Research of Impact:

CRC Annual Report 2023-24

University of Pittsburgh

Table of Contents

1 CRC Users Lead Research at Pitt

Creating a New Advisory Committee Structure

2 CRC Genomics Resources Meet Challenges of AI and Data Science

3 New Model of Community-owned Research Computing and Data Infrastructure

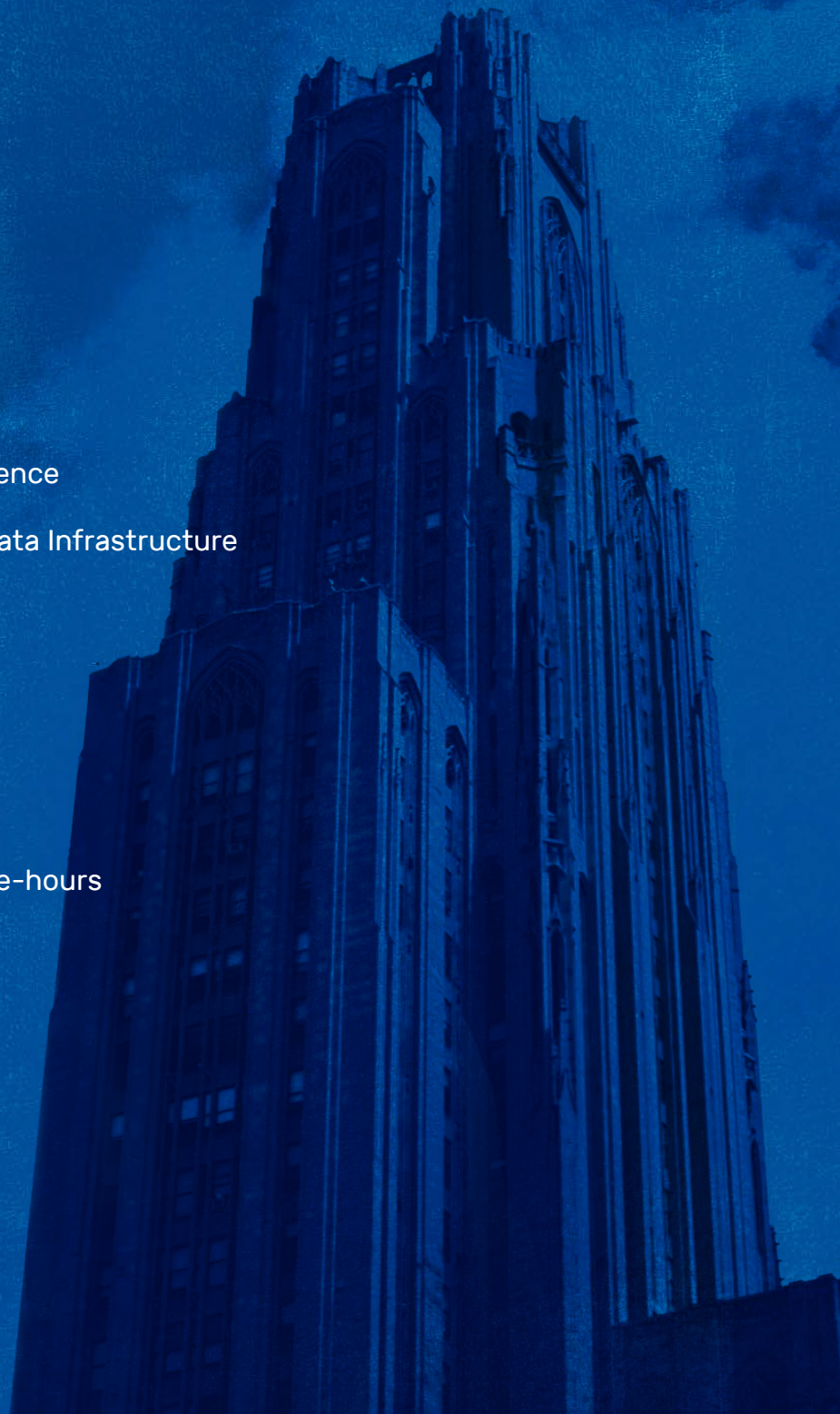
4 Improvements to User Training

Filling the Gap: A Dedicated Teaching Cluster

5 Researchers of Impact

7 User Survey Reveals Satisfaction and Areas to Improve

8 Cluster Utilization and Top 10 Departments with the Most Core-hours



CRC Users Lead Research at Pitt

The University of Pittsburgh Center for Research Computing (CRC) is happy to report that it continued to support an impressive range of funded research leading to prominent publications in fiscal year 2024. That is not a new situation: We have documented for years that CRC resources contribute to a vast range of work within Pitt's research enterprise.

137
GRANTS

$88M
RESEARCH EXPENDITURES

In FY24, CRC enabled research that was supported by 137 grants, representing $88M in research expenditures. CRC users published more than 240 papers and book chapters.

Creating a New Advisory Committee Structure

In 2024, CRC inaugurated a new advisory committee made up of 12 researchers from disciplines across the University who collaborate with CRC. The committee met in February and May 2024. At the February meeting, CRC Codirectors Kim Wong and Adam Hobaugh and Vice Chancellor for Research Computing Barr von Oehsen presented the CRC's SWOT (strengths, weaknesses, opportunities, and threats) Analysis.

Wong presented a forecast about where research computing and data are heading, what challenges and opportunities lie before us, and how CRC can leverage these opportunities. Wong described how high-performance computing, data-intensive computing, and artificial intelligence and ML are converging into a superstorm. Quantum computing is another superstorm brewing on the horizon. Fundamentally, CRC's role is to provide infrastructure to support our researchers in weathering this storm, with user support, training, and outreach playing complementary roles in its practical delivery.

CRC Genomics Resources Meet Challenges of AI and Data Science

Artificial intelligence (AI) and data science driven by high-performance computing continue to make a growing impact in health care and health sciences research. CRC has been a driver behind providing the tools and skills that create the genomics data that are the analysis targets for so many AI projects.

Here are just a few projects CRC currently supports that are integrating AI in medicine and health sciences, all of which rely on CRC's high-throughput computing (HTC) resources:

- Heterogeneous Data Integration and Precision Medicine via Machine Learning
- Developing Artificial Intelligence Tools for Advancing Precision Oncology
- Genome-wide Rare Variant Association Analysis of Cognitive Aging in Older Adults
- Genome Sequence Analysis of Candida Species Clinical Isolates

CRC collaborator Adrian Lee, director of the Institute for Precision Medicine, describes the impact of CRC on Pitt's capabilities:
"The University and Institute for Precision Medicine invested in some of CRC's HTC computational resources that are specifically designed for genomics applications. But the thing that is particularly important about CRC is the faculty support, particularly [of] Fangping Mu, who is invaluable in teaching and coordinating training for investigators.

"A large number of labs are now generating HTC data and using high-performance computing in a routine manner. Many of the experiments we do, single-cell RNA seq DNA seq, all have to be done using high-performance computing. You can't do it on your local desktop. The work is going on exponentially. Building the infrastructure has been key to all of our success. Probably [more than] 100 labs have used it. It's a central component for many labs now."

CRC collaborator Uma Chandran, research professor in the School of Medicine's Department of Biomedical Informatics and director of the Genomics Analysis Core, explains the growth driving CRC's continuing efforts to expand resources:

"In the past, you could figure out what genes were expressed in a tumor. But a tumor consists of a lot of different cell types. Five to 10 years ago, you couldn't ask about the composition of that tumor. As of a few years ago, you could try to pull out what cell types are in that tumor using bioinformatics methods. But it was still one big homogeneous mass when the data comes out, because you can't separate out those different cells in the data analysis.

"Now you can look at that tumor in space and map where the different cell types are in relation to the tumor and what percentage of those other cell types are in the tumor. You may be able to predict how a particular person might respond to therapy.

"Incorporating machine learning and AI for the analysis makes these tools powerful. The demand for genomics on campus—and the demand for computin g resources and support—will continue to go through the roof."

New Model of Community-owned Research Computing and Data Infrastructure

CRC is developing a community-owned research computing model, in which individually owned research computing clusters are centrally managed and shared among the community.

Researchers can become owners under the community model by purchasing capital equipment for computational resources that is integrated into the central cluster. There are two levels of access to the community-owned clusters: general and owner. General access is available to the University community on a first-come, first-served basis, and those users' jobs will be preempted when owners need the resource. Owners have the highest priority and are guaranteed on-demand access to resources that are proportional to their investment.

Pitt provides subsidies to keep computing resources affordable. These subsidies cover the costs of supporting infrastructure, power and cooling, data center space, system support, security monitoring, and hardware and software updates.

CRC leaders are available to consult on hardware purchases based on owners' needs, budget, and compatibility with CRC's ecosystem. CRC will order, install, and configure equipment purchased by the owners within a month of delivery to the Pitt data center. In addition to supporting computing hardware purchases, CRC can provide owners with access to the CRC data storage ecosystem.

Support from the CRC team is central to the community model. The team is committed to responding to all hardware and software issues within one business day, with an estimated resolution time, which may vary depending on the complexity of the issue and other factors.

- **Faculty and departments can buy into the CRC ecosystem with exclusive access to resources on demand.**
- **When the resources are not in use, they are offered to the community for high-throughput computing jobs.**
- **The model is a win-win for all involved:**
  - **CRC administers the hardware and the software stack.**
  - **Principal investigators get access on demand without waiting in the queue.**
  - **The broader community can leverage the idle resources when they are not in use.**

Improvements to User Training

In addition to the CRC on-ramp workshop offered every term, we continue to offer annual introductory workshops on Linux and Bash scripting. We offer a wide range of rotating workshops on topics such as R for data science, Python, and machine learning.

We are offering a new weekly one-hour hybrid office hour beginning Sept. 6, 2024. Rotating CRC team members will be available in person at our office in Schenley Place or virtually on Zoom. For the office hours, no help tickets or email are necessary to get immediate, live access to the deep experience of the CRC team. If one consultant doesn't have the answer, they will put the user directly in touch with a consultant who does.

Filling the Gap: A Dedicated Teaching Cluster

Deploying dedicated resources for teaching will guarantee quality service to both instructors and students alike. CRC has created a student-facing portal, JupyterHub, that we have integrated with powerful dedicated backend resources.

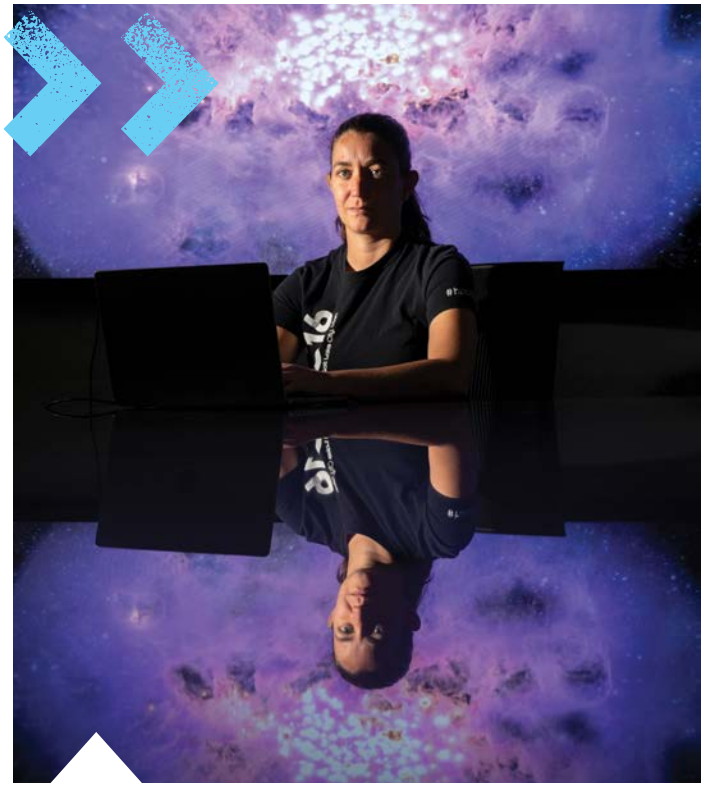
- 14x GPU nodes with 4x GTX 1080 per node
- 45x CPU nodes with dual-socket Intel Xeon Gold 6126 (Ice Lake); 24 cores/node; 192GB RAM

Researchers of Impact

Using AI to Study AI

Morgan Frank, assistant professor in the School of Computing and Information's Department of Informatics and Networked Systems, works with CRC resources to create maps to identify and quantify which skills thrive and decline in the face of technological change and which areas face the biggest impacts.

"The worst possibility is mass technology unemployment," he says. "We hope to prevent that by looking at the probabilities of automation in different jobs."



Modeling Galactic Formations

In 2022, **Evan Schneider**, assistant professor of physics and astronomy, became the third Pitt faculty member and first woman to be named a Packard fellow. Her hydrodynamics code, called Cholla, models galaxy evolution. Helped by CRC resources, she is working to ensure that her code can take advantage of—and keep up with—Frontier, the country's first and fastest exascale supercomputer, located at the Oak Ridge National Laboratory. Cholla will model star formation, supernova explosions, and other astrophysical phenomena to better understand galaxy formation, growth, and evolution.

SHREC Meets the Challenges of Space Computing

Computing in space is not like computing on Earth, according to **Alan George**, professor in and chair of the Department of Electrical and Computer Engineering in the Swanson School of Engineering, and founder and director of the National Science Foundation Center for Space, High-performance, and Resilient Computing (SHREC), led by Pitt. Challenges include resource constraints and environmental hazards—size, weight, power, severe shock, wide temperature swings, and ionizing radiation.

Graduate students in SHREC collaborate closely with CRC for unique computing resources exploring communication patterns, graph analytics problems, benchmarking systems, path planning, and machine learning. CRC offers hardware and, importantly, one-on-one support.

"Those applications cover graph processing, machine learning, [and] bioinformatics, among many other domains, to stress different parts of these novel systems," says Luke Kljucaricm. the graph analytics team's reconfigurable systems lead (he graduated this Spring). "[Kim] Wong, Sasha Kessler, and everyone at the Network Operations Center have provided us with tremendous support, enabling us to focus on our research and shorten development time."

User Survey Reveals Satisfaction and Areas to Improve

In early 2024, we circulated a survey to all CRC users asking how well we were serving them and how we could do better. The survey provided us with valuable feedback on our users' areas of satisfaction and dissatisfaction. Some areas of dissatisfaction were expected (more storage), and some were not (wait times). While our focus is on addressing the issues where we need improvement, we were also pleased to see that **more than 80%** of CRC users reported themselves as satisfied or highly satisfied with our overall service.

What Clusters They Use

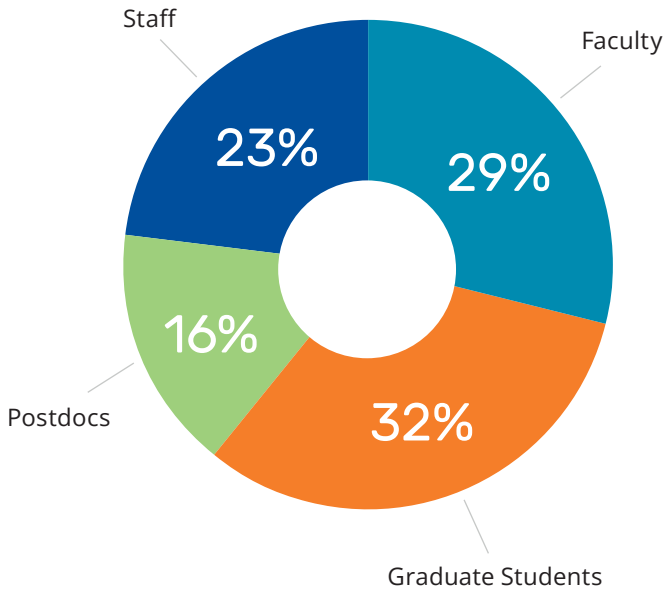
26% mpi

35% smp

55% htc

32% gpu

Our Users

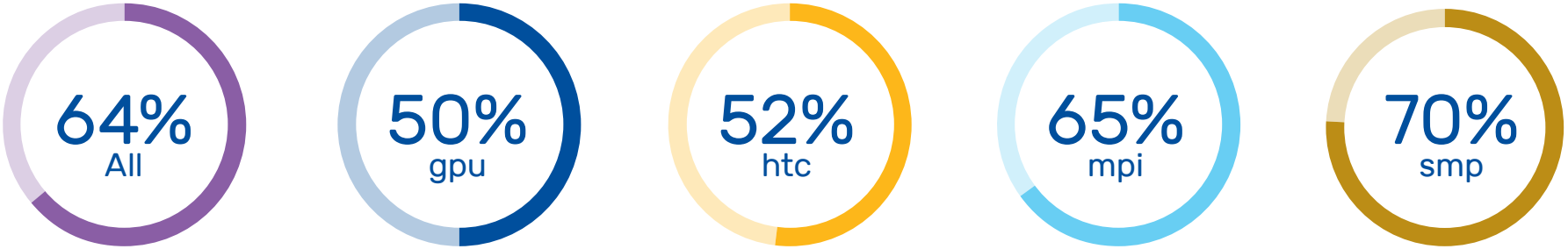


Are you satisfied with the overall service of CRC?

Responses are rated from 1-5, with 1 representing the lowest level of satisfaction and 5 the highest.

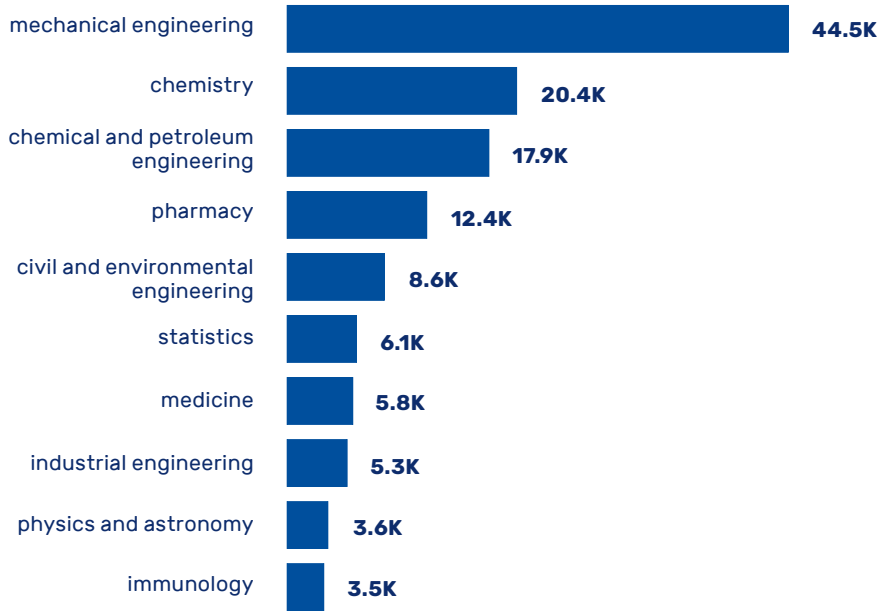
1	2	3	4	5
3%	3%	10%	29%	55%

Cluster Utilization and Top 10 Departments with the Most Core-hours



Responding to User Help Tickets

In February 2024, CRC switched out the help ticket system from our existing internal system to the TDX system maintained by Pitt IT, creating possibilities for streamlining and efficiency and relieving CRC consultants of the task of maintaining the system in addition to responding to the help tickets. The CRC consultant team resolved a total of 3,084 tickets from Jan. 1, 2023, to May 31, 2024, 2,000 on the old system and 1,084 on the TDX system.



(out of a total **114** departments)

University of Pittsburgh

Center for Research Computing

312 Schenley Place
4420 Bayard Street
Pittsburgh, PA 15260

412-648-3094

crc.pitt.edu

114680-0724