What to do when your data are **large** but not `big`

Dillon Niederhut

PyBay – the San Francisco Bay Area Python Conference

20 August 2016

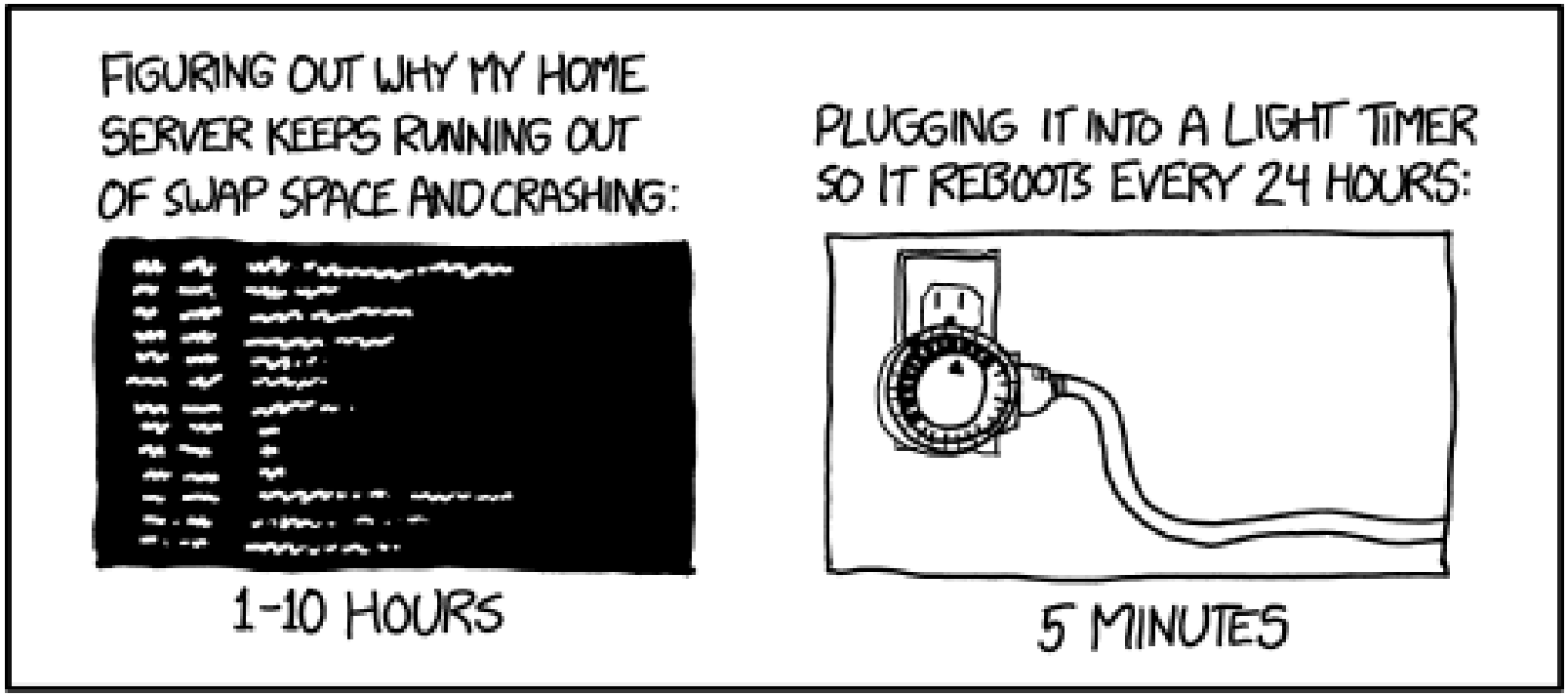
about this talk

- data at github.com/deniederhut/pybay_2016
- python libraries : `celery`, `h5py`, `numpy`, `pandas`, `pymongo`
- other libraries : `mongodb`, `rabbitmq`, `sqlite`

about me

- dlab.berkeley.edu
- @DLabAtBerkeley

size concerns



[1]

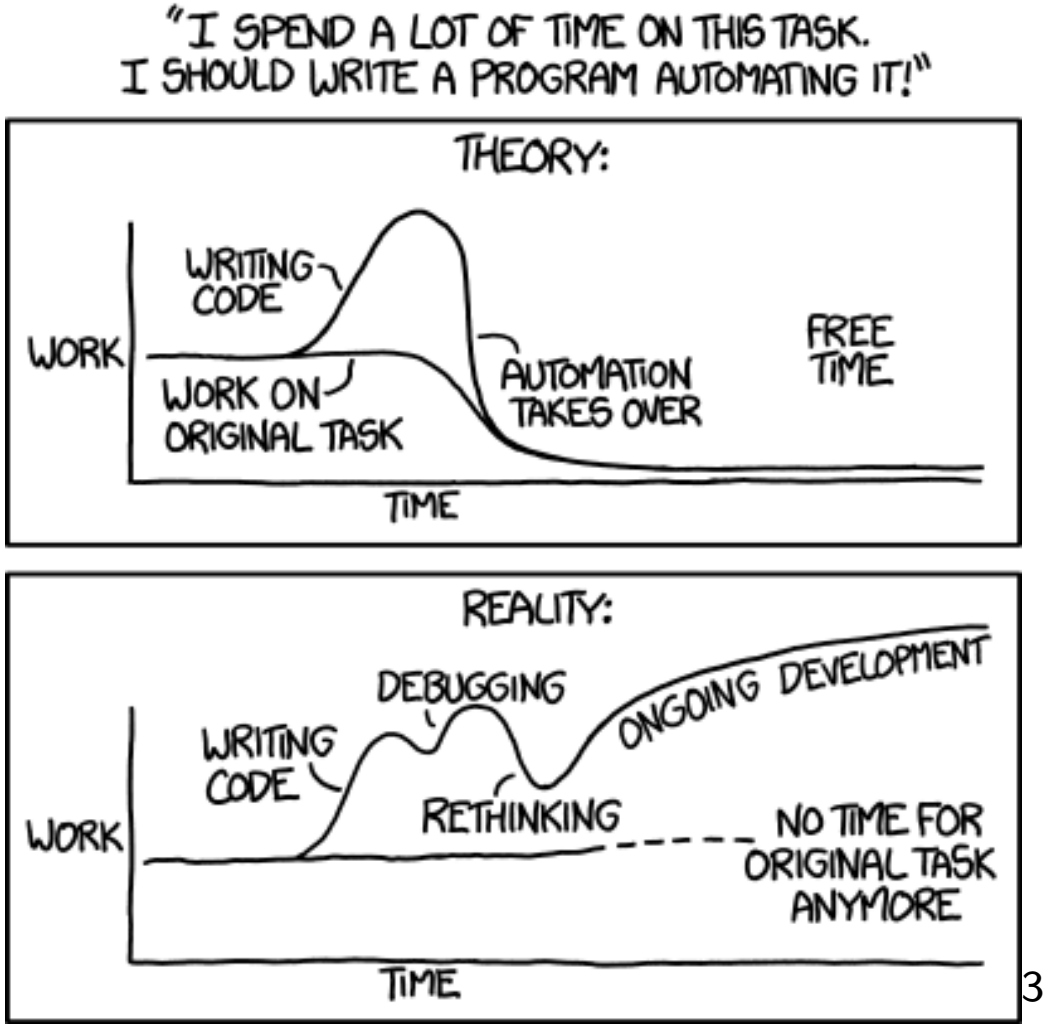
[1] from xkcd

time concerns

[2]

[2] always relevant

code concerns

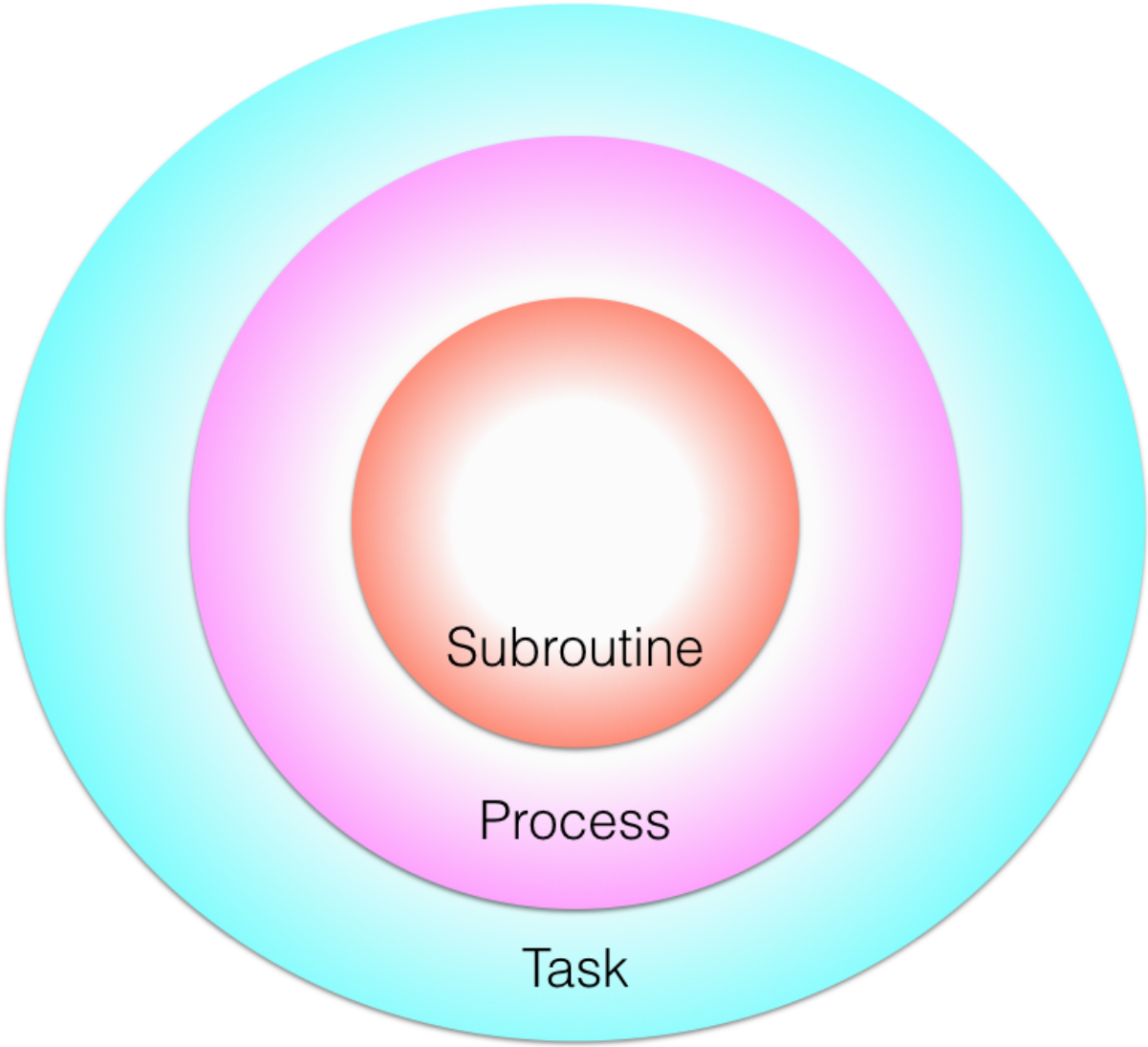


[3]

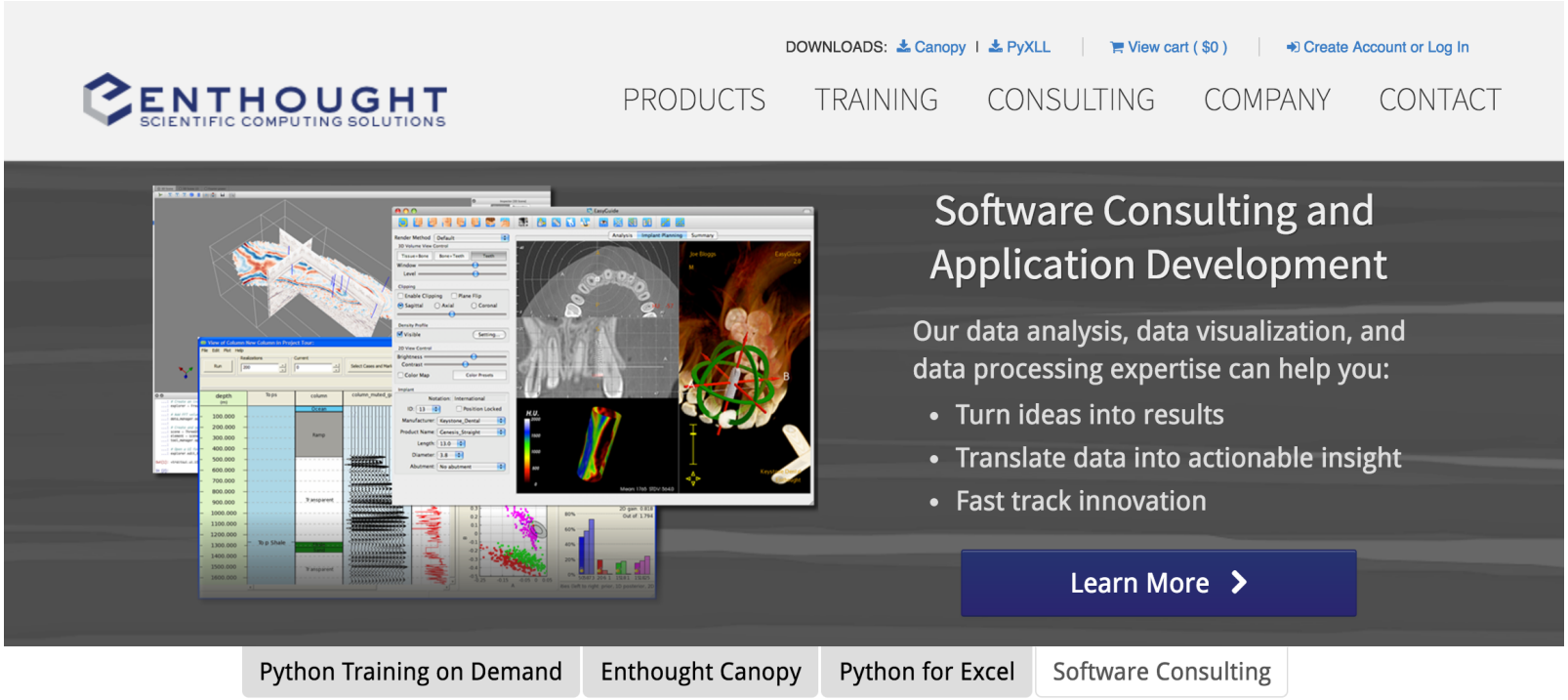
[3] thanks Randall!

sequential processing

parallel processing



contact



- dillon.niederhut.us
- @dillonniederhut