HARVARD
Faculty of Arts and Sciences
RESEARCH COMPUTING

Research Computing and Informatics Services:

Gaining a Competitive Edge

Bob Freeman, PhD
Research Computing Facilitator
XSEDE Campus Champion

robert_freeman@harvard.edu
@DevBioInfoGuy

Goals

- Help you to understand the complexities of our services
- Help frame our role in the services landscape
- Demonstrate leadership in advanced technologies
- Foster collaborations across institutional boundaries

- “Work smarter, better, faster”
- ... and to Think Differently
- Enable groups to be successful with your research!

People of Research Computing

HARVARD
Faculty of Arts and Sciences
RESEARCH COMPUTING

- 14 people (leadership, cloud, sysadmin, client support)
- Strong overlap with Informatics group (10 in leadership, web, science, sci/code)
- Ph.D.s in biology and physics, w/ roles as sysadmins and developers
- Supporting ALL sciences & major HU computing projects (BICEP2, ATLAS, etc.)

Informatics Group

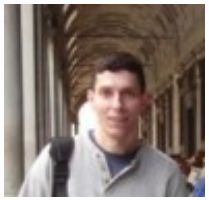
HARVARD
Faculty of Arts and Sciences
RESEARCH COMPUTING

Director: Michele Clamp
mclamp@g.harvard.edu

Senior Web Developers

Eric Mattison
mattison@g.harvard.edu

Reuven Koblick
reuven_koblick@harvard.edu



Bioinformatics

George Marnellos
gmarnellos@fas.harvard.edu

Chris Williams
williams03@g.harvard.edu

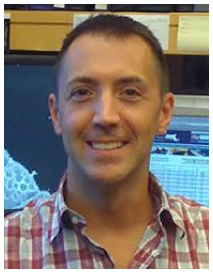
Post-Docs and Junior Fellows

Tim Sackton (with Edwards lab)
mattison@g.harvard.edu

Jacob Sniff (with Conte Center)
reuven_koblick@harvard.edu

RC and Honorary Informatics

Aaron Kitzmiller
akitzmiller@g.harvard.edu



Bob Freeman
robertfreeman@g.harvard.edu

RC Services

HARVARD
Faculty of Arts and Sciences
RESEARCH COMPUTING

Odyssey (cluster)	55K+ load-balanced cores (CPUs) and increasing *high-throughput / high-performance computing*
Storage	14 PB private, public, and scratch storage; and increasing *Lab and personal data; scratch/work temporary space*
Visualization	65+ GPGPUs and increasing *high-end visualization; real-time interactive rendering*
Virtual Machines	~600+ KVM images *Web portals, license or server-based resources; restricted datasets*
Apps/Libraries	>3000, in chemistry, biology, statistics, social sciences, & others *Predominantly cluster-based & open source; some desktop/licensed*
Other	Application licensing Code Optimization Instrumentation Core facilities support
HUIT	Networking, student computing, admin apps, & desktop support

Core Bioinformatics Services

HARVARD
Faculty of Arts and Sciences
RESEARCH COMPUTING

- Primary sequence analysis i.e. basecalling/multiplexing
- Sequence data QC/trimming
- Alignment to reference genome
- Bioinformatics software installation/support
- Consultation and recommendations for amount and type of sequencing
- Consultation and recommendations for sequence analysis

Non-Core Activities

These include:

- Transcriptome assembly and differential expression analysis
- Whole genome assembly and genome comparison
- Annotation of assembled transcriptomes
- Database construction
- SNP calling and filtering
- Custom script/algorithm writing
- Phylogenetic analysis

We are happy to meet and help people get up and running with these things.

High Performance Computing

HARVARD
Faculty of Arts and Sciences
RESEARCH COMPUTING

Old Way: Individual groups maintain their own resources

- Takes time away from the real research
- Much reinventing of the wheel
- Does not scale well

New Way: RC and HUIT provide resources

- Physical infrastructure: power, cooling, and space
- System administration, algorithm/programming support, & technical advice
- Consolidation of knowledge and expertise
- Economies of scale

Who can benefit from HPC?

- If your analysis takes longer than 10 min on your desktop/laptop
- If you're doing tens/hundreds of tasks
- Analyses using a GUI that could be done at the command line

High Performance Computing

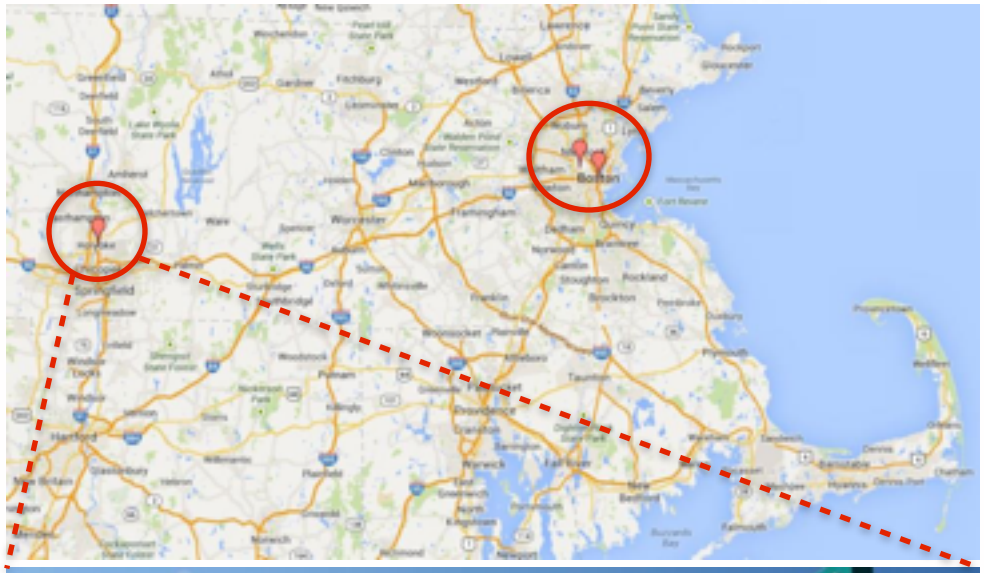
HARVARD
Faculty of Arts and Sciences
RESEARCH COMPUTING

Value to you??

- Groups can purchase dedicated hardware or use the general resources
- Leads to better and more efficient use of resources:
  - Easier entry into the world of HPC
  - Groups who own resources get priority on the resources they own
  - Others can use of the spare cycles when the resources are not in use

When dealing with clusters, fully utilized machines are a good thing!

What is Odyssey?



Compute nodes/disk are located in 3 data centers:

- 60 Oxford St
  - *Personal home folders*
  - *Legacy equipment*
- 1 Summer Street
  - *Lab disk shares*
  - *Compute nodes < 2012 (20K+ cores)*
- Holyoke, MA
  - *Compute nodes > 2012 (33K+ cores)*
  - *'regal', scratch filesystem*

Topology may effect the efficiency of work!

Beyond RC: XSEDE

HARVARD
Faculty of Arts and Sciences
RESEARCH COMPUTING

Extreme Science & Engineering Discovery Environment!

- Umbrella organization for collection of integrated advanced digital resources and services
- 5-year, $120m project funded by NSF & successor to TeraGrid
- Supports 16 supercomputers and high-end visualization and data analysis resources across the country
- Campus champion exists to help you apply for and transition to these off-campus resources
- XSEDE and on-campus resource use are not mutually exclusive!
- https://www.xsede.org/

https://rc.fas.harvard.edu/partnerships/xsede/

Consulting & Outreach

- Consulting

  Given appropriate expertise, some significant engagements to help with code and analysis optimizations are possible

- Adv CyberInfrastructure – Research & Education Facilitation

  Special 2-yr NSF grant for cluster computing education/facilitation

  Partner with you to enable your research: use compute resources, work more efficiently, and be more competitive

  Collaborations: Clemson U., U. Southern California, U. Utah, U. Wisconsin (Madison), U. Hawaii

  Bob Freeman, PhD & Aaron Kitzmiller, PhD

Workflow Use Cases

HARVARD
Faculty of Arts and Sciences
RESEARCH COMPUTING

- Transition to HPC from GUI
  - video processing workflow, all GUI driven
  - uses ImageJ and MATLAB, with 6 streams and ~10 steps
  - in progress, but down to 30 minutes from 12 hrs
- Video processing workflow
  - complex cluster pipeline for tracking animal movements
  - custom code developed by post-doc who left; no version control
  - in progress, adapting to current cluster, version control, & multicore processing
- Recoding of geophysical simulations
  - visco-elastic models for simulating earthquake movements
  - transitioned from MATLAB to fortran, & parallelized code via MPI
  - reduced compute time to 10 hrs from weeks

ACI-REF
Advanced CyberInfrastructure
Research & Education Facilitators

NSF

Accessing RC Resources

HARVARD
Faculty of Arts and Sciences
RESEARCH COMPUTING

- RC Website & Documentation (only authoritative source!)
  https://rc.fas.harvard.edu/

- OdyBot: for quick fix problems (free prizes for quiz!)
  https://odybot.org/

- Getting help
  - Submit a ticket via web form: https://portal.rc.fas.harvard.edu/
  - or email: rchelp@fas.harvard.edu

Best way to help us? Give us...

Description of problem:
Steps to Reproduce (1., 2., 3...)
Actual results
Expected results
Additional info (login/batch? partition? JobIDs?)

Training

- Training opportunities: New & Evolving!
  - Office Hours: every Wed 12 - 3 pm @ RC conference room
  - Intro to Odyssey & RC Services
  - Using HPC Resources Efficiently series
  - Highlighted topics in the first ½ hour of office hours
  - Guest lectures in courses
  - On demand

  - Parallel MATLAB (11/21)
  - Software Carpentry (http://software-carpentry.org/) (11/8-9)

Research Computing

HARVARD
Faculty of Arts and Sciences
RESEARCH COMPUTING

Please talk to your peers, and
We wish you success in your research!

http://rc.fas.harvard.edu
https://portal.rc.fas.harvard.edu
rchelp@fas.harvard.edu
@fasrc

Harvard Informatics
@harvardifx

robert_freeman@harvard.edu
@DevBioInfoGuy