OVERVIEW

- Allow users to compete for fastest time climbing a hill or sprinting down a straightaway and rank the leaders.
- Users define the starting line, path, and finish line for each competitive “segment”
- Match a user’s fitness activities to segments and rank the activity

CHALLENGES / OPPORTUNITIES

- Be able to match activities to segments within seconds
- Initial loading for 300+ million activities
- Data quality

SOLUTIONS

- Oracle Exadata Machine (Half Rack)
- Oracle Database 11g Enterprise Edition
  - Spatial Option with Linear Referencing System
  - Partitioning
  - Parallel Pipelined Functions

RESULTS

- Stores and simplifies processing of more than 5 billion miles of user activities in a 40TB database
- Able to process and match activities to segments in seconds and able to match new segments to 5+ TB of activity data in minutes
- Allows for additional real time features on Garmin devices (Edge 1000)
- Enables additional reporting features about the use of Garmin Fitness and Wellness products
May 21, 2014
Walter E. Washington Convention Center
Washington, DC  USA
Steve Mitchell
Garmin Software Engineer

Tim Gerber
Database Administrator
How Garmin Connect Manages and Analyzes 5 Billion Miles of Fitness GPS Data
Program Agenda

• Overview of Garmin’s use of Oracle Spatial
• Segments and Leader Boards
• Challenges and their Solutions
• Benefits
• Q&A
Garmin Segments and Leaderboards

Leveraging Oracle Spatial for Challenges

- Customers wanted a way to challenge one another.
- Segments provide a way to compete along a stretch of road or trail.
- Leaderboards rank user activities on a given challenge segment.
- Oracle Spatial and Graph plus Linear Referencing System are used to match user activities to segments and extract the elapsed time for ranking.
Garmin Segments and Leaderboards

Defining Segments from User Activities
Garmin Segments and Leaderboards
Ranking Users Activities
Meet the Edge 1000

GARMIN VIDEO
Garmin Segments and Leaderboards
Distilling Data for the Edge 1000

The file sent to an Edge 1000 is like one drop from a distillery.

There is a lot of energy spent behind the scenes to produce it.

Oracle Spatial drives the process.
Garmin Segments and Leaderboards
Matching with Oracle Spatial and Graph

Criteria:
- Start
- End
- Direction
- Path
Garmin Segments and Leaderboards

Segment Matching Challenges

GPS Variance:
- Buildings
- Trees
- Canyons
- Satellite drift
- Equipment
Garmin Segments and Leaderboards
Oracle Spatial and Graph to the Rescue

- SDO_ANYINTERACT casts a broad net for possible match candidates.

- SDO_GEOM.SDO_BUFFER defines the distance from the start & end within which a candidate must pass.

- PL/SQL loops through sections of path between the start and end to make sure the paths roughly match.
Garmin Segments and Leaderboards
Oracle Linear Reference System to the Rescue

• Needed a way to measure elapsed time.
• LRS time measure gives time entered & exited.
• Time comparison tests directionality.
• SDO_LRS.REDEFINE_GEOM_SEGMENT function allows switching to distance measure when necessary for other uses.
Garmin Segments and Leaderboards
Building an Oracle Spatial and LRS Solution

• Define data type (Oracle LRS w/time measure)
• Spatially index data
• Load test data.
• Validate data.
Garmin Segments and Leaderboards
Define USER_SDO_GEOM_METADATA

```
INSERT INTO USER_SDO_GEOM_METADATA(
    table_name, column_name, diminfo, srid
) VALUES (
    'SEGMENT', 'POLYLINE',
    SDO_DIM_ARRAY(
        SDO_DIM_ELEMENT('Longitude', -180, 180, 0.05),
        SDO_DIM_ELEMENT('Latitude', -90, 90, 0.05),
        SDO_DIM_ELEMENT('Time', 0, 0, 0.05)
    ), 8307);
```
new JGeometry(
    3302, // Line string; 3D (x, y, m), 3rd is dimension
    8307, // SRID for latitude/longitude (8307)
    {1,2,1}; // offset, line, straight,
    SDO.ordinates(geom)
);
SELECT COUNT(*) FROM GEO_ACTIVITY a WHERE SDO_GEOM.VALIDATE_GEOMETRY(a.polyline, .005) = 'FALSE';
UPDATE GEO_ACTIVITY a
SET a.polyline = SDO_UTIL.REMOVE_DUPLICATE_VERTICES (a.polyline, .005)
Garmin Segments and Leaderboards
Identifying User Data Problems

User paused multisport device at a triathlon in one city and then resumed it at home.
Garmin Segments and Leaderboards

Hardware Data Problems

**Summary**
- Distance: 13,343.84 mi
- Time: 0:00
- Avg Speed: -159.217.5 mph
- Elevation Gain: 425 ft
- Avg Temperature: -3,471,668.6 °F

**Details**

**Timing**
- Time: 0:00
- Moving Time: 596.31.24
- Elapsed Time: 0:00
- Avg Speed: -159.217.5 mph
- Avg Moving Speed: 2.5 mph
- Max Speed: 4.3 mph

**Elevation**
- Elevation Gain: 425 ft
Garmin Segments and Leaderboards

Export/Import Data Problems

20,696 mile polyline made up of 134 activities exported from Training Center desktop application.
Garmin Segments and Leaderboards
Identifying User Data Problems

Indoor treadmill activity with GPS enabled.
Programming Solution

Point Filter Class Diagram

- **ValidLatLonFilter**
  - double : MAX_LAT
  - double : MAX_LON
  - boolean : isValid(TrackContext)

- **TimestampFilter**
  - boolean : isValid(TrackContext)

- **MinimumDistanceFilter**
  - double : maxDistanceMeters
  - boolean : isValid(TrackContext)

- **MinimumSpeedFilter**
  - double : minSpeedMetersPerSecond
  - boolean : isValid(TrackContext)

- **MaximumElapsedTimeFilter**
  - double : maxElapsedTime
  - boolean : isValid(TrackContext)

- **MaximumDistanceFilter**
  - double : maxDistanceMeters
  - boolean : isValid(TrackContext)

- **MaximumSpeedFilter**
  - double : maxSpeedMetersPerSecond
  - boolean : isValid(TrackContext)

- **PointFilter**
  - boolean : isValid(TrackContext)

- **PointContext**
  - Coordinate[] : track
  - List<Coordinate>() : validatedTracks
  - int : pointIndex
  - int : trackIndex
  - Coordinate : previousPoint
  - Coordinate : currentPoint
  - Coordinate : nextPoint
  - double : accumulatedDistance
Garmin Segments and Leaderboards

Segment Summary

- Segments give users a way to compete with each other.
- Oracle Spatial and Graph allows us to match user activities to segments.
- Oracle LRS allows us to measure performance.
- The difficult part was provisioning clean data and overcoming differences in GPS data.
The World of Partitioning
Divide and Conquer the Data

• The Garmin Connect database is 40 TB and growing at a pace of 2 TB per month.
• Of this 40 TB, 5 TB is spatial data used for segments and leader boards.
• Putting this data into a single non-partitioned table is unrealistic.
• To get the best performance, we needed to partition the tables.
The World of Partitioning

Our Partitioning Strategy

- All of our queries will be based around activity type (i.e. running and cycling) and date.
- Spatial indexing allows for partitioning but not sub-partitioning.
- Therefore we chose to use a composite key of activity_type, date with range partitioning.
Running tests with different degrees of parallelism helped drive out the optimal number of parallel processes for the segment matching.
## Power of Parallelism

<table>
<thead>
<tr>
<th>SEGMENT NAME</th>
<th>PARALLEL 2</th>
<th>PARALLEL 4</th>
<th>PARALLEL 8</th>
<th>PARALLEL 10</th>
<th>PARALLEL 12</th>
<th>PARALLEL 16</th>
</tr>
</thead>
<tbody>
<tr>
<td>tour de longchamps</td>
<td>108123</td>
<td>44082</td>
<td>21977</td>
<td>17825</td>
<td>16254</td>
<td>17041</td>
</tr>
<tr>
<td>Garmin Velothon Berlin</td>
<td>963</td>
<td>539</td>
<td>315</td>
<td>273</td>
<td>256</td>
<td>260</td>
</tr>
<tr>
<td>Bugaksan Mountain</td>
<td>6809</td>
<td>5137</td>
<td>1624</td>
<td>1380</td>
<td>1270</td>
<td>1255</td>
</tr>
</tbody>
</table>
Power of Parallelism
Parallel Pipelined Functions

- All of our segment matching is using a PL/SQL package.
- To help performance within the package, we have taken advantage of parallel pipelined functions.

```sql
FUNCTION get_sections_for_segment (segment_pk_v NUMBER,
                                   buffer_dist NUMBER,
                                   interval_in_meters NUMBER,
                                   max_skip_percent NUMBER,
                                   source_table_cursor IN for_segment_cursor_type)
RETURN activity_segment_table_type DETERMINISTIC
PIPELINED PARALLEL_ENABLE
(PARTITION source_table_cursor BY HASH (geo_activity_pk))
```
The Data Load Challenge

The Problem

How do you process, analyze and load 300 million fitness activities in a minimal time frame all while new activities are being created?
The Data Load Challenge
The Chosen Solution

• We were able to utilize several queues to place current activities being created in one queue and the historical activities in another queue.
• Processed both queues continuously until geo activities for all activities were created.
• We were able to process, analyze, load all 300 Million fitness activities in the database in under 20 days.
The Load Process
Benefits
The Results are In!

• Provides “clean” versions of our data.
• Able to provide users results within seconds.
• Significantly less time and resources consumed to get to production over writing custom code.
• Allows real time features for Garmin products.
• Enables additional reporting information about Garmin products.
Q&A

View the Video of the Garmin Edge 1000 in action.