PFLOCK Report

Andres Calderon

University of California, Riverside

June 26, 2020

Implementation done

- The workflow was divided in two big jobs:
  1. Finding pairs, centers and disks.
  2. Pruning redundant and duplicate disks (it required repartition).
- Run validations using LA_50K and $\epsilon = 5m$.
- Average time = 2.96s

Candidate disks finding

Task duration distribution

Summary Metrics for 928 Completed Tasks

Metric	Min	25th percentile	Median	75th percentile	Max
Duration	4 ms	10 ms	15 ms	26 ms	0.7 s
Scheduler Delay	4 ms	7 ms	14 ms	37 ms	0.1 s
Task Deserialization Time	3 ms	5 ms	6 ms	24 ms	1 s
GC Time	0 ms	0 ms	0 ms	0 ms	0.1 s
Result Serialization Time	0 ms	0 ms	0 ms	0 ms	3 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	68 ms
Shuffle Read Size / Records	0.0 B / 0	417.0 B / 14	929.0 B / 35	1910.0 B / 76	9.3 KB / 426
Shuffle Remote Reads	0.0 B	0.0 B	216.0 B	1157.0 B	9.3 KB

Candidate disks finding

Node load distribution

▾ Aggregated Metrics by Executor

Executor ID ▲	Address	Task Time	Total Tasks	Failed Tasks	Killed Tasks	Succeeded Tasks	Shuffle Read Size / Records
1 stdout stderr	mr-11:37775	14 s	32	0	0	32	38.3 KB / 1554
10 stdout stderr	mr-08:40613	14 s	398	0	0	398	502.5 KB / 20532
11 stdout stderr	mr-04:34243	14 s	36	0	0	36	41.6 KB / 1687
12 stdout stderr	mr-02:40285	14 s	46	0	0	46	63.7 KB / 2632
2 stdout stderr	mr-06:36319	13 s	34	0	0	34	44.4 KB / 1841
3 stdout stderr	mr-09:37074	14 s	30	0	0	30	47.2 KB / 1976
4 stdout stderr	mr-03:37754	14 s	36	0	0	36	43.8 KB / 1801
5 stdout stderr	mr-10:44889	14 s	183	0	0	183	263.3 KB / 10902
6 stdout stderr	mr-01:36598	14 s	41	0	0	41	53.9 KB / 2205
7 stdout stderr	mr-12:38327	13 s	31	0	0	31	52.2 KB / 2210
8 stdout stderr	mr-05:45464	14 s	32	0	0	32	32.9 KB / 1321
9 stdout stderr	mr-07:35560	13 s	29	0	0	29	42.2 KB / 1758

LCM pruning

Task duration distribution

Summary Metrics for 928 Completed Tasks

Metric	Min	25th percentile	Median	75th percentile	Max
Duration	6 ms	25 ms	77 ms	0.3 s	2 s
Scheduler Delay	2 ms	6 ms	8 ms	13 ms	0.6 s
Task Deserialization Time	3 ms	6 ms	6 ms	8 ms	0.4 s
GC Time	0 ms	0 ms	0 ms	0.2 s	0.9 s
Result Serialization Time	0 ms	0 ms	0 ms	0 ms	24 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	20 ms	0.7 s
Shuffle Read Size / Records	0.0 B / 0	938.0 B / 30	2000.0 B / 72	4.0 KB / 156	19.5 KB / 852
Shuffle Remote Reads	0.0 B	358.0 B	1506.0 B	3.5 KB	19.5 KB

LCM pruning

Node load distribution

▾ Aggregated Metrics by Executor

Executor ID ▲	Address	Task Time	Total Tasks	Failed Tasks	Killed Tasks	Succeeded Tasks	Shuffle Read Size / Records
1 stdout stderr	mr-11:37775	18 s	69	0	0	69	189.0 KB / 7384
10 stdout stderr	mr-08:40613	18 s	172	0	0	172	511.8 KB / 20214
11 stdout stderr	mr-04:34243	18 s	69	0	0	69	209.3 KB / 8256
12 stdout stderr	mr-02:40285	19 s	40	0	0	40	114.2 KB / 4562
2 stdout stderr	mr-06:36319	18 s	55	0	0	55	135.5 KB / 5236
3 stdout stderr	mr-09:37074	17 s	74	0	0	74	232.4 KB / 9222
4 stdout stderr	mr-03:37754	18 s	64	0	0	64	177.6 KB / 6966
5 stdout stderr	mr-10:44889	18 s	94	0	0	94	264.2 KB / 10350
6 stdout stderr	mr-01:36598	18 s	53	0	0	53	157.0 KB / 6220
7 stdout stderr	mr-12:38327	19 s	59	0	0	59	141.4 KB / 5482
8 stdout stderr	mr-05:45464	18 s	80	0	0	80	214.9 KB / 8388
9 stdout stderr	mr-07:35560	18 s	99	0	0	99	311.7 KB / 12370

What's next

- Dealing with some minor bugs.
- Completing experiments and plot the results.