

Shared Research Computing Policy Advisory Committee (SRCPAC) Spring 2026 Meeting

Date: Friday, April 24, 2026

Time: 10:00 AM - 11:30 AM

Chair: Alex Urban, Associate Professor of Chemical Engineering

Meeting Minutes

Welcome & Introductions – Professor Urban, Chair of SRCPAC

Professor Urban called the meeting to order and welcomed attendees. He outlined the agenda and described the structure of shared research computing at Columbia, encompassing SRCPAC, the Shared Research Computing Facility, Research Computing Services (RCS), and Foundations for Research Computing (FoRC). He noted that bi-weekly operating meetings between CUIT and the EVPR address ongoing system and policy matters, and invited interested faculty to engage.

High-Performance Computing Updates - Max Shortte, Manager, CUIT High Performance Computing

HPC Infrastructure and Utilization

The Insomnia cluster, Columbia's perpetual shared HPC system, comprises 92 compute nodes, 14,480 CPUs, 60 TB of RAM, and 158 GPUs. From July 2025 through March 2026, average daily utilization reached 91.0% for Insomnia and 88.8% for Ginsburg, with Insomnia achieving near-100% uptime following storage integration.

Systems Biology represented the largest share of Insomnia users at approximately 21% (147 users) and also ranked first in total wall time. Social Science CC accounted for 8% (59 users), followed by Statistics (7%, 51 users) and Biomedical Engineering (7%, 49 users). Chemistry represented 6% (45 users), while Chemical Engineering, Biostatistics, and Mailman SPH each contributed ~5% (37, 36, and 36 users, respectively). Astronomy & Astrophysics (34 users, 4.9%), Computer Science (33 users, 4.7%), and IEOR (30 users, 4.3%) followed, with Physics at 2.9% (20 users). The remaining 17.6% (123 users) fall under Other departments.

In total, Insomnia supports approximately 700 user accounts, reflecting broad and well-distributed adoption across the university.

Queue Policy Updates

Shortte reviewed updates to the queue structure, including the introduction of a new **burst** queue supporting wall times of up to **14 days** and a **test** queue with a **six-hour** limit running on

dedicated hardware. Over the past six months, the cluster processed approximately 2.95 million jobs and 1.31 million CPU hours. Group partitions accounted for ~54% of total CPU consumption, followed by Short queues at ~40%, with Burst contributing ~5% and HPC_Test ~1%. This distribution reflects strong utilization of primary queues while also demonstrating effective use of opportunistic capacity, with fair-share policies continuing to prioritize node ownership as intended.

The discussion also addressed whether CPU and GPU hours consumed in the preemption-based burst queue should be discounted to encourage use of idle capacity. For example, a 99% discount factor would allow a 100-hour job to count as only one allocation hour against fair-share.

Free / Edu Tier Updates

The free and educational tiers currently support approximately 1,100 accounts, consisting of 472 educational accounts and 628 free-tier accounts, operating on 2 dedicated nodes. Sustained by annual node contributions from A&S, SEAS, and the EVPR, these tiers continue to expand in line with growing demand across the research community.

These tiers play a critical role in onboarding new users, supporting instructional use, and enabling early-stage research and experimentation. By lowering barriers to access, they provide a scalable entry point into the HPC environment and help drive broader adoption across the cluster.

Research Computing Services Update – Axinia Radeva, Senior Manager, CUIT Research Computing Services; presented by Halayn Hescock, Senior Director, CUIT Research Services

Portfolio and New Services

Hescock presented an overview of the RCS portfolio, highlighting a new Data Security Officer (DSO)-as-a-Service for the Secure Data Enclave (SDE). Research teams that cannot identify a qualified DSO within their own staff may now engage a trained CUIT IT staff member to fulfill that compliance role.

Globus: High-Speed Data Transfer Service

CUIT's managed Globus endpoint supports 16 active collections and 57 users. In total, these researchers have transferred 112.56 TB of data since the service's inception a year ago. 370 users holding Globus subscription access more broadly; discussion clarified that the CUIT-managed endpoint is a centrally operated resource for those without dedicated infrastructure, while Globus itself is open to any user. Researchers requiring a new collection may contact globus@columbia.edu.

Research Software

RCS's discounted software program, which aggregates requests to negotiate group licensing, has

generated approximately \$478,000 in savings across more than 200 labs. Software ordering is now streamlined and automated via the iLab platform.

LabArchives: Electronic Research Notebook

- LabArchives is approaching its ten-year milestone.
- Growth and Collaboration: The service saw 10% growth in 2025, with over 3,000 external collaborators (e.g., Duke, Mt. Siani) active in Columbia Notebooks.
- Trends: January is the peak month for new notebook creation, coinciding with “Love Data Week” and data management initiatives.

Foundations for Research Computing Update – Daniel Woulfin, Computational Research Instruction Librarian; presented by Jessica Eaton, Research Analyst, CUIT Research Computing Services

Eaton reported that Foundations has delivered 41 trainings and workshops since last fall, reaching 615 attendees — an even split between CUIT and Libraries' events — with over a quarter returning for additional sessions. Graduate students constitute the majority of participants. Training spans three categories:

- Fundamentals – programming languages and introductory HPC concepts
- Applications – practical skills including cluster usage and an overview of Columbia AI tools for research
- Requested and guest-led workshops – department-targeted sessions for groups such as the School of Social Work, Climate School, and SIPA

Systematic feedback collection is informing a prioritized list of requested workshops for the coming year. Professor Urban proposed establishing periodic faculty consultation sessions to align Foundation programming with current research training needs. Participants expressed support.

External Research Computing Resources - Halayn Hescoek, Senior Director, CUIT Research Services

Empire AI

Columbia participates in Empire AI, a New York State initiative launched by Governor Kathy Hochul in April 2024 whose ten-member consortium includes Columbia, NYU, CUNY, SUNY, Cornell, the University of Rochester, RIT, Rensselaer Polytechnic Institute, Mount Sinai, and the Flatiron Institute. CUIT's HPC and RCS teams support Columbia researchers on the current ALPHA system — 24 HGX nodes with H100 and H200 GPUs — which has completed 344 jobs totaling 4,428 hours of run time, with 3,731 GPUs allocated across 43 accounts, 14 projects, and 12 principal investigators. CUIT developed a billing and allocation solution adopted by the Empire AI Consortium, and Empire AI has adopted CUIT's SLURM queue architecture, reducing the transition

burden for Columbia researchers.

The Beta system is now expected to be operational in May 2026, revised from an earlier spring projection. Future phases include Gamma facility construction in early 2027 and full Gamma system operation by early 2028. Empire AI at the Beta phase operates on a usage-based model with university subsidy and is not free, in contrast to NSF ACCESS.

NSF ACCESS

ACCESS (Advanced Cyberinfrastructure Coordination Ecosystem: Services and Support) is the NSF program providing national HPC resources at no cost through a proposal-based allocation process. RCS supports researchers with onboarding through the Columbia Discover Allocation and assists in transitioning to sustained, scaled-up access. Columbia now serves 114 active ACCESS users, up from 87 in 2024 and 54 in 2023, with over 2 million ACCESS credits and continued growth in both CPU and GPU usage.

In discussion of Empire AI access timing, researchers noted the value of the rapid-access model available through NSF ACCESS for small jobs and testing and suggested that Empire AI consider lower-barrier entry points where feasible.

NAIRR

NAIRR (the National Artificial Intelligence Research Resource) is a U.S. government initiative to democratize access to high-end AI computing for researchers. Columbia has obtained three research allocations through the NAIRR 2024 Pilot. Hescock clarified that NAIRR access does not require an associated NSF grant.

Claude AI at Columbia

Discussion arose regarding institutional access to Anthropic's Claude AI. Darcy Peterka reported that at least 20 Zuckerman Institute principal investigators are prepared to begin using the tool, noting that some members of the research community have independently begun accessing Claude, with leadership advising caution regarding data sensitivity. Several participants confirmed the breadth of demand and the urgency of a university-wide deployment.

CUIT confirmed that Columbia has executed a contract with Anthropic for Claude access. The service will be made available to the full community in early May 2026. Researchers seeking early access may submit a chartstring by email for provisioning.

Closing Remarks - Professor Urban, Chair of SRCPAC

Professor Urban invited continued engagement with SRCPAC, including participation in bi-weekly discussions and feedback on research computing operations, services, and training.

Professor Urban noted that a resource comparison chart covering on-premises HPC, Empire AI, ACCESS, and NAIRR — including scope, eligibility, cost, and support for each platform — would be posted online. He encouraged members to share it with their departments as a practical reference for identifying the appropriate computing environment for different research needs.

Professor Urban thanked all presenters and participants and closed the meeting.

Slides provided on: <https://research.columbia.edu/content/srcpac>

Attendees

Jessica Eaton

R. Yamir Gomez Carrasco

Halayn Hescocock

Darcy Peterka

Axinia Radeva

Sharon Sputz

Max Shortte (Zoom)

Alexander Urban

Daniel Woulfin (Zoom)

+ 6 more in-person attendees

+ 22 more virtual attendees