Oracle Spatial Summit 2015

Performance, Performance, Performance Exadata For Massive Spatial Workloads

Daniel Geringer
Oracle Corporation

ORACLE

Oracle 12c
Much Faster Spatial Algorithms

SPATIAL_VECTOR_ACCELERATION
Oracle 12c Initialization Parameter

- New faster algorithms for spatial operators and functions (up to 350x)
- Metadata caching increases performance:
  - For all spatial operators and functions
  - For all DML operations (INSERT, UPDATE, DELETE)
- Recommended for any application with mission critical spatial query performance requirements.
- Requires Oracle Spatial and Graph License
  - ALTER SYSTEM SET SPATIAL_VECTOR_ACCELERATION = TRUE
  - ALTER SESSION SET SPATIAL_VECTOR_ACCELERATION = TRUE

Oracle Exadata Database Machine
Engineered System

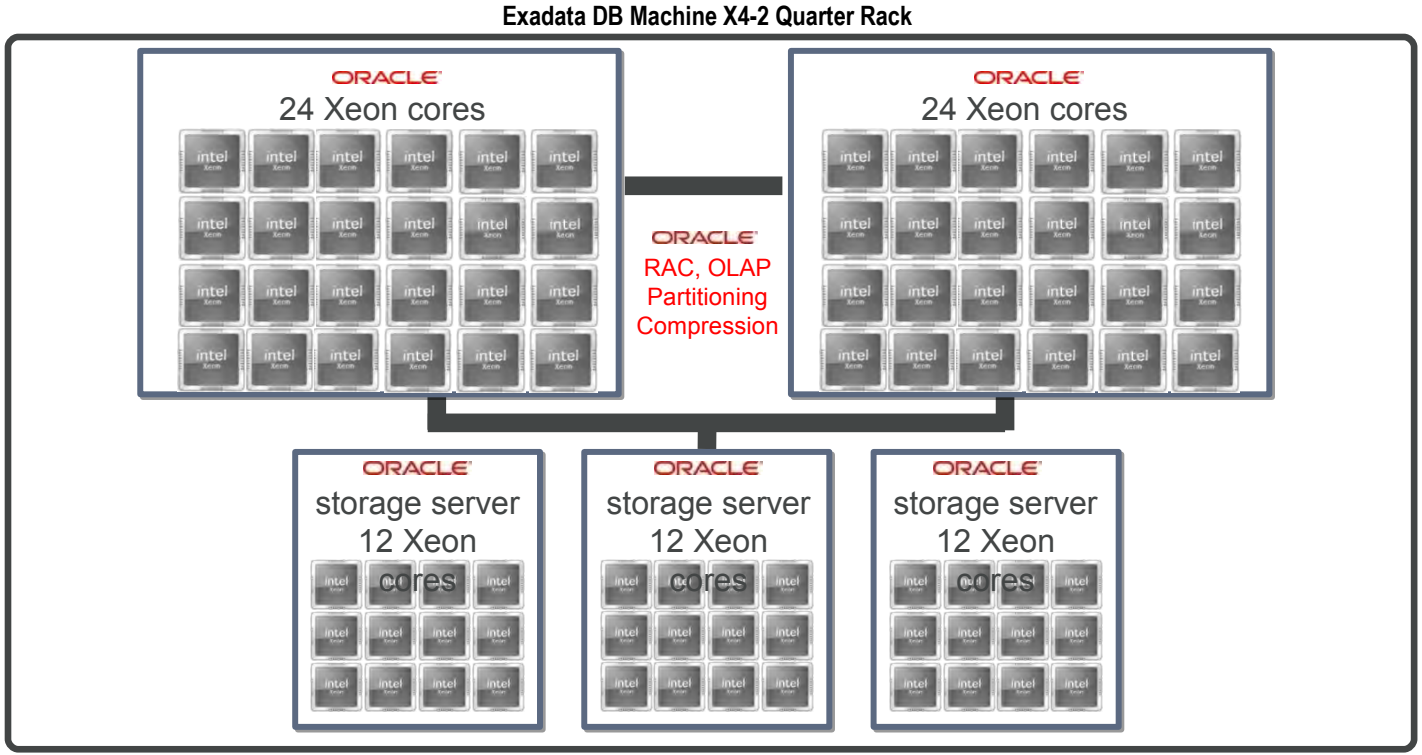
What Is the Oracle Exadata Database Machine?

- Oracle SUN hardware uniquely engineered to work together with Oracle database software
- X4-2 Key features:
  - Database Grid – Up to 192 Intel cores connected by 40 Gb/second InfiniBand fabric, for massive parallel query processing.
  - Raw Disk – Up to 300 TB of uncompressed storage (high performance or high capacity)
  - Memory – Up to 4 TB
  - Hybrid Columnar Compression (HCC) – Query and archive modes available. 3x to 30x compression.
  - Storage Servers – Up to 14 storage servers (168 Intel cores) that can perform massive parallel smart scans. Smart scans offloads SQL predicate filtering to the raw data blocks. Results in much less data transferred, and dramatically improved performance.
  - Flash memory – Up to 44 TB

Exadata Configurations

- X4-2
  - Eight Rack – 24 cores
  - Quarter Rack – 48 cores
  - Half Rack – 96 cores
  - Full Rack -192 cores

Exadata X4-2 Quarter Rack Diagram



Parallel Spatial Operators And Functions
Key Differentiator

Parallel Query and Spatial

US Rail Application

Parallel Query And Spatial Operators

US Rail Application

- Requirement
  - GPS locations for each train collected throughout the day
  - Each location has other attributes (time, speed, and more)
  - GPS locations have a degree of error, so they don't always fall on a track.
  - Bulk nearest neighbor queries to find closest track, and project reported train positions onto tracks
- This information is used for:
  - Tracking trains
  - Analysis for maintenance, ensure engineers are within parameters, etc…

Parallel Query And Spatial Operators

What we tested

- 45,158,800 GPS train positions.
- For each train position:
  - Find the closest track to the train (with SDO_NN)
  - Then calculate the position on the track closest to the train

Parallel Query and Spatial Operators

US Rail Application

```
CREATE TABLE results PARALLEL 72 NOLOGGING AS
SELECT a.locomotive_id,
            sdo_lrs.find_measure (b.track_geom, a.locomotive_pos)
FROM locomotives a,
          tracks b
WHERE sdo_nn (b.track_geom, a.locomotive_pos,
                      'sdo_num_res=1')  = 'TRUE';
```

Parallel Query And Spatial Operators

Exadata Results

- On Exadata X2-2 Half RAC:
  - 34.75 hours serially vs. 44.1 minutes in parallel
  - 48 database cores - 47x faster
- On Exadata X3-2 Full Rack
  - 128 database cores – about 125x faster
  - About 16.6 minutes in parallel
- X4-2 Full Rack (192 cores) even faster

Parallel Query and Spatial

Government Sponsored Enterprise Validation of Home Appraisals

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

Parallel Enabled Geocoding

Oracle Spatial & Graph - Geocoder

- Geocoder is included in your Oracle Spatial and Graph license.
- Open data model for Geocoder reference data
- If you have reference data, you can populate the data model yourself
- If you don't have the reference data, Oracle Partners sell it in Transportable Tablespace format (plug and play data).
  - HERE
  - Tom Tom
  - ADCI
  - others

Oracle Spatial and Graph Geocoder

- Forward / Reverse / Street Centerline / Rooftop (point based) support
- In database geocoding –
  - PL/SQL APIs
  - Optimal for parallel enabled batch geocoding
  - For batch processing, leverage parallel enabled pipeline table functions
- Web service based geocoding
  - Java servlet based with XML geocoding APIs
  - Deployed in J2EE container
  - Optimal for non-batch request in web based applications.
  - Can perform batch processing too

Geocode Times On Exadata X4-2 1/2 RAC

- X4-2 with 96 cores
- Geocoded 77216 addresses in 3.32 seconds
- 23,257 geocodes per second

Oracle Spatial and Graph
Massive Point 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

Oracle's Newest Massive Point 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 Terabytes
  - HCC Compressed For Query High on Exadata - 2.24 Terabytes

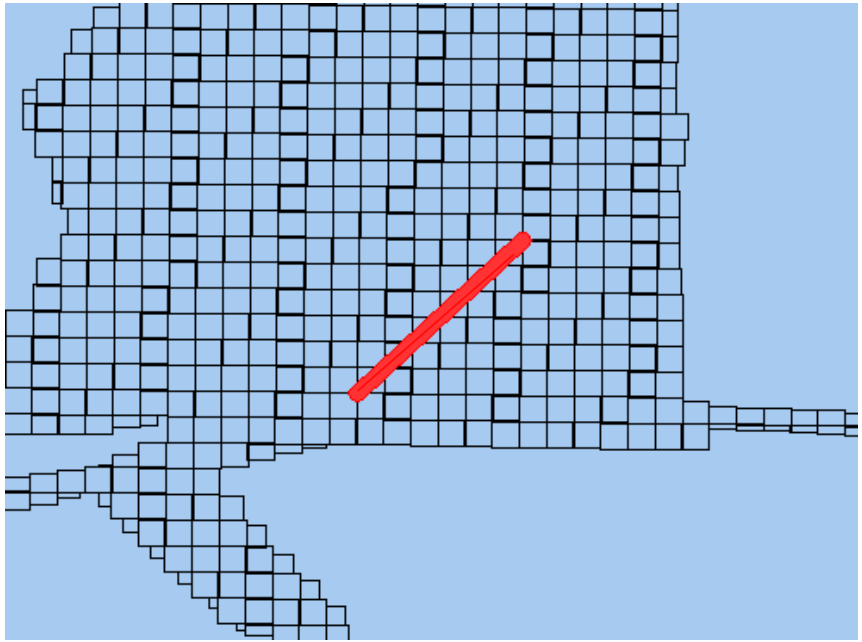
X4-2 Full Rack – Massive Parallel Load

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 Terabytes)
- 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

HCC and Spatial

Hybrid Columnar Compression and Spatial

HCC and Spatial

- Point, Line and Polygon geometries can all benefit from HCC
- Lines and Polygons, they must be stored inline (less than 4K in size).
- Options include:
  - COMPRESS FOR QUERY LOW
  - COMPRESS FOR QUERY HIGH
  - COMPRESS FOR ARCHIVE LOW
  - COMPRESS FOR ARCHIVE HIGH

HCC and Spatial

- Two ways to compress:
  - Create Table As Select
  - Direct Path Inserts

1. Create Table As Select

```
CREATE TABLE edges_compressed
COMPRESS FOR QUERY LOW
NOLOGGING AS SELECT * FROM edges;
```

EHCC and Spatial

2. Direct Path Inserts (full code example in presentation appendix)

```
-- PL/SQL Example with append_values hint.
DECLARE
  id_tab       ID_TAB_TYPE;
  edge_tab  GEOM_TAB_TYPE;
BEGIN
   -- Population of id_tab and edge_tab shown in presentation appendix
   FORALL i IN edge_tab.first .. edge_tab.last
       INSERT /*+ append_values */ INTO edge_ql VALUES (id_tab(i), edge_tab(i));
   COMMIT;
```

HCC and Spatial – Uniform Geometries

Strategy For Much Higher Compression Rates

- Uniform geometries spatial layers have the same number of coordinates in every row.
- Some examples:
  - Point data (x NUMBER, y NUMBER)
  - Box polygon (lx NUMBER, ly NUMBER, ul NUMBER, uy NUMBER)
  - Two point line (x1 NUMBER, y1 NUMBER, x2 NUMBER, y2 NUMBER)
  - Four point polygon (x1 NUMBER, y1 NUMBER, …, x5 NUMBER, y5 NUMBER)
- For much higher compression rates, store uniform geometries as a series of NUMBER columns instead of SDO_GEOMETRY

HCC and Spatial – “Uniform” Geometries

Box Polygon With Function Based Index - Example

- Create a function based index on uniform geometries to perform spatial queries
- The following ANYINTERACT queries were run on a 116 million row table
- Query High compression – 18.17x… queries still very fast.

Anyinteract Query	Uncompressed	Query Low 3.92x comp	Query High 18.17x comp	Archive High 21.57x comp
10 acre polygon (487739 rows returned)	1.86 sec	2.02 sec (1.08x perf)	2.7 sec (1.45x perf)	12.75 sec (6.85x perf)

EHCC – Non Uniform Geometries

- Non-uniform geometries layers can have a different number of coordinates in every row.
- Some examples:
  - Zip code polygons
  - County polygons
  - Road line strings
- Use SDO_GEOMETRY for non-uniform geometry columns

EHCC and INSERT /*+ append_values */

- Currently, INSERT /*+ append_values */ can EHCC compress points, lines and polygons
- Lines and Polygons must be less than 4K (stored inline)
- INSERT /*+ append_values */ does not compress if the column contains a spatial index.
  - The lifting of this restriction is under investigation.

The Spatial & Graph SIG User Group

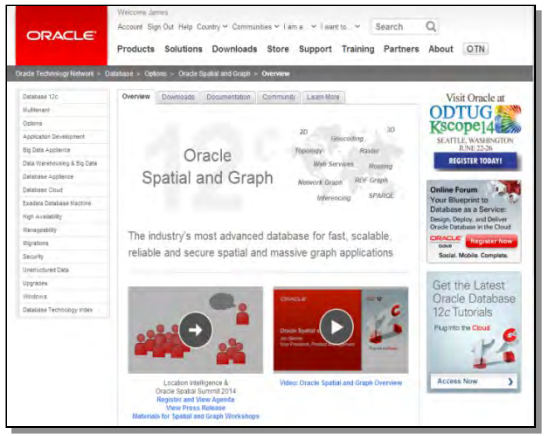
- The SIG promotes interaction and communication that can drive the market for spatial technology and data
- Members connect and exchange knowledge via online communities and at annual conferences and events

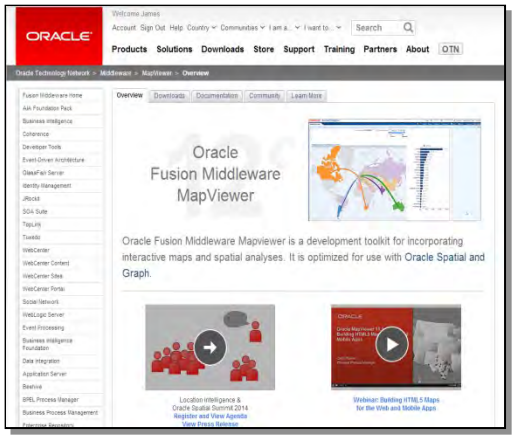
- **Meet us here at the Summit**

Morning Reception	Social Hours
Tuesday and Wednesday 7:45 to 8:30 a.m. Registration Area	Tuesday and Wednesday 6 to 7 p.m. Registration Area

- **Join us online**
  - **LinkedIn (search for "LinkedIn Oracle Spatial")**
  - **Google+ (search for "Google+ Oracle Spatial")**
  - **IOUG SIG (sign up for free membership through www.ioug.org)**
  - **OTN Spatial – Communities (search for "Oracle Spatial and Graph Community")**
- **Contact the Board at oraclespatialsig@gmail.com**

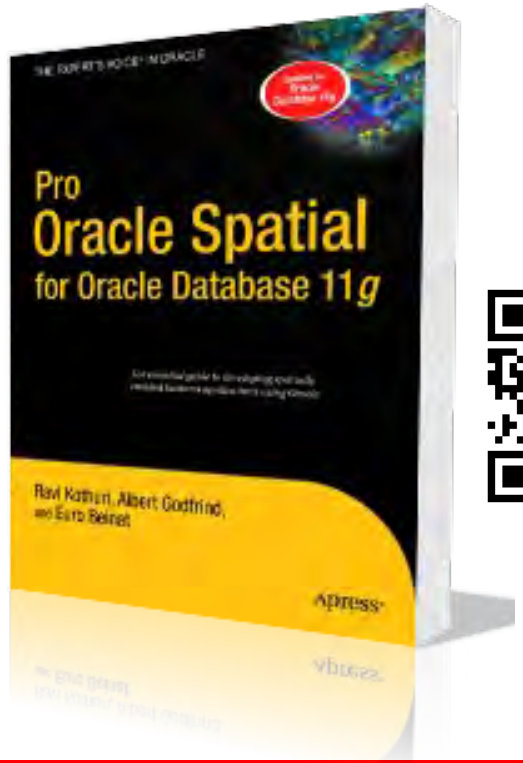
Resources: Oracle Technology Network

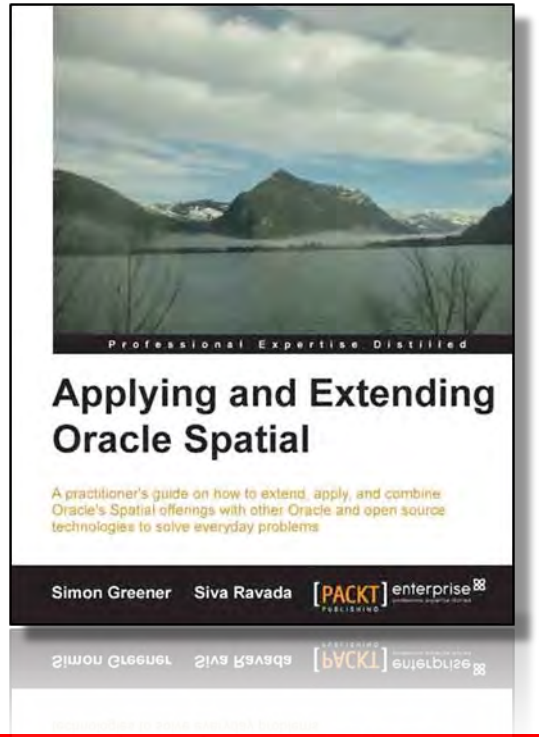




- www.oracle.com/technetwork/database/options/spatialandgraph
- www.oracle.com/technetwork/middleware/mapviewer
- blogs.oracle.com ➔ oraclespatial ➔ oracle_maps_blog

More Resources

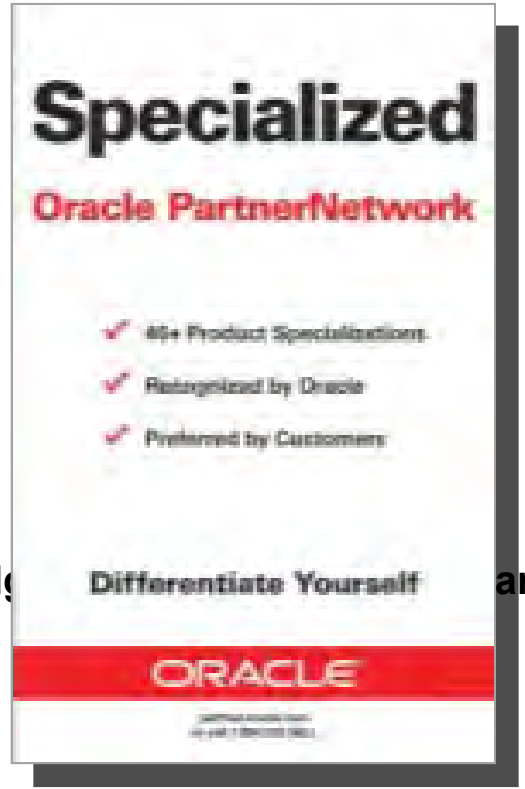




Certification

- **Individual Certification, Partner Specialization**
  - Credentials for individuals with Spatial implementation expertise
  - OPN Specialization – differentiates partner organizations delivering Spatial services
  - Study materials, exam information, program guidelines are available at **www.oracle.com/technetwork/database/options/spatialandg artne**

  - Talk to Oracle team this week



Q & A