Research Technologies, Computing, and Library services

June 9, 2022

NORTHEASTERN

Agenda

- Introduction to Research Technologies
- Research Computing and the Discovery cluster
- Storage on Discovery
- Library services – Storage options
- Library services – Data management
- Security

Who is Research Technologies?

- Northeastern's Research Technologies (RT) team is a cross-functional technology support group
- Provide assistance to academic research groups and research services
- Support through guidance on hardware, software, workflows, and secure data solutions.
- Our mission is to offers support for current technology needs and provide consultations for enhanced technology integration to expand the resources available to you at the university.

Research Application Support Team

11 new dedicated roles over the past 10 months:

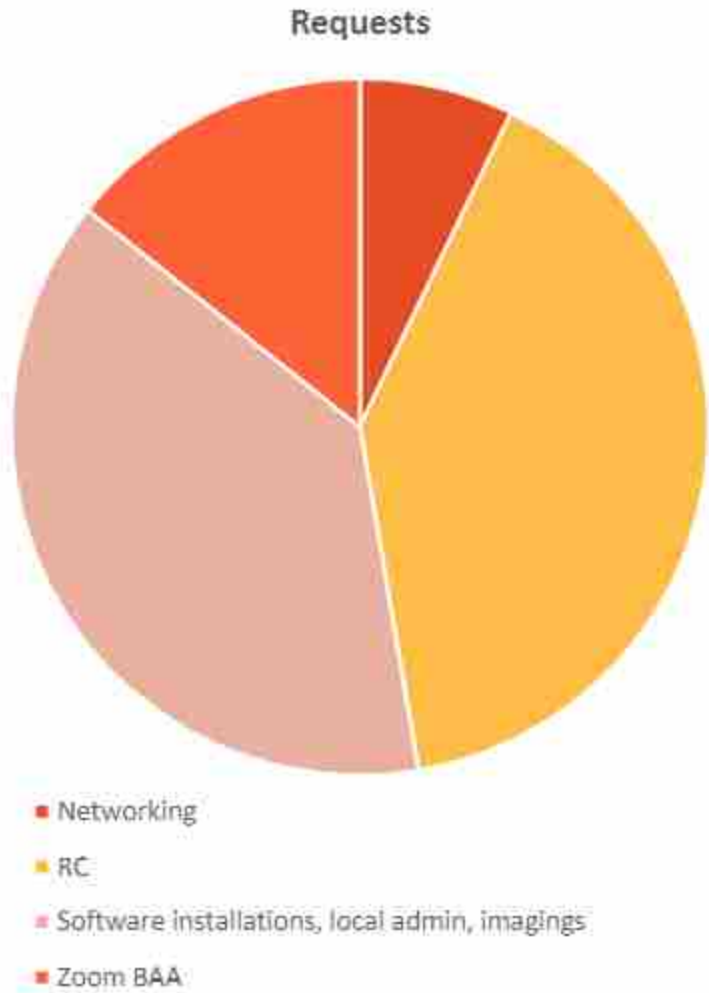
- Security
- Developer-Operations
- CMMC
- Networking
- Customer Experience
- Research Computing

Research support offered through:

- Consultations
- Training/Documentation
- Compliance
- Data Integrity/Backup/Encryption
- University/NU-RES Standards

Our Day to Day

1. CX
   - Lab equipment troubleshooting
   - HIPAA Zoom
   - Consultations (storage)
   - Onboarding
   - Software installations
2. Research Computing
   - Discovery
   - MGHPCC (server)
   - Globus
   - Secure data enclave
   - Software installations
3. Networking
   - Static IP
   - MGHPCC
4. Enterprise Applications and Support
   - Github etc
   - ServiceNow
   - Epaws and eclaws

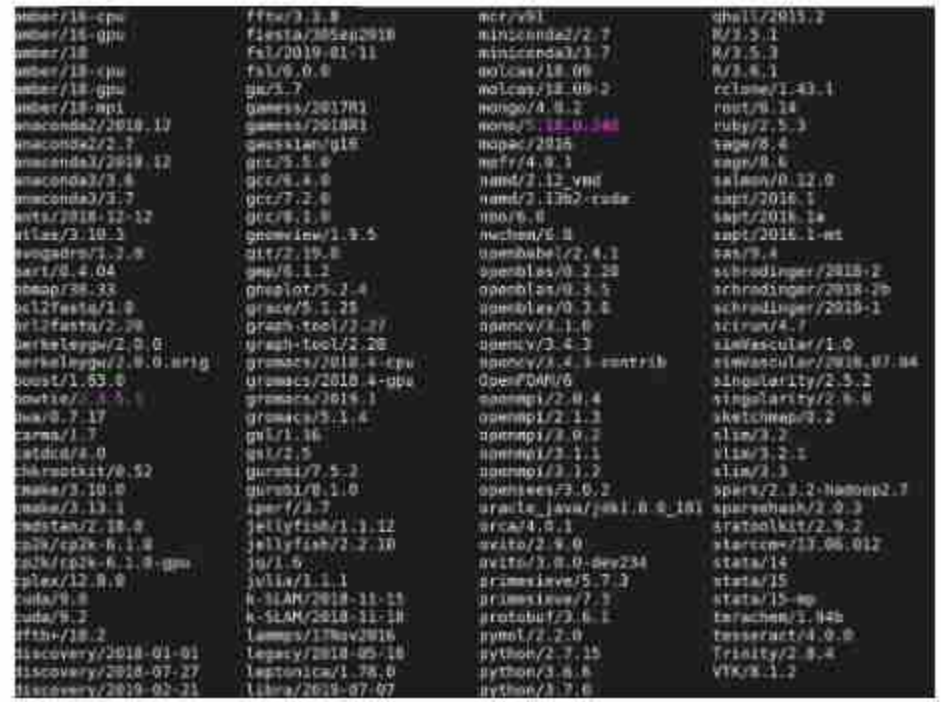


Discovery Cluster

- Discovery Cluster in MGHPCC (Holyoke, MA).
- 25k+ public CPU cores, over 200 GPUs, & 1200 nodes.
- Connected to Northeastern via 10 Gbps Ethernet (GbE) for high-speed data transfer.
- Performant, archive, encrypted, and cloud storage solutions.
- Discovery Hardware

Software on Discovery

- Popular software packages available as modules.
- Install software to local directory.
- Request software installation through ServiceNow.
- Discovery software
- Discovery shell environment

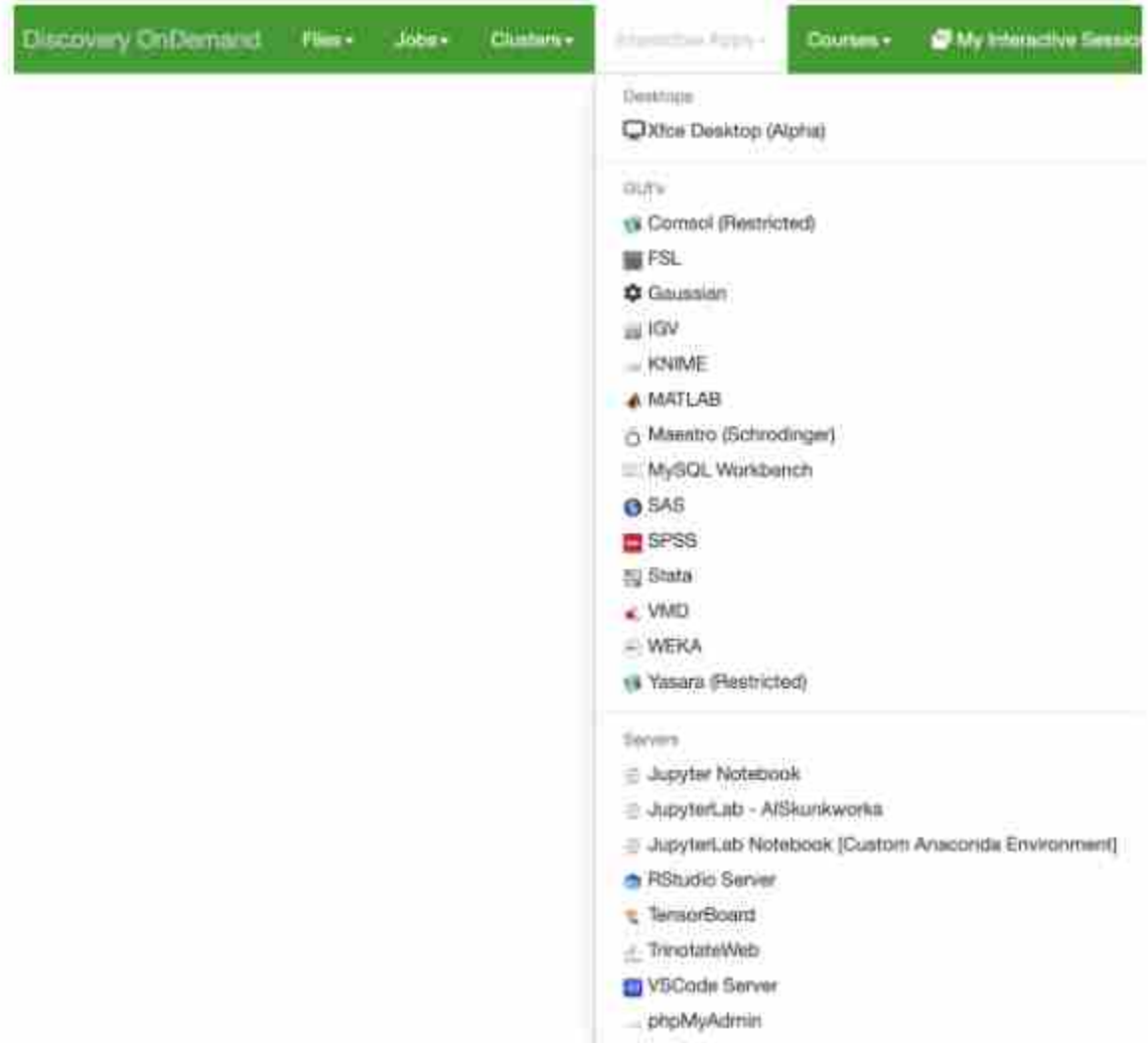


GPUs on Discovery

- All NVIDIA GPUs on Discovery.
- Range from V100 to A100.
- Access gpu partition using ***--gres=*** flag while running jobs interactively (**srun)** or in the background (**sbatch)**.
- GPUs on Discovery

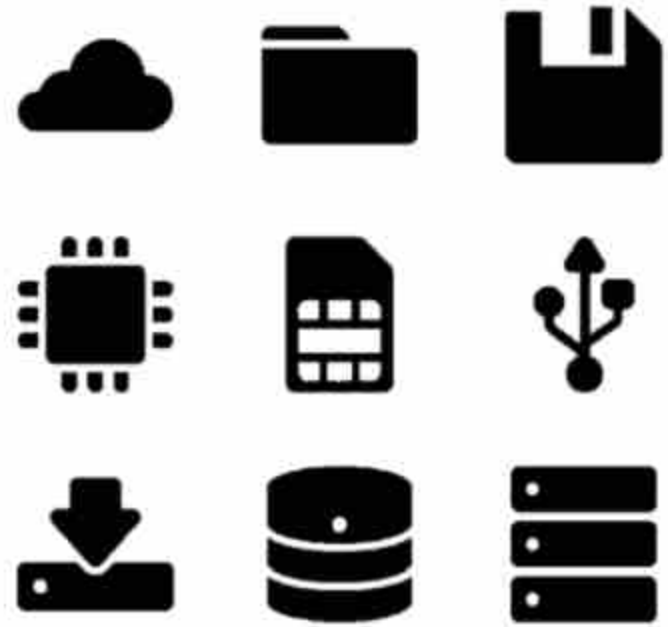
Accessing Discovery Through OOD

- Open OnDemand (OOD) – Interactive HPC via the Web
- Web portal to Discovery cluster.
- Various Interactive Apps available including Jupyter, JupyterLab, VSCode, etc.
- ***Any OS***: Use a web browser and go to
- File Explorer – Manage files on Discovery using OOD



Storage on Discovery

- All storage systems secure
- Performant – NU's HPC cluster
- Encrypted – centralized service for faculty & researchers working with sensitive data with encryption
- Archival – NESE (RC) & DRS (Library)
- Free quotas across all tiers for RC systems (including NESE) – 35 TB
- One Drive (O365) - 5 TB



Storage on Discovery

Home

- Permanent & Backed up
- Limited size (75 GB)
- Not performant

Scratch

- Performant
- Total capacity: 3 PB
- Temporary storage
- Not backed up
- 28-day purge policy

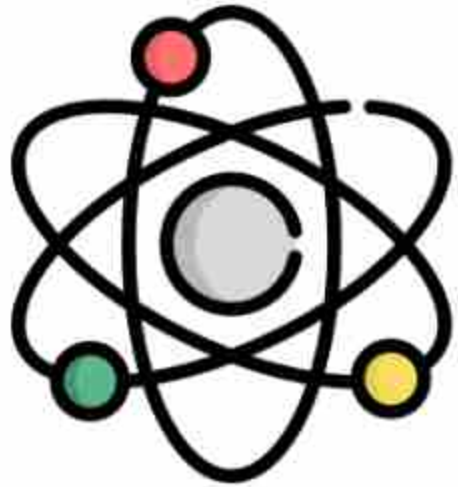
Data Transfer on Discovery

MAC	Using Terminal
	Using SSHFS
	Using Globus (recommended for transferring large files)
Windows	Using MobaXterm
	Using FileZilla
	Using Globus (recommended for transferring large files)

- Transferring data on Discovery
- Details about Globus are hyperlinked here

Research Computing (RC) Resources

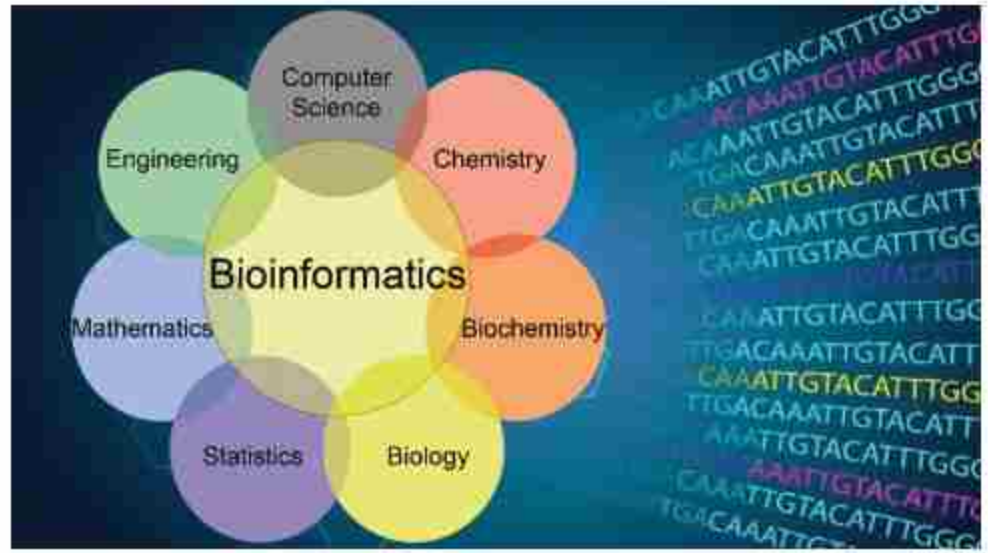
OPEN OnDemand



Bioinformatics, Computational Science, & Data Science

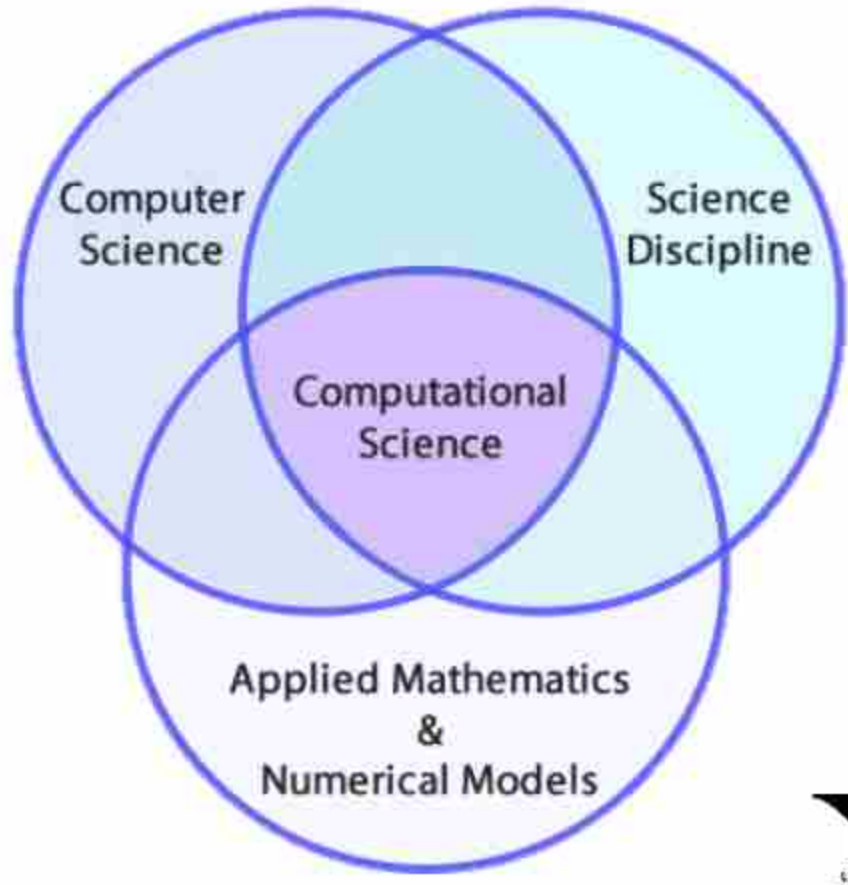
Bioinformatics

- Next generation sequencing analysis pipelines
  - Whole Genome seq
  - Bulk RNA-seq
  - Single cell RNA-seq
- Workflow frameworks (e.g., Nextflow)
- Bioinformatics containers
- Software Installation (bioconda, system wide modules)
- Custom scripts and optimization
- Research & Collaboration



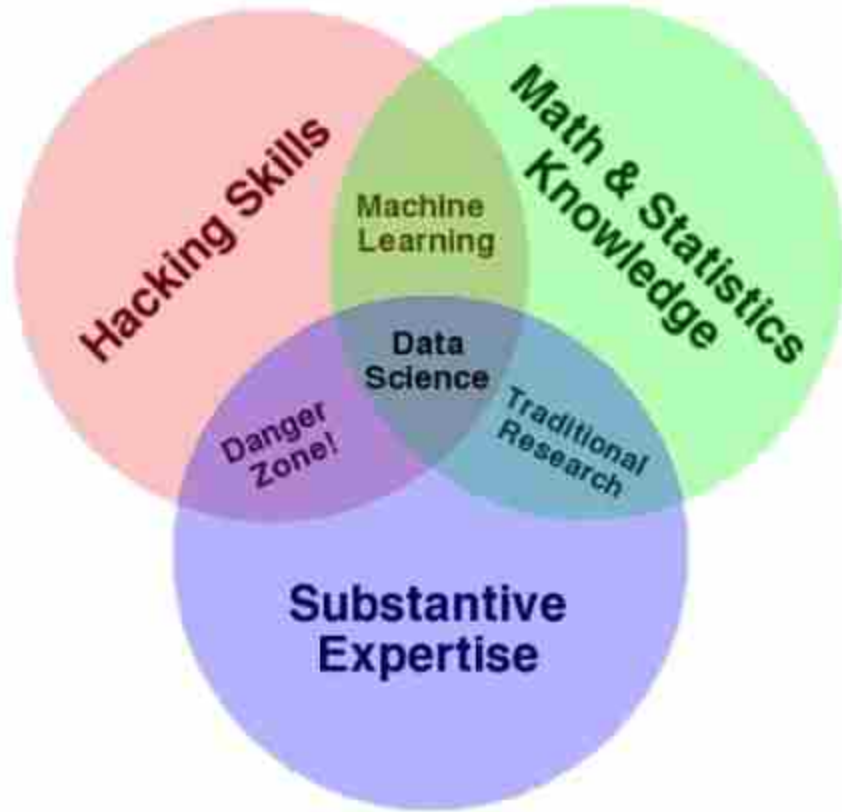
Computational Science

- Parallelization
- Benchmarking, scaling, & performance tuning
- Code Facilitation & Optimization
- Resilient Workflow & Checkpointing
- Computational Physics & Chemistry workflows
- Research & Collaboration



Data Science

- ML Platforms – KNIME, WEKA, RAPIDS
- Debugging Tools – PyCharm, VSCode
- Software packages – Jupyter notebook, PyTorch, Tensorflow, Plotly
- ML datasets on shared location
- Code Facilitation, Parallelization, & Optimization
- Research & Collaboration



Library Storage and Data Management Options for Researchers

- The Digital Repository Service (DRS)
- Data Management and Sharing Plans (DMPs)

Northeastern University

What is the Digital Repository Service?

The DRS was developed by Northeastern University Library as a tool for University faculty and staff to

- protect the valuable information and data that has been created as part of the University's research and instructional mission.
- provides long-term security for the files it stores to help ensure that data is as accessible and usable in the present and the future.

Digital Repository Service (DRS)

- Store datasets under 1TB per file
- Compatible with any file type
- DRS users can deposit files up to 1GB unaided. Larger files require library staff assistance
- Files may be restricted to NU only or made available to the public

Data suitable for DRS

Data that is fixed or in a publishable state

Data that does not contain personal identity (PII), health (PHI), or financial (PFI) information

Data that does not require encryption

Data that is in compressed format packages

Data that has documentation, like a README or other guide to the data

How can I add my work to the DRS?

NORTHEASTERN FACULTY AND STAFF CAN SIGN IN AND START DEPOSITING RIGHT AWAY

MEMBERS OF THE NORTHEASTERN COMMUNITY CAN CONTACT DRS STAFF TO GET PERMISSION TO START DEPOSITING ON BEHALF OF THEIR DEPARTMENT OR RESEARCH GROUP

LIBRARY STAFF CAN PROCESS AND DEPOSIT BATCHES OF MORE THAN 20 FILES ON BEHALF OF NORTHEASTERN FACULTY, STAFF, AND STUDENTS

Data management & sharing plans

Also known as DMPs and DMSPs

Plans for managing data during a research project

Plans for ensuring data remains accessible and usable once the project is over

Why manage your research data?

- To fulfill grant requirements
- To organize large sets of data
- To preserve data
- To make data accessible and usable after project completion
- To improve project efficiency

DMPTool

DMPTool a free, open-source, online application that helps researchers create data management plans (DMPs).

- DMP and DMSP proposal templates for funding
- Many publicly available resources to know the funders are looking to fund which helps with the proposal

Learn more and get help with DMPs and DMSPs

Contact Jen Ferguson

- j.ferguson@northeastern.edu
- or through the Data Management for Research page
  - Self-service guides materials
  - Direct link to DMPTool
  - Contact info and appointment scheduler

Security – Who We Are

The Office of Information Security (OIS) is committed to protecting university data and personal privacy. Researchers share these same goals, and by working together we can implement appropriate safeguards for your project.

The Research Technologies Security Analyst works in OIS and has access to a wealth of security subject matter experts.

For more information about OIS, visit security.its.northeastern.edu

Security – Why it Matters

				
Increased exposure on research universities	Sharing vs Security	Multitude of systems, devices, and people	Phishing, Ransomware	Remote/Virtual Collaboration
				
Data Sensitivity	Legal Compliance	Contractual Obligations	University Policy	Accountability

Security – Partnering with You

- Assisting with security questionnaires or other necessary requirements to access data
- Meeting with you and your data provider to implement a security plan
- Coordinating with Northeastern teams (NU RES, OGC, RC, CX) to streamline the experience
- Suggesting compliant tools and/or environments for your specific project
- Developing a security/data management plan together
- Providing secure architecture and tools for your projects
- Collaborative Projects such as CMMC certification and the NU Data Classification Policy
- Standard and customized training
- Interactive events such as Data Destruction Day and Cybersecurity Awareness Month

Security - Services

Consulting
- Security assessment of research agreements
- Questions, incorporating security, best practices
- Build/review data management plan

Training
- Security Awareness and Data Classification
- Specific compliance training
- Custom training

Documentation
- Knowledge articles and communication
- University policies
- Security by NU data classification

Our Contact Information

researchtechnologies@northeastern.edu