Agenda

- The Ecosystem for Research Networking (ERN)

- ERN Broadening the Reach (ERN BTR) Working Group

- Leveraging the Cloud for Research: Findings and Recommendations

- Opportunities

Ecosystem for Research Networking (ERN)
(formerly the Eastern Regional Network)

ERN Vision: Simplify, support, catalyze, and foster multi-campus collaborations and partnerships between academic institutions of all types and sizes across the U.S. that advance the frontiers of research, pedagogy, and innovation.

Mission: To achieve the vision through a consortium of academic institutions, research facilities, core service providers, network providers, and industry partners, both public and private, **organized around a shared interest in supporting and enabling collaborative data and computation-enabled science by providing standards, blueprints, policies, and training associated with the design and implementation of an infrastructure to access data and research instruments,** a distributed federated environment designed to simplify, support, and encourage collaborative science, scholarship, and education.

To realize the mission and vision, ERN will enable collaborations for democratization of access to research instruments, technical expertise, infrastructure, services, and resources to lower barriers to participation for scientists engaged in collaborative research across institutional and disciplinary boundaries.

ERN Broadening The Reach (BTR) Working Group - Goals

- Engage/support small/medium under-resourced campuses: MSIs, HSIs, HBCUs, EPSCoR
- Learn directly from the institutions in the region on how best to support them
- Explore the role of regionals as facilitator and user support for these smaller institutions within the ERN
- Build & leveraging a highly skilled, diverse workforce to support advanced CI
- Identify potential collaboration opportunities
- Develop white papers to be presented at conferences and used as input to other potential funding opportunities

ERN BTR Working Group - Activities

- Working group meets monthly (we welcome your participation)
- A series of workshops
  - Identify the needs of the community
  - Raise awareness to existing regional and national resources, and funding opportunities
  - Identify existing best practices and models
- Share findings among the community at conferences and through publications
- Recommendations for Funding Agencies, ERN, and Institutions
- Research Collaboration Opportunities

Workshop - Leveraging the Cloud for Research December 2021

A workshop to bring together the community of researchers and information technology community representing R1's through the small/medium under- resourced campuses, to learn directly from them challenges and opportunities in leveraging the cloud for research and how the ERN can support them

- Goals
  - Introduction to the Cloud
  - Challenges and opportunities in leveraging the Cloud in research projects
    - Featured the NSF-funded Exploring Clouds for Acceleration of Science (E-CAS) projects
  - Institutional support and on-boarding for research in the cloud
  - Tools, resources, and training for cloud
- One-day workshop included two sessions and six presentations
- Over 140 registrants, 88 attendees and 45 schools represented across the US and beyond. In addition to academic institutions, attendees also represented RENS, National Labs, and Industry members
- Pre-workshop survey provided a baseline of the issues most important to the community

Traditional (On-premise) High Performance Computing (HPC)

Challenges

- High cost of acquisition

- Lack of specialized expertise to support equipment

- Researcher and graduate student valuable time spent on support

- In many cases, no appropriate place to house

- One-time grant-funded equipment becomes obsolete

- Institutional support unavailable to sustain support

- Without administrative and technical support, departmental resources are under-utilized

Power of the Cloud

- Unlimited access to computing resources for research creating a more leveling playing field
- Easy access to data and computation at arbitrary scale
- Power to democratize research by making computing power available to vast majority of researchers
- Address massive challenge of sharing, mining, searching, and analyzing large datasets in real-time with collaborators across disciplines and in disparate locations
- Scales and provision resources quickly as needed, run very large simulations that use thousands of cores
- Ability to use pre-built models and functions without technical knowledge

Cloud-based HPC Challenges

- Funding
  - Cloud-based HPC requires ongoing financial support
- Access to expertise
- Network bandwidth
- Complexity of cloud management and integration
- Lack of adequate training and understanding of how to use the cloud
- Information security considerations

NSF-Funded Exploring the Cloud for Acceleration of Science (E-CAS) Program

Why ECAS

- Illustrate the viability of cloud platforms to do academic research
- Identify gaps
- Document the findings for the research community

Acceleration of Science:

- Achieve the best time-to-solution for scientific applications and workflows using cloud computing
- Number of concurrent simulations or workflows
- The ability to process near real-time streaming data

Innovation:

- Explore the use of heterogeneous hardware resources such as CPUs, GPUs, and FPGAs to support and extend application workflows
- Integrate and optimize the cloud services for research

NSF-Funded E-CAS
Accelerating the Research and Technology

- **E-CAS - Deciphering the brain's neural code through large-scale simulations of cortical circuits Project** (SUNY: Salvador Dura-Bernal)
    - Researchers were awarded significant cloud credits access to unlimited number of nodes
    - No queues
    - Ability to scale up resources as necessary
    - A fully customized SLURM Cluster
    - Kubernetes is used to deploy the tool for the wider community
    - Developed AI methods along with improved brain modeling tools and brain simulation workflows
    - E-CAS accelerated the research and technology - "Resulted in large scale models of areas of the brain that would have taken years!"

NSF-Funded E-CAS
Accelerating the Research and Technology

- **E-CAS –Particle Physics at the Largest Scale in Google Cloud Project**
  - Modern particle physics relies on powerful, large scale computing resources, with increasing emphasis on co-processors, such as GPUs.
    - Resources often expensive and inefficient
  - Cloud provides flexible, customizable, and scalable access to these resources
- Challenges
  - Access to expertise
  - Managing the cloud and the integration complexity
  - Network bandwidth

A Promising Strategy - Hybrid Environments

- Use both local resources and cloud resources as appropriate
- Cloud as an R&D testbed for groups that may be planning to purchase larger-scale, on-premises resources
- Tools such as Jupyter Notebook, Docker, and Kubernetes, make executing the same code in the local environment and in the cloud easy
- Researchers can develop and test their code using in-house equipment, and run final experiments in the cloud for access to additional processor cores and memory

Institutional Perspective

- Todd Raeker (University of Michigan), Brian Pasquini (University of Pittsburgh), Jeff Whitworth (UNC Greensboro)
- Most researchers use the cloud on their own with no centralized approach
- Large scalability requires institutional backing and negotiation on restrictions, and proper elasticity and cost management
- Some institutions are working proactively to organize the approach across the institution
  - Changing business operating models to be more agile, innovative, and adaptive through cloud tools and services

Institutional Perspective (cont'd)

Challenges

- Budget and financial

- Hidden costs associated with the cloud

- Cybersecurity management

- Procurement

- Network bandwidth

- Integration complexity

- Skilled workforce

- Lack of standardization for cloud management tools and configuration

- Data governance and management

Services, Training, and Tools

- Commercial cloud providers offer sophisticated tools for estimating and controlling costs, as well as training, and simplified interfaces
- External cloud service providers are being used for managing services
- Services for improving network access to the big three public cloud providers
- RENS can be leveraged to build relationships with cloud providers and cloud commercial partner solutions
- Tools and resources for navigating the cloud from organizations such as Internet2 and Omnibond
- Training programs available for the enterprise, however, limited options for the research community

  – Costs prohibitive for less-resourced institutions
- Cloudbank (NSF) and NIH STRIDES

Research Community is Relatively New to the Cloud

- Best practices are needed for getting started with the cloud
- Guidelines and templates are needed on how institutions are handling cloud security policy issues and procedures
- Grant overhead for cloud services continues to be a challenge
- **CloudBank** provides managed services to simplify cloud access for research and education, and resources are free of indirect costs for a subset of NSF solicitations that are CloudBank-eligible
- Gradual adoption of the technology through hybrid models, keeping data stored in a local environment and using the cloud for computation

Implementation of the Cloud
Critical Success Factors

- Data security
- Availability and reliability
- Customizable service level agreement
- Network bandwidth
- Compatibility
- Technical support
- Management support
- Human and resource readiness
- Managing complexity
- Cost flexibility
- Ease of use

Recommendations

- Recommendation to Funding Agencies
  - Expand funding support for programs to lower the barriers for cloud adoption
  - Investments in workforce development projects that address the expertise gaps
- Recommendations to institutions
  - Support participation of the campus community in conferences, and seminars on cloud computing. Organize information sharing through seminars and lunch and learns on cloud computing.
  - Establish and document institutional policy guidelines for the cloud implementation.
  - Provide cloud migration consultation services to the research community to help overcome obstacles and make their path more about science and less about purchase orders.
- Recommendations to the ERN
  - Design a framework for leveraging the cloud that includes templates for plans, programs, policies, guidelines, and best practices to address and overcome the major challenges in cloud adoption, while balancing the requirements of the enterprise and the research community.

Summary: Work-in-Progress

Lowering the barrier for providing access to advanced cyberinfrastructure and for leveraging the cloud in research is complex and requires:

- Funding support
- Campus cultural transformation
- Standards
- Policy requirements and guidelines
- Ease of use
- Expertise
- Communities for information sharing and collaboration
- Education and Outreach, including awareness campaign and professional development

Acknowledgements

ERN Broadening The Reach
Working Group Members

- Michael Ahn
- Galen Collier
- Maureen Dougherty
- Scott Frees
- Forough Ghahramani
- John Goodhue
- Ventsi Gotov
- John Hicks
- John Huffman
- Jim Kyriannis
- Derek Leydig

- Dave Marble
- Julie Ma
- John McNutt
- Linh Ngo
- Todd Price
- Barr von Oehsen
- Jennifer Oxenford
- Andy Sherman
- Marc Sugarman
- Scott Valcourt

ERN Steering Committee

John Barden

Forough Ghahramani

John Goodhue

Jim Griffioen

Vasant Honavar

Florence Hudson

Jim Kyriannis

David Marble

John Moore
Emeritus

Sharon Pitt

Carrie Ramp

Bruce Segee

Barr von Oehsen

Additional Research Support for smaller institutions

- Internet2 - CLASS
- CaRCC RCD Capabilities model Focused Tools

Additional Research Support for smaller institutions

- Internet2 - CLASS
- CaRCC RCD Capabilities model Focused Tools

What is CLASS?

CLASS: Cloud Learning and Skills Sessions
is Internet2's Learning Program
to accelerate cloud adoption and support workforce development
within the research and higher education (R&E) community.

Contact: class@internet2.edu

Internet2 CLASS program

Learning Modules

Essentials
- "Cloud 101", vendor-neutral covers AWS, GCP and Azure

Advanced
- Build on previous CLASS advanced curriculum
- Certifications paths (eg. Solutions Architect Associates Certification Prep)

Incubators
- Hands-on learning experiences in partnership with external community organizations and cloud vendors to accelerate cloud adoption on nationwide campuses

Community of practice
- A supportive network to share best practices around cloud computing adoption in R&E

CaRCC RCD CAPS (Capabilities) model focused tools

https://carcc.org/rcdcm/focused-tools/

Brief history of the full RCD CAPS model

- Workshop Dec 2018
  - Internet2, The Quilt, & Educause
- Five (CaRCC) facings
  - Data facing
  - Researcher facing
  - Systems facing
  - Software facing
  - Strategy and Policy facing

- 150 questions

Motivation for creating RCD CAPS model focused tools for smaller/under resourced institutions?

- The RCD CAPS Model allows institutions to assess support for computationally and data-intensive research, to identify gaps, and to benchmark vs. peers. It provides a great community dataset.
- However, it can be overwhelming for small institutions and emerging RCD programs.
- RCD CAPS Model focused tools will help emerging programs identify the necessary elements to build a research program from scratch.
- 2022 focus groups and workshop
  - Spring focus group - NJEdge, GPN (also BOF at GPN annual meeting)
  - Summer workshop - Nexus Day (day before PEARC)

Some of our findings

- Many programs struggled to complete the RCD CM, but they really wanted to see all the questions anyway
- Others were completely overwhelmed and confused by the tool
    - Some questions are way outside of what they can even consider
    - IT leadership may not know who researchers are, what they need
    - Assessment is time consuming; hard to carve out time from an already full plate
- We need to help early-stage programs understand the value
    - …of convincing c-level leaders to be an RCD advocate
    - …of having an assessment to make their arguments
- They need help building impact stories, especially in support of shared resources and services
    - E.g., campus cluster vs. cloud costs, training, …, etc.
- Need a way to "write your own journey"
    - Start with a conversation to begin the journey
    - Help them understand why it can be worth the effort to do the assessment.

Current thinking …

1. A *Conversation template or script* that helps frame a discussion, and gets them thinking about key aspects of RCD support

2. A means of working with the RCD CD assessment tool that allows them to assess a focused set of topics, deferring/ignoring others that are way out of scope for them at that point.

Acknowledgements

Thank you!

We greatly appreciate your time today.
Contact:
Forough.ghahramani@njedge.net
jhicks@internet2.edu

ERN: https://ernrp.org, info@ernrp.org
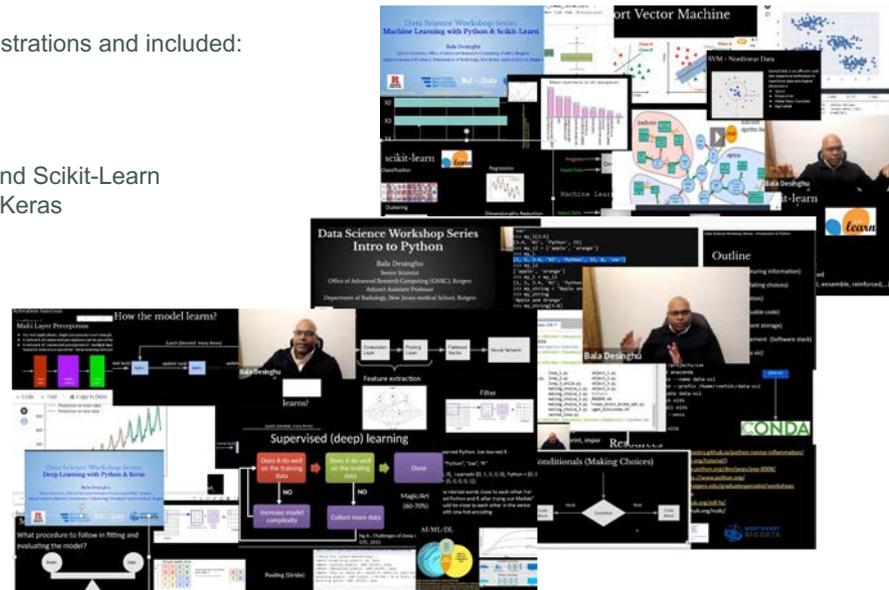
Q&A and Discussion!

Data Science Workshop Series

Data Science Workshop Series is a collaboration between the ERN Broadening the Reach and the Structural Biology Working Groups partnering with Rutgers University Office of Advanced Research Computing(OARC), the New Jersey Big Data Alliance (NJBDA), and the Northeast Big Data Innovation Hub. The workshops provide on-hands training of various software applications focusing on the support of Machine Learning, Deep Learning and Artificial Intelligence workflow methodology. The workshops are open to anyone in the research community with emphasis on the MSIs, HSIs, HBUCs and EPSCoR institutions. Compute resources necessary for the on-hand efforts are made available to all participants and moderating support personnel address questions or issues.

Previous workshops averaged 225 registrations and included:

- Introduction to Python
- Python for Big Data Analytics
- Machine Learning with Python and Scikit-Learn
- Deep Learning with Python and Keras

ERN All Hands Meeting

Tentative March 23-24, 2023

The All Hands Meeting (AHM) gives us an opportunity to come together and discuss current and future ERN projects, collaborations and partnerships, the five ERN Working Groups' efforts, future funding opportunities, future workshops, and other topics. As with previous AHMs we plan to invite several NSF Program Officers as well as industry partners who have been involved with several of the ERN activities. Plans are for an in-person meeting with the tentative schedule below. Finalized information will be posted on our website and invitations sent to the ERN community.

Thursday March 23:
Opening Session
Working Group Updates

      Structural Biology/Cryo-EM

      Materials Discovery

      Broadening the Reach

      Architecture/Federation/Policies
Discussion Sessions

Friday March 24:
Future Plans and Next Steps

ERN PEARC'22 Short Papers

July 13, 2022

11:15 - 11:30   **Broadening the Reach for Access to Advanced Computing: Leveraging the Cloud for Research**

Workforce Track
The Loft

Many smaller, mid-sized and under-resourced campuses, including MSIs, HSIs, HBCUs and EPSCoR institutions, have compelling science research and education activities along with an awareness of the benefits associated with better access to cyberinfrastructure (CI) resources. These schools can benefit greatly from resources and expertise for cloud adoption for research to augment their in-house efforts. The Ecosystem for Research Networking (ERN), formerly the Eastern Regional Network, Broadening the Reach (BTR) working group is addressing this by focusing on learning directly from the institutions on how best to support them. ERN BTR findings and recommendations will be shared based on engagement with the community, including results of workshops and surveys related to challenges and opportunities as institutions are evaluating using the cloud for research and education. This work is NSF-sponsored CC*CRIA: OAC-2018927.

ERN PEARC'22 Short Papers

July 13, 2022

3:30 - 3:45   **The ERN Cryo-EM Federated Instrument Pilot Project**
Best Short Paper
Systems Track
Studio 2
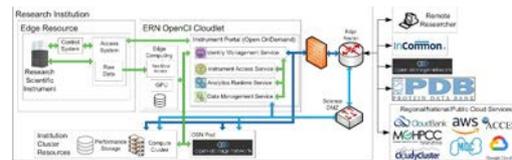
Feedback and survey data collected from hundreds of participants of the Ecosystem for Research Networking (formerly Eastern Regional Network) series of NSF (OAC-2018927) funded community outreach meetings and workshops revealed that Structural Biology Instrument driven science is being forced to transition from self-contained islands to federated wide-area internet accessible instruments. This paper discusses phase 1 of the active ERN CryoEM Federated Instrument Pilot project whose goal is to facilitate inter-institutional collaboration at the interface of computing and electron microscopy through the implementation of the ERN Federated OpenCI Lab's Instrument CI Cloudlet design. The conclusion will be a web-based portal leveraging federated access to the instrument, workflows utilizing edge computing in conjunction with cloud computing, along with real-time monitoring for experimental parameter adjustments and decisions. The intention is to foster team science and scientific innovation, with emphasis on under-represented and under-resourced institutions, through the democratization of these scientific instruments.

ERN PEARC'22 Short Papers

July 13, 2022

3:45 - 4:00 **Federating CI Policy in Support of Multi-Institutional Research: Lessons from the Ecosystem for Research Networking**
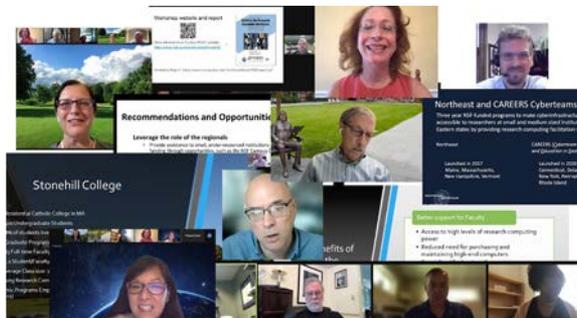
Systems Track
Studio 1

The ERN (Ecosystem for Research Networking) works to address challenges that researchers face when participating in multi-campus team science projects. There are a variety of technical and collaborative coordination problems associated with shared access to research computing and data located across the national cyberinfrastructure ecosystem. One of these problems is the need to develop organizational policy that can work in parallel with policies at at different institutions or facilities. Generally, universities are not set up to support science teams that are distributed across many locations, making policy alignment an even more complex issue. We describe some of the work of the ERN Policy Working Group, and introduce some key issues that surfaced while developing a guiding policy framework.
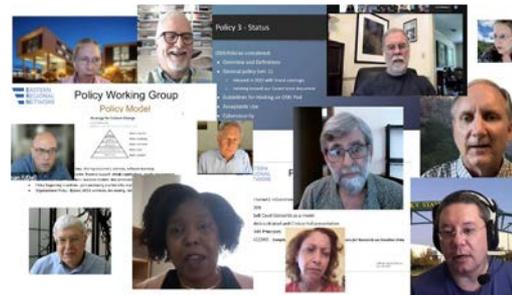
ERN Broadening The Reach Workshop

Spring 2023