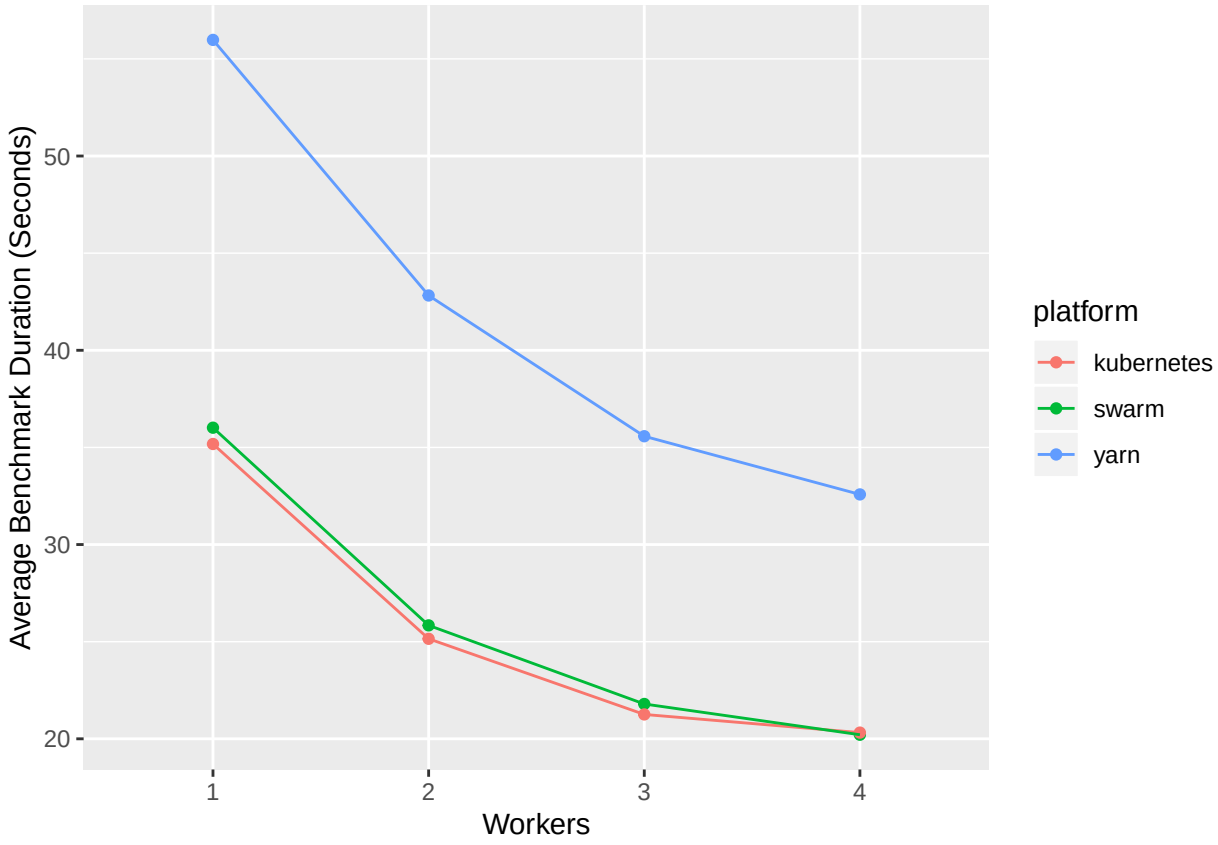
PySpark-Benchmark

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




```
##                  benchmark workers   platform   seconds
##  1:                Group By       4      swarm 32.132405
##  2:             Repartition       4      swarm 36.495715
##  3:              Inner join       4      swarm 14.228287
##  4:   Broadcast inner join       4      swarm 12.084065
##  5:                 SHA-512       4      swarm 21.801709
##  6:           Calculate Pi       4      swarm 18.127979
##  7: Calculate Pi dataframe       4      swarm  6.599748
##  8:                Group By       3      swarm 35.301522
##  9:             Repartition       3      swarm 36.700244
## 10:              Inner join       3      swarm 16.799081
## 11:   Broadcast inner join       3      swarm 15.560974
## 12:                 SHA-512       3      swarm 24.090170
## 13:           Calculate Pi       3      swarm 17.859944
## 14: Calculate Pi dataframe       3      swarm  6.239339
## 15:                Group By       2      swarm 42.362474
## 16:             Repartition       2      swarm 40.441895
## 17:              Inner join       2      swarm 23.612223
## 18:   Broadcast inner join       2      swarm 19.866129
## 19:                 SHA-512       2      swarm 30.357701
## 20:           Calculate Pi       2      swarm 17.769533
## 21: Calculate Pi dataframe       2      swarm  6.464988
## 22:                Group By       1      swarm 57.123716
## 23:             Repartition       1      swarm 47.359194
## 24:              Inner join       1      swarm 40.434331
## 25:   Broadcast inner join       1      swarm 34.599507
## 26:                 SHA-512       1      swarm 47.925313
```

```
## 27:           Calculate Pi       1      swarm 18.286174
## 28: Calculate Pi dataframe       1      swarm  6.381473
## 29:               Group By       4       yarn 51.800268
## 30:            Repartition       4       yarn 20.342610
## 31:             Inner join       4       yarn 40.568093
## 32:   Broadcast inner join       4       yarn 25.523503
## 33:               SHA-512       4       yarn 36.856140
## 34:           Calculate Pi       4       yarn 42.880719
## 35: Calculate Pi dataframe       4       yarn 10.091344
## 36:               Group By       3       yarn 62.135808
## 37:            Repartition       3       yarn 24.255938
## 38:             Inner join       3       yarn 31.669951
## 39:   Broadcast inner join       3       yarn 25.277192
## 40:               SHA-512       3       yarn 48.053262
## 41:           Calculate Pi       3       yarn 47.879126
## 42: Calculate Pi dataframe       3       yarn  9.755213
## 43:               Group By       2       yarn 81.188789
## 44:            Repartition       2       yarn 30.677414
## 45:             Inner join       2       yarn 37.473158
## 46:   Broadcast inner join       2       yarn 33.213614
## 47:               SHA-512       2       yarn 60.826048
## 48:           Calculate Pi       2       yarn 47.831836
## 49: Calculate Pi dataframe       2       yarn  8.532428
## 50:               Group By       1       yarn 83.665880
## 51:            Repartition       1       yarn 49.218491
## 52:             Inner join       1       yarn 67.929862
## 53:   Broadcast inner join       1       yarn 54.122219
## 54:               SHA-512       1       yarn 63.573516
## 55:           Calculate Pi       1       yarn 62.505460
## 56: Calculate Pi dataframe       1       yarn 10.823174
## 57:               Group By       4 kubernetes 31.591970
## 58:            Repartition       4 kubernetes 39.915214
## 59:             Inner join       4 kubernetes 14.597409
## 60:   Broadcast inner join       4 kubernetes 12.629441
## 61:               SHA-512       4 kubernetes 19.390661
## 62:           Calculate Pi       4 kubernetes 17.705973
## 63: Calculate Pi dataframe       4 kubernetes  6.404747
## 64:               Group By       3 kubernetes 33.951357
## 65:            Repartition       3 kubernetes 34.955504
## 66:             Inner join       3 kubernetes 17.008884
## 67:   Broadcast inner join       3 kubernetes 15.167445
## 68:               SHA-512       3 kubernetes 22.715305
## 69:           Calculate Pi       3 kubernetes 18.489418
## 70: Calculate Pi dataframe       3 kubernetes  6.498523
## 71:               Group By       2 kubernetes 40.487864
## 72:            Repartition       2 kubernetes 39.740077
## 73:             Inner join       2 kubernetes 22.544939
## 74:   Broadcast inner join       2 kubernetes 19.715776
## 75:               SHA-512       2 kubernetes 28.908532
## 76:           Calculate Pi       2 kubernetes 18.209500
## 77: Calculate Pi dataframe       2 kubernetes  6.435347
## 78:               Group By       1 kubernetes 52.695373
## 79:            Repartition       1 kubernetes 48.875651
## 80:             Inner join       1 kubernetes 40.975831
```

```
## 81:   Broadcast inner join       1 kubernetes 32.836951
## 82:                SHA-512       1 kubernetes 45.326952
## 83:          Calculate Pi       1 kubernetes 19.018687
## 84: Calculate Pi dataframe       1 kubernetes  6.502409
##                  benchmark workers   platform   seconds
```