

PFLOCK Report

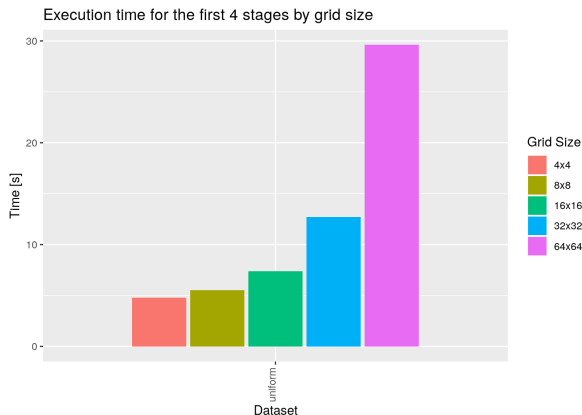
Andres Calderon

University of California, Riverside

July 19, 2019

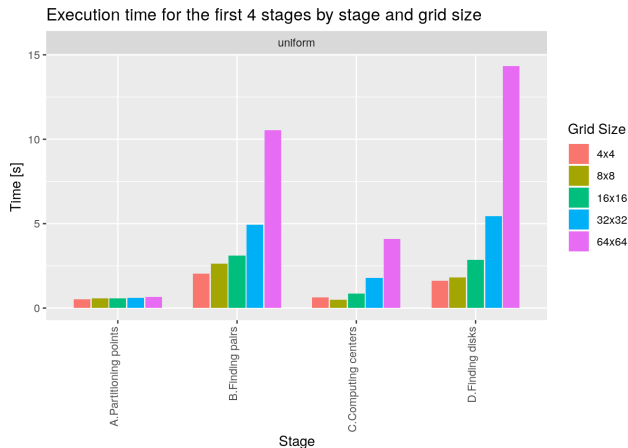
Testing MF Algorithm (Uniform dataset)

- ▶ Test MF algorithm in a uniform distributed dataset:
 - ▶ 100K points distributed in the same LA region.
 - ▶ Using grid partitioning.
 - ▶ $\varepsilon = 25$, $\mu = 5$.



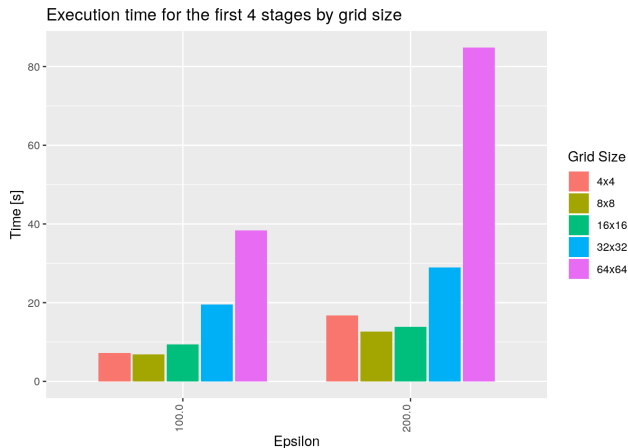
Testing MF Algorithm (Uniform dataset)

- Break down uniform distributed dataset by stage (first 4 stages for now)



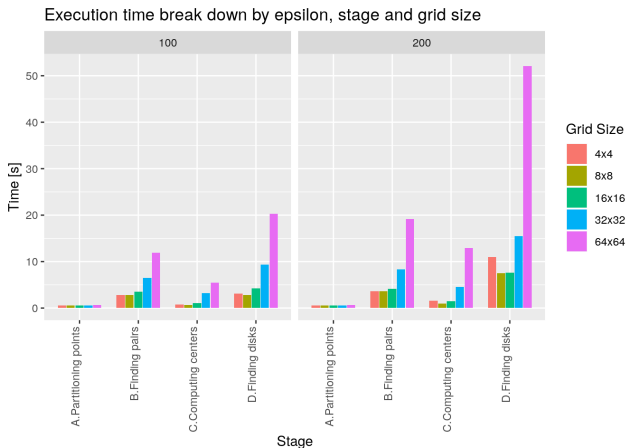
Testing MF Algorithm (Uniform dataset)

- Indeed, in uniform dataset is possible to reach larger epsilon values ($\mu = 5$):



Testing MF Algorithm (Uniform dataset)

- Indeed, in uniform dataset is possible to reach larger epsilon values ($\mu = 5$):



What to do with skew datasets

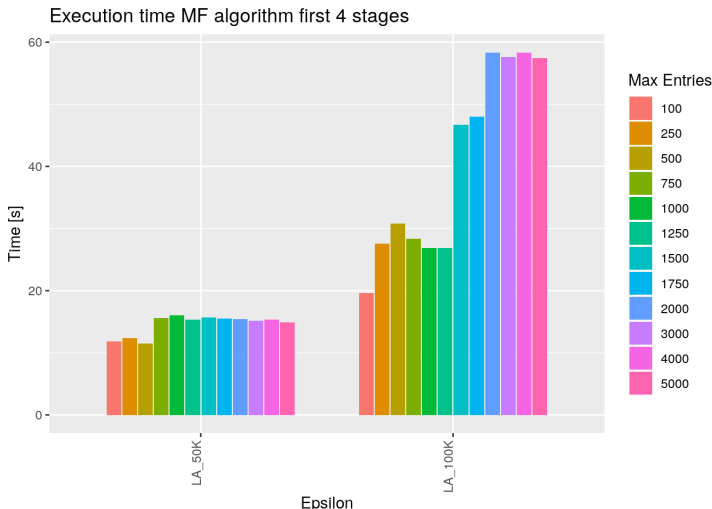
- ▶ That is promising but uniform datasets are quite unrealistic.
- ▶ Also, MF should be able to deal with some kind of skew data.
- ▶ So, let's explore a Quadtree partitioning: a coarse quadtree for the first 4 stages and a finer one for the last 2.

Quadtree in GeoSpark

- ▶ Hacking a bit the GeoSpark Quadtree partitioner to have access to the main parameters:
 - ▶ **MaxLevel**: Maximum level of the tree. It controls how deep will be the tree. Keeping this value high will allow finer partitioning.
 - ▶ **MaxEntries**: Maximum number of entries (capacity) of each cell. If a cell reaches the maximum capacity it will split.
- ▶ So, keeping a relatively high value of **MaxLevel**, varying the value of **MaxEntries** will control the number of partitions in the Quadtree.

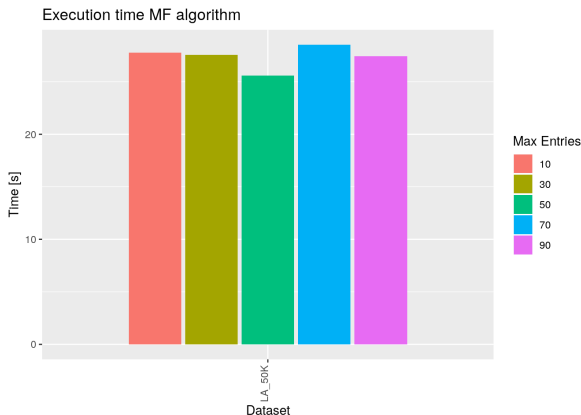
Testing MF Algorithm (Quadtree partitioning)

- ▶ Finding the **MaxEntries** optimal value for the first 4 stages.
- ▶ Testing LA_50K and LA_100K datasets.



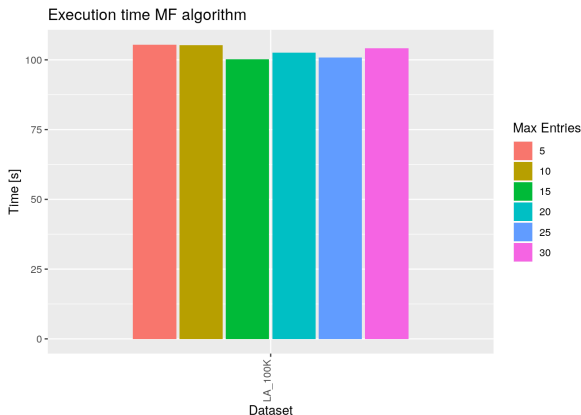
Testing MF Algorithm (Quadtree partitioning)

- ▶ Finding the **MaxEntries** optimal value for the last 2 stages.
- ▶ Testing LA_50K. Using **MaxEntries** = 500 for the first quadtree.
- ▶ $\varepsilon = 10$, $\mu = 5$.



Testing MF Algorithm (Quadtree partitioning)

- ▶ Finding the MaxEntries optimal value for the last 2 stages.
- ▶ Testing LA_100K. Using **MaxEntries** = 100 for the first quadtree.
- ▶ $\varepsilon = 7$, $\mu = 5$.



[Bonus] MF Algorithm (Uniform dataset + Quadtree partitioning)

- ▶ Using first quadtree `MaxEntries` = 1000.
- ▶ $\varepsilon = 100$, $\mu = 5$.

