RIDIR Report

Andres Calderon

University of California, Riverside

October 9, 2019

Working on partitioning issue

▶ GeoSpark partitioners retrive a small number of partitions when deal with polygons

▶ I have tried some alternatives:

  ▶ StandardQuadTree: the low-level data structure in GeoSpark. Allows to set parameters as maxItemsPerZone and maxLevel but even setting minimal values the number of partitions is small.

  ▶ Simba, Stark and JTS Quadtree: There are no direct access to the geometry of the Partitioner's cells. We should understand and modify the source code.

  ▶ Workaround: Force a grow up on those partitions of the GeoSpark's quadtree with a large number of edges. I already have a prototype but it still need to test the performance.
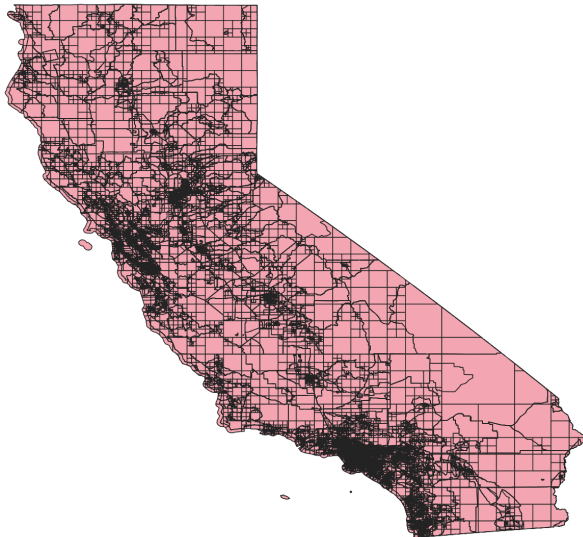
Test - CA_districts 2010 dataset
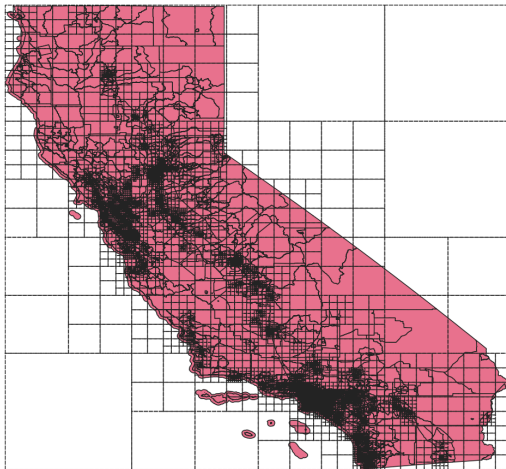
Focus on bug during integration and additional tests

- ▶ It is related to the precision of Polygon's coordinates. Some polygon's sections are not touching by a very small margin.
- ▶ Fixed on Phili datasets by truncating to 6 decimal places.
- ▶ Run experiments with CA_districts datasets 2000 and 2010. Merged DCEL is done but the overlay operations still fails for errors during the merging stage.
- ▶ Checking the correct coordinate system and precision.

Test - CA_districs dataset - 2000 vs 2010

Output -CA_districs dataset - 2000 vs 2010

What is next?

▶ Decide and integrate the partition strategy.
▶ Work on CA_districts datasets 2000 vs 2010 issues.
▶ Check support for multipolygons.