This report is for performance comparison between standard MangoDB implementation and speed up version applying multiprocessing.

For multiprocessing the concurrent.futures.ProcessPoolExecutor is used.

Using build-in “Time” module and “time.perf\_counter” function, time of function’s execution is measured.

Data from different function was collected, see table below.

Same .csv files were used. Same user/status were added, modified and so on, for precise comparison.

There are 12 cpu at the local machine. For better time performance the following parameters were chosen:

For Users DB (2000 total number of users):

* chunk size is 300
* numbers of cpu: 6

For Status\_udates DB (200000 total number of statuses):

* chunk size is 25000
* numbers of cpu: 8

Also there is reduced set f chunk size and number of processors for each DB for deep comparison.



Column “MONGO/csv” – represent version where .csv are uploaded to MONGO db thru the .csv reader. All other options are using Panda module.

There is no significant difference in performance for relatively small “Users” data base. But for large “Statuses” database we can see differences in “loading” only. Differences depends on size of chunk and number of cpus.

Summary: Without any additional words, the performance increased significantly for large “Statuses” database (about 2-2.5 times). But remain the same or close for smaller DB – “Users”.