AN OVERVIEW OF RCAC COMMUNITY CLUSTERS AND STORAGE RESOURCES

Amiya K Maji

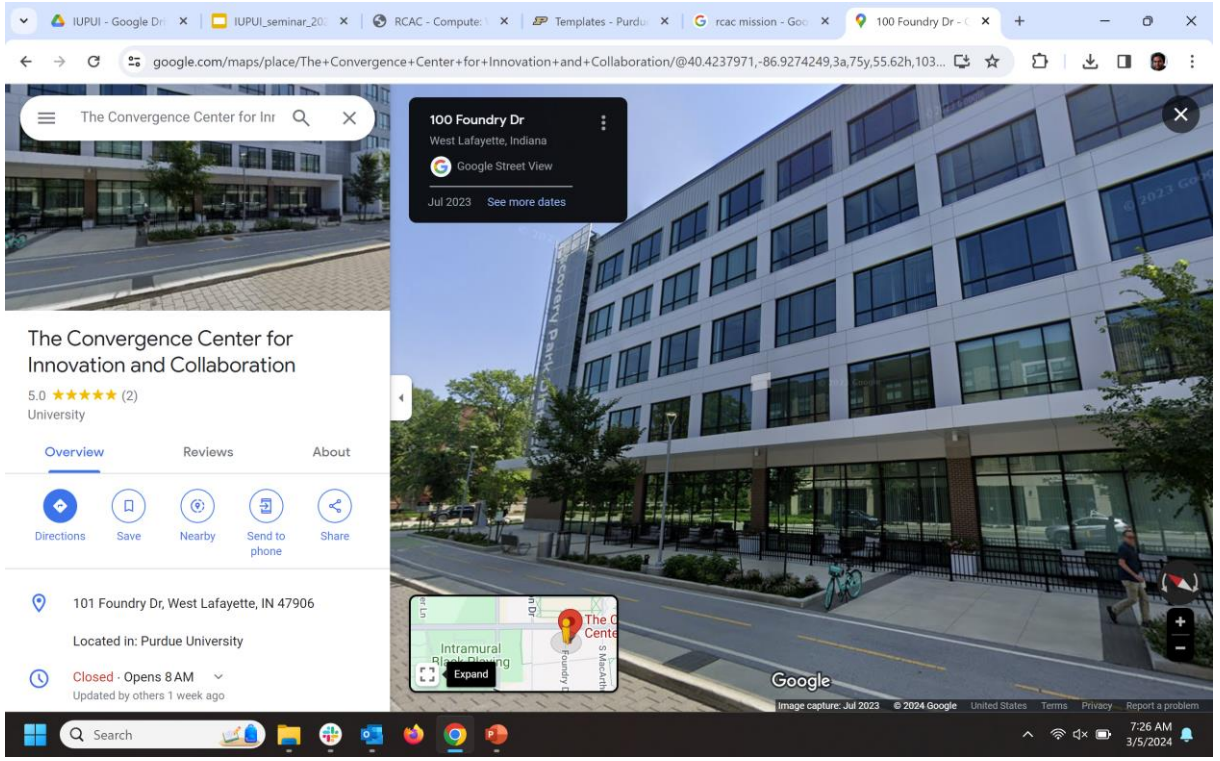
Purdue Indianapolis Seminar

Mar 5, 2024

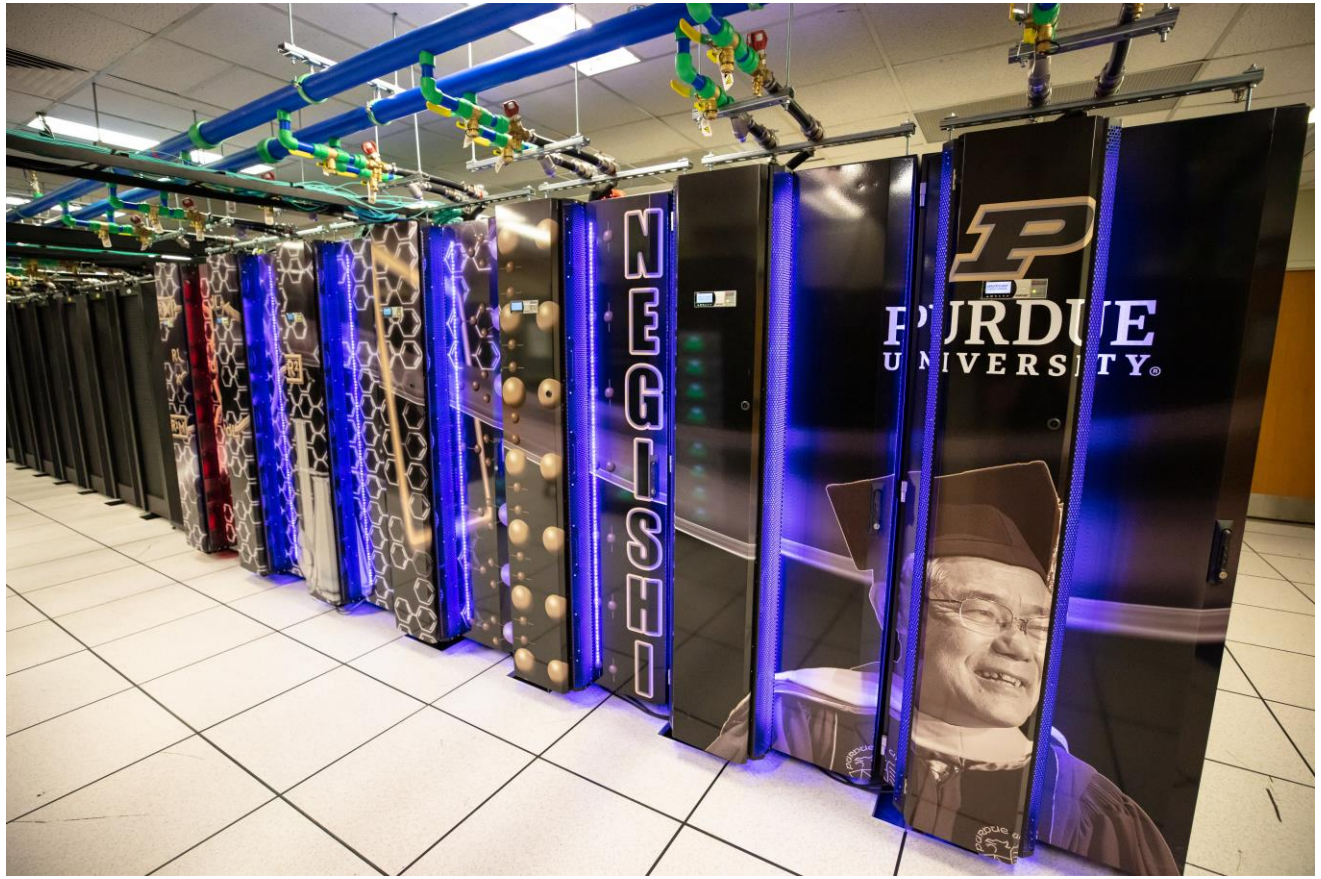
PURDUE UNIVERSITY | Rosen Center for Advanced Computing

Rosen Center for Advanced Computing

- To support and advance research discoveries at Purdue through partnerships with faculty and research groups.



Negishi Ribbon Cutting



RCAC Services

- **Compute**
  - Community clusters
  - Kubernetes cloud
  - Anvil
- **Storage**
  - Network storage
  - Archive
- **Visualization**
  - AR, VR, XR
  - Digital twins
- **Grant collaboration**
  - Grant preparation
  - Software development
  - Data science
- **Training**

Community Cluster

- Faculty A needs 10 x 64-core nodes
- Faculty B needs 5 x 64-core nodes
- Faculty C needs 2 x 64-core nodes
- … … …
- Build a 100-node cluster for all the faculties
  - Ease of maintenance
  - Cost effective
  - Node failures do not lead to work stoppage
  - Use additional burst capacity when others are not using their nodes
  - Faculties buy “shares” on the cluster

Community cluster

- Each PI and their students get a dedicated “queue” for submitting jobs
- Shared “standby” queue for burst capacity
  - Limited to 4 hours per job
- Students can be added/removed from self-service website
- Cluster storage
  - Home: 25GB per user (persistent)
  - Scratch: 200TB per user (temporary)
  - Depot: On Demand (persistent)
- All community clusters use Linux and the Slurm scheduler

Current Community Clusters

- Negishi
  - 450-node CPU cluster for computational research
- Gilbreth
  - Nvidia GPU cluster for accelerated applications (DL, MD, Life Sciences)
- Scholar
  - Teaching cluster for classroom use
  - Free
- Bell
  - Sold out
- Hammer
  - High-energy physics (CMS)
- Anvil
  - NSF-funded
  - Allocated via ACCESS

Technical Specifications

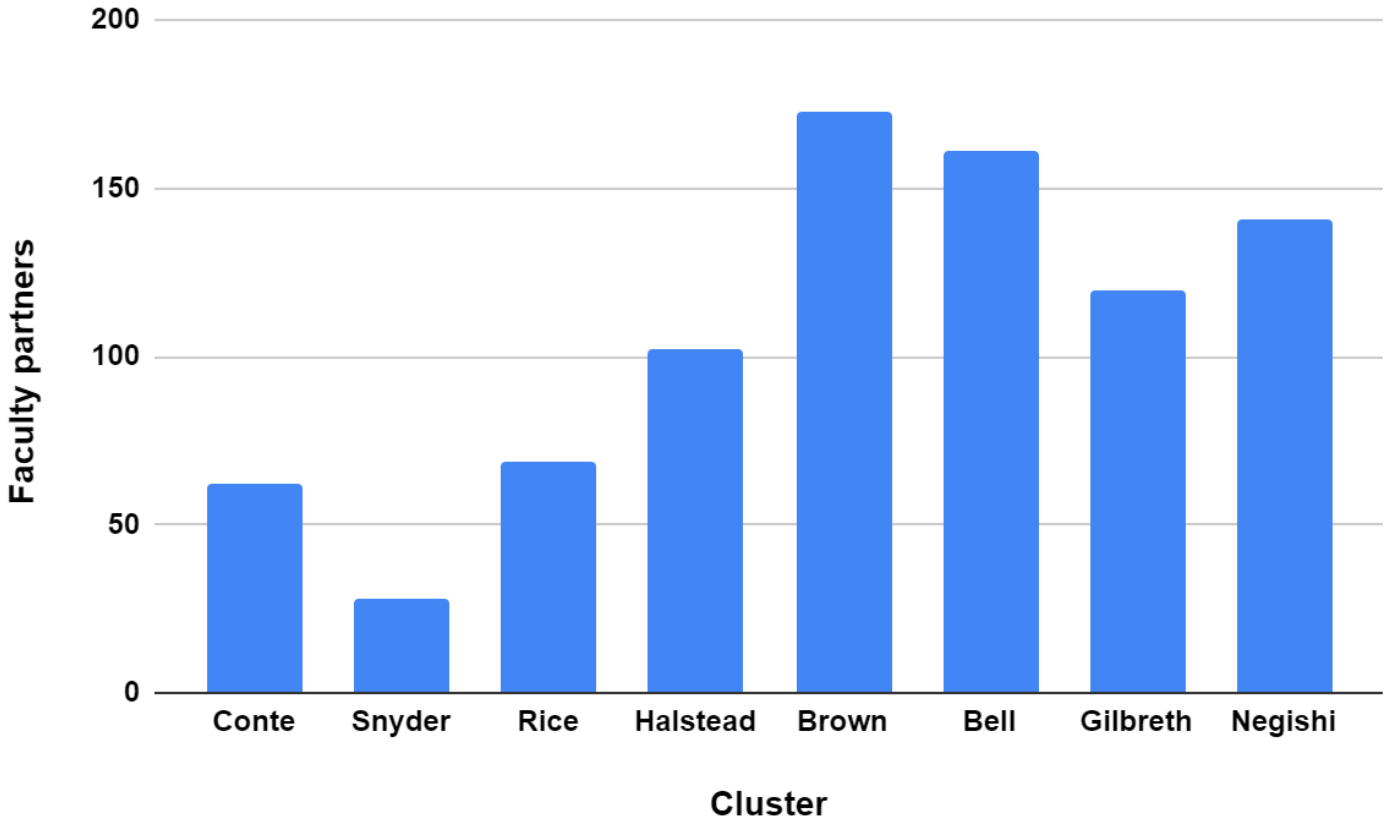
Negishi

- 450+ nodes
  - 2 x 64-core AMD Milan processors
  - 256 GB memory
- 100 Gbps infiniband interconnect
- 6 x 1TB nodes
- 5 x 3 AMD MI210 GPUs
- 6.7 PB scratch storage

Gilbreth

- Heterogeneous cluster
- 100 Gbps infiniband interconnect
- 4.5 PB scratch storage
- Generations of Nvidia GPUs
  - V100
  - A100
  - H100
  - A10
  - A30

Community Cluster Usage



Scholar

- Dedicated HPC cluster for teaching
- Free for all Purdue faculties
- GPUs are available for classes
- Most popular use-cases
  - OnDemand Desktop
  - JupyterHub
  - Rstudio
  - Bioinfo tools

Scientific Applications

- Compilers: GCC, Intel, AMD, Nvidia
- MPI libraries: OpenMPI, Intel, MVAPICH2
- Numerical libraries
- Data formats
- Popular applications
  - Chemistry, Physics, Biology, Statistics, etc.
  - ~280 application modules
- 600+ biocontainers
- Most engineering applications
  - Matlab, Ansys, Abaqus, Tecplot, Comsol, etc.
- https://www.rcac.purdue.edu/knowledge/applications

Storage Services

- Per cluster storage services
  - Home
  - Scratch
- Global
  - Data Depot
- Archival storage
  - Fortress tape archive
  - Free

Data Depot

- 6.5 PB GPFS file system
- Shared workspace for research groups
- **Data owned by faculty/PI**
- **Fault tolerant**
  - All data duplicated at independent “sites”
- Regular snapshots for recovering old files
- Accessible from all clusters
- Use Globus for bulk data transfers
- Can be mounted as a network drive on laptop
- $70/TB per year

Fortress Data Archive

- 200 PB HPSS tape archive
- Free for all Purdue researchers
- Practically unlimited storage
- Not for interactive work
- Data can be transferred using Globus or HSI/HTAR

Regulated Data and Compute

- Export controlled cluster
  - Weber
- Cloud storage for sensitive data
  - Box
- Protected health data
  - dbGaP compliant storage attached to Negishi

Data Publication

- Purdue University Research Repository (PURR)
  - Managed by Purdue Libraries
  - https://purr.purdue.edu/
  - Assistance for Data Management Plan

Grant Application Resources

- **RCAC facilities document**
  - https://docs.lib.purdue.edu/gendes/4/
- **Data management plan**
  - Visit https://purr.purdue.edu
  - purr@purdue.edu
- **Computation and storage**
  - rcac-help@purdue.edu
- **Software development**
  - rcac-help@purdue.edu
- **Visualization**
  - envision@purdue.edu

Community Cluster Pricing

As of March 2024

- **Negishi**
  - $3360 for access to 64-cores
  - End of Life: Fall 2028
- **Gilbreth**
  - One A100 GPU: $2200/year, $7000/5yr
  - One A10 GPU: $1400/year, $4000/5yr
  - One A30 GPU: $1700/year, $5000/5yr
  - One H100 GPU: $27000/5yr
- RCAC will provide complimentary access for Purdue Indianapolis faculties transitioning from IU clusters

On the Horizon

- 2024 Community Cluster
  - Gautschi (named after Purdue CS faculty)
  - Arriving in Fall 2024
- More GPUs …
- Experimental systems
  - Grace-Hopper

Resources

- RCAC website: www.rcac.purdue.edu
- Cluster user guides: https://www.rcac.purdue.edu/knowledge
- Trainings: https://www.rcac.purdue.edu/training
- Coffee hour consultations: https://www.rcac.purdue.edu/coffee
- Purchase: https://www.rcac.purdue.edu/purchase
- User management: https://www.rcac.purdue.edu/account/groups
- PURR: https://purr.purdue.edu
- RCAC facilities document: https://docs.lib.purdue.edu/gendes/4/

Contacts

- **Amiya Maji**
  - Computer Science
  - amaji@purdue.edu
- **Sarah Rodenbeck**
  - Data Science
  - srodenb@purdue.edu
- **Ryan DeRue**
  - Computer Science
  - rderue@purdue.edu
- **RCAC user support**
  - rcac-help@purdue.edu

Questions

THANK YOU