Research&Innovation

Center for Advanced Computing: Services

Rich Knepper, CAC Director, rich.knepper@cornell.edu, 607-255-0313 - Research Computing l Consulting l Training & Education

www.cac.cornell.edu

Accelerate Discovery

Do you need to achieve your research goals faster and more efficiently using AI/ML, Cloud Computing, or HPC?

AI/ML & Data-Intensive Computing

CAC services empower researchers to tackle complex AI/ML and data-intensive computing challenges, accelerating discovery. We facilitate access to and scaling on a range of computing resources, including HPC clusters, Cornell's Red Cloud, public clouds, and national computing resources.

For projects requiring custom high-performance computing clusters with GPUs and large memory, we architect, build, and maintain these systems at the University data center. Our high-capacity storage solutions and data management tools enable researchers to organize, store, and access research data efficiently.

AI/ML model preparation and data processing can be performed on Red Cloud CPUs/GPUs. For example, WCM's Shuibing Chen Laboratory develops on Red Cloud, including TensorFlow and PyTorch library testing, and then scales to the Cayuga Cluster.

We also assist researchers in utilizing public clouds and national resources such as Jetstream2, Frontera, and the planned NSF Leadership-Class Computing Facility supercomputer (LCCF). We were selected to provide national user training for the LCCF.

Our team of PhD scientists, computational, and data analysis consultants provide timely support, including AI/ML programming; the provisioning of instances; preprocessing, Extract Transform Load, and data injection; and, server, container, and library environment building.

Data visualization support is available for TensorBoard and Bokeh, Plotly, and Matplotlib. Colab is used as a Jupyter Notebook service.

Local & National Computing Resources

red cloud

- Create instances with up to 128 CPU cores and 240GB RAM that deploy in seconds.
- NVIDIA GPU instances are available for AI/ML testing.
- Adding H-100 GPUs.
- Ceph cluster with 1.9PB of raw capacity provides persistent disk storage, object storage, and a CephFS distributed file system.
- Red Cloud Exploratory Accounts are free.

Jetstream2

- CAC was selected as a regional cloud site for the Jetstream2 national cloud system.
- 1,024 cores and 768TB storage are available through an ACCESS allocation by selecting the Cornell zone in the Exosphere dashboard.
- CAC helps Cornell researchers get ACCESS allocations to other national resources and answers questions about usage.

Core Services

Services	Capability
HPC Cluster Computing	Architect, build, and maintain HPC CPU/GPU Clusters
Red Cloud Computing	More than 3,000 CPUs plus NVIDIA H100, T4, V100, and A100 GPUs
Data Storage	Ceph, AWS S3 object store, archival
AWS, Google, Azure Computing	A to Z fast start, instance provisioning
National Computing	Help secure ACCESS allocations and facilitate use of national resources
AI/ML Programming	TensorFlow, PyTorch, Keras, JSL Healthcare AI, CryoSPARC, crewAI, Python, R, Java, C++, MATLAB, Shell
Large Language Models	Fine-tune local/publicly-hosted LLMs
Researcher Training & Education	Scientific Computing Series, Cornell Virtual Workshops, YouTube channel, eCornell certificates, guest lectures
Proposal Development	Strategy sessions, SWOT analysis, writing/editing, data management plans

Engagement Process

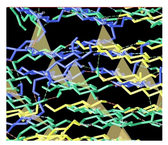
Analyze

CAC professionals work closely with Cornell faculty and researchers to analyze project requirements.

Plan

Faculty or research staff select levels of computing, storage, and/or hourly consulting services desired.

Deliver



Services delivered. CAC consulting in regular contact with PI or research team. Accounting available online.

Maintain

Quality, turnkey maintenance available for HPC clusters, storage systems, DBs, and workflows.

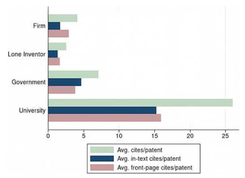
No Fee & Fee-Based Services

- Researchers who wish to try out Red Cloud may request a free exploratory account that includes 1 hour of consulting.
- Other professional computing and consulting services are fee-based. A Cornell account is required for chargeback.
- No fees are required for an initial requirements meeting; access to high-speed networks from CAC-supported services; up to 20 hours of consulting to architect an AI/ML/HPC cluster or storage system to be housed in the University's data center and maintained by our staff; training opportunities; and, proposal development services.

AI/ML Project Examples

AI in Veterinary Medicine – Parminder Basran has a keen interest in ML methods in radiation oncology. CAC prepared scripts and demoed how MATLAB PCT works on a local machine and Red Cloud, and provided workflow integration advice.

ML in Chemistry and Chemical Biology – Robert A. DiStasio, Jr. runs simulations and ML on molecular properties and chemical reactions using the POOL Cluster built and maintained by CAC.



ML at Dyson – Matt Marx uses CAC systems to link patents to academic articles to understand the scientific heritage of innovation. Hand-tuned heuristics and the GROBID ML package were combined to achieve higher performance than ML alone.

Research Grants

Awards/Subawards/Support	Role
Chishiki-AI	Build AI Tutor to deliver scalable learning via the Cornell Virtual Workshop platform for the national Civil Eng. community
HPC-ED	Pilot the discovery and sharing of training across a multitude of users and providers
TACC Frontera & LCCF	Training partner for Frontera and planned NSF Leadership-Class Computing Facility
Jetstream2 Cloud System	Jetstream2 Regional Cloud at Cornell
I-WRF Atmospheric Research	Develop an integrated, multi-node Weather Research & Forecasting Model and verification container framework
IRIS-HEP Physics at Info Frontier	Optimized parallel codes for the analysis of CERN detector data
SCiMMA	Chief architect for the Multi-Messenger Astrophysics framework
NANOGrav Physics Center	Design infrastructure, software, and data workflows to detect gravitational waves

AI/ML Publications

CAC staff author papers on AI research and tools such as variational autoencoders to represent and interpret high-dimensional datasets and automated modeling using ML and Deep Learning.