Shared Research Computing Policy Advisory Committee (SRCPAC)

Spring 2025 Meeting

Alexander Urban, *Chair of SRCPAC*

Agenda

- Welcome & Introductions
- Empire AI Alpha Phase
- High-Performance Computing Update
  - Alexander Urban, Chair of SRCPAC
  - Max Shortte, Manager CUIT High Performance Computing
- Research Computing Services Update
  - Axinia Radeva, Sr. Manager of CUIT Research Services
- Foundations for Research Computing Update
  - Marc Spiegelman, Chair of the FoRC Advisory Committee
  - Jonathan O. Cain CUL
  - Future of Training in Research Computing
- Other Business & Closing Remarks

COLUMBIA|RESEARCH

Introductions

- Welcome & **Introductions**
- Empire AI Alpha Phase
- High-Performance Computing Update
  - Alexander Urban, Chair of SRCPAC
  - Max Shortte, Manager CUIT High Performance Computing
- Research Computing Services Update
  - Axinia Radeva, Sr. Manager of CUIT Research Services
- Foundations for Research Computing Update
  - Marc Spiegelman, Chair of the FoRC Advisory Committee
  - Jonathan O. Cain CUL
  - Future of Training in Research Computing
- Other Business & Closing Remarks

Shared Research Computing at Columbia

Faculty-led

Shared Research Computing Policy Advisory Committee (SRCPAC)

FoRC Advisory Committee

EVPR Office of Research Initiatives and Development (ORID)

CUIT

Shared Research Computing Facility

- High-Performance Computing (HPC)
- Procurement & acquisition of computer hardware
- Maintenance
- Access management (free, rental, buy-in, education)

Research Computing Services (RCS)

- Embedded Support
- Globus Data Transfer
- Secure Data Enclave
- Workshops
- IT/Cloud Consulting
- Software (LabArchives Overleaf Pro, GraphPad Prism & SnapGene)
- NSF ACCESS allocations

Libraries

Foundations for Research Computing (FoRC)

- Training: core languages & libraries, software development tools, best practices, and computational problem-solving
- Map resources for research computing

COLUMBIA|RESEARCH

Launched by Governor Kathy Hochul in April 2024 in partnership with New York's leading public and private institutions, Empire AI is a NYS initiative to advance artificial intelligence.

The Empire AI ALPHA system consists of 13 HGX Nodes, 8 H100 80GB GPUs per node and uses fair-share scheduling on a rotating priority access system to ensure equal allocation of time.

CUIT's High Performance Computing (HPC) and Research Computing Services (RCS) teams provide support.

Columbia University Empire AI Metrics:
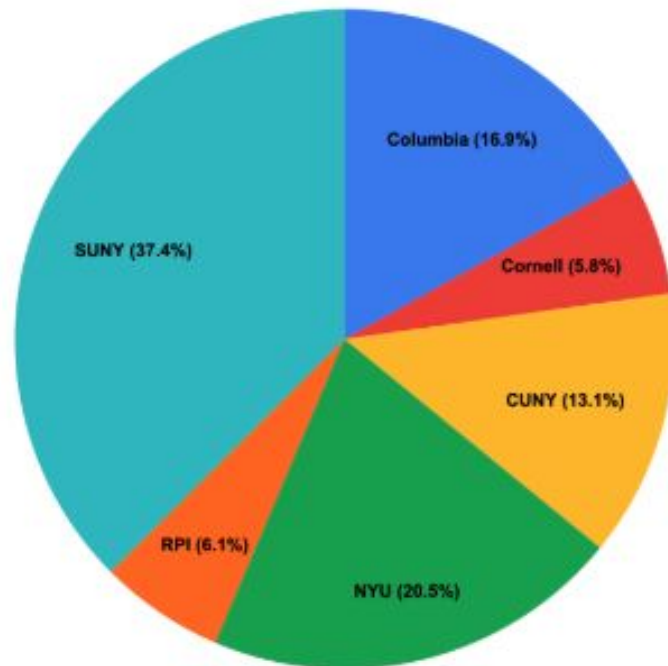Completed Job Count: 344
Completed Job Run Count: 4428:48:38
Allocated GPUs Used: 3,731
Number of Accounts: 43
Number of Projects: 14
Number of PIs: 12

Pie chart: SUNY (37.4%), Columbia (16.9%), Cornell (5.8%), CUNY (13.1%), NYU (20.5%), RPI (6.1%)

High Performance Computing Updates

Alex Urban
Chair, Shared Research Computing Policy Advisory Committee (SRCPAC)

Halayn Hescock
Sr. Director, CUIT Research Services

Max Shortte
Manager, CUIT High Performance Computing

Current HPC Footprint

Terremoto Phase [1-2]

137 nodes:
- 111 Standard nodes (192 GB)
- 14 High Memory nodes (768 GB)
- 12 GPU nodes with NVIDIA **V100** GPUs

Manitou - GPU Cluster

15 GPU nodes:
- 13 nodes with 1TB of memory 96 cores and 8 **A6000** GPUs with **NVLink**
- 2 nodes with 256G of memory 32 cores and 8 **A6000** GPUs

Ginsburg Phases [1-3]

286 nodes:
Total of 9,152 cores (32 cores per node)
- 191 Standard Nodes (192 GB)
- 56 High Memory Nodes (768 GB)
- 18 GPU 2x **RTX 8000** GPU modules
- 4 GPU 2x **V100S** GPU modules
- 9 GPU 2x **A40** GPU modules
- 8 GPU 2x **A100** GPU modules

Insomnia

60 nodes:
Total of 4,800 cores (80 cores per node)
- 31 Standard Nodes (192 GB)
- 15 High Memory Nodes (768 GB)
- 3 GPU 2 x **L40**
- 7 GPU 2 x **L40S**
- 1 GPU 4 x **Intel**
- 3 GPU 2 x **H100**

Current Considerations with HPC

- **Storage**

- **Retirement Policy**

- **Bright Management**

Consolidation Project

What?
- Consolidate storage for all clusters into one device

Why?
- Move off of problematic `DDN/Lustre` storage
- Over `$375,000+` saved annually on Bright licenses.
- Allow continued usage of retired equipment
- Achieve `90-95%` resource utilization efficiency across all HPC clusters.

How?
- Expand current Kalray storage on Insomnia
- Connect all clusters to new storage
- Keep `Terremoto` and `Ginsburg` queues separate

When?
📌 Target completion by the end of `Q2 2025` (exact date to be determined).

COLUMBIA|RESEARCH

New HPC Hardware Retirement Policy

Compute nodes are purchased for a five-year life. We aim to keep the nodes available until one of the following conditions is met:

- **Hardware Failure:** Nodes will be disabled if they experience hardware failure. Since they are no longer under maintenance contracts or warranties, repair is not guaranteed.
- **Space for New Hardware:**
  - The oldest nodes will be decommissioned when space is needed to accommodate new hardware installations.
  - Affected users will be informed at least 6 months in advance.

Job Access & Fairshare on Warranty-Expired Clusters

- **5-Day Jobs** – Researchers can run **5-day** jobs on older clusters if they have at least one working node.
- **Free Partition Fallback** – If all nodes in a partition fail, research groups **retain storage** access and can submit **12-hour** jobs via the **free** partition.
- **Fairshare Calculation** – Fairshare is proportional to the number of nodes a research group owns.
- **Operating Committee meeting will be set to discuss**

Moving Away From Legacy Storage

- As part of ongoing efforts to optimize storage performance, the `Terremoto` and `Ginsburg` clusters will transition away from `legacy storage` to Kalray storage solutions.
- `Manitou` cluster will be `integrated` into `Insomnia`, reducing the number of clusters from `four` to `three`.
- `Manitou details`:
  - The Manitou integration will add `120 GPUs` to Insomnia's GPU pool.

COLUMBIA|RESEARCH

Moving Away From Legacy Storage

TERREMOTO　　**GINSBURG**　　**INSOMNIA**

MANITOU

KALRAY STORAGE
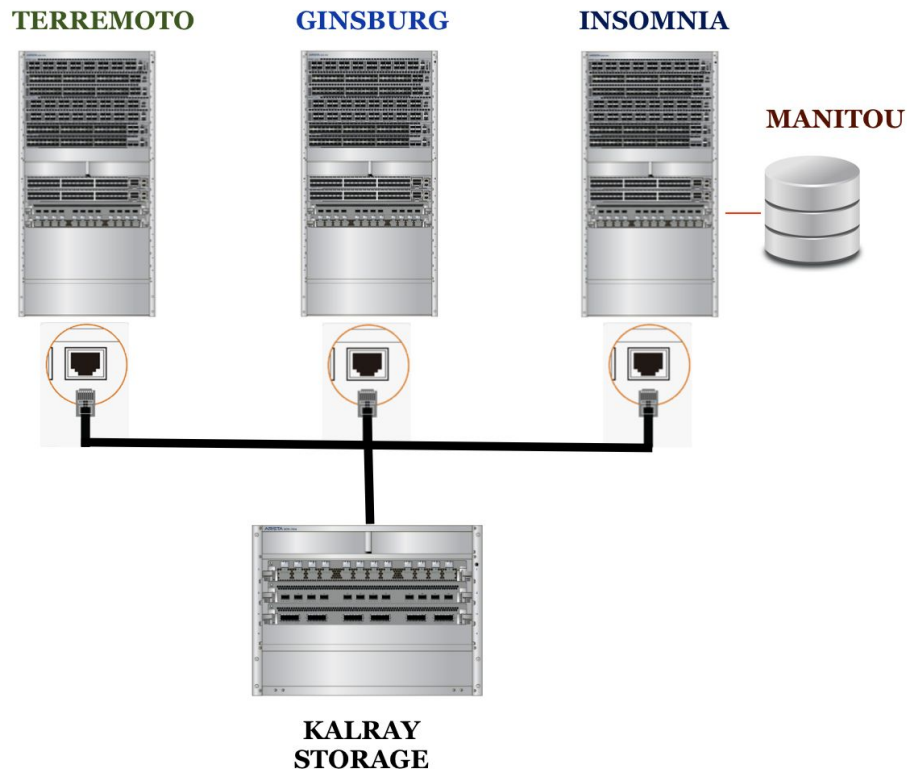
Performance Improvements:

Metric	Difference
Read Speed *(GB/s)*	**+32%**
Write Speed *(GB/s)*	**+23.5%**
IOPS *(K)*	**+21%**
Latency *(ms)*	**-13%**

COLUMBIA|RESEARCH

Manitou GPUs will become available to Insomnia

Overview of Combined Resources Post Manitou Integration

- **Total Nodes**: 75
- **Total CPUs**: 8,664 CPUs
- **Total RAM:** 44.2 TiB
- **Total GPUs:** 144 GPUs

SLURM FairShare, Queues, and Priorities

- **SLURM fairshare** will **remain unchanged** across all clusters.
- **Each cluster** will preserve:
  - Current fairshare allocation
  - Queues
  - Priorities
- **Only the storage device** is changing (DDN → Kalray).
- **No impact** on:
  - Job distribution
  - Scheduling
  - Prioritization

HPC Statistics | Calendar Year 2024

Ginsburg Statistics | 2024

Ginsburg Cluster Usage Over Time (With Downtime Factored In)
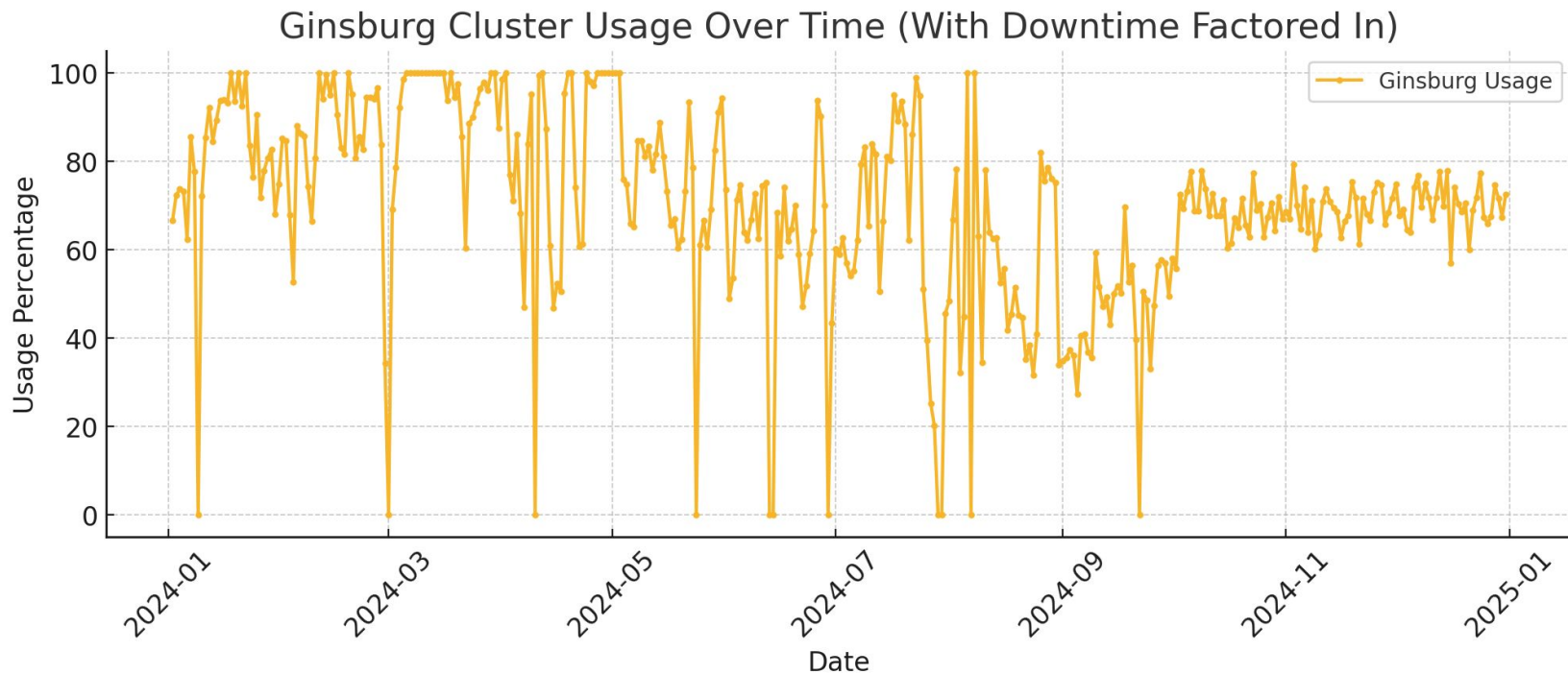
Insomnia Statistics | 2024

Insomnia Cluster Usage Over Time (With Downtime Factored In)

COLUMBIA|RESEARCH
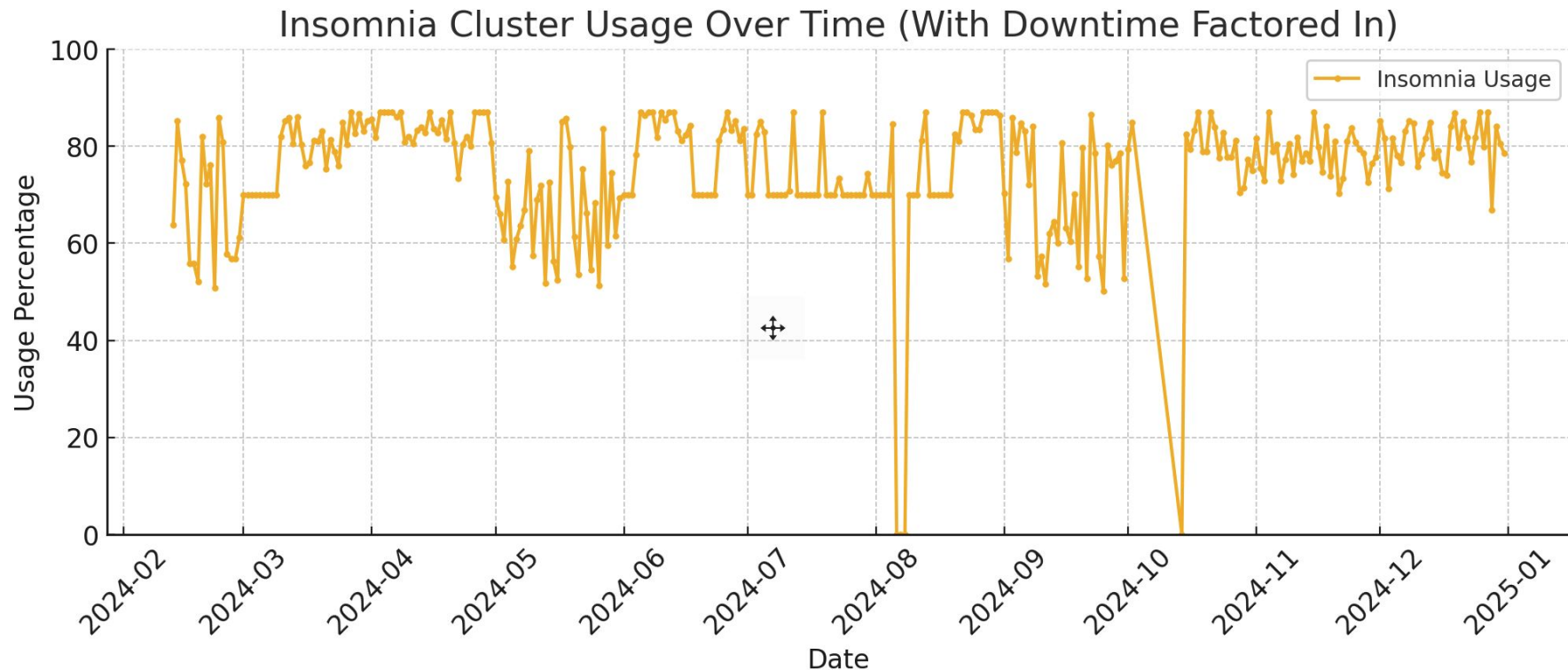
Free/Edu Tier Updates

A new Free/Education tier has been established on Insomnia
- Contributed by A&S, SEAS, and EVPR
- Each year, they will purchase a new node to add to the free tier
- Users have been migrated from the retired equipment to Insomnia

Free/Edu Tier Statistics | 2024

Total Accounts:

Partition	Accounts
Edu accounts	56
Free accounts	160
	216

Total Completed Jobs:

Partition	Completed Jobs
Total Completed Jobs	**887,669**

♔ COLUMBIA|RESEARCH

HPC Support Services

- **Email**
  - hpc-support@columbia.edu - all HPC questions
- **Group Information Sessions**
  - HPC support staff meet with your group, upon request
- **Training Workshops every semester (Online)**
  - Introduction to Linux
  - Introduction to Bash/Scripting
  - Introduction to High Performance Computing
  - Introduction to Python in HPC Environments (coming soon)

Research Computing Services Updates

Axinia Radeva, *Sr. Manager, Research Computing Services, CUIT*

Research Computing Services   cuit.columbia.edu/rcs

⎯ 🖥️ **Embedded Research Computing Support** – Dedicated support for research computing across campuses.

⎯ 🔐 **Secure Data Enclave (SDE)** – Virtual platform for secure data analysis.

⎯ 📓 **Electronic Research LabArchives Notebooks** – Organize, store, and share research data with backups & audit trails.

⎯ 🔄 **Globus Research Data Transfer** – Secure, high-speed data transfers.

⎯ ☁️ **Cloud Research Consulting** – Expert guidance on computing resources & onboarding.

⎯ 🏛️ **ACCESS National HPC Resources** – Try Columbia's Discover allocation & apply for national HPC resources.

⎯ 📊 **GraphPad** & 🧬**SnapGene Discounts** – Statistical analysis and molecular biology software at reduced rates.

⎯ ✍️ **Overleaf Professional** – Collaborative LaTeX & Rich Text writing tool for scientific publishing.

COLUMBIA|RESEARCH

Improved SRCPAC website https://research.columbia.edu/content/srcpac

Research Services: Evolving Website and Expanding Video Library

Find your CUIT research resources in one place:
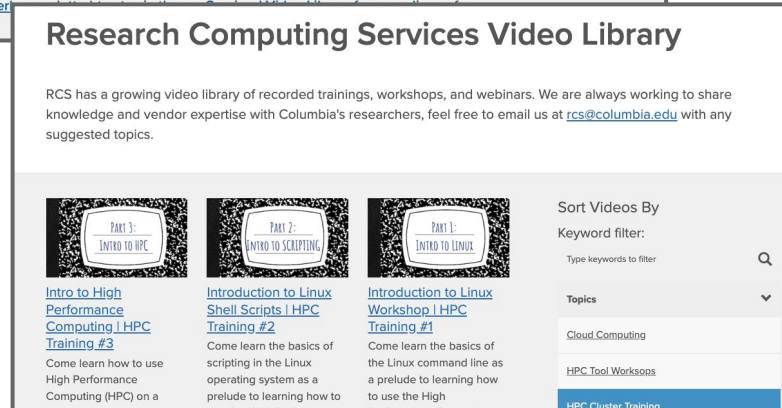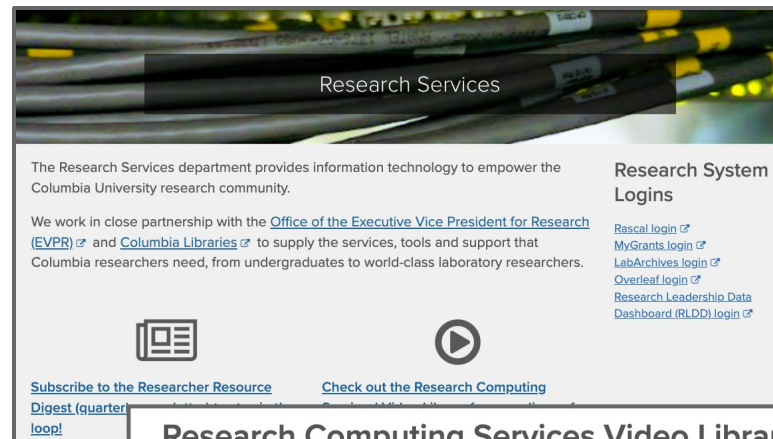**https://www.cuit.columbia.edu/research**

- Research system login links
- Highlights of newly-offered tools
- Research Services portfolio overview
- FAQ

A calendar of research events across Columbia & a growing repository of training recordings is available:
https://www.cuit.columbia.edu/rcs/training

- HPC cluster trainings
- Intel & NVIDIA HPC tool workshops
- Cloud computing overviews
- SnapGene webinars and more

Research Services

The Research Services department provides information technology to empower the Columbia University research community.

We work in close partnership with the Office of the Executive Vice President for Research (EVPR) and Columbia Libraries to supply the services, tools and support that Columbia researchers need, from undergraduates to world-class laboratory researchers.

Research System Logins

- Rascal login
- MyGrants login
- LabArchives login
- Overleaf login
- Research Leadership Data Dashboard (RLDD) login

Subscribe to the Researcher Resource Digest (quarterly...) loop!

Check out the Research Computing...

Research Computing Services Video Library

RCS has a growing video library of recorded trainings, workshops, and webinars. We are always working to share knowledge and vendor expertise with Columbia's researchers, feel free to email us at rcs@columbia.edu with any suggested topics.

PART 3: INTRO TO HPC

Intro to High Performance Computing | HPC Training #3

Come learn how to use High Performance Computing (HPC) on a

PART 2: INTRO TO SCRIPTING

Introduction to Linux Shell Scripts | HPC Training #2

Come learn the basics of scripting in the Linux operating system as a prelude to learning how to

PART 1: INTRO TO LINUX

Introduction to Linux Workshop | HPC Training #1

Come learn the basics of the Linux command line as a prelude to learning how to use the High

Sort Videos By

Keyword filter:

Type keywords to filter

Topics

- Cloud Computing
- HPC Tool Worksops
- HPC Cluster Training

COLUMBIA|RESEARCH

Research Computing Trainings & Workshops (Since April 2024)

15 Workshops: 6 held in-person, with over 188 attendees (400+ registrations)*

◆ **Trainings by RCS/HPC Staff (8 Sessions)** – Hands-on sessions focused on Columbia's research computing services:
- **Intro to HPC Series** (Intro to Linux, scripting, and using on-prem HPC)
- Secure Data Enclave
- ACCESS HPC offerings from NSF
- Globus data transfer
- Columbia Data Platform for storage & analysis
- Researcher Resources at Columbia overview
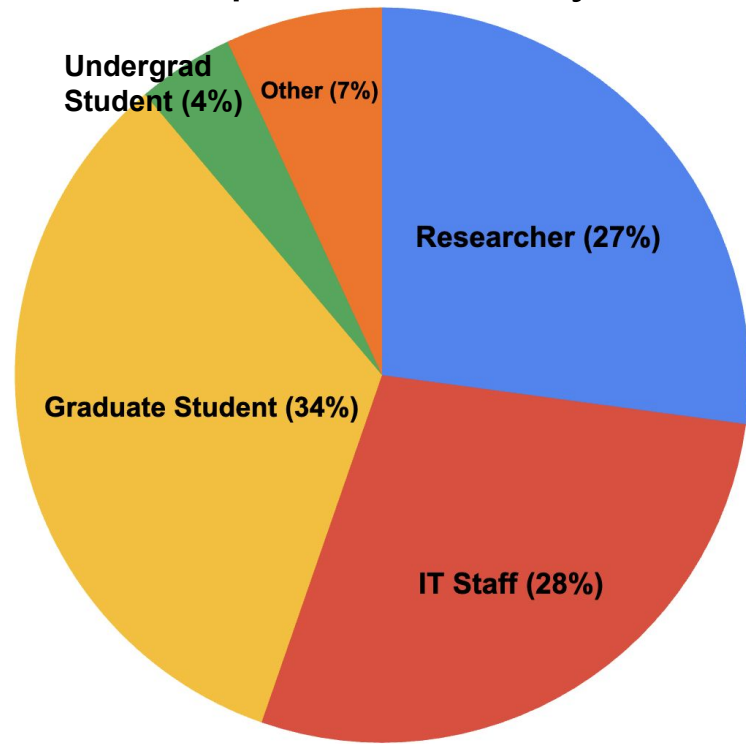- Using Jupyter Notebooks on HPC

◆ **Vendor-Led Workshops (7 Sessions)** – Covering the latest capabilities (e.g., AI, GPUs) of industry tools and platforms:
- Google Cloud Platform
- Intel
- NVIDIA (3 sessions)
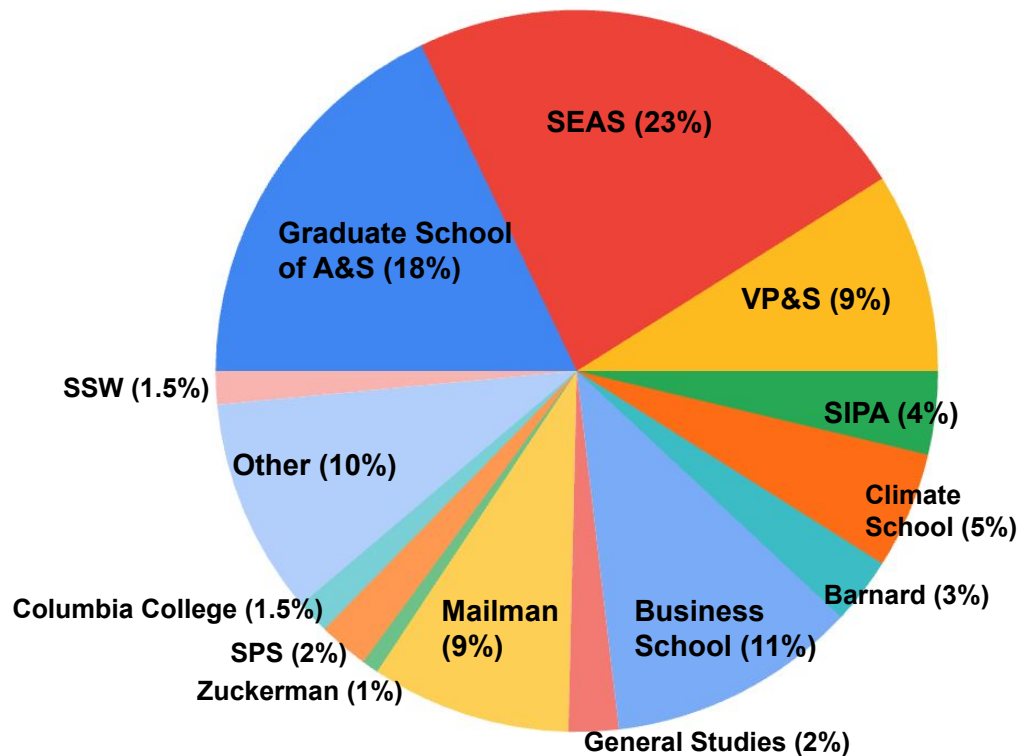- SnapGene Molecular Biology Software (2 sessions)

Attendance data not validated for 3 events

COLUMBIA|RESEARCH

Research Computing Trainings & Workshops Metrics

Workshop attendees by role

- Researcher (27%)
- IT Staff (28%)
- Graduate Student (34%)
- Undergrad Student (4%)
- Other (7%)

Workshop attendees by school

- SEAS (23%)
- Graduate School of A&S (18%)
- VP&S (9%)
- SIPA (4%)
- Climate School (5%)
- Barnard (3%)
- Business School (11%)
- General Studies (2%)
- Mailman (9%)
- Zuckerman (1%)
- SPS (2%)
- Columbia College (1.5%)
- Other (10%)
- SSW (1.5%)

COLUMBIA|RESEARCH

Globus High Speed Data Transfer Service

2019
Globus Standard launched with Google Drive Connector:
Non-sensitive data transfer between personal/lab computers/servers and HPC clusters

2023
Globus Open Access added: Globus connectors enable seamless integration with various storage systems

ActiveScale	
AWS S3	
Azure Blob	
BlackPearl	
Box	
Ceph	
Dropbox	
Google Cloud	
Google Drive	
iRODS	
OneDrive	
POSIX	
POSIX (Staging)	

2024
Globus BAA signed and High Assurance subscription established: Enables sensitive data (PHI, PII, RHI) to be transferred.

CUIT and CUIMC now provide Globus High Assurance for all Columbia users.

NEW: Globus as a service

ACCESS Roadmap at Columbia

January 2017
Columbia's Test Allocation was launched on XSEDE

September 2022
ACCESS succeeded XSEDE

Our existing Allocation was renamed to the Columbia Discover Allocation

September 2023
ACCESS approved a 750,000 supplement to our Discover Allocation

Total of 1,500,000 ACCESS credits on the Columbia Discover Allocation

March 2024
RCS launched the National High-Performance Computing – ACCESS service page

August 2024
Columbia Discover Allocation was renewed for August 2025

Launched the National High-Performance Computing ACCESS Request Service Intake Form

COLUMBIA|RESEARCH

ACCESS Metrics

2023
- 32 Allocations
- Active Users: 54
- Active PIs: 28
- CPU Hours: 62,793,892.7
- Jobs Executed: 94,163

2024
- **241%** ⇧ **109** Allocations
- **61%** ⇧ Active Users: **87**
- **3.5%** ⇧ Active PIs: **29**
- **151%** ⇧ CPU Hours: **157,378,768.6**
- **258%** ⇧ Jobs Executed: **337,305**

ACCESS CPU Utilization at Columbia University
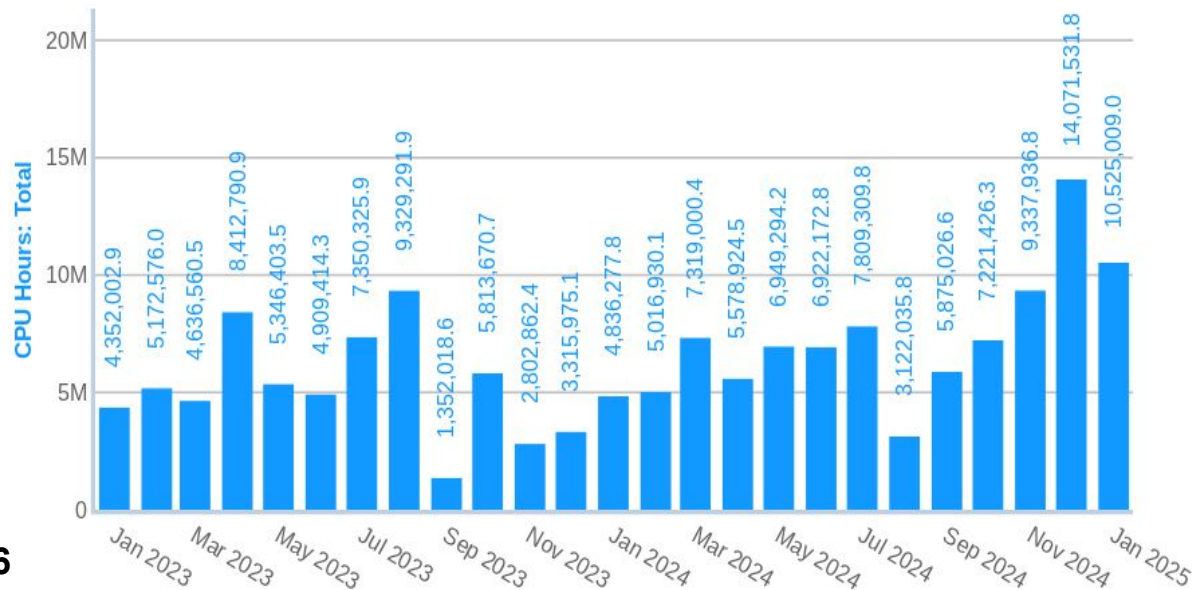
2023-01-01 to 2025-01-31 Src: XDCDB. Powered by XDMoD/Plotly JS

Added two new resources to our Columbia Discover Allocation: NCSA DeltaAI and NCAR Derecho-GPU

COLUMBIA|RESEARCH

Research Computing Services (RCS) is available to discuss your research technology needs:

Email: rcs@columbia.edu

RCS: cuit.columbia.edu/rcs

Research Services: cuit.columbia.edu/research

COLUMBIA|RESEARCH

Foundations for Research Computing Update

Marc Spiegelman, *Foundations Advisory Committee*
Jonathan Cain, *Columbia Libraries*
Daniel Woulfin, *Columbia Libraries*

Outline

- Purpose and Design of foundations.
- Current state and issues
- Goals and demand have not changed…but computation has
- Maybe it's time to re-evaluate

Foundations Mission

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

Purpose:

- To provide the investment in people and computational skills required to complement our investment in hardware, software and systems administration
- Provide a structured *pathway* for onboarding students and researchers into modern *research computing* at Columbia.

Initial Design of Foundations

- **Novice Level**
  - Institutional Partnership with Software Carpentry
  - Software Carpentry Workshops
- **Intermediate Level**
  - CUIT/RCS HPC Training
  - Topical workshops and intensives (tensorflow, nlp, etc)
  - Python User Group/Python Club
  - Integration with Departmental Training (e.g. MechE)
  - Other modes (Distinguished Lecture series, CIG)
- **Advanced level**
  - Coordination with departmental curriculum

Current Research Computing Training Landscape

Foundations for Research Computing
2x/year Foundational Workshops

Training: core languages & libraries, software development tools, best practices, and computational problem-solving

CUIT
Sample of trainings related to RCS/HPC Services

- **Intro to HPC Series** (Intro to Linux, scripting, and using on-prem HPC
- Secure Data Enclave
- ACCESS HPC offerings from NSF
- Globus data transfer
- Columbia Data Platform for storage & analysis
- Researcher Resources at Columbia overview
- Using Jupyter Notebooks on HPC
- Vendor-led training

Libraries
Computational Research Instruction

- Open workshops on computational and algorithmic literacy
- Introductions to programming
- Classroom instruction
- Requested/Invited workshops
- Individual consultations

COLUMBIA|RESEARCH

Carpentries Workshop *Applicants* over Time

COLUMBIA UNIVERSITY
Foundations for Research Computing
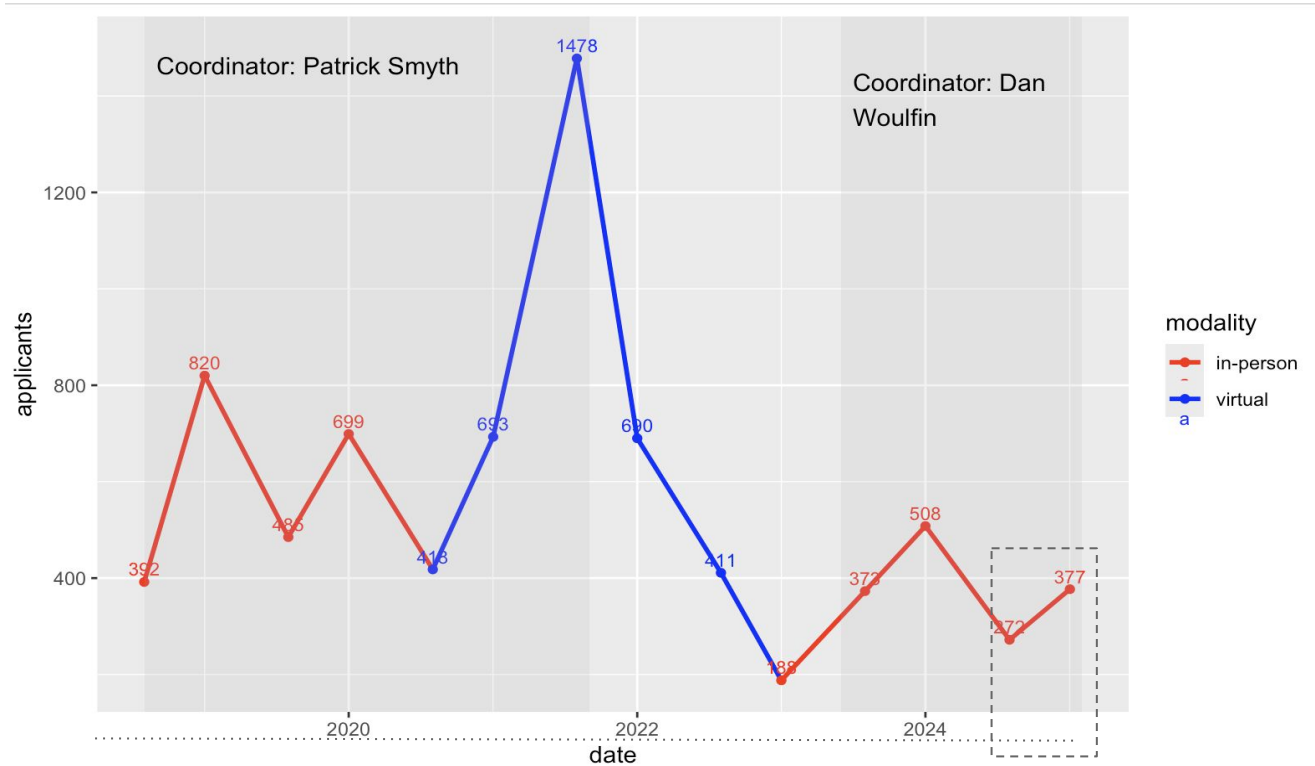
Appropriate Demand (FY 2025)

Category	Applicants	Appropriate for the curriculum - first pass	Accepted	RSVP'd	
Postdocs	84	38 (45%)	30 (79%)	23 (77%)	
PhDs	119	47 (39%)	34 (72%)	23 (68%)	
ARS	37	21 (57%)	10 (48%)	9 (90%)	
Masters	394	222 (56%)	72 (32%)	56 (78%)	
Total	649	336 (52%)	166 (46%)	113 (68%)	**Day 1 count - 100 (88%)**

Some Issues for consideration

- Carpentries Model for Beginner training doesn't scale, and is difficult to sustain
- Hierarchical structure of Foundations has reverted
- Loss of faculty guidance/coordination
- Audience and relationship to SRCPAC is less clear
- Curriculum and methods have not changed much (while the world of computation has)

Nevertheless:

- Foundations != Carpentries
- Clear demand for informal training remains strong across the university
- We need to rethink how the University meets this demand

Discussion on a Changing Framework

Addressing unmet needs by developing a more sustainable, expanded Foundations program that is data-informed, user-centered, and more modular.

Some current directions include:

- Re-engaging the Foundations Faculty Advisory Group
- Revising the Foundations Introductory curriculum and Models
  - Developing and curating resources for self-paced asynchronous learning
  - Exploration of AI Tutors (e.g. "Course Assistant", LLM's)
  - Explore peer to peer mentoring among researchers/graduate students
- Research Computing Services (CUIT) continues to provide training for support of research computing resources
- Centering computational literacy as a structuring focus in the Library

Current Research Computing Training Landscape

Foundations for Research Computing

~~2x/year Foundational Workshops~~

Becomes a website showing repository of training

CUIT

Sample of trainings related to RCS/HPC Services

- **Intro to HPC Series** (Intro to Linux, scripting, and using on-prem HPC
- Secure Data Enclave
- ACCESS HPC offerings from NSF
- Globus data transfer
- Columbia Data Platform for storage & analysis
- Researcher Resources at Columbia overview
- Using Jupyter Notebooks on HPC
- Vendor-led training

Libraries

Computational Research Instruction

- Open workshops on computational and algorithmic literacy
- Introductions to programming
- Classroom instruction
- Requested/Invited workshops
- Individual consultations

COLUMBIA|RESEARCH

Research Computing Training Landscape going Forward

Foundations for Research Computing

Provide a coordinated resource for foundational computational training for the University

CUIT/RCS

Direct support for SRCPAC HPC Services

- **Intro to HPC Series** (Intro to Linux, scripting, and using on-prem HPC
- Containerization
- Secure Data Enclave
- ACCESS HPC offerings from NSF
- Globus data transfer
- Columbia Data Platform for storage & analysis
- Researcher Resources at Columbia overview
- Using Jupyter Notebooks on HPC
- Vendor-led training

Libraries

Computational Research Instruction

- **Basic beginner training for all**
- **Better integration of AI in curriculum/Training**
- Better coordination with Faculty/ Formal Curriculum
- Curated Self-learning materials
- More focus on learning to learn

COLUMBIA|RESEARCH

Nominate yourself or a colleague as
SRCPAC Co-Chair

or

Offer to chair or join a temporary *ad-hoc*
subcommittee on entry-level training.

COLUMBIA|RESEARCH

Time for your
questions and suggestions

COLUMBIA|RESEARCH