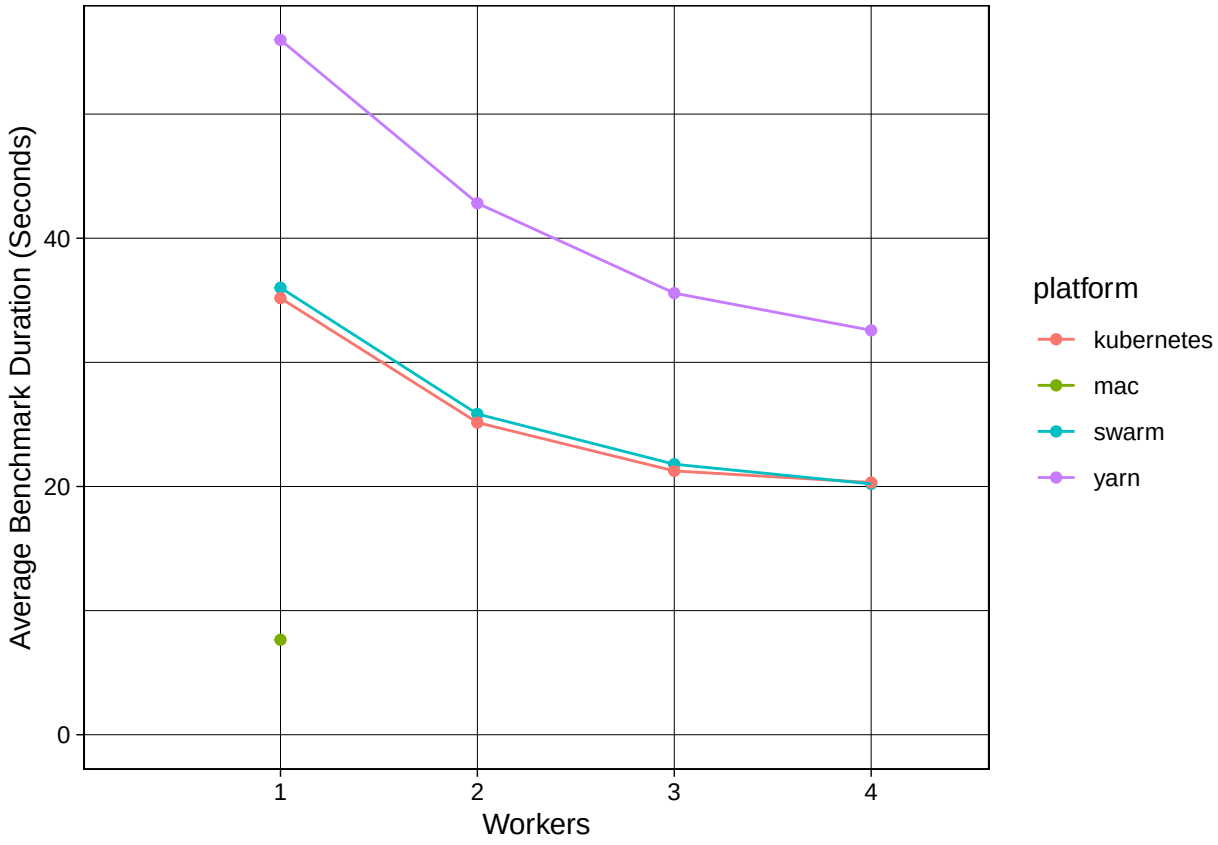
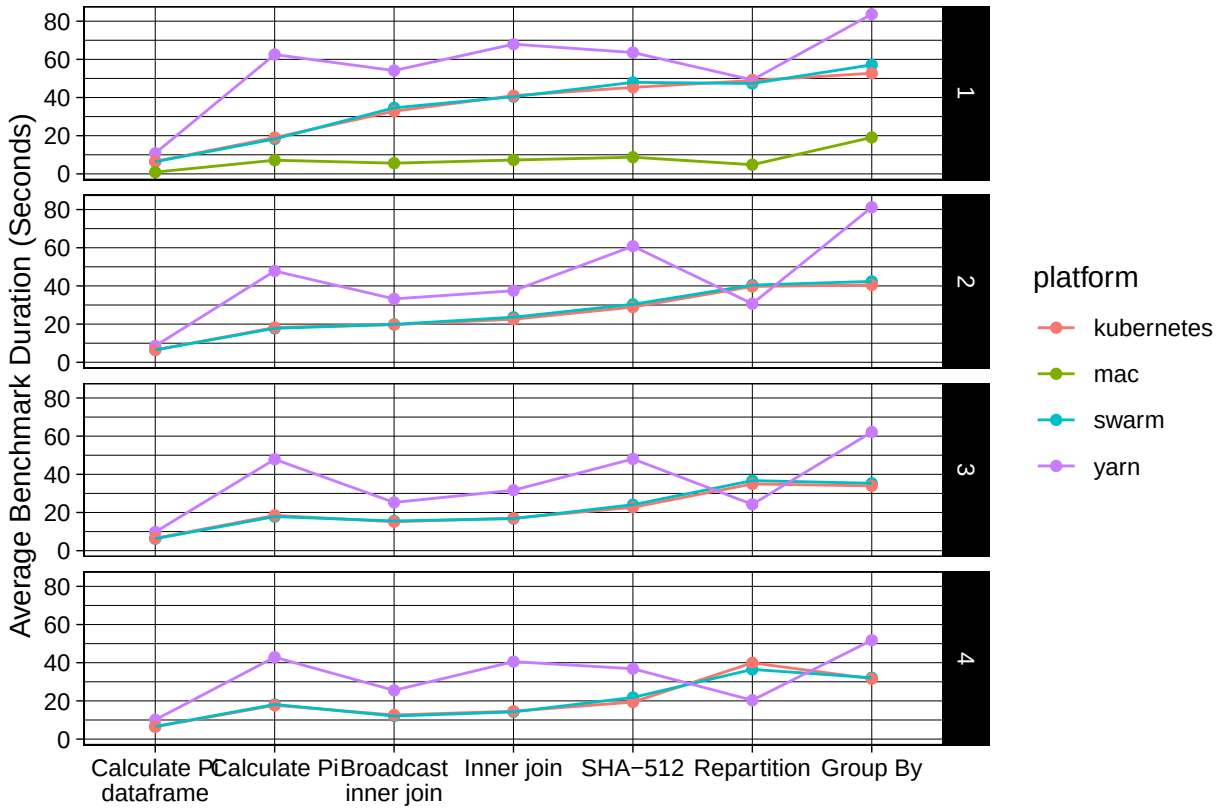
PySpark-Benchmark

```
##
## Attaching package: 'scales'
## The following object is masked from 'package:readr':
##
##     col_factor
```






```
##                  benchmark workers   platform    seconds
##  1:               Group By       4      swarm 32.1324053
##  2:             Repartition      4      swarm 36.4957152
##  3:              Inner join      4      swarm 14.2282874
##  4:   Broadcast inner join       4      swarm 12.0840647
##  5:                 SHA-512      4      swarm 21.8017089
##  6:           Calculate Pi       4      swarm 18.1279794
##  7: Calculate Pi dataframe       4      swarm  6.5997477
##  8:               Group By       3      swarm 35.3015223
##  9:             Repartition      3      swarm 36.7002437
## 10:              Inner join      3      swarm 16.7990808
## 11:   Broadcast inner join       3      swarm 15.5609736
## 12:                 SHA-512      3      swarm 24.0901705
## 13:           Calculate Pi       3      swarm 17.8599437
## 14: Calculate Pi dataframe       3      swarm  6.2393394
## 15:               Group By       2      swarm 42.3624745
## 16:             Repartition      2      swarm 40.4418950
## 17:              Inner join      2      swarm 23.6122225
## 18:   Broadcast inner join       2      swarm 19.8661287
## 19:                 SHA-512      2      swarm 30.3577006
## 20:           Calculate Pi       2      swarm 17.7695325
## 21: Calculate Pi dataframe       2      swarm  6.4649885
## 22:               Group By       1      swarm 57.1237157
## 23:             Repartition      1      swarm 47.3591939
## 24:              Inner join      1      swarm 40.4343307
## 25:   Broadcast inner join       1      swarm 34.5995072
## 26:                 SHA-512      1      swarm 47.9253133
```

```
## 27:           Calculate Pi       1      swarm 18.2861741
## 28: Calculate Pi dataframe       1      swarm  6.3814731
## 29:               Group By       4       yarn 51.8002683
## 30:            Repartition       4       yarn 20.3426104
## 31:             Inner join       4       yarn 40.5680934
## 32:   Broadcast inner join       4       yarn 25.5235029
## 33:               SHA-512       4       yarn 36.8561395
## 34:           Calculate Pi       4       yarn 42.8807191
## 35: Calculate Pi dataframe       4       yarn 10.0913443
## 36:               Group By       3       yarn 62.1358084
## 37:            Repartition       3       yarn 24.2559383
## 38:             Inner join       3       yarn 31.6699510
## 39:   Broadcast inner join       3       yarn 25.2771924
## 40:               SHA-512       3       yarn 48.0532618
## 41:           Calculate Pi       3       yarn 47.8791256
## 42: Calculate Pi dataframe       3       yarn  9.7552129
## 43:               Group By       2       yarn 81.1887893
## 44:            Repartition       2       yarn 30.6774137
## 45:             Inner join       2       yarn 37.4731580
## 46:   Broadcast inner join       2       yarn 33.2136141
## 47:               SHA-512       2       yarn 60.8260476
## 48:           Calculate Pi       2       yarn 47.8318355
## 49: Calculate Pi dataframe       2       yarn  8.5324283
## 50:               Group By       1       yarn 83.6658797
## 51:            Repartition       1       yarn 49.2184908
## 52:             Inner join       1       yarn 67.9298620
## 53:   Broadcast inner join       1       yarn 54.1222193
## 54:               SHA-512       1       yarn 63.5735157
## 55:           Calculate Pi       1       yarn 62.5054601
## 56: Calculate Pi dataframe       1       yarn 10.8231742
## 57:               Group By       4 kubernetes 31.5919700
## 58:            Repartition       4 kubernetes 39.9152137
## 59:             Inner join       4 kubernetes 14.5974094
## 60:   Broadcast inner join       4 kubernetes 12.6294408
## 61:               SHA-512       4 kubernetes 19.3906608
## 62:           Calculate Pi       4 kubernetes 17.7059726
## 63: Calculate Pi dataframe       4 kubernetes  6.4047467
## 64:               Group By       3 kubernetes 33.9513567
## 65:            Repartition       3 kubernetes 34.9555043
## 66:             Inner join       3 kubernetes 17.0088838
## 67:   Broadcast inner join       3 kubernetes 15.1674451
## 68:               SHA-512       3 kubernetes 22.7153051
## 69:           Calculate Pi       3 kubernetes 18.4894181
## 70: Calculate Pi dataframe       3 kubernetes  6.4985226
## 71:               Group By       2 kubernetes 40.4878643
## 72:            Repartition       2 kubernetes 39.7400768
## 73:             Inner join       2 kubernetes 22.5449391
## 74:   Broadcast inner join       2 kubernetes 19.7157764
## 75:               SHA-512       2 kubernetes 28.9085322
## 76:           Calculate Pi       2 kubernetes 18.2094996
## 77: Calculate Pi dataframe       2 kubernetes  6.4353475
## 78:               Group By       1 kubernetes 52.6953728
## 79:            Repartition       1 kubernetes 48.8756509
## 80:             Inner join       1 kubernetes 40.9758309
```

```
## 81:   Broadcast inner join       1 kubernetes 32.8369509
## 82:                SHA-512       1 kubernetes 45.3269522
## 83:          Calculate Pi       1 kubernetes 19.0186871
## 84: Calculate Pi dataframe       1 kubernetes  6.5024090
## 85:               Group By       1        mac 19.0837340
## 86:            Repartition       1        mac  4.8080520
## 87:             Inner join       1        mac  7.2605653
## 88:   Broadcast inner join       1        mac  5.6180644
## 89:                SHA-512       1        mac  8.7335966
## 90:          Calculate Pi       1        mac  7.1601150
## 91: Calculate Pi dataframe       1        mac  0.8973699
##                  benchmark workers   platform    seconds
```