

High Performance Computing Homework 5

Team 7: Dai Nam Nguyen and Sara Restrepo Velasquez

The following table displays the file types that provide the smallest size, fastest write and load CPU time for each of the matrices.

Table 1. Smallest file size, fastest write and load CPU times for each matrix.

Matrix	Parameter	Value	File type
A	Smallest File Size [MB]	34	HDF5
	Fastest Write CPU Time [s]	0.118072	NPY
	Fastest Load CPU Time [s]	0.069288	NPY
B	Smallest File Size [MB]	24	NPY
	Fastest Write CPU Time [s]	0.015547	NPY
	Fastest Load CPU Time [s]	0.010096	NPY
C	Smallest File Size [MB]	34	HDF5
	Fastest Write CPU Time [s]	0.114305	NPY
	Fastest Load CPU Time [s]	0.066683	NPY
D	Smallest File Size [MB]	0.004	CSV and NPY
	Fastest Write CPU Time [s]	0.000348	HDF5
	Fastest Load CPU Time [s]	0.000087	HDF5
E	Smallest File Size [MB]	0.004	CSV and NPY
	Fastest Write CPU Time [s]	0.000623	HDF5
	Fastest Load CPU Time [s]	0.000065	HDF5

It is important to clarify that the HDF5 file size contains all the matrices. It is possible that measuring each matrix separately may change the results seen in the table above.

```
[[login001: HW5]$ du -b *
50000000    A.csv
200000128   A.npy
100000000   B.csv
25000128    B.npy
625000000   C.csv
200000128   C.npy
500         D.csv
328         D.npy
56          E.csv
144         E.npy
4423        generate.py
3281        loadFiles.py
34605706    matrix_db.hdf5
511         results_generate_py.txt
715         results_loadFiles_py.txt
```

Also note that the command **du -h** seems to round the file size numbers, and the smallest file size that it can display seems to be 4 KB. This is likely the reason why in some cases the same file size was obtained for two different formats (e.g. CSV and NPY for matrix D and E). Using the command **du -b** outputs more accurate results in bytes, as shown in

Figure 1. Results from **du -b** command.. Again, if this command were used, the results from Table 1 would change.

Figure 1. Results from du -b command.