10 Awesome Tricks for Numerical Researchers

Matthew R. Goodman

1 3Scan Biodata Nerd – San Francisco
matt@3scan.com

July 14, 2011
Scipy 2011
1.) numpy/scipy/matplotlib

DUH
2.) iPython

- Interactive namespace
- Intelligent debugging
- New parallel tools and kernel wrangling
- Web-notebook?

If you are not using this tool you are DOING IT WRONG
3.) numexpr/theano

- Kill those pesky intermediate evaluations
- Less malloc and dealloc thrashing etc.
- numexpr ← super easy
- theano ← comes w/ GPU magic (more complicated)
Transmitting data from memory to CPU (and back) faster than a plain `memcpy()`

Stolen from project page. The above graphic should offend you . . .
5.) `mayavi2` / `mlab`
6.) Cython

- Autogen/Compile Python-esque code into C
- Easy to use for wrapping existing code
- EPIC Speedups
7.) Your Local HPC

- HPC admins get graded on machine utilization and assisted publications.
- They WANT to give you hours
- They WANT to help you setup tools
  - Give them a good excuse!
- Check out Teragrid and your local facilities.
8.) Amazon EC2

- **FREE TO RESEARCHERS!**
  (google for “EC2 research” or see links)
- 4000 character proposal and and .edu email get you $7500 in EC2 credit . . . every 6 months.
- Do you need four cores 24 hours a day or 100 for 4 hours a day? GPU instances? 16 cores and 60GB or RAM? Clusters w/ 10Gig-E Interconnects. All possible.
- **Can mail-in hard drives for big data projects**
9.) Starcluster

- Tool that spins up EC2 instances
- Image with most of tools previously mentioned already installed
- Manages permanent drives and data movement
10.) Trilinos/PyTrilinos

- Parallel linear algebra system
- Huge number of linear/non-linear solvers/optimizers
- Harsh learning curve, but EXTREMELY scalable
Numpy/Scipy(+scikits)/Matplotlib

iPython – Use this.

Numexpr/Theano – Easy speed

pytables (+blosc) – Easy fast IO

mayavi2 and mlab – Beautiful Visualizations
6–10

- Cython – Awesome fast. . .
- HPC Admins – Talk to these people!
- Amazon EC2 – Free Cycles for Researchers/Educators
- Starcluster – Awesome simple configuration of above
- Trilinos/PyTrilinos – DOF > 100M? Use this.
List-o-links

- numpy/scipy – http://numpy.scipy.org/
- ipython – http://ipython.scipy.org/moin/
- pytables – http://www.pytables.org/moin
- Teragrid – https://www.teragrid.org/
- EC2 Researchers Link –
  http://aws.amazon.com/education/
- Starcluster –
  http://aws.amazon.com/education/
- Trilinos/Pytrilinos –
  http://trilinos.sandia.gov/
Thanks!

Enjoy your afternoon!