAR Data

Laser Imaging Detection and Ranging (LIDAR) Data

• Optical remote sensors that collect millions of 3D points per second, along with numeric attributes associated with each point.

• Sensors targeted at the ground to generate precise elevation data models, sometimes called point clouds.

• Provides accurate representation of:
  – Railway infrastructure
  – Highways, roads, buildings, bridges
  – Forestry terrain
  – Bathymetry (sea floor elevations) and more…
LIDAR Data – Major Challenges

• Storage
  – Where do I put all this data? Archive?
  – Compress?
  – Optimal format for analysis? LAS, CSV, Proprietary

• Analysis
  – Derivative product generation (TINs, Contours, DEMs)
  – Spatial queries
Oracle Engineered Systems – Exadata and SPARC Supercluster

Massive Parallel Architecture - Unique capabilities

- **Hybrid Columnar Compression – HCC**
  - Four levels of compression. Data searched in compressed form.
    - COMPRESS FOR QUERY LOW
    - COMPRESS FOR QUERY HIGH
    - COMPRESS FOR ARCHIVE LOW
    - COMPRESS FOR ARCHIVE HIGH

- **Smart Scan**
  - Storage servers in addition to traditional compute node servers
  - Smart scans offload SQL predicate filtering to the raw data blocks
  - Results in much less data transferred, and dramatically improved performance
  - No index searches
Exadata X4-2 Configurations

- Eighth Rack – 24 cores
- Quarter Rack – 48 cores
- Half Rack – 96 cores
- Full Rack – 192 cores
Exadata X4-2 Quarter Rack Diagram

Exadata DB Machine X4-2 Quarter Rack

storage server
12 Xeon cores

RAC, OLAP Partitioning Compression

24 Xeon cores

storage server
12 Xeon cores

storage server
12 Xeon cores
Oracle’s Newest LIDAR Data Model

• “New paradigm” for LIDAR data, optimized for Oracle engineered systems.

• On Exadata and SPARC Supercluster (SSC), leverages Oracle’s unique:
  – HCC compression technique, for extremely high compression rates
  – Parallel Enabled Smart Scan for extreme performance, including spatial queries.
  – **Spatial queries with no spatial/non-spatial indexes**, against LIDAR data in compressed form
  – Any polygon shape
Oracle’s Newest LIDAR Data Model
(continued...)

• “Back to basics... a simple flat relational model
• X,Y,Z and attributes stored as ordinary Oracle NUMBER columns
• Table can contain other columns data types too.
• Works with any point table/view, not just LIDAR data
What we tested

• 639+ Billion Points (639,478,217,460 rows)
  – 60,185 LAZ formatted files
  – Uncompressed as LAS – 11.63 Terrabytes
  – HCC Compressed For Query High on Exadata - 2.24 Terrabytes
X4-2 Full Rack – Massive Parallel Load

Stream LAZ, LAS or CSV Formatted Files

- 639+ Billion Points (639,478,217,460 rows)
- 60,185 LAZ formatted files streamed directly into Oracle
- Utilized External Table preprocessor with las2txt open source ETL tool
- QUERY HIGH compression (2.24 Terrabytes)
- 4 Hours 39 Minutes (or 38,193,765 rows per second)
Polygon Query Example

- 300 meter buffer around a 9.44 km diagonal line
- Each cell contains approximately 3 million points
- Query returned 20,122,526 points
- 1,166,523 points/second
- Compressed For Query High
SPATIAL QUERY IN COMPRESSED FORM

QUERY RATES OVER 2,000,000 POINTS/SECOND

SPATIAL SEARCH – ANY POLYGON SHAPE

NO INDEXES
Parallel Enabled
Spatial Operators and Functions
Validation Of Home Appraisals

Exadata Results

• Validate home appraisals for a Government Sponsored Enterprise (GSE)
• Requirement - Find all the parcels touching parcels to validate appraisals
• Processed 2,018,429 parcels
  – Exadata X2-2 ½ RAC:
    • Serially – **38.25 minutes**
    • Parallel - 48 cores (45x faster) - **50 seconds**
  – Exadata X3-2 Full RAC (128 cores) about 120x faster
  – Exadata X4-2 (192 cores) even faster
Q & A

More Spatial & Graph at OOW...

• Spatial & Graph Sessions/Activities List: www.tinyurl.com/spatial2014

• Talk to us at
  • Oracle Spatial & Graph SIG Meetup, today at 4-5pm, OTN Lounge (Moscone South Lobby Mezzanine)
  • Spatial demo booth – in the Database DEMOgrounds – Moscone South Left, pod SLD-114
## Spatial and Graph at OOW 2014

View [this list at tinyurl.com/spatial2014](http://tinyurl.com/spatial2014)

### Sessions

<table>
<thead>
<tr>
<th>Date/Time</th>
<th>Title</th>
<th>Location</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Wednesday, Oct 1</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12:45 PM – 1:30 PM</td>
<td>PL/SQL: The Scripting Language Liberator (with Steve Feuerstein)</td>
<td>Moscone South - 307</td>
</tr>
<tr>
<td>2:00 PM – 2:45 PM</td>
<td>Big Data Spatial Performance with Oracle Database 12c</td>
<td>Moscone North - 130</td>
</tr>
<tr>
<td>2:00 PM – 2:45 PM</td>
<td>How Oracle Real-Time Decisions Meets the Challenges of the Italian Ministry of Labor</td>
<td>Moscone West - 311</td>
</tr>
<tr>
<td><strong>Thursday, Oct 2</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9:00 AM – 9:45 AM</td>
<td>Managing Linear Assets with Oracle Enterprise Asset Management</td>
<td>Westin Market Street - City</td>
</tr>
<tr>
<td>12:00 PM – 12:45 PM</td>
<td>Garmin Uses Oracle Exadata and Oracle Spatial and Graph to Manage Five Billion Miles of Data</td>
<td>Moscone South - 306</td>
</tr>
<tr>
<td>2:30 PM - 3:15 PM</td>
<td>Big, Fast Graph Analysis and Data Management for Hadoop</td>
<td>Moscone North - 131</td>
</tr>
</tbody>
</table>
Oracle BIWA Summit 2015
Now including Oracle Spatial Summit

• BIWA SIG joins hands with Spatial SIG and NoCOUG for the Oracle BIWA and Spatial Summits

• Jan. 27-29, Oracle HQ in Redwood Shores, CA

• Keynotes by Ken Inman, Neustar; Balaji Yelamanchili, Oracle

• Educational Sessions, Hands On Labs, Networking

• Abstract submissions still being accepted – submit soon!

• www.biwasummit.org
View preliminary agenda, submit session proposals
User Groups

• **Oracle Spatial & Graph Special Interest Group:**
  Connect and exchange knowledge with the user community

  **Join the Spatial & Graph Meetup:**
  **Wed 10/1, 4:00-5:00pm**
  **OTN Lounge, Moscone South Lobby Mezzanine**
  Share experiences, ask your questions to SIG & Oracle Spatial experts

  – Join our groups: LinkedIn, Google+, IOUG SIG
  – Visit OTN Spatial – Community
  Search online for “Oracle Spatial and Graph Community”
  – Email oraclespatialsig@gmail.com

• **Individual Certification, Partner Specialization**
  - Talk to Oracle team this week at OPN Lounge or Oracle Spatial & Graph demopod in Moscone South
Resources: Oracle Technology Network

• www.oracle.com/technetwork/database/options/spatialandgraph
• www.oracle.com/technetwork/middleware/mapviewer
• blogs.oracle.com ➔ oraclespatial ➔ oracle_maps_blog
More Resources
More Spatial & Graph at OOW

• Spatial & Graph Sessions/Activities List:  www.tinyurl.com/spatial2014

• Talk to us at
  • Oracle Spatial & Graph SIG Meetup, today at 4-5pm, OTN Lounge (Moscone South Lobby Mezzanine)
  • Spatial demo booth – in the Database DEMOgrounds – Moscone South Left, pod SLD-114