

Research IT

Presentation to Joint BCB Group

2015-12-02

Agenda

- Introduction
- Pricing model
- Resources
- Consulting
- Some examples
- Websites
- Other Services
- Questions

What / Who is ResearchIT at Iowa State?

- History:
 - Group formed as a collaboration between LAS & CALS in summer of 2014
 - Initially led by Michael Hofmockel
 - Originally focused on Biology areas, has since expanded
 - Currently working with Physics, Chem, CS, etc. and have the support of ITS
- What do we do?
 - Collaborate with researchers to solve complex technology problems
 - Allow researchers to focus effort on their core competencies
 - Provide access to computational hardware, software, storage, consulting
 - Focus our time on tasks that scale up
- Who ? Collaboration with LAS, CALS, ITS
 - www.biology-it.iastate.edu
 - Lots of involvement from other IT people in college, departments and ITS
 - researchit@iastate.edu

Pricing Model

- We do not charge for our time
 - Our salaries are paid through the colleges, departments, etc. that we support
- We do not charge per CPU cycle
- We want to collaborate with you on grant proposals to include funding for access to shared infrastructure
 - Work closely with LASERS team
- We will do our best to find spare cycles to run your research if you don't have funding
- Offloading liability

Data Storage

- Have invested in high performance, highly available, network attached storage
- Name: my.files.iastate.edu (aka new Research Files)
 - Device: EMC Isilon
 - Located in Durham / ASB
 - Multiple nodes in each location
 - File snapshots, hourly for 7 days, daily for 4 months, weekly for a year
 - Fast network access (10Gbps)
- Providing each PI 2TB of free storage
 - College is subsidizing each additional TB, PI cost is \$400 / TB / Year
 - Will take old storage on trade through end of year: <http://it.las.iastate.edu/storage-trade-program>
- Currently working on other long term lower performance storage options (e.g. Box, Glacier, maybe tape)

Computational Resources

- [Full List of hardware](#)
- BigRam:
 - 1.5TB of RAM
 - for when your whole dataset needs to fit in RAM (assembly)
- BioCrunch:
 - 80 cores
 - for multithreaded, non MPI software (mapping, alignment, etc.)
- Speedy / Speedy2:
 - 24/32 Fewer faster cores (3.2 / 3.4GHz)
 - for single threaded software
- Condo HPC Cluster:
 - 168 Nodes (16 cores/128GB RAM each) + 1 Large Mem (1TB) + 1 Huge Mem (2TB)
 - for MPI capable software
- LAS-Win-01:
 - When software will only run on Windows
- Adding more:
 - Rebuilding Baker Ctr. machines (two Dell servers + two small clusters)

Software

- RISA: 119 software titles
 - <https://git.its.iastate.edu/projects/RIT/repos/modules>
 - Can run this on any RHEL7 (Cent7) machine
 - We have converted from RPMs to this model
 - Currently bigram and speedy2 are using this new software repository, rest will be converted soon
- Condo: 81 software titles
 - Login to `condo.its.iastate.edu`
 - `$ module avail`

Consulting

- Do not charge for time, but also not your dedicated programmer
- Can advise on computational performance of software
- Can help you identify optimal software for your pipeline
- Help identify ways to optimize pipeline
 - Hardware
 - Different software
 - Parallelization wrappers (e.g. gnu parallels)
- Our goal is to work on tasks that can benefit others through packaged software, documentation, collaboration, etc. and that help you overcome technical challenges that moves your science forward
 - Faster iterations

Examples

- Hufford lab cut genome assembly time by running on BigRAM
- Helped Nikolau lab migrate from MEGA to RAxML
- Sivasankar lab cut MD (gromacs) job run time, and costs by moving off Amazon EC2 to Condo
- Helped Windus lab compile GAMESS on Condo
- Helped Dobbs lab create web interface for rpimotif tool
- Helped Wolter lab parallelize matlab code

Websites

- Lab Websites
- Custom web based research tools
- Secure virtual web environments
- Code review
- Our platform: http://www.biology-it.iastate.edu/luggage_doc/
- Need to publish a new service, or tool on the web ?
 - <http://it.las.iastate.edu/public-ip-or-dns-request>

Other Services

- FTP server to host genome files for gbrowse:
 - gbrowse.las.iastate.edu
 - tested with : <http://genome.ucsc.edu/cgi-bin/hgCustom>
- Version Control:
 - Can help you start tracking code in git, several options for hosting including on-site: <http://it.las.iastate.edu/version-control>
- Backups:
 - We can set you up with backup software to protect data on your local disk: <http://it.las.iastate.edu/computer-backup>

Questions ?

- Contact Info:
 - email: researchit@iastate.edu
 - web: www.biology-it.iastate.edu
 - web form to engage for help: <http://www.biology-it.iastate.edu/pipeline-or-computational-run-questionnaire>
 - IRC: #bitcom