

PFLOCK Report

Andres Calderon

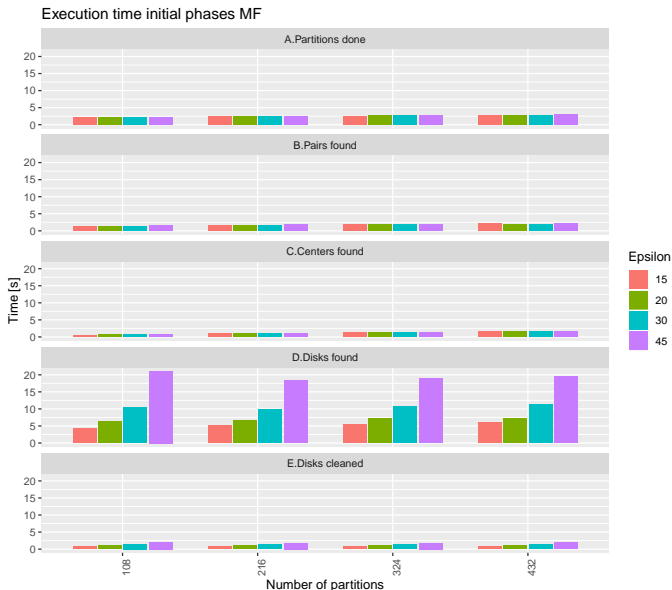
University of California, Riverside

February 28, 2020

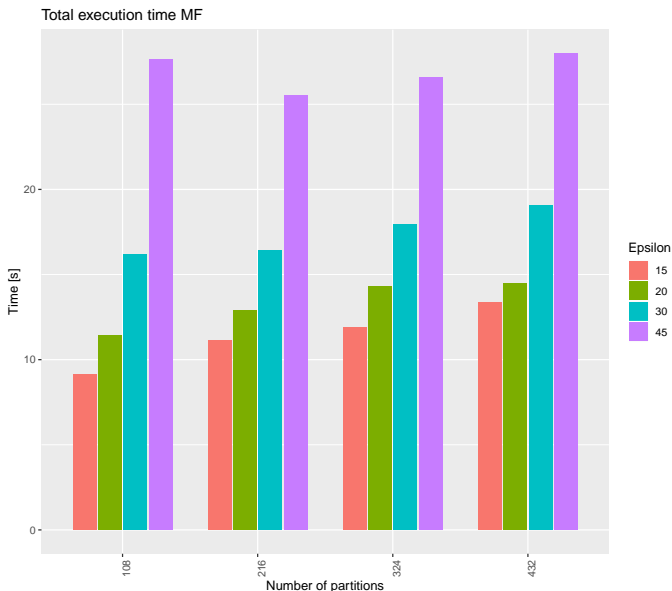
Experiment settings...

- ▶ LA_50K dataset, time instant 320, 50419 points.
- ▶ $\mu = 3$, ε between 15 and 45.
- ▶ 12 executors, 9 cores each (108 cores total).
- ▶ Average time of 3 runs.
- ▶ Varying values for partitions in quadtree (1 to 4 times number of available cores).
- ▶ Varying values for parallelism (1 to 4 times number of available cores).

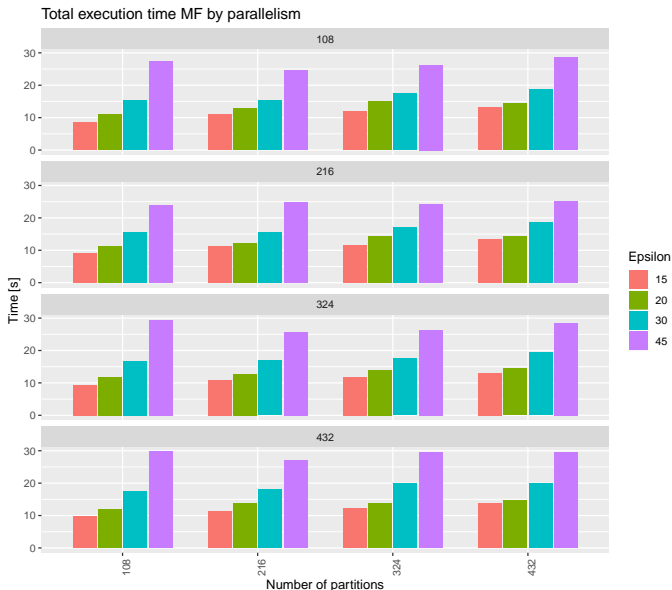
Partitions in quadtre and epsilon...



Partitions in quadtre and epsilon...



Parallelism...



An issue with repartition between Spark stages...

Spark 3.3.3.1.0.0-79 Jobs Stages Storage Environment Executors GeoTesterRDD: epalizer=43.6 grid+... application UI

Stages for All Jobs

Completed Stages: 14
Skipped Stages: 0

Completed Stages (14)

Stage Id	Description	Submitted	Duration	Tasks: Succeeded/Total	Input	Output	Shuffle Read	Shuffle Write
31	count at GeoTesterRDD scala 193	+details 2020/02/26 19:41:54	0.7 s	216/216			85.7 MB	
30	distinct at GeoTesterRDD scala 192	+details 2020/02/26 19:41:53	0.9 s	216/216	3.9 GB			85.7 MB
24	count at GeoTesterRDD scala 178	+details 2020/02/26 19:41:43	10 s	216/216			74.8 MB	
23	mapToPair at JoinQuery.java 429	+details 2020/02/26 19:41:39	4 s	853/853	6.6 MB		17.1 MB	74.8 MB
22	flatMapToPair at SpatialRDD.java 334	+details 2020/02/26 19:41:38	0.7 s	1106/1106	30.7 MB			17.1 MB
18	aggregate at SpatialRDD.java 502	+details 2020/02/26 19:41:37	0.9 s	1106/1106	30.5 MB			
14	count at GeoTesterRDD scala 163	+details 2020/02/26 19:41:36	1 s	1106/1106	112.0 MB			
10	count at GeoTesterRDD scala 132	+details 2020/02/26 19:41:34	1 s	853/853	26.5 MB		3.3 MB	
8	flatMapToPair at SpatialRDD.java 334	+details 2020/02/26 19:41:33	0.4 s	216/216	5.8 MB			3.3 MB
6	count at GeoTesterRDD scala 116	+details 2020/02/26 19:41:32	1 s	853/853		5.0 MB		
5	flatMapToPair at SpatialRDD.java 334	+details 2020/02/26 19:41:31	0.6 s	216/216	4.5 MB			5.0 MB
3	collect at SpatialRDD.java 240	+details 2020/02/26 19:41:31	0.3 s	216/216	4.3 MB			
1	count at GeoTesterRDD scala 86	+details 2020/02/26 19:41:29	2 s	216/216			1100.1 KB	
0	rdt at GeoTesterRDD scala 78	+details 2020/02/26 19:41:25	3 s	0/1	1721.7 KB			1100.1 KB

Task summary...

Summary Metrics for 216 Completed Tasks

Metric	Min	25th percentile	Median	75th percentile	Max
Duration	5 ms	55 ms	5 s	7 s	9 s
Scheduler Delay	3 ms	5 ms	13 ms	20 ms	43 ms
Task Deserialization Time	4 ms	5 ms	15 ms	16 ms	42 ms
GC Time	0 ms	0 ms	0.5 s	0.9 s	2 s
Result Serialization Time	0 ms	0 ms	0 ms	0 ms	1 ms
Getting Result Time	0 ms	0 ms	0 ms	0 ms	0 ms
Peak Execution Memory	0.0 B	0.0 B	0.0 B	0.0 B	0.0 B
Shuffle Read Blocked Time	0 ms	0 ms	0 ms	0 ms	0.3 s
Shuffle Read Size / Records	0.0 B / 0	0.0 B / 0	675.4 KB / 7674	708.6 KB / 8252	738.5 KB / 8614
Shuffle Remote Reads	0.0 B	0.0 B	559.3 KB	659.3 KB	722.8 KB

- Aggregated Metrics by Executor

Executor ID #	Address	Task Time	Total Tasks	Failed Tasks	Killed Tasks	Succeeded Tasks	Shuffle Read Size / Records	Blacklisted
1	ssbul-06:42420 ssbcr	55 s	9	0	0	9	6.2 MB / 74491	false
10	ssbul-06:12:35488 ssbcr	1.4 min	31	0	0	31	6.2 MB / 74500	false
11	ssbul-06:05:41885 ssbcr	48 s	9	0	0	9	6.2 MB / 74367	false
12	ssbul-06:10:37177 ssbcr	1.2 min	24	0	0	24	6.2 MB / 74221	false
2	ssbul-06:04:46771 ssbcr	1.0 min	9	0	0	9	6.2 MB / 74426	false
3	ssbul-06:02:45889 ssbcr	53 s	9	0	0	9	6.3 MB / 74948	false
4	ssbul-06:11:45603 ssbcr	1.2 min	28	0	0	28	6.2 MB / 73879	false
5	ssbul-06:06:36433 ssbcr	80 s	26	0	0	26	6.2 MB / 74030	false
6	ssbul-06:09:43199 ssbcr	55 s	9	0	0	9	6.3 MB / 75136	false
7	ssbul-06:03:35658 ssbcr	1.1 min	26	0	0	26	6.2 MB / 74124	false
8	ssbul-06:01:41117 ssbcr	48 s	9	0	0	9	6.2 MB / 73878	false
9	ssbul-06:10:38772 ssbcr	1.2 min	27	0	0	27	6.2 MB / 74066	false