DS-GA-3001.001/.002 – Advanced Python for Data Science

Assignment 2

Due date: 02/16/2018

An astrophysicist colleague was recently complaining about how long it was taking to run an N-body simulation. "It's really just a simple calculation, and I'm only simulating four planets, but it takes a few minutes to run one simulation." You kindly offer to take a look at code to see if it is possible to speed it up. Your colleague provides you the code "nbody.py", which can downloaded from NYU Classes.

Although your colleague said the code was simple, it is still fairly complex, so you decide to tackle the problem in stages. A first scan of the code reveals a number of potential areas that could be improved. These include:

1. Reducing function call overhead
2. Using alternatives to membership testing of lists
3. Using local rather than global variables
4. Using data aggregation to reduce loop overheads

As you're a cautious programmer, you decide to address each of these in turn. This will ensure that it is possible to check the program is still working correctly after each change, and to assess the performance improvement that the change achieved.

For each of the possible areas of improvement, create a new version of nbody.py calling them nbody_1.py, nbody_2.py, nbody_3.py, nbody_4.py and commit them to the repository. You must also add a file called nbody_opt.py that contains all of the optimizations together.

Finally, generate a .pdf file with the logs produced by the commands:

python3 -m cProfile -o timeStats.profile <code.py>
python3 -m pstats timeStats.profile
timeStats.profile% sort cumulative
timeStats.profile% stats

for code.py in [nbody.py, nbody1.py, nbody2.py, nbody3.py, nbody4.py, nbody_opt.py]

How much speedup do you get (time_original/time_optimized)?