

Shared Research Computing Policy Advisory Committee (SRCPAC)

Spring 2022 Meeting

Meeting Called By: **Chris Marianetti**, *Chair of SRCPAC*



SRCPAC Agenda

- **Welcome & Introductions**
 - Chris Marianetti, Chair of SRCPAC
- **High-Performance Computing Update**
 - Kyle Mandli, Chair of the HPC Operating Committee
 - Cesar Arias, Manager of High Performance Computing, CUIT
- **Research Computing Services Update**
 - Axinia Radeva, Manager of CUIT Research Services
- **Foundations for Research Computing Update**
 - Marc Spiegelman, Chair of the FoRC Advisory Committee
 - Jonathan Cain, Associate University Librarian for Research and Learning
- **Other Business & Closing Remarks**
 - Chris Marianetti, Chair of SRCPAC



High Performance Computing Update

Kyle Mandli, Chair of the HPC Operating Committee

Cesar Arias, Manager of High Performance Computing, CUIT



High Performance Computing Updates

- NIH G20 Final Site Visit (Virtual) - May 2022
- It's been 10 years since we received the 10M grant that made the Shared Research Computing Facility (SRCF) possible.

Columbia University's Shared Research Computing Facility is supported by NIH Research Facility Improvement Grant 1G20RR030893-01, and associated funds from the New York State Empire State Development, Division of Science Technology and Innovation (NYSTAR) Contract C090171, both awarded April 15, 2010.



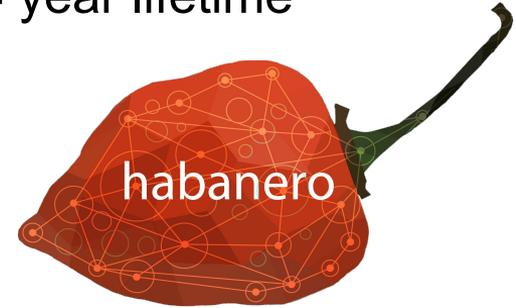
HPC Governance

- HPC operations are governed by the faculty-led **HPC Operating Committee**, chaired by **Kyle Mandli**.
- The operating committee reports to SRCPAC and reviews business and usage rules in open, semiannual meetings
- The last meeting was held on March 26, 2020 and the next one will be in Fall 2022.
- **All HPC Users (Ginsburg, Terremoto, Habanero) are invited to participate.**



Habanero Retirement

- Launched in 2016, expanded in 2017 with a 4 year lifetime
- Phase 1 will be retired at the end May 2022
- Phase 1 consisted of 222 nodes
 - 176 Standard
 - 32 High Memory
 - 14 GPU
- Habanero hardware is no longer under warranty and nodes will be disposed of as they die. We already have a number of nodes out of service.

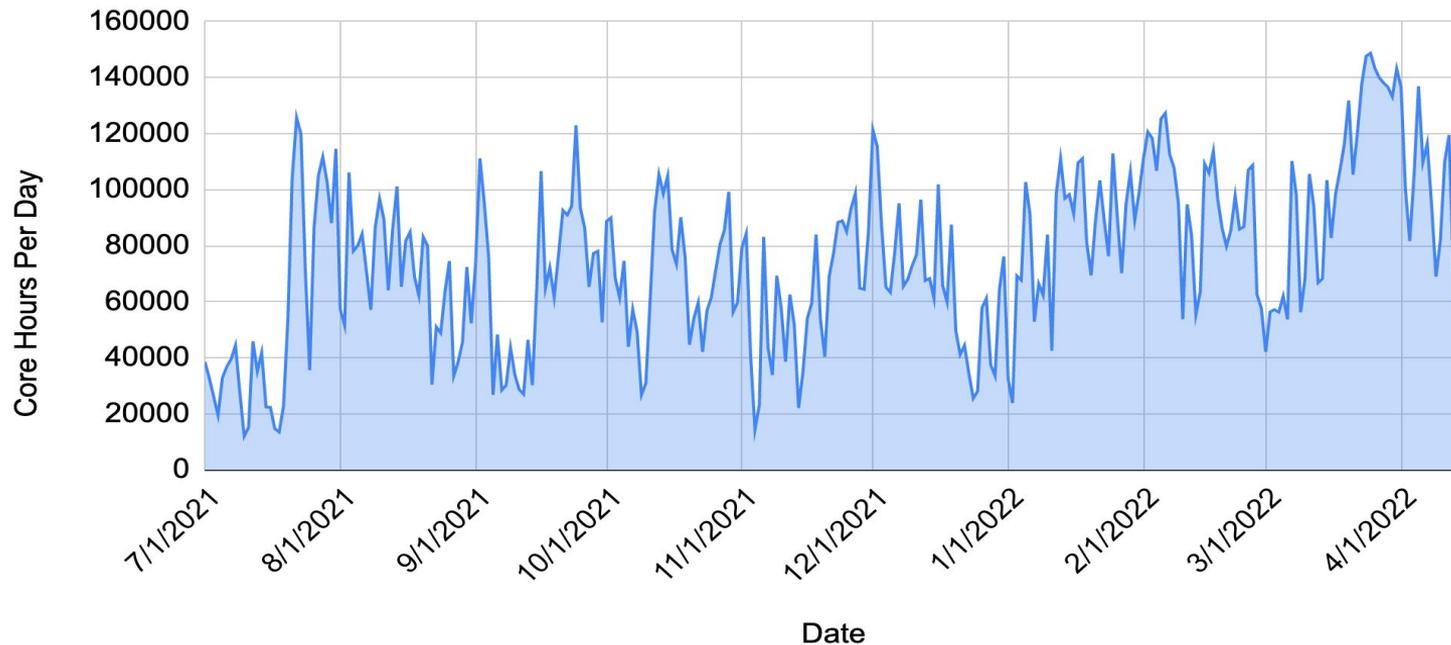


Habanero Retirement

- A portion (up to $\frac{1}{3}$) of Habanero Phase 1 will be available for free and edu users.
- At some point in the next 9 months, we will have to relocate the free nodes to make room for new equipment.
- Free users:
 - ~~Low priority, 6 hours, 4 nodes max~~
 - Increasing to 12 hour jobs, 8 nodes max
 - Users who do not have resources on other clusters.
 - Support limited to on-line documentation only
 - Major software packages will supported
- Phase 2 will retire in April of 2023.



Habanero - Cluster Usage in Core Hours

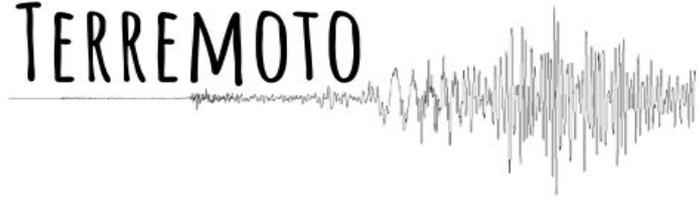


Max Core Hours Per Day =
148,698

Total core hours in the past 12 months =
28 million



TERREMOTO

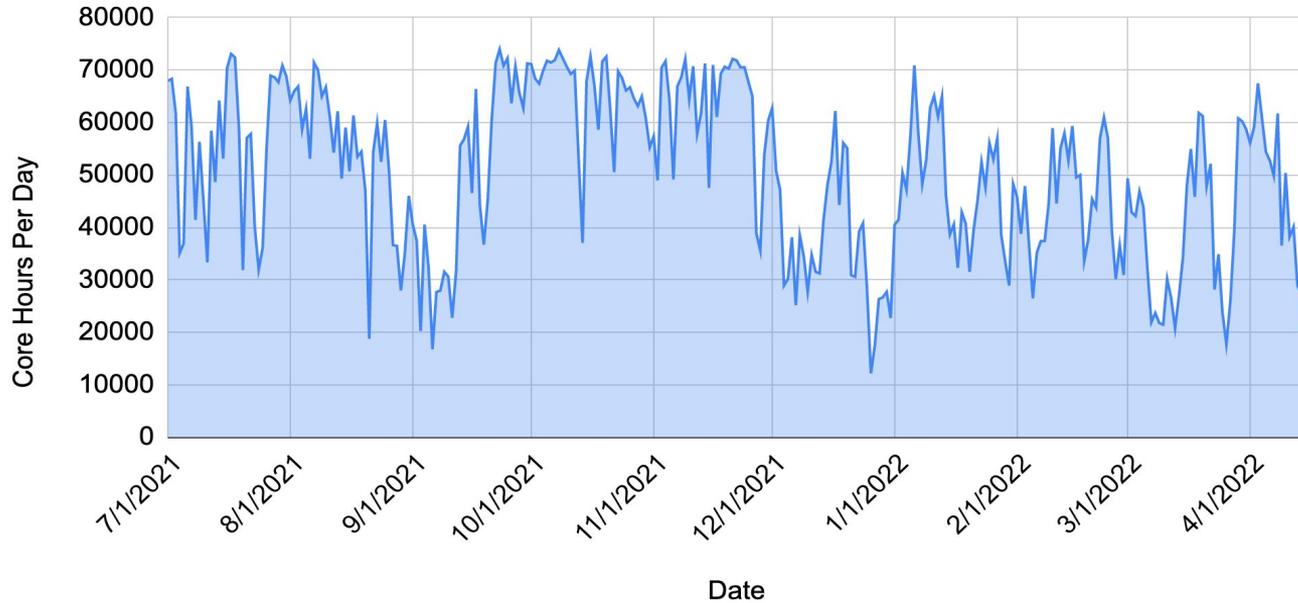


- Launched in December 2018
- Expanded in December 2019
- 5 year lifetime

- **Phase 1 retirement - December 2023**
- **Phase 2 retirement - December 2024**



Terremoto - Cluster Usage in Core Hours



*Max Core Hours Per Day = **73963***

*Total core hours in the past 12 months = **18 million***



Ginsburg

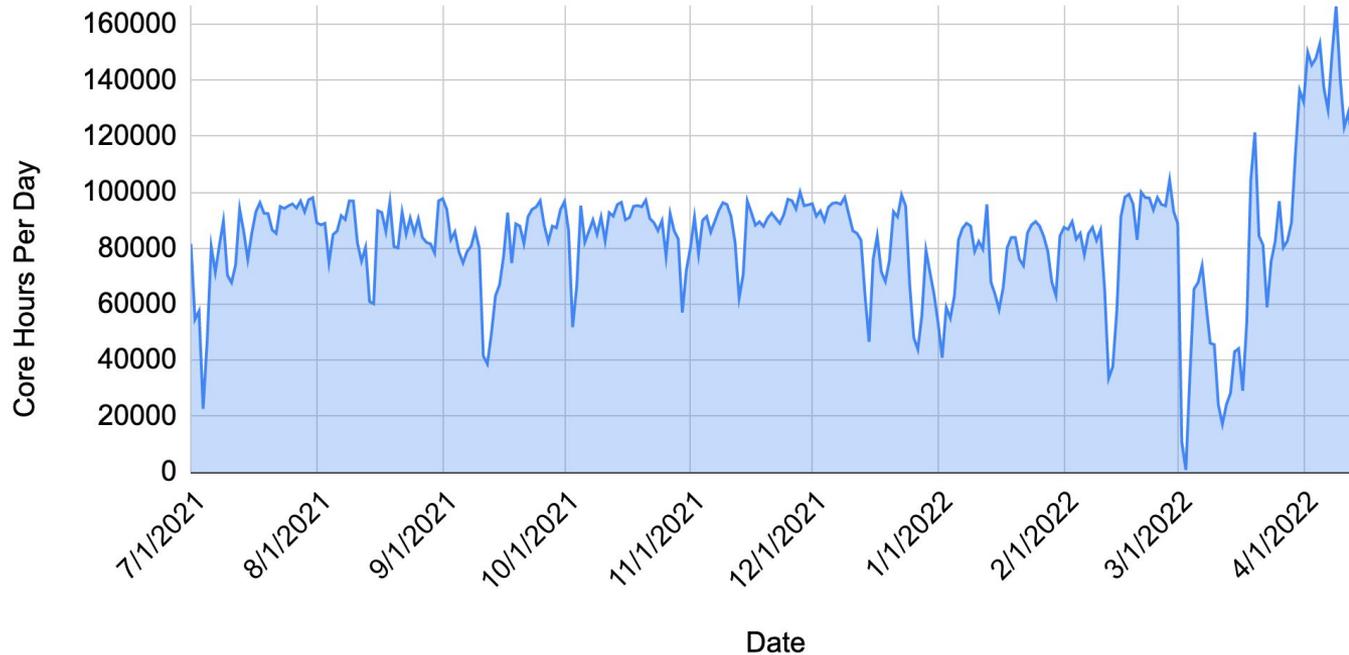
- Ginsburg went live in February of 2021 with a \$1.4 million joint purchase by 33 research groups and departments, consisting of 139 nodes with a total of 4448 cores.
- Ginsburg Expansion went live in March 2022 with over a million-dollars, expanding the cluster by 99 nodes with a total 3168 cores this Spring.
- The system now consists of 238 nodes with a total of 7616 cores (32 cores per node), including 33 GPU hardware accelerated systems.



Ginsburg



Ginsburg - Cluster Usage in Core Hours



*Max Core Hours Per Day
Pre-Expansion = **104,397***

*Max Core Hours Per Day
Post-Expansion = **166,468***

*Total core hours used since
launch: **32 million***



Shared High Performance Computing

Providing Shared Compute
Since 2012

Since 2012, more than

- 18 Million jobs run
- 314 Million core hours of compute provided
- 350 Peer-reviewed publications

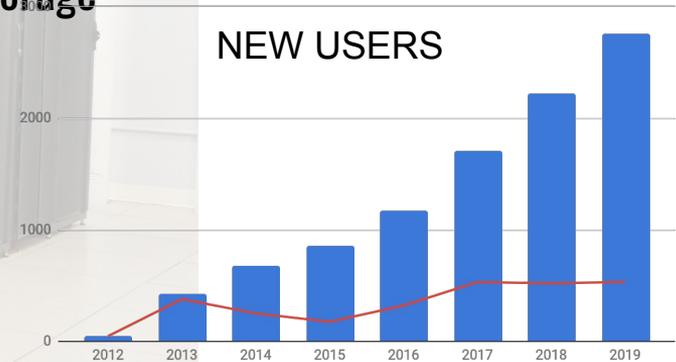
Currently more than

- 677 Compute Nodes
- 18,176 Cores
- 1236 TFlops
- 2.1 Petabytes of Storage

Faculty-led Governance

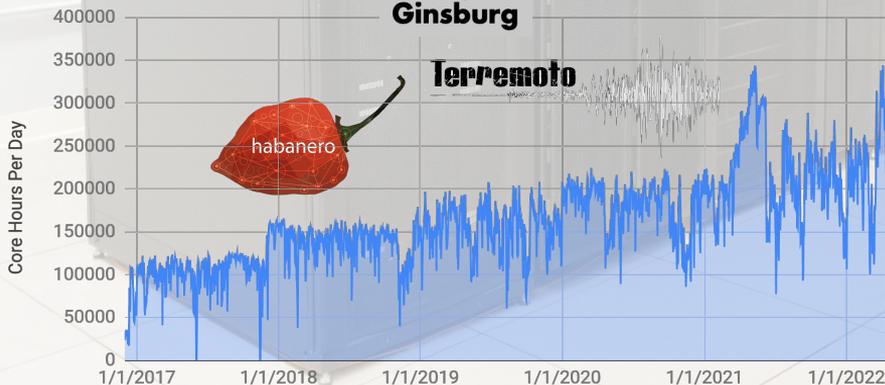
More than
70 Groups and Departments

■ New Users (Cumulative) ■ New Users by Year



NEW USERS

Core Hours Usage 2017 - Present



Introductory training offered



COLUMBIA UNIVERSITY
INFORMATION TECHNOLOGY

HPC Purchase Round - Spring 2022

- Announcement of buy-in opportunity was sent out in mid-April
- We will expand the Ginsburg cluster and expand the new GPU cluster.
- Purchase round to commence in mid-May and be open through mid-June 2022
- Go-live of new equipment in late Fall 2022

If you are aware of potential demand, including new faculty recruits who may be interested, please contact us at rsc@columbia.edu



GPU Cluster

- We have placed an order to initiate a mid-range GPU cluster
- ETA is 6 months
- We will be taking orders in this purchase round to join in
 - Mercury GPU424 5U Server:
 - CPU: (2) EPYC 7643 2.3GHz 48-core
 - Memory 1TB DDR4-3200 16x64GB
 - GPU: (8) A6000 + (4) NVLink



HPC Purchase Round 2022 - Menu Options and Specs

All servers come with Cascade Lake Intel 6226R cpus (2 per server)

- 2.9 GHz speed, 32 cores per server
- Servers offer over 50% more total CPU performance than prior generation Terremoto servers

HPC Purchase Options

- Standard server (192 GB)
- High Memory server (768 GB)
- GPU server (2 x Nvidia A40)
- GPU server (2 x Nvidia A100)



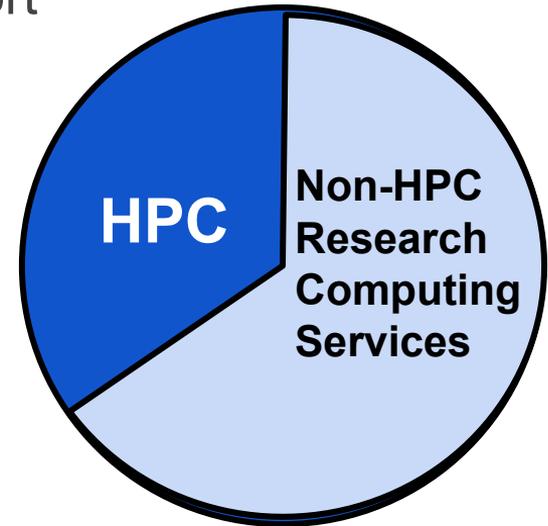
HPC Support Services

- **Email**
 - hpc-support@columbia.edu
- **Office Hours (Online)**
 - Speak with HPC support staff via Zoom from 3pm – 5pm on 1st Monday of month: [Registration required](#)
- **Group Information Sessions**
 - HPC support staff meet with your group, upon request
- **Training Workshops every semester (Online)**
 - Introduction to Linux
 - Introduction to Scripting
 - Introduction to High Performance Computing
- **Cloud Computing Consulting**
 - Complimentary assistance moving HPC workloads to the cloud



Research Computing Services Update

- **Research Computing Services Overview**
- **Research Computing Services Update**
 - Embedded Research Computing Support
 - Secure Data Enclave on-prem
 - Secure Data Enclave in the cloud



Research Computing Services

Embedded Research Computing Support

We provide embedded research computing support to CPRC, SSW, DSI, Stats, and other affiliates on the Morningside and Medical Center campuses.

Secure Data Enclave (SDE)

A virtual platform used for working with restricted data sets.



Electronic Research Notebooks with LabArchives

This service helps organize and store research data, provides information sharing, and enables collaboration, all with automated backups and a comprehensive audit trail. Enterprise license is covered by CUIT and the Libraries. Approved for RHI and PHI.

Globus

Our enterprise Globus subscription helps you efficiently, securely, and reliably transfer data directly between systems.



Cloud Research Computing Consulting

Looking to utilize the Cloud to further your research efforts? Our team can help you determine the best resources and configurations to support your needs and assist with onboarding.



XSEDE National HPC Campus Contact

Columbia researchers can try out a test allocation and receive guidance for applying for free XSEDE national HPC resources.

SnapGene

A molecular biology software that allows users to plan, visualize, and document molecular biology procedures. Beginning last year, CUIT has offered our researchers the opportunity to purchase an annual SnapGene license at the **reduced price** through the University's multi-seat standard license.



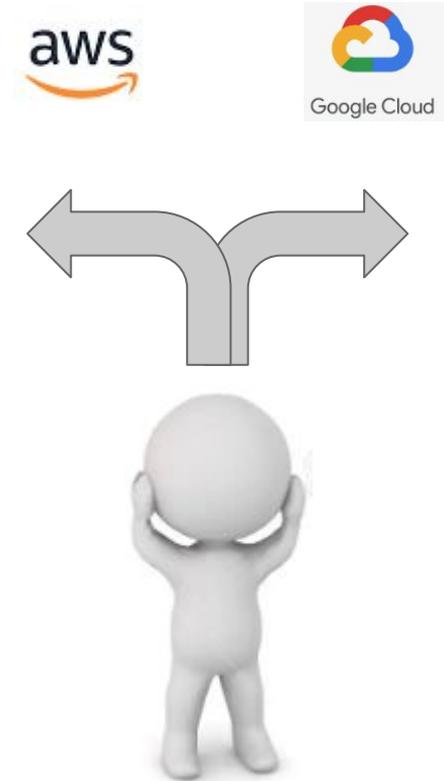
Research Computing Services Update

- **Secure Data Enclave - on prem**
 - Virtual cold room for researchers to analyze and collaborate on projects with restricted data sets
 - 23 projects and 50 active users
 - Automated Project Onboarding Process
 - Upgrade of the current on-prem SDE environment with the latest blades with the maximum RAM
 - New delivery groups are created and licenced software is installed only for authorized groups



Research Computing Services Update

- **Secure Data Enclave in the cloud**
 - Columbia University has a Business Associate Agreement (BAA) in place for GCP and AWS.
 - Exploring GCP and AWS environments to leverage improved storage size, scalability, compute power and memory capabilities.
 - Testing our ideas, assumptions and costs in isolation on a smaller scale to verify the functionality of deploying SDE to the cloud.



Research Computing Services

Research Computing Services support is available to discuss your research technology needs by emailing rsc@columbia.edu.



SRCPAC Spring 2022 Update

April 20th, 2022

Foundations Mission

Foundations for Research Computing provides **informal training** for Columbia University graduate students and postdoctoral scholars to develop fundamental skills for harnessing computation: core languages and libraries, software development tools, best practices, and computational problem-solving.

Foundations Primary Activities

- **Novice boot camps:** 2 day training based on Software Carpentry curriculum for novice learners
- **Python User Group:** twice-monthly meeting for those using Python in their research or who are curious about the Python programming language
- **Intermediate intensives:** 1 day training for intermediate learners
- **Workshops:** 1.5 - 2 hour training opportunity to advance computational skills in a group setting. Workshops are often led by partners including CUIT and the Libraries

Foundations for Research Computing: Highlights

Year 4 Goals:

- Continue to enhance & develop remote programming
- Continue through the Pandemic
- Provide a variety of offering targeting different experience levels
 - Boot camps trained 155 people
 - Python User Group 210 attendees
 - Workshops 290 attendees

Foundations response to Covid

The challenges presented by COVID-19 required changes to the format of Foundations instruction, including the bootcamps. The shift to remote teaching provided some positive outcomes, including:

- Zoom format allowed helpers to address questions quickly in bootcamps
- Zoom format allowed for recordings & sharing with participants for all programming
- Able to support same number of workshop participants with a smaller number of workshop helpers
- Remote allows for greater participation in Python User Group
- Able to identify more specific needs for training by the way that researchers attended bootcamps

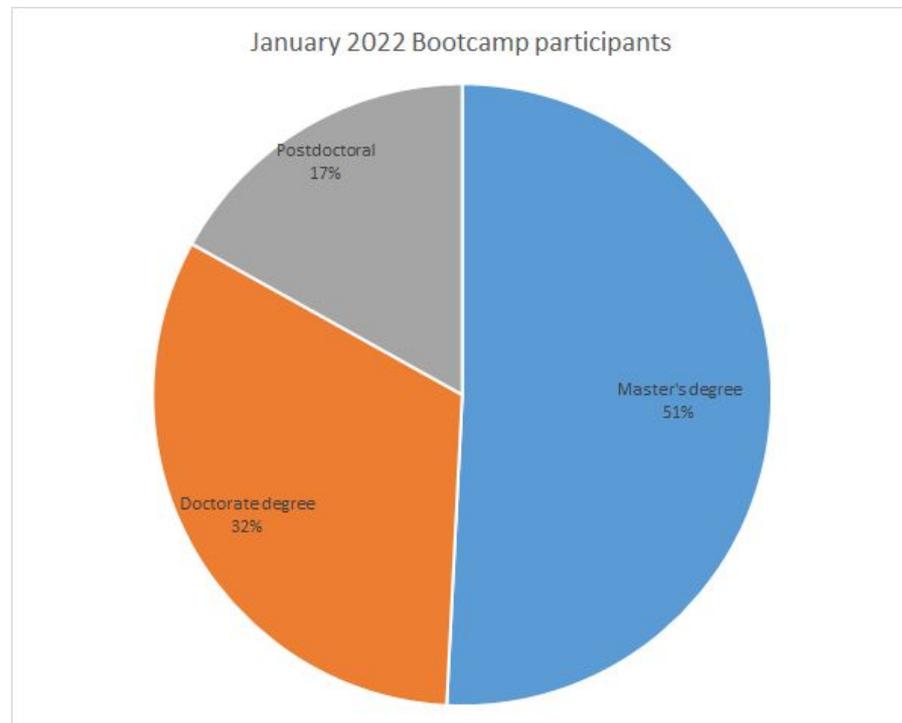
However there are some downsides to the transition as well:

- Software Carpentry is focused on in-person pedagogy
- Zoom fatigue
- Fewer (volunteer) instructors are comfortable with teaching in the zoom environment

Spring 2022 Bootcamp

January 2022 bootcamp offered remotely

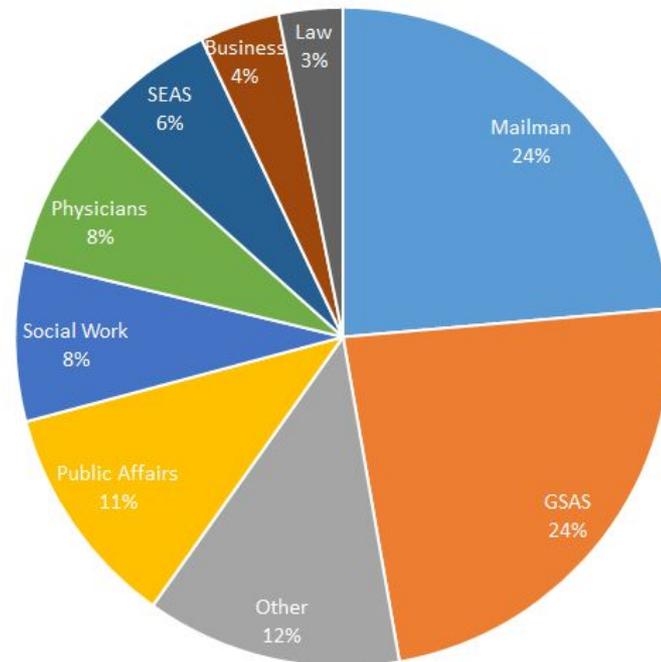
- 950 Unique applicants
- Trained 127 researchers
- Collaborated w/office of Compliance to also include 28 PhD students in prep for a stats course taught by Art Palmer



Foundations for Research Computing: Highlights

January 2022 Bootcamp Participants

- Worked with the Libraries Assessment team to simplify and create a better mechanism for student selection for the program



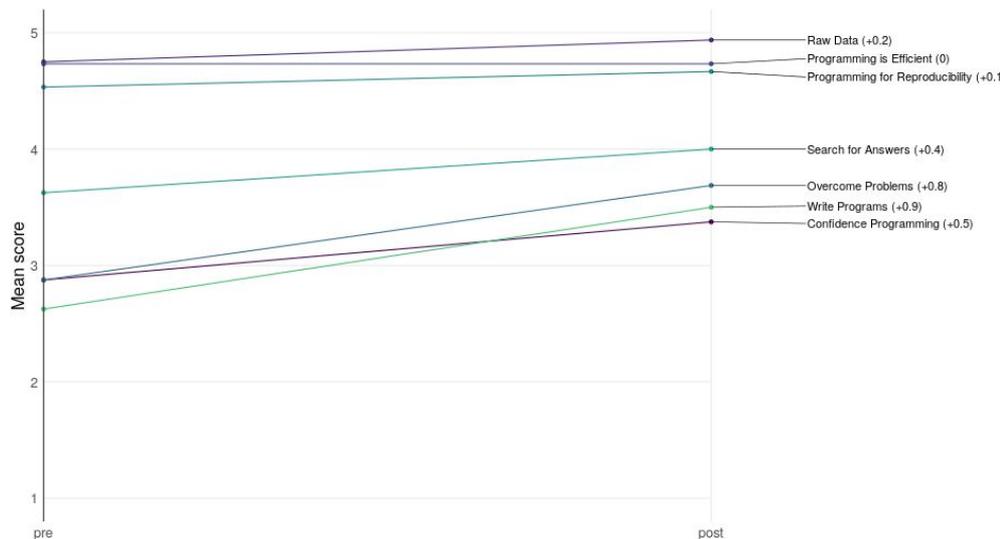
Spring 2022 Bootcamps

“I thought the instructors were extremely thorough and friendly. I was so impressed and happy with the amount of information that I learned.”

“It was a very good introduction to the concepts of Git, python and unix. The instructors explained everything in detail and I was always helped when I was stuck.”

“Clear syllabus and well-documented examples; patient instructors”

Change in self-reported skill and confidence levels pre- and post-workshop (n = 16)



Next Bootcamp, Fall 2022

Fall 2022 bootcamp

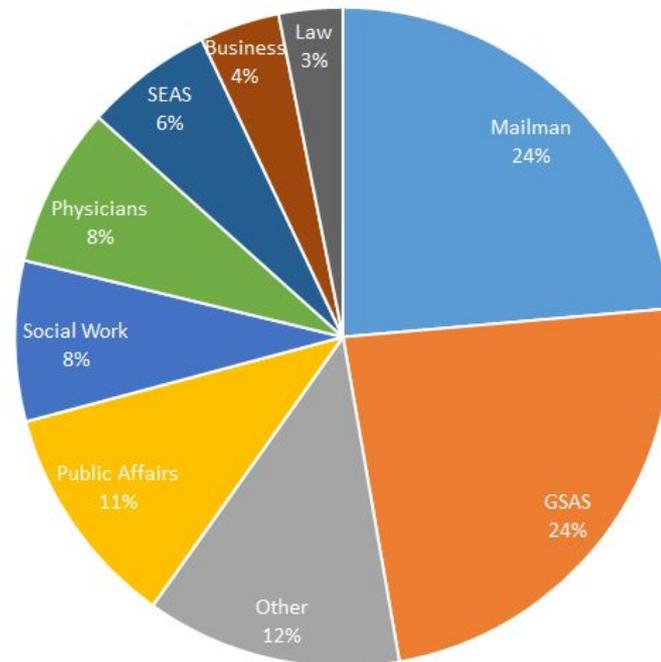
- Aim: return to in-person
- Planning is underway
- Looking for ready made space!



Foundations for Research Computing: Highlights

- Worked with the Libraries Assessment team to simplify and create a better mechanism for student selection for the program

January 2022 Bootcamp Participants



Python User Group

Python User Group - meets twice-monthly, hired two graduate students who work closely with a member of the Libraries Research Data Services to identify topics and develop curriculum.

- Scaling Python Analytics with Dask
- General Intro to Pandas & EDA
- Intro to Julia for Python Users
- Geopandas & You
- Intro to Text Analysis in Python
- Test-Driven Development & ML Audio Transcription

Foundations for Research Computing: Highlights

Instructors

- Trained 9 new instructors last year
- Currently recruiting new instructor for spring cohort using the revised interview process designed last year.
- Underwritten the Software Carpentry Training for 46 instructors over time
 - 85% of the instructors have participated in Foundations offerings
 - 45% have participated in more than 1 bootcamp
- Current instructor pool at 29 Instructors

Foundations for Research Computing: Demand

Demand has always been higher than what Foundations can provide.

- **Fall Bootcamp:** 1,500 unique applicants for 120 spots.
- **Spring Bootcamp:** 950 applicants for 120 spots

Continuing questions:

- How to scale to meet demand?
- Who should Foundations serve?
- How to articulate different training needs among these applicants?

Mechanical Engineering

- Ran a Software Carpentry bootcamp in mid August for incoming masters students
- Adapted Python portion to be discipline specific
- Intends to offer a full week bootcamp next year
- Successfully running smoothly, minimal central resources
- Calls upon Foundations trained instructors

Models for Expansion: Supporting Center Grant Proposals

Center for Learning the Earth with Artificial Intelligence & Physics (LEAP)

- Participated in successful proposal for a \$25 million NSF Science & Technology Center award
- Currently working with Tian Zheng, *Chair, Department of Statistics & Education Director for LEAP* to train students as Software Carpentry instructors, 1-2 per semester
- Depending on timing, they will participate as helpers for the regular bootcamps, before being instructors for the LEAP bootcamps
- Those students will join the growing Foundations instructor community on campus

Future goals for the program

- Recruit a new Foundations Program Manager
- Increase the capacity of the program to try and meet demand
- Continue to explore alternative models/partnerships (e.g. MechE Orientation, LEAP)
- Further develop the Python User Group community
- Develop a more robust assessment plan
- Seeking input/new partners in guiding future directions of Foundations
- Happy to take any questions

Other Business and Closing Remarks

- Consumer-grade GPU cluster

Contact rcs@columbia.edu

Fall 2021 Bootcamps

Quotes:

The overall breadth of the material, as well as the UNIX and Git training, were excellent

I think the ability to follow step by step what the instructor was doing (doing this on Zoom probably helped), also helpers were extremely efficient and resolving questions. Additionally, I felt the material was made very amenable for complete beginners.

Having the helpers answer chat questions, the explanations of everything were helpful and no question was too simple, everyone was made to feel comfortable, I have never felt more comfortable in a coding course environment, all of the instructors were great, the setup information/resources given before the bootcamp started were very helpful.