

SUMMARY: KEY PROCESSING TIME DIFFERENCES

Different commands	Dask	Spark	Panda
Reading single file	16 milliseconds	2.3 seconds	2.37 seconds
Reading multiple files in one go (here 3 files)	61.8 milliseconds	3.97 seconds	6.27 seconds
Timestamp Disintegration Command	14 milliseconds	53.8 milliseconds	4.62 seconds
Loading a Data-frame	Immediate	150 milliseconds (By default, shows first 20 rows only)	immediate
Any Group-By Usage using a particular column	36.9 milliseconds	24.9 milliseconds	263 milliseconds
2 files merge time (Both files > 100mb)	21.6 milliseconds	48.9 milliseconds	Unable to process i.e., system hangs
2 files merge time (One file > 200mb, Other < 20mb)	8.02 milliseconds	17 milliseconds	Unable to process i.e., system hangs
3 files merger time. (Adding 3 rd file to merger of 2)	11 milliseconds	16 milliseconds	Unable to process i.e., system hangs
Mean Calculation and standard calculation time on merged files	2.22 seconds, 2.72 seconds (With computation time recognized in total calculated time)	2.76 seconds, 1.05 seconds	-