

RIDIR Report

Andres Calderon

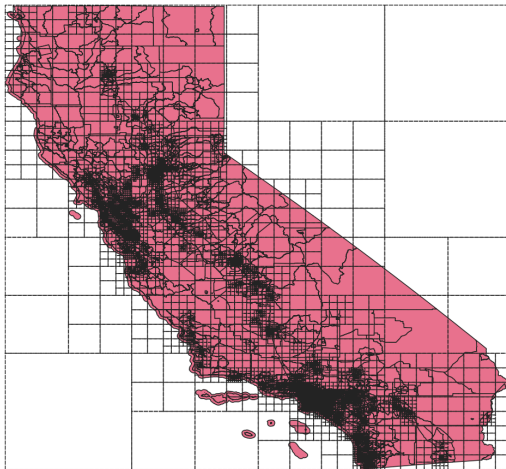
University of California, Riverside

October 9, 2019

Working on partitioning issue

- ▶ GeoSpark partitioners retrieve 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.

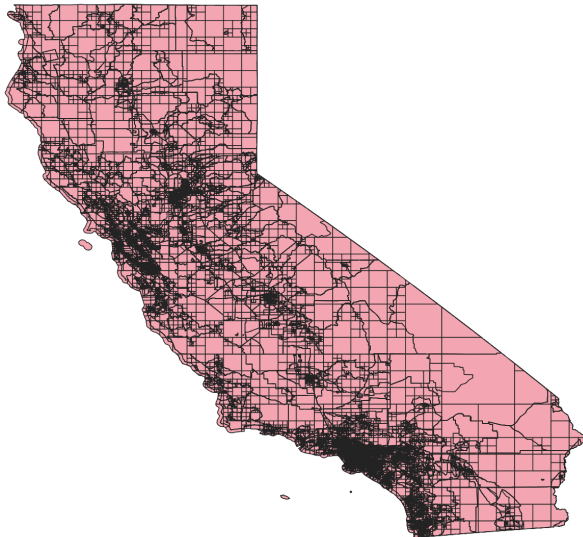
Test - CA_districts 2010 dataset



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_districts dataset - 2000 vs 2010

```
[Root] [scala@h180 ~]$ bin/tools/spark-submit --master spark://h180:7071 --num-executors 30 --executor-cores 4 --class DCELMerger /home/scala@h180/RIDIR/Code/DCEL/target/scala-2.11/dcel_2.11-0.1.jar --input1 ~/dataset/ny/CA/cal12010_polygons6414.tsv --offset1 2 --input2 ~/dataset/ny/CA/cal12000_polygons6414.tsv --offset2 2 --partitions 7500 --debug
```

2019-10-09 10:05:16.183[DCEL][1570640716.18]	Session started	0.00 0[START
2019-10-09 10:05:17.822[spark://h180:7071]		
2019-10-09 10:05:17.823[DCEL][1570640717.82]	Session started	1.04 0[END
2019-10-09 10:05:17.823[DCEL][1570640717.82]	Polygons A read	0.00 0[START
2019-10-09 10:05:29.525[DCEL][1570640729.53]	Polygons A read [dataset is merged]	12.10 8047[END
2019-10-09 10:05:29.525[DCEL][1570640729.53]	Polygons B read	0.00 0[START
2019-10-09 10:05:35.042[DCEL][1570640735.04]	Polygons B read	5.12 7020[END
2019-10-09 10:05:35.042[DCEL][1570640735.04]	Partitioning polygons	0.00 0[START
2019-10-09 10:05:36.501[DCEL][1570640736.50]	Partitioning polygons	1.46 22279[END
2019-10-09 10:05:37.374[DCEL][1570640737.37]	Extracting clipped polygons A	0.00 0[START
2019-10-09 10:06:05.493[DCEL][1570640765.49]	Extracting clipped polygons A	28.12 62770[END
2019-10-09 10:06:05.493[DCEL][1570640765.49]	Extracting clipped polygons B	0.00 0[START
2019-10-09 10:06:25.680[DCEL][1570640785.61]	Extracting clipped polygons B	20.12 60153[END
2019-10-09 10:06:25.680[DCEL][1570640785.61]	Building local DCEL in A	0.00 0[START
2019-10-09 10:07:22.241[DCEL][1570640842.24]	Building local DCEL in A	56.63 22279[END
2019-10-09 10:07:22.241[DCEL][1570640842.24]	Building local DCEL in B	0.00 0[START
2019-10-09 10:07:40.712[DCEL][1570640886.71]	Building local DCEL in B	18.47 22279[END
2019-10-09 10:07:40.712[DCEL][1570640886.71]	Merging DCEL A and B	0.00 0[START
2019-10-09 10:08:00.366[DCEL][1570640886.37]	Merging DCEL A and B	25.65 22279[END
2019-10-09 10:08:31.261[Saved edges.wkt [2514843 records]		
2019-10-09 10:08:57.725[Saved source.wkt [5029086 records]		
2019-10-09 10:09:34.626[Saved vertices.wkt [5029086 records]		
2019-10-09 10:10:23.354[Saved hedges.wkt [5029086 records]		
2019-10-09 10:10:49.276[Saved faces.wkt [1089471 records]		



View Log File

Previous LaTeX Error

Next LaTeX Error

Previous LaTeX Warning

Next LaTeX Warning

Previous LaTeX Building

Next LaTeX Building

What is next?

What is next?

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