

- A Advanced Research Computing Environments
 N Application Software Support Outcomes
 B BOF
- R Break
 C Co-located Event
 L Lunch Break
 M ML/AI
 P Panel
 Y Plenary
- V Poster/VIS Reception
 S Student Program
 T Tutorials
- W Workforce development Diversity Professionalization
 K Workshops

JULY 27 • MONDAY

- 8:00am – 12:00pm **T** **An Introduction to Advanced Features in MPI** Brella
Speakers: Amit Ruhela, Victor Eijkhout
 The MPI library is now in version 3, but most programmers use mechanisms from MPI-1 or 2 at best. This half-day tutorial, aimed at current MPI programmers, will discuss a number of MPI-3 features that offer more flexibility, a more elegant expression of algorithms, and higher performance. There will be lab sessions to exercise the material.
-
- 8:00am – 12:00pm **T** **Empowering Research Computing at Your Campus Through the Open Science Grid** Brella
Speakers: Christina Koch, Carrie Brown, Mats Rynge, Emelie Fuchs, Lauren Michael
 This training will provide researcher-facing cyberinfrastructure professionals with the information and hands-on skills they need to introduce their local researchers to the Open Science Grid. Attendees will survey existing research examples that use the OSG and work collaboratively to build an individual plan for increasing OSG awareness on their campus. By sharing their personal experiences and learning more about the OSG, we hope to empower researcher-facing cyberinfrastructure professionals by expanding their toolkit for engaging their campus community.
-
- 8:00am – 12:00pm **T** **TUTORIAL FULL - Customizing OpenHPC: Integrating Additional Software and Provisioning New Services including Open OnDemand** Brella
Speakers: Christopher Simmons, Derek Simmel, Karl Schulz
 Since its inception as a Linux Foundation project in 2015, the OpenHPC project (<https://openhpc.community>) has steadily grown to provide a modern, consistent, reference collection of HPC cluster provisioning tools, together with a curated repository of common cluster management software, I/O clients, advanced computational science libraries and software development tools, container-based execution facilities, and application performance profiling tools. Although the primary focus of OpenHPC remains enabling the on-premises deployment of new HPC clusters rapidly, the OpenHPC software repository itself serves as a reliable, portable, integrated collection of software, libraries, tools and user environment that can be employed in containers and VMs in the cloud. This can be particularly useful for those environments that one to either augment their on premises HPC resources with additional cloud-based resources (a hybrid cloud model) or deploy a full cloud solution. The goal of this tutorial is to help HPC cluster operators, users, and computational science application developers gain a better understanding of OpenHPC, and how it can be customized and extended to provide additional services that are commonly requested by our community including adding Open OnDemand support. This tutorial will cover extending an OpenHPC cluster to support additional software and features for both the on-premise and cloud models.
-
- 8:00am – 12:00pm **T** **TUTORIAL FULL - Introduction to NumPy** Brella
Speakers: Wim Cardoen
 We propose a half-day hands-on tutorial (i. e. 3 hour) "Introduction to NumPy" (Intermediate Level) to be held at the PEARC20 Conference taking place in Portland, OR (July 26-30, 2020).
-
- 8:00am – 12:00pm **K** **ACM SIGHPC SYSPROS Symposium 2020** Brella
Speakers: John Blaas, Jenett Tillotson
 In order to meet the demands of high-performance computing (HPC) researchers, large-scale computational and storage machines require many staff members who design, install, and maintain these systems. These HPC

systems professionals include system engineers, system administrators, network administrators, storage administrators and operations staff who face problems that are unique to high performance computing systems. While many conferences exist for the HPC field and the system administration field, none exist that focus on the needs of HPC systems professionals. Support resources can be difficult to find to help with the issues encountered in this specialized field. Often times, systems staff turn to the community as a support resource and opportunities to strengthen and grow those relationships are highly beneficial. This Workshop is designed to share solutions to common problems, provide a platform to discuss upcoming technologies, and to present state of the practice techniques so that HPC centers will get a better return on their investment, increase performance and reliability of systems, and researchers will be more productive. Additionally, this Workshop is affiliated with the systems professionals' chapter of the ACM SIGHPC (SIGHPC SYSPROS Virtual ACM Chapter). This session would serve as an opportunity for chapter members to meet face-to-face, discuss the chapter's yearly workshop held at SC, and continue building our community's shared knowledge base.

8:00am –
12:00pm

K Supporting Computational and Data-intensive Research through Research Computing and Library Collaboration

Brella

Speakers: Robert M Freeman Jr, Shelley Knuth, Christina Maimone, Amy Neeser, Anna Sackmann, Claire Mizumoto, Jonathon Anderson, Nickoal Eichmann-Kalwara

This workshop brings together perspectives from six institutions where research computing and libraries have successfully partnered to improve data and computing support. It combines a series of case study presentations from existing collaborations at a variety of institutions across the country with opportunities for participants to discuss and share next steps for collaboration at their own institutions. Participants will leave with concrete suggestions for how libraries and research computing groups can work together to support the full research lifecycle. This workshop will enable attendees to: 1) Share both best practices and lessons learned from collaborations between library and research computing groups from several institutions. 2) Allow participants to reflect upon and discuss the next steps for collaboration between research computing and library groups at their own institutions. 3) Build and sustain partnerships among providers of advanced research computing from varied institutional organizations including libraries, research computing, and business/social science computing centers. 4) Contribute to and support the further development of a network of researcher-facing professionals from both research computing and library organizations who are connected virtually through the CaRCC Data-Facing and Researcher-Facing Tracks

8:00am –
12:00pm

K Tapis Early Adopter

Brella

Speakers: Joe Stubbs, Gwen Jacobs, Sean Cleveland, Maytal Dahan

Tapis is an NSF-funded web-based API framework for securely managing computational workloads across infrastructures and institutions, so that experts can focus on their research instead of the technology needed to accomplish it. Tapis offers a set of authentication, authorization, data transfer, job management, and execution services that can span multiple data centers and manage batch, interactive and streaming jobs. More than 15 active, funded projects across a wide range of domains of science and engineering rely on Tapis to run 100,000s of computational jobs for 1000s of users to accomplish their research objectives. As part of work funded by the National Science Foundation starting in 2019, Tapis is delivering a version 3 ("V3") of its platform with several new capabilities, including a multi-site Security Kernel, Streaming Data APIs, and first-class support for containerized applications. The Tapis project is looking to reach the wider cyberinfrastructure community to share information and gather feedback on how best to address the community's needs. The Tapis Early Adopter workshop is aimed at presenting current and potential Tapis community members with the latest features and future roadmap of the Tapis Framework. Attendees will learn about Tapis services, focusing on the changes from V2 to V3 and the latest updates to existing services and new service offerings such as the distributed security kernel, streaming service and support for containerized applications. An in depth presentation and discussion of the capabilities and Tapis V3 design will occur to provide transparency into the Tapis offerings. Selected attendees will be offered an opportunity and are encouraged to present their use cases to the Tapis development team to acquire a better understanding of how Tapis might be leveraged for projects, including support for integrating Machine Learning workflows. Further, throughout the workshop attendees will have the opportunity to discuss feature requests and suggestions for the Tapis Framework as they network with the Tapis development team and other community members.

8:00am –
12:00pm

K Trustworthy Scientific Cyberinfrastructure (TrustedCI@PEARC20)

Brella

Speakers: Jim Basney

The Fourth Workshop on Trustworthy Scientific Cyberinfrastructure (TrustedCI@PEARC20) provides an opportunity for sharing experiences, recommendations, and solutions for addressing cybersecurity challenges in research computing. The workshop provides a forum for information sharing and discussion among a broad

range of attendees, including cyberinfrastructure operators, developers, and users. Please visit <https://www.trustedci.org/pearc20-workshop> for workshop agenda and materials.

8:00am – 5:00pm	T Open OnDemand, XDMoD, and ColdFront: an HPC center management toolset <i>Speakers: Dori Sajdak, Eric Franz, Ben Plessinger</i>	Brella
	<p>The University at Buffalo Center for Computational Research (UB CCR) and Ohio Supercomputer Center (OSC) team up to offer HPC systems personnel a tutorial for installing, configuring and using what many HPC centers now consider vital software products for managing and enabling access to their resources. UB CCR offers two open source products - an allocations management system, ColdFront, and an HPC metrics & data analytics tool, XDMoD. OSC provides the open source OnDemand portal for easy, seamless web-based access for users to HPC resources. These three tools have been designed to work together to provide a full package of HPC center management and access products. In this tutorial the system administrators and software developers from OSC and UB CCR will demonstrate the installation and configuration of each of these software packages. We'll show how to use these three products in conjunction with each other and the Slurm job scheduler. We will begin the tutorial with a short overview of each software product and how they tie together to provide seamless management of an HPC center. After demonstrating how the three software products are installed and configured, OSC staff will walk participants through the configuration of interactive apps in OnDemand. The tutorial will conclude with instructions on how to configure the latest version of OnDemand to display graphics of XDMoD job metrics.</p> <p>Tutorial participants will be provided with a Docker "cluster in a container" setup to allow them to follow along with installation and configuration steps as presented in the tutorial. Participants will need Docker installed on their local system to participate fully. We will use Slack for communication and present the tutorial via Zoom.</p>	
8:00am – 5:00pm	K The Important Role People Play in Advancing the Use of Cloud Cyberinfrastructure in Research <i>Speakers: Brian Voss, Joel Peterson, Nick Weber, Rick McMullen, Shawn Slavin</i>	Brella
	<p>As a sequel to the successful workshop delivered at PEARC19 in Chicago, we will continue to explore the impact that people bring to advancing the use of commercial cloud computing in research. This year, the workshop will provide insights into not only the Humans Advancing Research in the Cloud (HARC) project from the Pervasive Technology Institute at Indiana University, but also the National Institutes of Health Science and Technology Research Infrastructure for Discovery, Experimentation, and Sustainability (STRIDES) Initiative. Representatives from both programs will present sessions to attendees about the initiatives, including description of offerings, experiences, and perspectives relevant to the community of cloud research support personnel. Papers and presentations will be solicited from the broader PEARC community (as was done in last year's workshop) to add to the agenda, offering perspectives and experiences from those who provide the human support behind use of cloud offerings to advance research.</p>	
10:00am – 10:30am	R Break and Networking	Brella
10:00am – 12:00pm	C Hackathon Kick-off Meeting	Brella
11:00am – 12:00pm	S Student Welcome	Brella
12:00pm – 1:00pm	L Lunch Break and Networking	Brella
12:00pm – 1:00pm	S Student Mentor/Mentee Meet and Greet Student Program: Student Mentor/Mentee Meet and Greet	Brella
1:00pm – 3:00pm	S Speed Networking Onboarding	Brella
1:00pm – 5:00pm	T CANCELED SESSION - Democratizing/Demystifying Advanced Cyberinfrastructure for Deep Learning <i>Speakers: Paola A. Buitrago, Nick Nystrom, Tom Maiden, John Urbanic</i>	Brella
	<p>To harness deep learning's true potentials in research it is necessary to train our cyberinfrastructure community in making the most out of the resources made available through federally funded initiatives. The necessary training includes gaining the skills to (1)~characterize scientific use cases suitable for deep learning (DL),</p>	

(2)~characterize and compare different hardware solutions or platforms in the light of DL requirements, and (3)~optimally deploy a deep learning solution for a scientific use case in a suitable large scale hardware platform for enhanced productivity. To our knowledge, there are many related educational offerings, but not one that focuses on the relation of performance to hardware architecture, measuring performance, selecting architectures, maximizing performance, and scaling deep learning from an user perspective accessible to researchers with varying levels of hardware expertise. We propose the three-hour hands-on tutorial \textit{Democratizing/Demystifying Advanced Cyberinfrastructure for Deep Learning} to bridge this important gap.

1:00pm –
5:00pm

T **TUTORIAL FULL - Deep Dive into Constructing Containers for Scientific Computing and Gateways**

Brella

Speakers: Eric Coulter, Jeremy Fischer, Stephen Bird, Sanjana Sudarshan, Peter Vaillancourt, Suresh Maru

In recent years, using containers has been rapidly gaining traction as a solution to lower the barriers to using more software on HPC and cloud resources. However, significant barriers still exist to actually doing this in practice, particularly for well-established community codes which expect to run on a particular operating system version or resource. Additional barriers exist for researchers unfamiliar with containerization technologies. While many beginner tutorials are available for building containers, they often stop short of covering the complexities that can arise when containerizing scientific computing software. The goal of this full-day tutorial is to demonstrate and work through building and running non-trivial containers with users. We will containerize community scientific software, exhibit how to share with a larger community via a container registry, and then run on a completely separate HPC resource, with and without the use of a Science Gateway. The subject matter will be approachable for intermediate to advanced users, and is expected to be of interest to a diverse audience including researchers, support staff, and teams building science gateways.

1:00pm –
5:00pm

T **TUTORIAL FULL - Python 201: Building Better Scientific Software in Python**

Brella

Speakers: Geoffrey Lentner, Lev Gorenstein, Amiya Maji

Scientists and researchers are increasingly put in the position of also being software developers. Writing code in the pursuit of science and data analysis brings with it the challenge of making that code deployable and accessible to collaborators. Novice programmers often find themselves writing difficult to maintain, difficult to manage code that their peers and collaborators have trouble using. The goal of this tutorial is to expose researchers to several best practices in scientific software engineering that may otherwise take several years to become acquainted with. Though the implementation of these lessons is Python specific, the essential ideas can be exported to other languages or platforms. The tutorial is hands-on with full examples outlined in a "readthedocs" style website of materials. Participants are expected to already be familiar with the Python language to the extent that they understand what Python is, how to write code files in a text editor that they are comfortable with, and run that code at the command line on their platform of choice (Linux, macOS, Windows). Topics covered: Python packaging, automated testing, documentation management, logging, command line interfaces, performance profiling and optimization.

1:00pm –
5:00pm

T **TUTORIAL FULL - The Streetwise Guide to Jupyter Security**

Brella

Speakers: Rick Wagner, Kay Avila

Jupyter? is software that is designed to allow the running of arbitrary code within a web page and facilitates the sharing of so called Jupyter notebooks for the purpose of allowing code to be shared and easily run. Understandably, this can be of some concern for security staff. In the past few years these Jupyter notebooks have become ubiquitous in the academic and research communities as a method for sharing scientific processes and data, leading to a potential source of conflict between those tasked with running them and those tasked with security at an organizational level. Yet the methods for securing a JupyterHub installation may not be reaching those who are installing and managing them. The presenter team has encountered several indications that security information for Jupyter is still difficult to find. A standing policy at a major national lab has been that Jupyter would not be allowed due to high risk. This workshop will provide an overview of Jupyter and how it is used before diving into hands-on exercises covering the current best practices for securing a JupyterHub installation. Along the way we will explain the security risks to running Jupyter and sharing Jupyter notebooks as well as how to mitigate those risks.

1:00pm –
5:00pm

K **Strategies for Enhancing HPC Education and Training (SEHET20)**

Brella

Speakers: Nitin Sukhija, Scott Lathrop, Julia Mullen, Kate Cahill, Dana Brunson, Susan Mehringer

High performance computing is becoming central for empowering scientific progress in the most fundamental research in various science and engineering, as well as society domains. It is remarkable to observe that the recent rapid advancement in the mainstream computing technology has facilitated the ability to solve complex

large-scale scientific applications that perform advanced simulations of the implementation of various numerical models corresponding to numerous complex phenomena pertaining to diverse scientific fields. The inherent wide distribution, heterogeneity, and dynamism of the today's and future computing and software environments provide both challenges and opportunities for cyberinfrastructure facilitators, trainers and educators to develop, deliver, support, and prepare a diverse community of students and professionals for careers that utilize high performance computing to advance discovery. The SEHET20 workshop is an ACM SIGHPC Education Chapter coordinated effort aimed at fostering collaborations among the practitioners from traditional and emerging fields to explore strategies to enhance computational, data-enabled and HPC educational needs. Attendees will discuss approaches for developing and deploying HPC training, as well as identifying new challenges and opportunities for keeping pace with the rapid pace of technological advances - from collaborative and online learning tools to new HPC platforms. The workshop will provide opportunities for: learning about methods for conducting effective HPC education and training; promoting collaborations among HPC educators, trainers and users; and for disseminating resources, materials, lessons learned and good/best practices.

3:00pm – 3:30pm	R Break and Networking	Brella
3:00pm – 4:15pm	C Hackathon Check-in	Brella
3:00pm – 5:00pm	S Exhibitor/Student Panel Discussion	Brella

- A Advanced Research Computing Environments
 N Application Software Support Outcomes
 B BOF
R Break
 C Co-located Event
 L Lunch Break
 M ML/AI
 P Panel
 Y Plenary
V Poster/VIS Reception
 S Student Program
 T Tutorials
W Workforce development Diversity Professionalization
 K Workshops

JULY 28 • TUESDAY

8:00am – 9:30am **Y** **Plenary I - Welcome and Keynote - Cherri Pancake, ACM President** Brella
Speakers: Gwen Jacobs, Cherri Pancake
 Title: Can We Change the Face of Computing?

The computing field has earned a bad reputation for its lack of diversity. That's a very serious problem, and not just in terms of equity or social justice. When the workplace is too homogeneous, a company or institution is constraining its potential for future innovation and growth. Even worse, computing's inhospitable culture is discouraging the very people who could bring the new perspectives that are so badly needed. I hope to convince you that each of us should take personal responsibility for changing the "culture of technology" before our field slowly strangles itself.

To change a culture, it's important to understand how group behavior and values evolve and be realistic about how they can be influenced. My talk will focus on lessons I have learned about cultural change, first as an anthropologist and then working my way into and through a field that is overwhelmingly dominated by white male thinking. I'll suggest a number of practical actions we can take, in our everyday roles and at any stage in our careers, to become an "agent of change" for our profession.

Biography

Dr. Cherri M. Pancake is the past President of the Association for Computing Machinery (ACM.org) and Director of the Northwest Alliance for Computational Research. She recently retired as Professor of Electrical Engineering & Computer Science and Intel Faculty Fellow at Oregon State University. Her background combines environmental design and anthropology with computer engineering. Pancake conducted much of the seminal work identifying how the needs of scientists differ from computer science and business communities. Over a period of 25 years, she served as PI or coPI on research grants totaling over \$160 million from industry, not-for-profits, NSF, and US Departments of Agriculture, Commerce, Defense, Education, Energy, and Interior. The user-centered design methods she developed were reflected in software products from Hewlett Packard, Convex, Intel, IBM, and Tektronix. Her current research focus is software systems to help resource managers access and analyze complex research data more effectively so their decisions can be science-based. A Fellow of ACM and IEEE, Pancake was the founding chair of SIGHPC, ACM's Special Interest Group on High Performance Computing.

9:30am – 9:45am **R** **Break and Networking** Brella

9:30am – 9:45am **S** **Volunteer Team Standup** Brella


9:45am – 10:45am **Y** **Plenary I - Exhibitor Forum** Brella
Speakers: Pranav Mehta, Luke Wilson

Pandemic Accelerates the Scale Out by Pranav Mehta, IntelThe current pandemic has accelerated the need for scientific and technological advances in certain end user visible applications and underlying cyberinfrastructures to support them. This talk will look at some of these vertical use cases and outline challenges and possible solution approaches, based on our internal and collaborative research results, to satisfy the critical societal needs.

HPC Gives Computers a Voice by Luke Wilson, Dell
The dream of having a conversation with your computer is fast becoming a reality. But how are these voices produced? Neural networks have transformed voice synthesis, replacing artificial sounding voice clips with smooth, natural voices generated entirely by computer. Scale-out parallelism and acceleration are driving down the time to create these voice models from months to hours, turning the dream of conversational computers into a reality even faster.

10:00am – 12:00pm **C Hackathon Check-in** Brella

10:45am – 11:05am **N VisQueue : An Analytical Dashboard for Science Exploration on HPC Systems** Brella
Speakers: Andrew Solis, Gregory Foss, Craig Jansen, Mathew Stelmaszek
Monitoring a supercomputer's performance is vital for high performance computing (HPC) support and development, often its users, and even appreciated by the general public. Identifying the effectiveness of a system is key for its host institution's confidence and credibility, and accounting for necessary operational funding. Tracking the research topics and sciences which dominate a system's time and CPU can help guide decisions regarding design and support for both current and future system resources. This information is also useful for those whose tasks are to inform users, collaborators, and the general public about a system's utilization. Applications that monitor and display system performance, often in substantial detail, are essential. Yet mining the components and quantity of performance statistics can be a challenge. Additionally, the data is usually not linked to available potentially useful stats from a systems' user database like project abstracts and principal investigator (PI). This paper introduces VisQueue, an interactive dashboard for exploring HPC systems. Its users initially see an array of the included supercomputers, each represented with an interactive sunburst chart and correlated data table. Each system's currently active projects are listed with science domain, running jobs, and utilized resources. User's can explore the projects at a deeper level, and also reference a system's specifications. These pages include graphs of available metrics, a table for exploring all projects that have utilized the system, and a map showing user locations and outreach. Our motivation was to create something not only useful for HPC maintainers, but easy for the general public to understand the impact and importance of HPC resources. Careful consideration was taken to include only the information that was necessary and not overwhelm the user with too many choices or details.

10:45am – 11:05am **M Exploring collections of research publications with human steerable AI.**  Brella
Speakers: Alberto González Martínez, Billy Troy Wooton, Nurit Kirshenbaum, Dylan Kobayashi, Jason Leigh

 **Best Paper in “Trending now – machine learning and artificial intelligence” Track**

Understanding highly-dimensional data sets is a complex task. Traditionally, this problem has been tackled with linear pipelines that rely on mathematical models and algorithms to summarize relationships and structure, producing a visual representation of the data in a collapsed, low-dimensional form. The main issue with these traditional pipelines is that they are driven solely by algorithms or models, and without a human in the loop, they can potentially limit sense-making by masking expected or known structure in the data. Textual data, such as that contained in research publications, is one example of unstructured highly dimensional data, wherein the raw data must be converted to an abstract numeric representation that is highly dimensional. In recent years, Semantic Interaction has become an interesting approach to enabling model steering in Visual Analytics systems, as it provides mechanisms with which to adjust the parameter space, explore data, and test hypotheses. In order to facilitate this interaction modality, Semantic Interaction systems need to invert the computation of one or more mathematical models to support a bidirectional structure within their pipelines. Most examples of Semantic Interaction systems are limited to linear models to allow for this bidirectionality. In this paper we propose an inexpensive neural encoder approach to performing backward and forward computations within semantic interaction pipelines for analyzing textual data. We show that this approach allows for the efficient "merging" of new instances into a previously trained model without retraining. It also provides a reverse link, allowing the parameters of a trained model to be affected by user interactions with the visual representation of data. To demonstrate the usefulness of this approach we present the Zexplorer system, a tool for exploring Large Document Collections of Research papers with Semantic Interaction. The Zexplorer system is built as an extension to Zotero, a widely used open source bibliography system.

11:05am – 12:00pm **L Lunch Break and Networking** Brella

12:00pm – 1:20pm	N	Building Science Gateways for Humanities <i>Speakers: Jun Zhou, Karen Smith, Greg Wilsbacher, Paul Sagona, David Reddy, Ben Torkian</i> Building science gateways for humanities poses new challenges in the science gateway community. Compared to science gateways devoted to scientific content, humanities-related projects usually require 1) processing data in various formats, such as text, image, video, etc., 2) constant public access from a large audience, and 3) reliable security upgrade and low maintenance. Most traditional science gateways are monolithic in design, which is easy to write, but they can be computationally inefficient when integrated with numerous scientific packages for data capture and pipeline processing. Since these applications tend to be single-threaded or nonmodular, they can create traffic bottlenecks when processing large numbers of requests. Moreover, these science gateways are usually difficult to be backed up due to long gap between funding periods and age of applications. In this paper, we study the problem of building science gateways for humanities projects by developing service-based architecture and present two such science gateways: the Moving Image Research Collections (MIRC) - a science gateway focusing on image analysis for digital surrogates of historical motion picture film, and SnowVision - a science gateway for studying pottery fragments in southeastern North America. For each science gateway, we present an overview of the background, some unique challenges, design, and implementation. These two science gateways are deployed on XSEDE's Jetstream academic clouding computing resource and are accessed through a web interface. Apache Airavata middleware is used to manage the interactions between the web interface and deep-learning-based (DL) backend service running on the Bridges graphics processing unit (GPU) cluster.	Brella
12:00pm – 1:20pm	N	Implementing a Prototype System for 3D Reconstruction of Compressible Flow <i>Speakers: Christiaan Gribble, Victor Eijkhout, Paul Navratil</i> PowerFlow3D is a prototype system for acquiring, reconstructing, and visualizing three-dimensional structure of complex flows around objects in wind tunnel test procedures. PowerFlow3d combines modern high-performance computing (HPC) with existing acquisition, reconstruction, and visualization methods to provide a foundational capability that helps to reveal critical information about the underlying structure of unknown flows. We describe the implementation of our system, focusing on tomographic reconstruction, in particular, and highlight the practical challenges encountered throughout our initial research and development (R&D) process. The resulting prototype achieves both reasonable performance and fidelity and provides opportunities for enhanced performance, fidelity, and scale. The results of this initial R&D effort thus enable continued progress toward a scalable HPC-accelerated system for guiding real-time decisions during wind tunnel tests.	Brella
12:00pm – 1:20pm	N	PEGR: a management platform for ChIP-based next generation sequencing pipelines <i>Speakers: Danying Shao, Gretta Kellogg, Shaun Mahony, William Lai, B. Franklin Pugh</i> There has been a rapid development in genome sequencing, including high-throughput next generation sequencing (NGS) technologies, automation in biological experiments, new bioinformatics tools and utilization of high-performance computing and cloud computing. ChIP-based NGS technologies, e.g. ChIP-seq and ChIP-exo, are widely used to detect the binding sites of DNA-interacting proteins in the genome and help us to have a deeper mechanistic understanding of genomic regulation. As sequencing data is generated at an unprecedented pace from the ChIP-based NGS pipelines, there is an urgent need for a metadata management system. To meet this need, we developed the Platform for Eukaryotic Genomic Regulation (PEGR), a web service platform that logs metadata for samples and sequencing experiments, manages the data processing workflows, and provides reporting and visualization. PEGR links together people, samples, protocols, DNA sequencers and bioinformatics computation. With the help of PEGR, scientists can have a more integrated understanding of the sequencing data and better understand the scientific mechanisms of genomic regulation. In this paper, we present the architecture and the major functionalities of PEGR. We also share our experience in developing this application and discuss the future directions.	Brella
12:00pm – 1:20pm	N	Tailoring Data Visualization to Diversely Informed End Users <i>Speakers: Ryan Theriot, James Hutchison, Nurit Kirshenbaum, Jason Leigh</i> Visualization is invaluable for communicating complex data, and it is becoming more important when researchers seek insights, or when they need to share their insights with end-users with diverse level of expertise such as policy-makers or the general public. In this paper, we detail how we approach this issue using the HAVEN and ProjecTable systems we developed. These systems share data of projected renewable energy supply based on several possible scenarios that should help Hawaii achieve 100% renewable energy by 2045. However, each system offers different features directed at end-users interested in different levels of detail. We describe here the two systems and their use, both separately and in conjunction.	Brella
12:00pm –	C	CaRCC Town Hall	Brella

1:20pm

All are welcome! The Campus Research Computing Consortium (CaRCC) invites all to discuss and provide feedback on the CaRCC roadmap, and in particular to provide the community's initial conceptualization for possible models of sustainability and collaboration for the future of CaRCC and related or alternative communities as we move to support, advocate for, and enhance research computing and data, research IT, or research cyberinfrastructure on our campuses, in our regions and in the nation. CaRCC - the Campus Research Computing Consortium - is an organization of dedicated professionals developing, advocating for, and advancing campus research computing and data and associated professions. Through various working groups and based on community input via the 2017 stakeholder survey, members of the CaRCC community have developed a robust structure to connect professionals in the field, referred to as the People Network with tracks for researcher-facing, system-facing, and data-facing professionals, and one for emerging centers; a research computing and data job families guide; a document articulating the value propositions for campus research computing; brought together representatives from many communities of CI professionals in an ecosystem awareness workshop; and in collaboration with Internet2 and EDUCAUSE continue the development of a campus research computing and data capabilities model. CaRCC's origins are in the NSF award OAC-1620695 "RCN: Advancing Research and Education through a national network of campus research computing infrastructures - The CaRC Consortium." For this meeting we welcome all who are interested in helping formulate future directions of CaRCC-facilitated (or potentially other community's-facilitated) activities to serve the broader campus research computing community and ways to sustain those activities beyond the current NSF grant funding. Following this discussion at PEARC20, interested participants will be encouraged to join in further activities to refine and implement the ideas generated. If you are interested in participating in this but cannot make the meeting, please write to info@carcc.org For more information about CaRCC please visit <https://carcc.org/>

12:00pm –
1:20pm

M **A Pilot Benchmarking Study of Deep Neural Network Performance on Low Magnification Pathology ROIs**

Brella

Speakers: Christopher Pondoc, Yanling Liu, Alexander Plisov

Deep Neural Networks (DNNs) have successfully demonstrated superior overall performance in many image classification and recognition tasks on H&E histology images. Reported studies typically utilize high quality (20x or 40x) Whole Slide Images (WSIs) in order to deliver optimal performance. However, it remains uncertain how well DNNs can perform on lower quality region-of-interest (ROI) histology images in real life scenarios. The NCI Patient Derived Models Repository (PDMR) database hosts a catalog of low magnification (4x) ROIs of tissue histology images across a total of 60 cancer models, providing an ideal test case for evaluating DNNs performance in real life scenarios. In this study, using 5 pre-trained models, we have benchmarked the NCI PDMR database ROIs on a selected set of popular DNN classifiers. Overall, on the binary carcinoma vs. sarcoma classification test, we have reached 89.57% accuracy on 4x ROIs using our downsizing models and 84.18% accuracy on 4x ROIs using our patch-based models. On the multi-class carcinoma classification test, we have reached 72.06% top-2 accuracy on 4x ROIs using our downsizing models and 78.07% top-2 accuracy on 4x ROIs using our patch-based models. With such accuracies, we can utilize our DNNs to perform crucial tele-pathological tasks in underdeveloped countries and rural areas, enabling scientists to utilize histology images acquired from mobile devices for rapid screening in remote areas.

12:00pm –
1:20pm

M **Evaluation of Clustering Techniques for GPS Phenotyping Using Mobile Sensor Data** 🏆

Brella

Speakers: Karl Schulz, Zachary Tschirhart

🏆 **Best Student Paper in “Trending now – machine learning and artificial intelligence” Track**

With the ubiquitousness of mobile smart phones, health researchers are increasingly interested in leveraging these commonplace devices as data collection instruments for near real-time data to aid in remote monitoring, and to support analysis and detection of patterns associated with a variety of health-related outcomes. As such, this work focuses on the analysis of GPS data collected through an open-source mobile platform over two months in support of a larger study being undertaken to develop a digital phenotype for pregnancy using smart phone data. An exploration of a variety of off-the-shelf clustering methods was completed to assess accuracy and runtime performance for a modest time-series of 292K non-uniform samples on the Stampede2 system at TACC. Motivated by phenotyping needs to not-only assess the physical coordinates of GPS clusters, but also the accumulated time spent at high-interest locations, two additional approaches were implemented to facilitate cluster time accumulation using a pre-processing step that was also crucial in improving clustering accuracy and scalability.

12:00pm –
1:20pm

M **Extracting and Analyzing Deep Learning Features for Discriminating Historical Art**

Brella

Speakers: Paul Rodriguez, Alan B. Craig, Alison Langmead, Christopher Nygren

This goal of this research was use advanced computing to investigate a particular method for making stylistic attributions of old master paintings that was developed by Giovanni Morelli in the late nineteenth century. The "Morelli Machine" refers to an algorithmic approach based on analyzing fine details and style of minor items. However, the notion of "style" has been constructed and reconstructed over the century and a half since Morelli began publishing his findings. Instead, we asked what could computers contribute/reveal about a humanistic method that is fully determined, but that is also so clearly aligned with computation. In addition, given the recent interest and success in deep convolution network we wanted to explore what features are present in these networks that have some discriminatory power for distinguishing paintings. We extracted features from a large scale convolutional network (VGG19)

12:00pm – 1:20pm	M On Multiview Robustness of 3D Adversarial Attacks	Brella
	<i>Speakers: Philip Yao, Tingting Chen, Hao Ji, rew So</i> Nowadays deep neural networks have been applied widely in many applications of computer vision including medical diagnosis and self-driving cars. However, deep neural networks are threatened by adversarial examples usually in which image pixels were perturbed unnoticeable to humans but enough to fool the deep networks. Compared to 2D image adversarial examples, 3D adversarial models are less invasive in the process of attacks, and thus more realistic. There have been many research works on generating 3D adversarial examples. In this paper, we study the robustness of 3D adversarial attacks when the victim camera is placed at different viewpoints. In particular, we find a method to create 3D adversarial examples that can achieve 100% attack success rate from all viewpoints with any integer spherical coordinates. Our method is simple as we only perturb the texture space. We create 3D models with realistic textures using 3D reconstruction from multiple uncalibrated images. With the help of a differentiable renderer, we then apply gradient based optimization to compute texture perturbations based on a set of rendered images, i.e., training dataset. Our extensive experiments show that even only including 1% of all possible rendered images in training, we can still achieve 99.9% attack success rate with the trained texture perturbations. Furthermore, our thorough experiments show high transferability of the multiview robustness of our 3D adversarial attacks across various state-of-the-art deep neural network models.	
1:20pm – 1:35pm	R Break and Networking	Brella
1:35pm – 2:35pm	N Building an Interactive Workbench Environment for Single Cell Genomics Applications	Brella
	<i>Speakers: D Balamurugan, Kristina Plazonic, Kevin Abbey, Seema Husain, Neeraja Syed</i> We discuss the procedure to build an interactive workbench environment for single cell genomic applications with the Open OnDemand (OOD) science gateway. In our approach, an end-user submits a complex single cell RNA sequencing (scRNA) pipeline, checks the status of the job, and visualizes the output results. All of these tasks are accomplished through a web browser, relieving the users from the complexities involved in developing and handling a large-scale workflow. Our approach helped researchers in processing several input data sets of scRNA in the campus HPC cluster. Although the current work is focused on scRNA analysis, the same approach can be extended for any workflow.	
1:35pm – 2:35pm	N GeoEDF: An Extensible Geospatial Data Framework for FAIR Science	Brella
	<i>Speakers: Rajesh Kalyanam, Lan Zhao, X. Carol Song, Venkatesh Merwade, Jian Jin, Uris Baldos, Jack Smith</i> Collaborative scientific research is now increasingly conducted online in web-based research platforms termed "science gateways". Most science gateways provide common capabilities including data management and sharing, scientific code development, high performance computing (HPC) integration, and scientific workflow execution of varying automation. Despite the availability of scientific workflow frameworks such as Pegasus and workflow definition languages such as the Common Workflow Language (CWL), in practice typical workflows on science gateways still involve a mix of non-reusable code, desktop tools, and intermediate data wrangling. With the growing emphasis on FAIR (Findable, Accessible, Interoperable, Reusable) science, such mixed workflows present a significant challenge to ensuring compliance to these principles. These challenges are further compounded in the earth sciences where researchers spend inordinate amounts of time manually acquiring, wrangling, and processing earth observation data from repositories managed by organizations such as NASA, USGS, etc. Our extensible geospatial data framework, GeoEDF is designed to address these challenges, making remote datasets directly usable in computational code and facilitating earth science workflows that execute entirely in a science gateway. In this paper we describe the design of GeoEDF, current implementation status, and future work.	

1:35pm –
2:35pm

N Log Discovery for Troubleshooting Open Distributed Systems with TLQ

Brella

Speakers: Nathaniel Kremer-Herman, Douglas Thain

Troubleshooting a distributed system can be incredibly difficult. It is rarely feasible to expect a user to know the fine-grained interactions between their system and the environment configuration of each machine used in the system. Because of this, work can grind to a halt when a seemingly trivial detail changes. To address this, there is a plethora of state-of-the-art log analysis tools, debuggers, and visualization suites. However, a user may be executing in an open distributed system where the placement of their components are not known before runtime. This makes the process of tracking debug logs almost as difficult as troubleshooting the failures these logs have recorded because the location of those logs is usually not transparent to the user (and by association the troubleshooting tools they are using). We present TLQ, a framework designed from first principles for log discovery to enable troubleshooting of open distributed systems. TLQ consists of a querying client and a set of servers which track relevant debug logs spread across an open distributed system. Through a series of examples, we demonstrate how TLQ enables users to discover the locations of their system's debug logs and in turn use well-defined troubleshooting tools upon those logs in a distributed fashion. Both of these tasks were previously impractical to ask of an open distributed system without significant a priori knowledge. We also concretely verify TLQ's effectiveness by way of a production system: a biodiversity scientific workflow. We note the potential storage and performance overheads of TLQ compared to a centralized, closed system approach.

1:35pm –
2:35pm

M Autonomous Vehicle Control Using a Deep Neural Network and Jetson Nano

Brella

Speakers: Rocco Febbo, Brendan Flood, Julian Halloy, Patrick Lau, Kwai Wong, Alan Ayala

The idea of a self-driving car is one which is actively studied and tested for use on the road. In addition, the machine learning tools required to create such a vehicle has become more and more available to the public as time goes on. With a number of different libraries and softwares available for free download to design and train neural networks and with affordable but powerful miniature computers on the market, one can explore the possibility of creating a self-driving vehicle. The goal of our project was to construct such a car on a small scale using parts and software that are accessible to anyone on an affordable budget (\$250), and to test the effectiveness of DNN software neural networks on training such a car. This project serves as a simple testbed for experimenting different ideas in self driving vehicle. Core ideas of autonomous vehicles are explored with machine learning in mind. The paper details the challenges and experience of project and is the result of an REU project support by the NSF.

1:35pm –
2:35pm

W CyberAmbassadors: Results from Pilot Testing a New Professional Skills Curriculum 🏆

Brella

Speakers: Astri Briliyanti, Julie Wilson Rojewski, Dirk Colbry, Katy Luchini-Colbry

🏆 **Best Paper in “People involved in research computing – workforce development, diversity, and professionalization” Track**

The CyberAmbassadors Project (NSF #1730137) is a training grant to develop professional skills curriculum (communications, teamwork, leadership) to build capacity in Cyber Infrastructure (CI) Professionals. CI Professionals are experts at high performance computing, data science algorithms, and/or supercomputing infrastructure; they are often called upon to work with experts from STEM (science, technology, engineering, mathematics) in multi-disciplinary teams to solve complex problems. The CyberAmbassadors training program seeks to improve the function of these teams by helping CI Professionals build and practice skills for effective communication, teamwork and leadership within the context of complex, multidisciplinary research. This paper summarizes the results of the pilot testing of the CyberAmbassadors curriculum, which was conducted at institutions across the United States using both in-person, online and hybrid delivery methods. A Kirkpatrick evaluation model was used to assess expectations and reasons for participation, as well as satisfaction with the training and impacts on participants' learning and behavior. The curriculum was revised based on these initial pilot tests, and 43 volunteers have participated in "train the trainers" workshops to prepare to facilitate this training on a larger scale during 2019-20.

1:35pm –
2:35pm

W Introducing Computing and Technology through Problem-Solving in Discrete Mathematics

Brella

Speakers: Robert Sinkovits, Osvaldo Soto

Mathematics is usually taught using a traditional pencil-and-paper approach and frequently lacks real-world contextual referents, leading to a sense of alienation among secondary students and missed opportunities to advance students' mathematical and computational thinking. In contrast, many real-world problems require mathematical solutions that are often of such a scale that they can only be solved using computational techniques, but require powerful ways of thinking that span the Common Core State Standards (CCSS) for Mathematical Practices (MP) and Computer Science Principles (CSP). We describe here our program that aims at advancing both MPs and CSPs through the introduction of technology and computing into the high school

discrete mathematics curriculum. Our goal in the first two years of this new program is to reach a cohort of 40 teachers and 72 students. Using lessons learned from this pilot, we will consider how our program can be scaled up to reach a much larger audience.

2:35pm – R **Break and Networking** Brella
3:00pm

2:35pm – B **Ask.CI, the Q&A site for Research Computing : Year 2 Lessons Learned, Plans for Year 3** Brella
3:50pm

Speakers: Julie Ma, Torey Battelle, Katia Bulekova, Aaron Culich, Jacob Pessin, Vanessa Sochat, Zoe Braiterman, Dana Brunson, Thomas Cheatham, Julia Cho, Christopher Coffey, Adrian Del Maestro, Shawn Doughty, John Goodhue, Brian Guilfoos, Chris Hill, Robert Knight, Timothy Middelkoop, Lauren Milechin, Kaylea Nelson, Wilbur Ouma, Annelie Rugg, Bruce Segee, Raminder Singh, Susan Sons, Mete Tunca, Scott Valcourt, Larry Whitsel, Scott Yockel

Ask.CI, a Q&A site for Research Computing, was launched at PEARC18. The goal of the site is to aggregate answers to a broad spectrum of questions that are commonly asked by the research computing community. As researchers, facilitators, staff, students, and other users ask and answer questions on Ask.CI, they are creating a shared knowledge base for the larger community and alleviating the burden of a central resource to provide support to a user base. For smaller institutions, this provides a wealth of knowledge that was not previously available in an easily searchable Q&A format. For larger institutions, this self-service model frees up time for facilitators and cyberinfrastructure engineers to focus on more advanced subject matter, thereby elevating the practice. Establishing a Q&A site of this nature requires some tenacity. In partnership with the Campus Champions, Ask.CI has gained traction, and continues to engage the broader community to establish the platform as a powerful tool for research computing. Since launch, Ask.CI has attracted over 250,000 page views (currently averaging nearly 5000 per week), over 400 contributors, hundreds of topics, and a broad audience that spans the US and parts of Europe and Asia. While Ask.CI has shown steady growth in both contributions and audience, it is still in its early stages. Finding ways to continue to grow audience participation through innovation and outreach is an ongoing focus of attention. In the past year, we introduced "Locales", which allow institutions to create subcategories on Ask.CI where they can experiment with posting institution-specific content and use of the site as a component of their user support strategy. At this BOF, the Ask.CI team and locale partners will discuss strategies, key accomplishments, challenges, and plans for the upcoming year.

2:35pm – B **Identifying Opportunities and Needs for Science Gateways in Education at Minority Serving Institutions** Brella
3:50pm

Speakers: Marlon Pierce, Linda Hayden, Michael Zentner, Linda Akli, Dinadayalane Tandabany, Raphael Isokpehi, Charles Dey, Ana Carmen Gonzalez

Science gateways and, generally, cyberinfrastructure are designed to democratize access to scientific resources such as scientific software, computing, and data collections. Additional effort is needed to identify opportunities for supporting Minority Serving Institutions (MSIs) and understanding their requirements. This Birds-of-a-Feather session will provide a forum for increased communication between educators, researchers, and science gateway providers.

2:35pm – B **Open OnDemand User Group Meeting** Brella
3:50pm

Speakers: Alan Chalker, David Hudak, Robert Deleon, Robert Settlege

Open OnDemand is an NSF-funded open-source HPC portal based on OSC's original OnDemand portal. The goal of Open OnDemand is to provide an easy way for system administrators to provide web access to their HPC resources, including, but not limited to: -Plugin-free web experience -Easy file management -Command-line shell access -Job management and monitoring across different batch servers and resource managers -Graphical desktop environments and desktop applications This BoF is meant to be an open discussion to guide the future roadmap for Open OnDemand in the near term, by getting feedback from the community on the prioritization of the various tasks planned for the next few years. As there are many people that attended this BOF at PEARC'18 and PEARC'19 and spoke highly of them, the session leaders intend to replicate the same BOF format (with appropriate updates regarding what has been done in the past year and the roadmap for the current NSF award) and anticipate more attendees than were present at the previous BOFs. A report summarizing the status of current installations and additional feature requests will be generated and distributed to anyone that has expressed interest in Open OnDemand.

2:35pm – B **Professional and Educational Development of High Performance Computing Systems Operations and Facilitator Professionals** Brella
3:50pm

Speakers: Stephen Harrell, Shelley Knuth, Joshua Baller, John Blaas, Thomas Hauser, Scott Lathrop,

Robert Thoenen

Systems professionals who are interested in high performance computing (HPC) are often left without curricula to study to advance their professional knowledge. Rarely there is a course at a university or college that is dedicated to the topics a professional would need to advance in their field. Available courses typically only teach system administration without an HPC treatment or teach HPC as a way to approach scientific problems and gloss over any information about the operation of HPC machines. As such, much learning is "on the job" and often in an apprenticeship learning model. Because of this, many entry-level computing professionals are not aware of the opportunities in the HPC field and may either not seek positions in this field or may feel overwhelmed with a lack of knowledge. When they do pursue these types of positions the on-ramp to being productive can be quite steep if they have no prior experience in HPC. High-throughput interconnects, parallel storage and issues of computing at scale are some of the barriers that must be overcome in addition to any basic system administration training they may need. This BoF will discuss efforts from a joint committee between two SIGHPC virtual chapters, the System Professionals (SYSPROS) chapter and the Education chapter, to promote the educational advancement of HPC systems professionals. While some of this BoF will be dedicated to reporting current efforts from the committee, much of the time will be dedicated to crowd-sourcing information on ways to develop, collect, and distribute information of interest to HPC system professionals.

3:00pm – 4:15pm	C	Hackathon Check-in	Brella
3:00pm – 5:00pm	V	Poster and Visualization Showcase	Brella

- A Advanced Research Computing Environments
 N Application Software Support Outcomes
 B BOF
R Break
 C Co-located Event
 L Lunch Break
 M ML/AI
 P Panel
 Y Plenary
V Poster/VIS Reception
 S Student Program
 T Tutorials
W Workforce development Diversity Professionalization
 K Workshops

JULY 29 • WEDNESDAY

8:00am – 9:30am Y **Plenary Panel: Introduction of the new NSF Innovative HPC systems** Brella
Speakers: Gwen Jacobs, X. Carol Song, William Gropp, David Y. Hancock, Paola A. Buitrago, Amit Majumdar

PEARC20 Panel: Introduction of the new NSF Innovative HPC systems Continuing to provide a comprehensive portfolio of innovative advanced computing resources to meet the evolving needs of the nation's science and engineering communities, NSF announced in May the latest round of awards for five new systems to be deployed and in production during 2020 – 2021. These systems will be available to U.S. researchers via XSEDE's resource allocations process.

These new systems are results of the NSF solicitation 19-587 for "Advanced Computing Systems & Services: Adapting to the Rapid Evolution of Science and Engineering Research", featuring two tracks:

- Category I, Capacity Systems: production computational resources maximizing the capacity provided to support the broad range of computation and data analytics needs in S&E research; and
- Category II, Innovative Prototypes/Testbeds: innovative forward-looking capabilities deploying novel technologies, architectures, usage modes, etc., and exploring new target applications, methods, and paradigms for S&E discoveries.

This session features presentations from the announced awardees of this NSF solicitation, i.e., three Category I systems and two Category II systems as follows:

- **Anvil**: Category I: A National Composable Advanced Computational Resource for the Future of Science and Engineering, Carol Song, RCAC, Purdue University
- **Delta**: Category I: Crossing the Divide Between Today's Practice and Tomorrow's Science, Bill Gropp, National Center for Supercomputing Applications, UIUC
- **Jetstream2**: Category I: Accelerating Science and Engineering on Demand, David Y. Hancock, Indiana University Pervasive Technology Institute
- **Neocortex**: Category II: Unlocking Interactive AI Development for Rapidly Evolving Research, Paola Buitrago, Pittsburgh Supercomputing Center
- **Voyager**: Category II: Exploring AI Processors in Science and Engineering, Amit Majumdar, San Diego Supercomputer Center

9:30am – 9:45am S **Volunteer Team Standup** Brella

9:30am – 10:15am R **Break and Networking** Brella

10:15am – 10:55am N **Atomic and Molecular Scattering Applications in an Apache Airavata Science Gateway** Brella
Speakers: Sudhakar Pamidighantam, Barry Schneider, Klaus Bartschat, Oleg Zatsarinny, Kathryn Hamilton, Igor Bray, Armin Scrinzi, Fernando Martin, Jesus Gonzalez Vasquez, Jonathan Tennyson, Jimena Gorfinkiel, Robert Lucchesse

In this paper we document recent progress made in the development and deployment of a science gateway for atomic and molecular physics (AMP) [10]. The molecular scattering applications supported in the gateway and the early phase of the project have been described in an earlier publication [33]. Our objective in this paper is to present recent advances in both the capabilities and the adoption of the platform for additional software suites

and new possibilities for further development. The applications being deployed provide users with a number of state-of-the-art computational techniques to treat electron scattering from atomic and molecular targets and the interaction of radiation with such systems. One may view all of these approaches as generalized close-coupling methods, where the inclusion of electron correlation is accomplished via the addition of generalized pseudostates. A number of the methods can also be employed to compute high-quality bound-state wavefunctions by closing the channels and imposing exponentially decaying boundary conditions. The application software suites are deployed on a number of NSF and DoE supercomputing systems. These deployments are brought to the user community through the science gateway with user interfaces, post-processing, and visualization tools. Below we outline our efforts in deploying the Django web framework for the AMPGateway using the Apache Airavata gateway middleware, discuss the new advanced capabilities available, and provide an outlook for future directions for the gateway and the AMP community.

10:15am – N **Toward a Data Lifecycle Model for NSF Large Facilities** Brella
10:55am

Speakers: Laura Christopherson, Anirban Mandal, Ilya Baldin, Erik Scott

National Science Foundation large facilities conduct large-scale physical and natural science research. They include telescopes that survey the sky, gravitational wave detectors that look deep into our universe's past, and sensor-driven field sites that collect a range of biological and environmental data. The Cyberinfrastructure Center for Excellence (CICoE) pilot aims to develop a model for a center that facilitates community building, fosters knowledge sharing, and applies best practices in consulting with large facilities about their cyberinfrastructure. To accomplish this goal, the pilot began an in-depth study of how large facilities manage their data. Large facilities are diverse and highly complex, from the types of data they capture, to the types of equipment they use, to the types of data processing and analysis they conduct, to their policies on data sharing and use. Because of this complexity, the pilot needed to find a single lens through which it could frame its growing understanding of large facilities and identify areas where it could best serve large facilities. As a result of the pilot's research into large facilities, common themes emerged that enabled the creation of a data lifecycle model that successfully captures the data management practices of large facilities. This model has enabled the pilot to organize its thinking about large facilities, and frame its support and consultation efforts around the cyberinfrastructure used during research. This paper describes the model and discusses how it was applied to disaster recovery planning for a representative large facility-IceCube.

10:15am – W **Cyberinfrastructure Facilitation Skills Training via the Virtual Residency Program** Brella
10:55am

Speakers: Henry Neeman, David Akin, Hussein Al-Azzawi, Kevin L. Brandt, Jamene Brooks Kieffer, Dana Brunson, Dirk Colbry, Sandra Gesing, Anna Klimaszewski-Patterson, Claire Mizumoto, Joy A. Pine-Thomas, Anita Z. Schwartz, Horst Severini, Mohammed Tanash, Daniel Voss

Cyberinfrastructure (CI) Facilitation is the process of helping researchers to use research computing systems and services to advance their computing/data-intensive research goals. The growing need for CI Facilitation isn't being met by traditional academic degree/certificate programs, so informal education is required. Since 2015, the Virtual Residency Program (VRP) has been teaching key CI Facilitation skills to pre-service and in-service CI Facilitators, at introductory, intermediate and advanced levels, via a combination of (a) workshops, (b) conference calls, and (c) apprenticeships. The demand for CI Facilitation has expanded exponentially, with the known CI Facilitator population growing from dozens in the late 2000s to over a thousand presently, so the need for this kind of training has become increasingly acute. The 2019 VRP workshop was presented at a mixed introductory/intermediate level, with content suitable for both new and experienced CI Facilitators, to attract both of these populations, in many cases presenting the same topic first at the introductory level and then intermediate, to maximize impact. Since 2015, the VRP has served 664 participants from 316 institutions in every US state and 3 US territories plus 8 other countries on 5 continents.

10:15am – W **NCSA Internship Program for Cyberinfrastructure Professionals** Brella
10:55am

Speakers: Daniel Lapine, Volodymyr Kindratenko, Luisa-Maria Rosu

In 2017, National Center for Supercomputing Applications (NCSA) at the University of Illinois at Urbana-Champaign (UIUC) established a pilot internship program for cyberinfrastructure (CI) professionals. The program, funded by NSF's Office of Advanced Cyberinfrastructure (OAC) (award #1730519), was designed to address the shortage of a workforce with the specialized skills needed to support advanced CI operations. The program was envisioned to provide internship opportunities for individuals who want to gain first-hand experience in the CI operations at a supercomputing center, and develop and refine instructional materials to serve as a template that is openly distributed for use by other centers and institutions to train CI professionals. Program interns are selected from a pool of applicants with the main selection criteria of having a completed classwork equivalent to an associate degree and a demonstrated interest in a career in CI operations. Interns work directly with a group of NCSA engineers in one of the areas of CI focus to gain hands-on experience in the deployment

and operation of high-performance computing (HPC) infrastructure at a leading HPC center. The expectation is that interns will enter a workforce that will develop, deploy, manage and support advanced CI at other universities, centers, and industry to meet the needs of the national computational science research community across academia and industry.

10:55am – 12:00pm	L	Lunch Break and Networking	Brella
10:55am – 12:00pm	S	Student Panel Discussion	Brella
12:00pm – 1:20pm	N	Cluster Usage Policy Enforcement Using Slurm Plugins and an HTTP API <i>Speakers: Matthew Li, Nicolas Chan, Viraat Chandra, Krishna Muriki</i> Managing and limiting cluster resource usage is a critical task for computing clusters with a large number of users. By enforcing usage limits, cluster managers are able to ensure fair availability for all users, bill users accordingly, and prevent the abuse of cluster resources. As this is such a common problem, there are naturally many existing solutions. However, to allow for greater control over usage accounting and submission behavior in Slurm, we present a system composed of: a web API which exposes accounting data; Slurm plugins that communicate with a REST-like HTTP implementation of that API; and client tools that use it to report usage. Key advantages of our system include a customizable resource accounting formula based on job parameters, preemptive blocking of user jobs at submission time, project-level and user-level resource limits, and support for the development of other web and command-line clients that query the extensible web API. We deployed this system on Berkeley Research Computing's institutional cluster, Savio, allowing us to automatically collect and store accounting data, and thereby easily enforce our cluster usage policy.	Brella
12:00pm – 1:20pm	N	ICEBERG: Imagery Cyber-infrastructure and Extensible Building blocks to Enhance Research in the Geosciences. (A Research Programmer's Perspective) <i>Speakers: Bradley Spitzbart, Heather Lynch, Shantenu Jha, Matteo Turilli</i> The ICEBERG (Imagery Cyber-infrastructure and Extensible Building blocks to Enhance Research in the Geosciences) project (NSF 1740595) aims to (1) develop open source image classification tools tailored to high-resolution satellite imagery of the Arctic and Antarctic to be used on HPDC resources, (2) create easy-to-use interfaces to facilitate the development and testing of algorithms for application to specific geoscience requirements, (3) apply these tools through use cases that span the biological, hydrological, and geoscience needs of the polar community, and (4) transfer these tools to the larger non-polar community. Here we report on the project status and lessons learned.	Brella
12:00pm – 1:20pm	N	Research Computing Infrastructure and Researcher Engagement: Notes from Neuroscience <i>Speakers: Rajendra Bose, Alexander Antoniadis, John Pellman</i> The advent of the Mortimer B. Zuckerman Mind Brain Behavior Institute (Zuckerman Institute) at Columbia University over a decade ago presented the opportunity to design a discipline-focused Research Computing (RC) group allowing for close collaboration with a relatively fixed number of neuroscience laboratories to enhance discovery. Experiences and observations related to tailoring Zuckerman computing infrastructure, creating "task-based" services and systems, and engaging with researchers in our Institute are shared to inform others about establishing discipline-focused research computing teams. Case studies related to providing a GPU cluster service tailored to Institute needs and the evolution of infrastructure choices to hybrid designs allowing bursting to vendor-provided cloud services are reviewed. Future directions involving research software engineering and sharing whole data analysis pipelines are noted.	Brella
12:00pm – 1:20pm	N	Using Containers to Create More Interactive Online Training and Education Materials <i>Speakers: Brandon Barker, Susan Mehringer</i> Containers are excellent hands-on learning environments for computing topics because they are customizable, portable, and reproducible. The Cornell University Center for Advanced Computing has developed the Cornell Virtual Workshop in high performance computing topics for many years, and we have always sought to make the materials as rich and interactive as possible. Toward the goal of building a more hands-on experimental learning experience directly into web-based online training environments, we developed the Cornell Container Runner Service, which allows online content developers to build container-based interactive edit and run commands directly into their web pages. Using containers along with CCRS has the potential to increase learner engagement and outcomes.	Brella

12:00pm –
1:20pm

C Campus Champions Fellow Presentation

Brella

The 2019-2020 XSEDE Campus Champions Fellows program partners Campus Champions with staff from XSEDE's Extended Collaborative Support Service (ECSS), Workforce Development, Cyberinfrastructure Resource Integration (CRI) and Cybersecurity groups to work side by side for a year on real-world science and engineering projects. Fellows will develop expertise within varied areas of cyberinfrastructure, and they are already well positioned to share their advanced knowledge through their roles as the established conduits to students, administrators, professional staff, and faculty on their campuses. In addition to the technical knowledge gleaned from their experiences, the individual Fellows will benefit from their personal interactions with the XSEDE staff and will acquire the skills necessary to manage similar user or research group project requests on their own campuses. The Campus Champions Fellows program is a unique opportunity for a select group of individuals to learn first-hand about the application of high-end cyberinfrastructure to challenging science and engineering problems. The 2019-20 Fellows who will be sharing their experiences over the past year are: Helen Kershaw, Brown University; Beau Christ, Wofford College; Mike Renfro, Tennessee Tech; and Kyle Hutson, Kansas State University.

12:00pm –
1:20pm

W Building a Gateway Infrastructure for Interactive Cyber Training and Workforce Development

Brella

Speakers: Lan Zhao, X. Carol Song, Larry Biehl, Venkatesh Merwade, Matthew Huber, Jing Liu, Uris Baldos, Ilya Shunko

Science gateways provide integrated data and computation support for research, education and online collaboration. To address the emerging shortage of adequately trained students and workforce in science and technology, more projects are utilizing gateways to disseminate training materials and online modules. However, the contents of these online learning modules are mostly limited to static materials, lacking support for dynamic and interactive learning. As part of the NSF funded GeoEDF project, we enhanced the HUBzero platform with novel capabilities to improve the online learning experience. These improvements include developing general purpose online tools for education purpose and implementing integrated data and tool functions in the HUBzero course module so that instructors can create online course outlines that seamlessly combine static teaching materials with dynamic data, tools, and interactive coding environments such as Jupyter Notebook and RStudio. In this paper we describe in detail the newly augmented HUBzero course environment implemented on MyGeoHub and showcase several projects that have been using these capabilities on MyGeoHub to support a broad spectrum of learning and training activities, including (1) development and delivery of adaptable cyber training modules for teaching undergraduate and graduate students on the FAIR (Findable, Accessible Interoperable and Reusable) science principles and applications in the fields of hydrology and climate sciences; (2) development and delivery of a short course and workshop tutorials on studying global sustainability using a gridded crop modeling system on MyGeoHub; (3) development of next-generation workforce through undergraduate internships in partnership with the Purdue Discovery Park Undergraduate Research Internship (DURI) program; and (4) outreach to middle school students entering 10th and 11th grade in the TOTAL (Turned Onto Technology and Leadership) summer camp, introducing them to geospatial data management and analysis using MyGeoHub cyberinfrastructure.

12:00pm –
1:20pm

W Fostering Collaboration Among Organizations in the Research Computing and Data Ecosystem

Brella

Speakers: Sharon Broude Geva, Dana Brunson, Thomas Cheatham, James Deaton, James Griffioen, Curt Hillegas, Douglas Jennewein, Gail Krovitz, Tobin Magle, Patrick Schmitz, Karen Tomko, James Wilgenbusch

The widespread application and success of computational and data intensive research approaches in every discipline represented on our campuses has resulted in a rapid proliferation of organizations, technologies, and professions affiliated in different ways with the support and advancement of activities related to research computing and data (RCD). While most agree that this growth is helping to advance numerous disciplines, the proliferation of organizations seeking to support, promote, and advance RCD has led to some challenges. Specifically, a lack of understanding and consensus concerning which organizations should be considered a part of research computing hampers our ability to encourage collaborations among its complementary constituents, leads to unneeded and redundant activities, and makes it difficult to identify strategic priorities and address gaps where specific needs are not being met to advance various disciplinary activities. In this paper we introduce the ecosystem metaphor to help characterize the rapidly changing relationships between the growing set of organizations that in some way support and enable activities related to RCD. The ecosystem concept lends itself well to describing the many entities related to RCD because it emphasizes the larger system over its individual component parts and highlights their interdependence, while explicitly expecting their change over time. Our work to characterize the current RCD ecosystem, while imperfect, will serve as a foundation and framework for the development of a more complete view of the ever-changing RCD ecosystem. A more complete view of the RCD ecosystem will in turn help to advance the broad goals of its members by helping to foster and accelerate new and meaningful collaborations among them.

12:00pm – 1:20pm	W From novice to expert: Supporting all levels of computational expertise in reproducible research methods Brella <i>Speakers: Kate Hertweck, Carly Strasser</i> Training and documentation for on-premises infrastructure represent the foundation of most institutional support for computational researchers. For most academic research institutions, however, these approaches fall short of meeting the needs of diverse researchers with different levels of experience with data-intensive research. We describe a framework for characterizing levels of computational expertise and relate this model to informational support provided for biomedical researchers at a non-profit/academic research center. Our model differentiates between novice, competent practitioner, and expert users of reproducible computational methods, and is related to the composition and needs of an entire research community. We specify methods best suited for researchers with different levels of expertise, including formally structured short courses, code examples/templates, and online wiki-style documentation. We provide recommendations to encourage the development and deployment of these resources, and suggest methods for assessing their effectiveness. Supporting multiple types of informational resources for researchers with different computational needs can be labor-intensive, but ideally increases computational ability for the entire institution.
12:00pm – 1:20pm	W Northeast Cyberteam: Workforce Development for Research Computing at Small and Mid-sized Institutions Brella <i>Speakers: John Goodhue, Julie Ma, Adrian Del Maestro, Sia Najafi, Bruce Segee, Scott Valcourt, Ralph Zottola</i> Computing has become an essential component of research and education for nearly every scientific discipline. Meeting the need for support staff who can help faculty make the best use of available computing resources is a significant challenge for small and mid-sized institutions. The NSF-sponsored Northeast Cyberteam is addressing this challenge by building a pool of research computing facilitators that can be shared across institutional boundaries while also developing self-service tools that reduce the support burden.
1:20pm – 1:35pm	R Break and Networking Brella
1:35pm – 3:35pm	A Bear: Cyberinfrastructure for Long-Tail Researchers at the Federal Reserve Bank of Kansas City Brella <i>Speakers: Bj Lougee, Michael Robinson, Chris Stackpole, Mark Watson, Greg Woodward</i> The Federal Reserve Bank of Kansas City has developed a new cyberinfrastructure environment uniquely tailored to the long-tail researchers in and around the field of economics. We based our design on our researchers' usage modalities and our experience delivering advanced research computing to domains that are now in more need of computational power and the ability to work with large datasets.
1:35pm – 3:35pm	A Demonstrating a Pre-Exascale, Cost-Effective Multi-Cloud Environment for Scientific Computing Brella <i>Speakers: Igor Sfiligoi, David Schultz, Benedikt Riedel, Frank Wuerthwein, Steve Barnett, Vladimir Brik</i> Scientific computing needs are growing dramatically with time and are expanding in science domains that were previously not compute intensive. When compute workflows spike well in excess of the capacity of their local compute resource, capacity should be temporarily provisioned from somewhere else to both meet deadlines and to increase scientific output. Public Clouds have become an attractive option due to their ability to be provisioned with minimal advance notice. The available capacity of cost-effective instances is not well understood. This paper presents expanding the IceCube's production HTCondor pool using cost-effective GPU instances in preemptible mode gathered from the three major Cloud providers, namely Amazon Web Services, Microsoft Azure and the Google Cloud Platform. Using this setup, we sustained for a whole workday about 15k GPUs, corresponding to around 170 PFLOP32s, integrating over one EFLOP32 hour worth of science output for a price tag of about \$60k. In this paper, we provide the reasoning behind Cloud instance selection, a description of the setup and an analysis of the provisioned resources, as well as a short description of the actual science output of the exercise.
1:35pm – 3:35pm	A Deploying large fixed files datasets with SquashFS and Singularity Brella <i>Speakers: Pierre Rioux, Gregory Kiar, Alexandre Hutton, Shawn T. Brown, Alan C. Evans</i> Shared high-performance computing (HPC) platforms, such as those provided by XSEDE and Compute Canada, enable researchers to carry out large-scale computational experiments at a fraction of the cost of the cloud. Most systems require the use of distributed filesystems (e.g. Lustre) for providing a highly multi- user, large capacity storage environment. These suffer performance penalties as the number of files increases due to network

contention and metadata performance. We demonstrate how a combination of two technologies, Singularity and SquashFS, can help developers, integrators, architects, and scientists deploy large datasets ($O(10M)$ files) on these shared systems with minimal performance limitations. The proposed integration enables more efficient access and indexing than normal file-based dataset installations, while providing transparent file access to users and processes. Furthermore, the approach does not require administrative privileges on the target system. While the examples studied here have been taken from the field of neuroimaging, the technologies adopted are not specific to that field. Currently, this solution is limited to read-only datasets. We propose the adoption of this technology for the consumption and dissemination of community datasets across shared computing resources.

1:35pm – A **Grendel: Bare Metal Provisioning System for High Performance Computing** Brella
3:35pm *Speakers: Andrew Bruno, Salvatore Guercio, Doris Sajdak, Matthew Jones, Tony Kew*

The Center for Computational Research (CCR) at the University at Buffalo has developed Grendel: a fast, easy to use, bare metal provisioning system for High Performance Computing (HPC). Grendel simplifies network booting racks of compute nodes by providing a robust PXE boot server, rest API, and node management in a single binary for easy installation. In this paper, we describe CCR's HPC network architecture and how Grendel was used to provision the center's Linux based compute clusters. We also present some modern features built into Grendel including automatic host discovery, deploying Live OS images to bare metal compute nodes, and delivering kernel, initramfs, and other provisioning assets using access tokens and trusted HTTPS.

1:35pm – A **Informing the on/o -prem cloud discussion in Higher Education** Brella
3:35pm *Speakers: Srijith Rajamohan, Robert Settlege*

The proliferation of external vendors providing public cloud services has resulted in a push to migrate away from on-premise compute and storage resources. The desire to move capital costs to operational costs, promises of lower maintenance and operational costs, lower failure rates, and ease of use have all played a part. However, for larger organizations with recurring and predictable usage patterns it needs to be investigated whether an on-premise and dedicated computational cluster might indeed remain more appropriate and cost-efficient.

1:35pm – A **Monitoring and Analysis of Power Consumption on HPC clusters using XDMoD** 🏆 Brella
3:35pm *Speakers: Joseph White, Martins Innus, Robert Deleon, Thomas Furlani, Matthew D. Jones*

🏆 **Phil Andrews Most Transformative Contribution award winner**
🏆 **Best Paper in “Advanced research computing environments – systems and system software” Track**

As part of the NSF funded XMS project we are developing tools and techniques for the audit and analysis of HPC infrastructure. This includes a suite of tools for the analysis of HPC jobs based on performance metrics collected from compute nodes. Although it may not be salient to the user, the energy consumption of an HPC system is an important part of the cost of maintenance and contributes a substantial fraction of the cost of calculations done with the system. We added support for energy usage analysis to the open-source XDMoD tool chain. This allows HPC centers to provide information directly to HPC stakeholders about the power consumption. This includes providing end users with energy usage information about their jobs as well as providing data to allow HPC center staff to analyze how the energy usage of the system is related to other system parameters. We explain how energy metrics were added to XDMoD and describe the issues we overcame in instrumenting a 1400 node academic HPC cluster. We present an analysis of 14 months of data collected on real jobs on the cluster. We performed a machine learning analysis of the data and show how energy usage is related to other system performance metrics.

1:35pm – N **A Science Gateway for Simulating the Economics of Carbon Sequestration Technologies: SimCCS2.0** Brella
3:35pm

Speakers: Sudhakar Pamidighantam, Marcus Christie, Jun Wang, Eroma Abeysinghe, Kevin Ellett, Ryan Kammer, Sean Yaw, Brendan Hoover, Richard Middleton

The SimCCS2.0 Gateway provides a science gateway for optimizing CO₂ capture, transport, and storage infrastructure. We describe the design, creation, and production deployment of this platform, which is based on an Apache Airavata gateway middleware framework. This gateway provides an integrated infrastructure for data, modeling, simulation, and visualization of carbon sequestration technologies and their economics. It does so through simple user interfaces to map and select input data, build models, set up, and execute simulations on high performance computing systems. Also feature are community case studies to use as reference sets for verifying reproducibility of published models and reusing their respective data for modified simulations. The portal addresses the needs of diverse international stakeholders and provides a platform for integrating novel and complex models for carbon sequestration technologies moving into the future.

1:35pm –
3:35pm

N **An extensible Django-based web portal for Apache Airavata**

Brella

Speakers: Marcus Christie, Eroma Abeysinghe, Suresh Marru, Sudhakar Pamidighantam, Marlon Pierce, Dimuthu Upeksha, Stephen Adithela, Eldho Mathulla, Aarushi Bisht, Shivam Rastogi

The Apache Airavata science gateway middleware project has developed a new web frontend for the middleware's API based on the Django web framework and the Vue.js JavaScript framework. This new frontend has been designed to be a framework, called the Airavata Django Portal Framework (ADPF) that science gateway developers can use to customize and extend the user interface to add domain specific UI metaphors and to add gateway-specific user workflows. There are three main modes of extensibility: 1) custom scientific application execution configuration, 2) custom application results analysis, and 3) wholly custom user workflows. These modes of extensibility come out of the project's experience working with science gateways over the years. This new framework has been put into production for the 30+ science gateways hosted by the Science Gateways as a Platform (SciGaP) project at Indiana University and several gateways have already made extensions using ADPF.

1:35pm –
3:35pm

N **Custos: Security Middleware for Science Gateways**

Brella

Speakers: Isuru Ranawaka, Suresh Marru, Juleen Graham, Aarushi Bisht, Jim Basney, Terry Fleury, Jeff Gaynor, Dimuthu Wannipurage, Marcus Christie, Alexandru Mahmoud, Marlon Pierce, Enis Afgan

Science gateways represent potential targets for cybersecurity threats to users, scientific research, and scientific resources. In this paper, we introduce Custos, a software framework that provides common security operations for science gateways, including user identity and access management, gateway tenant profile management, resource secrets management, and groups and sharing management. The goals of the Custos project are to provide these services to a wide range of science gateway frameworks, providing the community with an open source, transparent, and reviewed code base for common security operations; and to operate trustworthy security services for the science gateway community using this software base. To accomplish these goals, we implement Custos using a scalable microservice architecture that can provide highly available, fault tolerant operations. Custos exposes these services through a language-independent Application Programming Interface that encapsulates science gateway usage scenarios.

1:35pm –
3:35pm

N **FutureWater Indiana: A science gateway for spatio-temporal modeling of water in Wabash basin with climate change in focus**

Brella

Speakers: Sudhakar Pamidighantam, Jennifer Dierauer, Chen Zhu, Lei Gong, Alan Walsh, Eroma Abeysinghe, Marcus Christie, Jun Wang

In this manuscript, we describe the FutureWater Science Gateway that simulates regional watersheds spatially and temporally to derive hydrological changes due to changes in critical effectors such as climate, land use and management, and soil conditions. We also discuss the gateway design, creation, and production deployment and how the resulting data is organized and explored. The Future- Water gateway is built based on Apache Airavata gateway middleware framework and hosted under the SciGaP project at Indiana University. The gateway provides an integrated infrastructure for simulations based on parallelized Soil and Water Assessment Tool (SWAT) and SWAT-MODFLOW software execution on Extreme Science and Engineering Discovery Environment (XSEDE) and Indiana University's (IU's) HPC resources. It organizes data in optimized relational databases and enables intuitive simulation result data exploration. The visualization involves geographical map integration and dynamic data provisioning using the R-Shiny application deployed in the gateway. The gateway provides intuitively simple user interfaces for providing simulation input data and combines available model data; it makes it possible to set up and execute the he portal addresses the needs of diverse stakeholder communities for educational, research, exploration and planning in academic, governmental and NGO organizations.

1:35pm –
3:35pm

N **Reproducible and Portable Workflows for Scientific Computing and HPC in the Cloud**

Brella

Speakers: Peter Vaillancourt, Bennett Wineholt, Brandon Barker, Plato Deliyannis, Jackie Zheng, Akshay Suresh, Adam Brazier, Rich Knepper, Rich Wolski

The increasing availability of cloud computing services for science has changed the way scientific code can be developed, deployed, and run. Many modern scientific workflows are capable of running on cloud computing resources. Consequently, there is an increasing interest in the scientific computing community in methods, tools, and implementations that enable moving an application to the cloud and simplifying the process, and decreasing the time to meaningful scientific results. In this paper, we have applied the concepts of containerization for portability and multi-cloud automated deployment with industry-standard tools to three scientific workflows. We show how our implementations provide reduced complexity to portability of both the applications themselves, and their deployment across private and public clouds. Each application has been packaged in a Docker container with its dependencies and necessary environment setup for production runs. Terraform and Ansible have been used to automate the provisioning of compute resources and the deployment of each scientific application in a

Mult-VM cluster. Each application has been deployed on the AWS and Aristotle federated cloud platforms. Variation in data management constraints, Multi-VM MPI communication, and embarrassingly parallel instance deployments were all explored and reported on. We thus present a sample of scientific workflows that can be simplified using the tools and our proposed implementation to deploy and run in a variety of cloud environments.

1:35pm – 3:35pm N **Tapis API Development with Python: Best Practices In Scientific REST API Implementation - Experience implementing a distributed Stream API**  Brella

Speakers: Sean Cleveland, Anagha Jamthe, Smruti Padhy, Joe Stubbs, Julia Looney, Michale Packard, Steve Terry, Richard Cardone, Gwen Jacobs, Maytal Dahan

 **Best Paper in “Application Software, Support, and Outcomes” Track**

In the last decade, the rise of hosted Software-as-a-Service (SaaS) application programming interfaces (APIs) across both academia and industry has exploded, and simultaneously, microservice architectures have replaced monolithic application platforms for the flexibility and maintainability they offer. These SaaS APIs rely on small, independent and reusable microservices that can be assembled relatively easily into more complex applications. As a result, developers can focus on their own unique functionality and surround it with fully functional, distributed processes developed by other specialists, which they access through APIs. The Tapis framework, a NSF funded project, provides SaaS APIs to allow researchers to achieve faster scientific results, by eliminating the need to set up a complex infrastructure stack. In this paper, we describe the best practices followed to create Tapis APIs using Python and the Stream API as an example implementation illustrating authorization and authentication with the Tapis Security Kernel, Tenants and Tokens APIs, leveraging OpenAPI v3 specification for the API definitions and docker containerization. Finally, we discuss our deployment strategy with Kubernetes, which is an emerging orchestration technology and the early adopter use cases of the Streams API service.

3:00pm – 5:00pm C **Hackathon Final Presentations** Brella

3:35pm – 4:50pm B **Curating a Training Curriculum for Effective Use of Cloud Platforms in Research Workflows** Brella

Speakers: Dana Brunson, Patrick Schmitz, James von Oehsen, Erik Lundberg, Scott Yockel

Cloud computing is impacting every sector of computing and data as the scale and rapid deployment of new capabilities make these platforms more compelling, and yet the US Research and Education (R&E) community is not currently able to take full advantage of these platforms. A major impediment is the lack of suitable and sustainable training for the research support workforce. Navigating the multitude of training opportunities for cloud computing has proven challenging for institutions and research groups, both because current offerings are highly segmented and vendor specific, and as well because the materials are largely focused on enterprise IT staff and use-cases. We propose a Birds of a Feather session to bring together a community of research computing and data professionals to discuss how a community could gather and curate a shared set of training materials tailored to the R&E community, and develop online cohort-based training leveraging these materials.

3:35pm – 4:50pm B **How to Extend the Productive Lifetime of Scientific Computing Equipment** Brella

Speakers: Mark Dietrich, Robert Gardner, Andrew Chien, Hakizumwami Birali Runesha

Every year the scientific computing community decommissions \$100Ms of equipment (based on original cost). The reasons for this are varied -- limited machine room space, limited power and cooling, limited IT management staff, new incoming hardware, etc. But often this hardware is still useful scientifically -- and PIs may wish to continue to operate it. In short there is a loss of scientific capability, in addition to shutdown and disposal costs, and the generation of e-waste. Our vision is to create, foster, and share best practices for new pathways that extend the operating lifetime of scientific computing equipment.

3:35pm – 4:50pm B **Strategies and Tactics for Increasing Inclusivity and Diversity in Advanced Research Computing** Brella

Speakers: Amy Schuele, David Hart, Tabitha Samuel, Shava Smallen

Working in Advanced Research Computing (ARC) can be tough-we all face technical challenges each and every day. Unfortunately, other factors can make members of certain groups feel like they don't quite fit in, which makes the job even tougher. In this Birds of a Feather session, we will share the resources that we have discovered in our experiences in ARC and hear from our colleagues about what other activities exist in our community. We would then like to end with a call to action. What can YOU do to make this community more inclusive, more willing to listen and respect the views of others, to become a community that embraces our differences and thrives from the diversity of thought!

3:35pm –
4:50pm

B User Training and Engagement in Scientific Computing

Brella

Speakers: Arman Pazouki, Dhruva Chakravorty, Shelley Knuth

The primary objective of this Birds of a Feather (BoF) session is to facilitate knowledge transfer on researcher training approaches. The BoF will include participants from various stakeholder groups. The topics discussed during this session will include both the "learner's perspective" in the form of effective teaching and learning, and the "institutional perspective" in the form of support, technology, and collaborations. Models of metrics and surveys adopted to quantitatively measure the effectiveness of these approaches will also be discussed.



- A Advanced Research Computing Environments
 N Application Software Support Outcomes
 B BOF
- R Break
 C Co-located Event
 L Lunch Break
 M ML/AI
 P Panel
 Y Plenary
- V Poster/VIS Reception
 S Student Program
 T Tutorials
- W Workforce development Diversity Professionalization
 K Workshops

JULY 30 • THURSDAY

8:00am – 9:40am A **A Research Computing and Data Capabilities Model for Strategic Decision-Making** Brella
Speakers: Patrick Schmitz, Claire Mizumoto, John Hicks, Dana Brunson, Gail Krovitz, James Bottum, Joel Cutcher-Gershenfeld, Thomas Cheatham, Karen Wetzel
 Research Computing and Data is changing at an accelerating rate, while the range of fields and disciplines depending on the cyberinfrastructure is expanding and becoming increasingly diverse. This poses significant challenges to academic institutions as they try to effectively assess and plan for the necessary support infrastructure to keep pace with the needs of researchers. We present a Research Computing and Data Capabilities Model that identifies the range of relevant approaches to fully support and enable research computing and data on campuses. This model allows institutions to assess their current capabilities, and provides structured input into strategic decision making, using a shared community vocabulary. We describe the background of the Model, key concepts and features of the Model and an associated assessment tool, initial experience in the community and lessons learned, and a roadmap for further development.

8:00am – 9:40am A **Cloud and on-premises data center usage, expenditures, and approaches to ROI**{Cloud and on-premises data center usage, expenditures, and approaches to return on investment: A survey of academic research computing organizations} Brella
Speakers: Alan Chalker, Curtis Hillegas, Alan Sill, Sharon Broude Geva, Craig Stewart
 The landscape of research in science and engineering is largely reliant on computation and data processing. There is continued and expanded usage by disciplines that have historically used advanced computing resources, new usage by disciplines that have not traditionally used HPC, and new modalities for usage in emerging areas such as Data Science, Machine Learning, and other areas of AI. Along with these new patterns have come new advanced computing resource methods and approaches, including the availability of commercial cloud resources with capabilities comparable to on-premises facilities. The Coalition for Academic Scientific Computation (CASC) has long been an advocate representing the needs of academic researchers using computational resources, sharing best practices and offering advice to create a national cyberinfrastructure to meet US science, engineering, and other academic computing needs. CASC has begun what we intend to be an annual survey of academic cloud and data center usage and return on investment reporting practices to improve understanding of current patterns and changes over time. This paper presents the motivation for understanding this usage, gives an overview of interesting results from initial surveys, and outlines future steps towards developing a reliable longitudinal survey to assess trends in cloud and data center usage and return on investments in these areas. We note that on-premises computational resources, commercial cloud resources, and national cyberinfrastructure projects such as XSEDE (the eXtreme Science and Engineering Discovery Environment), OSG (Open Science Grid), and CaRCC (Campus Research Computing Consortium) all fall within the broad category of cyberinfrastructure (cite paper below). Cloud computing in particular is one form of, not something distinct from, cyberinfrastructure. Critically important findings from this first survey include the following: many of the respondents are engaged in some form of analysis of return in research computing investments, but only a minority currently report the results of such analyses to their upper-level administration. Most respondents are experimenting with use of commercial cloud resources but no respondent indicated that they have found use of commercial cloud services to create financial benefits compared to their current methods. There is clear correlation between levels of investment in research cyberinfrastructure and the scale of both cpu core-hours delivered and the financial level of supported research grants. Also interesting is that almost every respondent indicated that they participate in some sort of national cooperative or nationally provided research computing infrastructure project and most were involved in academic computing-related organizations, indicating a high degree of engagement by institutions of higher education in building and maintaining national research computing ecosystems. Institutions continue to evaluate cloud-based HPC service models, despite having generally concluded that so far cloud HPC is too expensive to use compared to their current methods.

- 8:00am – 9:40am A **Integrating Jupyter into Research Computing Ecosystems** Brella
Speakers: Joe Stubbs, Julia Looney, John Fonner, Gregory Zynda, Erik Ferlanti, Marjo Poindexter, Elias Chalhoub, Maytal Dahan, Matthew Vaughn
Beginning with the initial release of the DesignSafe JupyterHub in late 2015, TACC has been building and maintaining custom JupyterHub clusters for research groups across different domains of science and engineering. Today, TACC maintains five production JupyterHub systems utilizing over half a terabyte of memory and hundreds of CPU cores supporting nearly 1,600 unique users combined. In this paper, we describe our approach to utilizing JupyterHub in these cyberinfrastructure projects and our collaborative approach to integrating Jupyter into different research communities. For two such groups, we present an in-depth discussion of the science use cases and technical drivers that informed and evolved the design of our offering, and our outreach and engagement efforts to promote adoption. We discuss the implementation of custom features we have added to better integrate JupyterHub into other components of the cyberinfrastructure platforms, and we conclude with a description of our plans for future architecture and usage of JupyterHub at TACC.
-
- 8:00am – 9:40am A **SLATE: Monitoring Distributed Kubernetes Clusters** Brella
Speakers: Gabriele Carcassi, Joe Breen, Lincoln Bryant, Robert Gardner, Shawn Mckee, Christopher Weaver
The SLATE (Services Layer at the Edge) accelerates collaborative scientific computing through a secure container orchestration frame-work focused on the Science DMZ, enabling creation of advanced multi-institution platforms and novel science gateways. The goal of the SLATE project is to provide a secure federation platform to simplify deployment and operation of complex and often specialized services required by multi-institution scientific collaborations, utilizing where applicable open source, cloud native tooling such as Kubernetes. This paper outlines the design and operation of a monitoring infrastructure suitable for application developers and resource providers which gives visibility to resource utilization and service deployments across a network of independently managed Kubernetes clusters.
-
- 8:00am – 9:40am N **MetaFlowlmics: Scalable and Reproducible Nextflow Pipelines for the Analysis of Microbiome Marker Data** Brella
Speakers: Cedric Arisdakessian, Sean Cleveland, Mahdi Belcaid
Computational scalability has become an important requirement for processing the massive amounts of data generated in contemporary sequencing-based experiments. The availability of large computational resources through academic, regional or national cyber-infrastructure efforts, as well as through inexpensive cloud offerings, has shifted the bottleneck, which now lies in the extensive expertise necessary to create reproducible and scalable bioinformatics pipelines and deploy them to such diverse infrastructures. We present here MetaFlowlmics, a comprehensive pipeline for the analysis of microbiome marker data using reproducibility, best practices and state-of-the-art cyberinfrastructure standards. MetaFlowlmics provides seamless scalability and extensibility, allowing users to build and test their pipelines on a laptop with small datasets and to subsequently run them on large datasets on an HPC or on the Cloud with a change to a single line of code. Our framework is built on top of the Nextflow workflow management system and provides an interoperable architecture that leverages self-contained Docker and Singularity instances with all the dependencies and requirements needed to quickly deploy and use the pipeline.
-
- 8:00am – 9:40am N **NLP Workflows for Computational Social Science: Understanding Triggers of State-Led Mass Killings** Brella
Speakers: Tim Burley, Lorissa Humble, Charles Sleeper, Abigail Sticha, Angela Chesler, Patrick Regan, Paul Brenner, Ernesto Verdeja
We leverage statistical and natural language processing (NLP) tools for a systematic analysis of triggers of state-led mass killings. The work advances the application of statistics and NLP in the social sciences and also contributes to scholarly efforts by empirically identifying the prominent triggering events of civilian mass killings. More specifically we seek to understand the timing and dynamics of political violence escalation, by examining systematically how certain types of political events may generate a government's policy of mass killing of civilians. The project provides pathways for the general application of promising NLP and statistical methods to the analysis of social event triggers as gleaned from big data repositories. Key objectives include: 1) To develop open source natural language processing (NLP) dictionaries and inference engines for event identification from texts, which are especially valuable for the analysis of political conflict and 2) Construct and validate a computational workflow to machine code millions of news articles (via NLP) for event identification, from a volume of data orders of magnitude larger than could be manually coded by a team of human readers. Having made considerable progress over multiple semesters, we share the methods and tools that have enabled us to overcome significant computational data analytics challenges.

8:00am – 9:40am	<p>N Scientific Data Annotation and Dissemination: Using the 'Ike Wai Gateway to Manage Research Data Brella</p> <p><i>Speakers: Sean Cleveland, Gwen Jacobs, Jennifer Geis</i></p> <p>Granting agencies invest millions of dollars on the generation and analysis of data, making these products extremely valuable. However, without sufficient annotation of the methods used to collect and analyze the data, the ability to reproduce and reuse those products suffers. This lack of assurance of the quality and credibility of the data at the different stages in the research process essentially wastes much of the investment of time and funding and fails to drive research forward to the level of potential possible if everything was effectively annotated and disseminated to the wider research community. In order to address this issue for the Hawai'i Established Program to Stimulate Competitive Research (EPSCoR) project, a water science gateway was developed at the University of Hawai'i (UH), called the 'Ike Wai Gateway. In Hawaiian, 'Ike means knowledge and Wai means water. The gateway supports research in hydrology and water management by providing tools to address questions of water sustainability in Hawai'i. The gateway provides a framework for data acquisition, analysis, model integration, and display of data products. The gateway is intended to complement and integrate with the capabilities of the Consortium of Universities for the Advancement of Hydrologic Science's (CUAHSI) Hydroshare by providing sound data and metadata management capabilities for multi-domain field observations, analytical lab actions, and modeling outputs. Functionality provided by the gateway is supported by a subset of the CUAHSI's Observations Data Model (ODM) delivered as centralized web based user interfaces and APIs supporting multi-domain data management, computation, analysis, and visualization tools to support reproducible science, modeling, data discovery, and decision support for the Hawai'i EPSCoR 'Ike Wai research team and wider Hawai'i hydrology community. By leveraging the Tapis platform, UH has constructed a gateway that ties data and advanced computing resources together to support diverse research domains including microbiology, geochemistry, geophysics, economics, and humanities, coupled with computational and modeling workflows delivered in a user friendly web interface with workflows for effectively annotating the project data and products. Disseminating results for the 'Ike Wai project through the 'Ike Wai data gateway and Hydroshare makes the research products accessible and reusable.</p>
8:00am – 9:40am	<p>N VisSnippets: A Web-Based System for Impromptu Collaborative Data Exploration on Large Displays  Brella</p> <p><i>Speakers: Andrew Burks, Luc Renambot</i></p> <p> Student paper – Honorable Mention in “Application Software, Support, and Outcomes” Track</p> <p>The VisSnippets system is designed to facilitate effective collaborative data exploration. VisSnippets leverages SAGE2 middleware that enables users to manage the display of digital media content on large displays, thereby providing collaborators with a high-resolution common workspace. Based in JavaScript, VisSnippets provides users with the flexibility to implement and/or select visualization packages and to quickly access data in the cloud. By simplifying the development process, VisSnippets removes the need to scaffold and integrate interactive visualization applications by hand. Users write reusable blocks of code called "snippets" for data retrieval, transformation, and visualization. By composing dataflows from the group's collective snippet pool, users can quickly execute and explore complementary or contrasting analyses. By giving users the ability to explore alternative scenarios, VisSnippets facilitates parallel work for collaborative data exploration leveraging large-scale displays. We describe the system, its design and implementation, and showcase its flexibility through two example applications.</p>
9:40am – 10:00am	<p>R Break and Networking Brella</p>
9:40am – 10:00am	<p>S Volunteer Team Standup Brella</p>
10:00am – 11:00am	<p>A Cyberinfrastructure of a Multi-Petabyte Microscopy Resource for Neuroscience Research Brella</p> <p><i>Speakers: Kathy Benninger, Greg Hood, Derek Simmel, Luke Tuite, Arthur Wetzel, Simon Watkins, Alan Watson, Marcel Bruchez, Alexander Ropewski</i></p> <p>Advances in microscopy technologies and techniques are driving development of an integrated cyberinfrastructure to manage the vast amounts of image data being generated. To address this need, a team of neuroscientists and cyberinfrastructure engineers has integrated mass storage, networking, and high-performance computing components to create the Brain Image Library (BIL). BIL serves the neuroscience community by providing a persistent repository for the massive amounts of data being generated (multi-petabyte scale) and by offering brain researchers supercomputing-class processing and visualization capabilities for</p>

-
- 10:00am – 11:00am A **Frontera: The Evolution of Leadership Computing at the National Science Foundation** Brella
Speakers: John West, Dan Stanzione, Todd Evans, Tommy Minyard, Dhableswar Panda, Omar Ghattas
As part of the NSF's cyberinfrastructure vision for a robust mix of high capability and capacity HPC systems, Frontera represents the most recent evolution of trans-petascale resources available to all open science research projects in the U.S. Debuting as the fifth largest supercomputer in the world, Frontera represents a robust and well-balanced HPC system designed to enable large-scale, productive science on day one of operations. The system provides a primary compute capability of nearly 39PF, delivered completely via more than 8,000 dual-socket servers with conventional Intel 8280 ("Cascade Lake") processors. A unique configuration of both desktop GPUs and advanced floating units from NVIDIA enables both machine learning and scientific workloads, and the system delivers nearly 2TB/s of total filesystem bandwidth with 55 PB of usable Lustre disk-based storage and 3PB of all flash Lustre storage. A Mellanox InfiniBand (IB) interconnect provides very low latency with 100Gbps to each node, and 200Gbps between switches in a fat tree topology with minimal oversubscription for efficient communication, even in jobs that use the full system with complex communication patterns. The system hardware is complemented by a robust set of software services, including Application Programmer Interfaces (APIs) to support an evolving user base that increasingly demands productive access via science gateways and automated workflows, as well as a first-of-its-kind partnership with the three major cloud service providers to create a bridge between "traditional" HPC and the cloud infrastructure upon which research increasingly depends.
-
- 10:00am – 11:00am A **HAL: Computer System for Scalable Deep Learning** Brella
Speakers: Volodymyr Kindratenko, Dawei Mu, Yan Zhan, John Maloney, Sayed Hadi Hashemi, Ben Rabe, Ke Xu, Roy Campbell, William Gropp, Jian Peng
We describe design, deployment and operation of a computer system purposely built to efficiently run deep learning frameworks. The system consists of 16 IBM POWER9 servers with 4 NVIDIA V100 GPUs each, interconnected with Mellanox EDR InfiniBand fabric, and a DDN all-flash storage array. The system is tailored towards an efficient execution of IBM Watson Machine Learning enterprise software stack that combines popular open-source deep learning frameworks. We build a custom management software stack to enable an efficient use of the system by a diverse community of users and provide guides and recipes for running deep learning workloads at scale utilizing all available GPUs. We demonstrate scaling of a PyTorch based deep neural network to produce state-of-the-art performance results.
-
- 10:00am – 11:00am N **Design and Deployment of Photo2Building: A Cloud-based Procedural Modeling Tool as a Service** Brella
Speakers: Manush Bhatt, Rajesh Kalyanam, Gen Nishida, Liu He, Christopher May, Daniel Aliaga, Dev Niyogi
We present a Photo2Building tool to create a plausible 3D model of a building from only a single photograph. Our tool is based on a prior desktop version which, as described in this paper, is converted into a client-server model, with job queuing, web-page support, and support of concurrent usage. The reported cloud-based web-accessible tool can reconstruct a building in 40 seconds on average and costing only 0.60 USD with current pricing. This provides for an extremely scalable and possibly widespread tool for creating building models for use in urban design and planning applications. With the growing impact of rapid urbanization on weather and climate and resource availability, access to such a service is expected to help a wide variety of users such as city planners, urban meteorologists worldwide in the quest to improved prediction of urban weather and designing climate-resilient cities of the future.
-
- 10:00am – 11:00am N **Environmental Visualization: Moving Beyond the Rainbows** Brella
Speakers: Stephanie Zeller, Paul Navartil, Francesca Samsel
Pseudo-coloring is a well-established, fundamental tool for visualizing scientific data. As the size and density of data grows, increasingly more discriminatory power is required to extract optimum feature resolution. The environmental community, in particular, relies heavily on this technology to dissect and interpret a huge variety of visual data. These scientists often turn to traditional rainbow colormaps, despite their well-documented deficiencies in rendering dense detail. A popular default, the desaturated rainbow's non-monotonically varying luminance range misrepresents data. Despite increasing overall feature resolution, this variance creates hue simultaneity and vibration, introducing false artifacts, neutralizing swaths of data and impeding analysis. Drawing on artistic color theory, we hypothesized the desaturated rainbow could be improved by increasing luminance ranges, decreasing saturation, and employing hue-cycling to boost discriminatory power. These adjusted maps

exhibit algorithmically corroborated higher feature resolve, a primary objective of all scientists interviewed, without distorting data in discordant false coloring. Our studies indicate that our maps are preferred by these domain scientists, thereby providing a potential alternative for effective, human-centric colormapping.

10:00am – 11:00am	N	The Hawai'i Rainfall Analysis and Mapping Application (HI-RAMA): Decision Support and Data Visualization for Statewide Rainfall Data	Brella
		<i>Speakers: Jared H. McLean, Sean B. Cleveland, Matthew Lucas, Ryan Longman, Thomas W. Giambelluca, Gwen Jacobs, Jason Leigh</i>	
		This paper discusses the design and implementation of the Hawai'i Rainfall Analysis and Mapping Application (HI-RAMA) decision support tool, an application providing researchers and community stakeholders interactive access to and visualization of hosted historical and near-real-time monthly rainfall maps and aggregated rainfall station observational data for the State of Hawai'i. The University of Hawai'i Information Technology Services Cyberinfrastructure team in partnership with members of the the Hawai'i Established Program to Stimulate Competitive Research (EPSCoR) 'Ike Wai project team developed the HI-RAMA as part of the 'Ike Wai Gateway to support water sustainability research for the state of Hawai'i. This tool is designed to provide user-friendly access to the information that can reveal the impacts of climate changes related to precipitation so users can make data-driven decisions.	
10:00am – 11:00am	C	Hackathon Awards/Closing	Brella
11:00am – 12:00pm	L	Lunch Break and Networking	Brella
12:00pm – 1:20pm	A	An Open Ecosystem for Pervasive Use of Persistent Identifiers	Brella
		<i>Speakers: Rick Wagner, Mike D'Arcy, Kyle Chard, Rachana Ananthakrishnan, Ian Foster, Jim Pruyne, Brendan McCollam, Robert Schuler, Carl Kesselman, Philippe Rocca-Serra</i>	
		Persistent identifiers (PIDs) are essential for making data Findable, Accessible, Interoperable and Reusable, or FAIR. While the advantages of PIDs for data publication and citation are well understood, and Digital Object Identifiers (DOIs) for data are increasingly widely used, there are two gaps in the current identifier ecosystem: 1) a single interface that provides a consistent baseline of capabilities encompassing key aspects of the research data lifecycle, including canonical landing pages and machine-readable metadata via the same URL; and 2) support for identifiers for more ephemeral data, particularly as data move across system boundaries, such as during workflows. To address these gaps, we have implemented, and made freely available as open source software, the FAIR Research Identifiers broker service. This service supports multiple identifier providers (ARK, Handle, DOI via DataCite, etc.) and uses Globus Auth to authorize identifier creation for a given provider and account. This paper provides a summary of the current identifier ecosystem, presents recommendations for identifier users, and describes our FAIR Research Identifiers service.	
12:00pm – 1:20pm	A	JetLag: An Interactive, Asynchronous Array Computing Environment	Brella
		<i>Speakers: Steven R. Brandt, Alex Bigelow, Sayef Sakin, Katy Williams, Katherine E. Isaacs, Kevin Huck, Rod Tohid, Bibek Wagle, Hartmut Kaiser, Shahrzad Shirzad</i>	
		We describe an interactive computing environment called JetLag. JetLag implements the following features of Phylanx project: (1) Phylanx, a Python-based asynchronous array computing toolkit; (2) the APEX performance measurement library; (3) a performance visualization framework called Traveler; (4) the Tapis/Agave Science as a Service middleware; and (6) a container infrastructure that includes Docker-based Jupyter notebook for the client and a singularity image for the server. The running system starts with a user performing array computations on their workstation or laptop. If, at some point, the calculation the user is performing becomes sufficiently intensive or numerous, it can be packaged and sent to another machine where it will run (through the batch queue system if there is one), produce a result, and have that result sent back to the user's local interface. Whether the calculation is local or remote, the user will be able to use APEX and Traveler to diagnose and fix performance related problems. The JetLag system is suitable for a variety of array computational tasks, including machine learning and exploratory data analysis.	
12:00pm – 1:20pm	A	OAuth SSH with Globus Auth	Brella
		<i>Speakers: Kyle Chard, Jason Alt, Rachana Ananthakrishnan, Ryan Chard, Ian Foster, Lee Liming, Steven Tuecke</i>	
		The Secure Shell (SSH) protocol and its OpenSSH implementation are a cornerstone of modern scientific computing as they enable users to access remote systems, transfer data, and execute programs. We describe	

extensions to the popular OpenSSH software that enables authentication with OAuth tokens from Globus Auth, rather than passwords or keys. Integration with Globus Auth allows users to use hundreds of supported identity providers, and enables external applications and services to obtain short-term tokens on behalf of users for securely accessing remote systems.

12:00pm – A **Workflow Submit Nodes as a Service on Leadership Class Systems**  Brella
1:20pm *Speakers: George Papadimitriou, Karan Vahi, Jason Kincl, Valentine Anantharaj, Ewa Deelman, Jack Wells*

 **Best Student Paper in “Advanced research computing environments – systems and system software” Track**

DOE scientists, today, have access to high performance computing (HPC) facilities with very powerful systems that enable them to execute their computations faster, more efficiently, and at greater scales than ever before. To further their knowledge and produce new discoveries, scientists rely on workflows - sometimes very complex - that provide them with an easy way to automate, reproduce and verify their computations. However, historically, creating workflow submission environments in large HPC facilities has been cumbersome, requires expertise and many man-hours of effort due to the peculiarities, policies, and the restrictions that these systems present. In this paper we discuss the approach a large DOE facility (OLCF) is taking in order to provide containers as a service to its users. This capability is used to create Pegasus workflow management system submit nodes as a service (WSaaS) at the Oak Ridge Leadership Computing Facilities (OLCF), targeting the Summit supercomputer. This deployment builds upon the Kubernetes/OpenShift cluster (Slate) that exists within OLCF's DMZ and its automation triggers. Additionally, we evaluate our approach's overhead and effort to deploy the solution as compared to previous solutions, such as setting up a Pegasus submission environment on OLCF's login nodes or submitting jobs remotely via the rvGAHP.

12:00pm – N **Accelerated Real-time Network Monitoring and Profiling at Scale using OSU INAM** Brella
1:20pm *Speakers: Pouya Kousha, Kamal Raj Sankarapandian Dayala Ganesh Ram, Hari Subramoni, Dhabaleswar Panda, Heechang Na, Trey Dockendorf, Karen Tomko*

Designing a scalable real-time monitoring and profiling tool with low overhead for network analysis and introspection capable of capturing all relevant network events is a challenging task. Newer set of challenges come out as HPC systems are becoming larger and users are expecting to have better capabilities like real-time profiling at fine granularity. We take up this challenge by redesigning OSU INAM and making it capable to gather, store, retrieve, visualize, and analyze network metrics for large and complex HPC clusters. The enhanced OSU INAM tool provides scalability, low overhead and fined-granularity InfiniBand port counter inquiry and fabric discovery for HPC users, system administrators, and HPC developers. Our experiments show that, for a cluster of 1,428 nodes and 114 switches, the proposed design can gather fabric metrics at very fine (sub-second) granularity and discovers the complete network topology in approximately 5 minutes. The proposed design has been released publicly as a part of OSU INAM Tool and is available for free download and use from the project website.

12:00pm – N **Implementing a Loosely-Coupled Integrated Assessment Model in the Pegasus Workflow Management System** Brella
1:20pm *Speakers: Patrick Clemins, Scott Turnbull, Morgan Rodgers, Asim Zia*

Integrated assessment models (IAMs) are commonly used to explore the interactions between different modeled components of socio-environmental systems (SES). Most IAMs are built in a tightly-coupled framework so that the complex interactions between the models can be efficiently implemented within the framework in a straight forward manner. However, tightly-coupled frameworks make it more difficult to change individual models within the IAM because of the high level of integration between the models. Prioritizing flexibility over computational efficiency, the IAM presented here is built using a loosely-coupled framework and implemented in the Pegasus Workflow Management System. The modular nature of loosely-coupled systems allows each component model within the IAM to be easily exchanged for another component model from the same domain assuming each provides the same input / output interface. This flexibility allows researchers to experiment with different models for each SES component and facilitates simple upgrades between each version of the independently developed component models.

12:00pm – W **Best Practices in Project Management in a Large, Distributed Organization: Lessons Learned from XSEDE** Brella
1:20pm *Speakers: Leslie Froeschl, Marques Bland, Scott Wells*

Having professional project management support has proven essential to the success and effectiveness of the Extreme Science and Engineering Discovery Environment (XSEDE) project. The value and level of investment needed for project management support in a large, complex, and highly distributed project should not be overlooked by others seeking to establish a distributed organization of this nature. This paper shares what the XSEDE Project Management & Reporting (PM&R) team believes are best practices for project management in a large, virtual organization such as XSEDE.

1:20pm – R **Break and Networking** Brella
2:20pm

1:35pm – Y **Plenary III - Dr. Amy Friedlander, (Acting) Office Director of the Office of Advanced
3:35pm Cyberinfrastructure Directorate of Computer and Information Science and Engineering, National
Science Foundation** Brella

Speakers: Amy Friedlander, Joseph Paris

Title: Envisioning the 21st Century Scientific Research Cyberinfrastructure: Stability, Innovation, and the Science of the Future

The central problem in technology-driven large-scale infrastructure systems is providing services that are predictable, reliable and broadly accessible without allowing them to become obsolete, stale, and under-resourced. For the cyberinfrastructure (CI), this problem is particularly acute as the technologies themselves advance very rapidly, user demand and expectations escalate, and the cyberinfrastructure is itself dynamic. This talk explores ways that the Office of Advanced Cyberinfrastructure (OAC) at the National Science Foundation manages this tension to provision the national scientific and engineering research CI while enabling innovation in the CI systems and services and nurturing the development of new capabilities to transform the next generation of scientific research.

Biography

Dr. Amy Friedlander is currently the Acting Office Director for the Office of Advanced Cyberinfrastructure, Directorate for Computer and Information Science and Engineering (CISE/OAC), where she has served as Acting Deputy Division Director and Deputy Office Director since November, 2014. Since joining NSF in 2010, she has led several strategic activities, including both initial coordination of the Public Access Initiative and the activities that culminated in the widely-distributed report *Rebuilding the Mosaic (2011)*. In addition to her position in OAC, Dr. Friedlander plays a central role in NSF's data management policies and activities that address the foundation's administrative data as well as the research data resulting from NSF's investments.

Prior to her NSF appointment, Dr. Friedlander held positions in the private non-profit and for-profit sectors. Among other projects, she participated in the Blue Ribbon Task Force on Sustainable Digital Preservation and Access, funded largely by NSF; led the initial strategic planning for the Library of Congress' National Digital Information Infrastructure and Preservation Program (NSIIPP); and served as editor-in-chief of the *ACM Journal on Computing and Cultural Heritage*. At the Corporation for National Research Initiatives (CNRI), she was the founding editor of *D-Lib Magazine* (www.dlib.org) and the author of a series of studies of the historical development large-scale technology infrastructures in the U.S.

Dr. Friedlander graduated from Vassar College, where she was elected to Phi Beta Kappa. She holds the M.A. and Ph.D. from Emory University and the M.S.L.I.S. from The Catholic University of America. She pursued postdoctoral work on quantitative methods and computer-assisted social science research at the Newberry Library in Chicago, IL.

Following Dr. Friedlander's plenary, the PEARC21 Chair, Joe Paris, will announce next year's PEARC conference.

3:30pm – S **Student Program Wrap-up** Brella
4:30pm Student Program Wrap-up

3:35pm – B **Bridging the Data Transfer Gap: An Open Discussion between Researchers, Administrators, and
3:50pm Network Engineers** Brella
Speakers: Richard Angeletti, Kathy Benninger, Tom Hutton, Chris Jordan, Matt Kollross, Michael

Lambert, Rick Mohr, Tabitha Samuel

Data transfers are a critical part of nearly every modern scientific workflow. The data sizes may range from small scripts written on a personal laptop to massive results generated by specialized scientific instruments. The transfer distances may span multiple networks across continents or simply across the street to another building. Regardless of the details, the efficiency of the workflow is inherently tied to the efficiency of the data transfer. The goal of this BoF is to open a dialog among all the stakeholders in these transfers: researchers, system administrators, network engineers, and other cyberinfrastructure personnel. Staff members of the XSEDE Data Transfer Services team will facilitate discussions about identifying and reporting data transfer issues, tools and services available to the community, and addressing data transfer performance issues.

3:35pm –
3:50pm

B COVID19: A Time of Trial and Unity in Research Computing

Brella

Speakers: Rebecca Belshe, Marisa Brazil, Christopher Coffey, Jeremy Frumkin, Douglas Jennewein, Chris Kurtz, Gil Speyer, Jason Yalim

This panel of tri-university speakers will discuss experiences and challenges faced due to the COVID-19 pandemic, and the collaborative efforts to address challenges as a team and a state-wide community through the Arizona Research Computing (ARC) initiative. The purpose of this panel is to discuss the insights gained by research computing institutions in this time of crisis, and the efforts made by those groups to resolve and overcome the associated hardships. We hope by sharing lessons learned from this experience, that this discussion will lead to greater knowledge within our community. This is an opportunity to share the lessons and challenges that have been experienced with the hope that broader discussions and collaborations will result.

3:35pm –
3:50pm

B Preparing EPSCoR institutions for Research Computing and Data Capabilities Model engagement

Brella

Speakers: Scotty Strachan, Venice Bayrd, Patrick Schmitz, Dana Brunson, Gwen Jacobs, Pips Veazey

This Birds of a Feather (BoF) group session is aligned with other PEARC20 workshops and sessions associated with the Research Computing and Data (RCD) Capabilities Model. This model allows institutions to assess their support for computationally- and data-intensive research, to identify potential areas for improvement, and to understand how the broader community views RCD support. Application of this model within National Science Foundation Established Program to Stimulate Competitive Research (NSF-EPSCoR) jurisdictions presents a set of potential challenges beyond typical expectations of model engagement and response quality. Institutions classified as NSF-EPSCoR-eligible are typically underfunded for research infrastructure and very often have minimal organized cyberinfrastructure support. As a result of this, participation in the RCD Capabilities Model assessment survey is likely to require additional preparation and outside assistance to ensure adequate survey engagement by institutional research and administrative personnel. This BoF, which is designed to kick off a series of EPSCoR-focused engagement events that will culminate in a full PEARC21 workshop, brings together RCD survey authors and research technology support personnel from NSF-EPSCoR institutions to review ongoing challenges, discuss current survey efforts, and provide feedback to the Model development process. Non-EPSCoR institutions with interest in the survey are welcome to participate.

3:35pm –
3:50pm

B The Cyberteam Portal: A Shared, yet Independent, Platform for Cyberteam Development

Brella

Speakers: Julie Ma, John Goodhue, Adrian Del Maestro, Shelley Knuth, Dhruva Chakravorty, Timothy Middelkoop, Sia Najafi, James Griffioen, Bruce Segee, Scott Valcourt, Ralph Zottola

The Cyberteam Portal was developed to support the management of project workflows and capture project results for a suite of over 40 projects launched as part of the Northeast Cyberteam Initiative, an NSF-sponsored program that aims to make regional and national cyberinfrastructure more readily accessible to researchers in small and mid-sized institutions. The portal performs four principal functions: manage project workflows and capture project results; manage cyberteam participants; match students and mentors with projects; and provide a front door to self-service training/learning resources. The portal was developed with an eye toward making it possible for other Cyberteam programs to adopt it without incurring significant cost, while maintaining their own identity and project workflows. The Great Plains CyberTeam, Rocky Mountain Advanced Computing Consortium (RMACC), SouthWest Expertise in Expanding, Training, Education and Research (SWEETER), and Kentucky Cyberteam programs have started exploring the use of the platform as a management tool for their related programs in a pilot program. We share our preliminary results at this BoF session and invite discussion and additional participation.

5:00pm –
6:00pm

R PEARC20 Virtual Happy Hour 🍷

Brella

Speakers: Gwen Jacobs

Please join PEARC20's General Chair, Gwen Jacobs, in celebration of our first virtual PEARC20 conference. Grab a drink of your choice and mingle via zoom with other conferece attendees, PEARC20 committee

members and PEARC21 committee members.

A Advanced Research Computing Environments **N** Application Software Support Outcomes **B** BOF

R Break **C** Co-located Event **L** Lunch Break **M** ML/AI **P** Panel **Y** Plenary

V Poster/VIS Reception **S** Student Program **T** Tutorials

W Workforce development Diversity Professionalization **K** Workshops

JULY 31 • FRIDAY

8:00am – **P** **FABRIC: Creating a Testbed of Testbeds** Brella

9:30am *Speakers: Ilya Baldin, James Griffioen, Inder Monga, Anita Nikolich, Kuang-Ching Wang*

Organizers:

Ilya Baldin (UNC RENCi)
James Griffioen (University of Kentucky)
Inder Monga (ESnet)
Anita Nikolich (University of Illinois)
Kuang-Ching Wang (Clemson University)

Speakers:

James Griffioen (University of Kentucky) -- FABRIC Overview
Ben Kirtman (University of Miami) -- Scientific Computing
Jason Mancuso (Cape Privacy) -- Distributed Machine Learning
Christos Papadopoulos (Colorado State) -- IoT and Scientific Devices
Ilkay Altintas (UC San Diego) -- Data Analysis and Workflows

Panel Abstract:

FABRIC is a nation-wide programmable network that will enable creation of a new, secure, flexible Internet. Running over the ESnet optical backbone with links up to 1 Tbps, it interconnects supercomputer facilities, cloud testbeds, NSF PAWR (wireless) testbeds, commercial clouds, the real Internet and powerful AI-enabled edge nodes at universities and labs. A key contribution will be FABRIC's programmability and Quality of Service (QoS).

Unlike equipment in the current Internet that simply forwards packets, nodes in FABRIC have significant computational capabilities that allow users to run computationally intensive programs throughout the network, applying application-specific routing, processing/transformation, and storing/caching of data in the network. Each node contains large amounts of compute, storage, and specialized devices such as GPUs, FPGAs, and network processors. The ability to design and deploy programs that run at any node in the network as opposed to only at the edge or in the cloud enables a completely new paradigm for developing HPC-related applications that don't rely solely on large centers.

This panel will provide an introduction to FABRIC. We will then present four example research projects in the areas of Scientific Computing, Distributed Machine/Deep Learning, IoT and Scientific Devices, and Data Analysis and Workflows. Panelists will describe ways to leverage FABRIC's features to significantly enhance their research. Attendees will be encouraged to think outside-the-box and brainstorm ways that FABRIC could be leveraged to advance their own research.

8:00am – **P** **Lessons from the Field, Resources to Advance Broadening Participation in Computing Efforts** Brella

9:30am *Speakers: Sarah Dunton, Carol Fletcher, John Goodhue, Jeffrey Forbes, Crystal Franklin*

Broadening participation in computing (BPC) poses a unique set of opportunities and challenges for research computing practitioners. Research computing is consistently on the cutting edge, and constantly evolving to meet new needs in research and education. As the field has grown, the demographic diversity of the students pursuing computer science, from K-12 to higher education and research or industry, continues to lag (Stanton, Goldsmith, Adrion, Dunton, Hendrickson, Peterfreund, Youngpradit, Zarch, Dounnay, 2017; Code.org, 2018; Code.org, CSTA, ECEP, 2019). Explicit attention to broadening participation in computing, at all levels of computer science and STEM education, is needed to ensure that the teams working on solutions to the major challenges of our

time represent the diversity of individuals who are impacted by them. This panel is designed for attendees seeking to develop BPC outreach efforts, write funding proposals, or who want to learn about BPC tools and resources. The panelists are leading and developing BPC efforts at the local, state and national level in a broad range of arenas, including K-12 education, advanced computing, higher education, informal education, and government policymaking. Panelists will highlight ways in which the advanced computing community can build on and integrate pre-existing BPC efforts, to bridge K-12 CS education and advocacy, CS outreach, and advanced computing. Each panelist will reflect on the trajectory of their BPC efforts, highlighting the networks in which they work, resources they have created, and the projects that they are currently leading. Panelists will outline best practices and recommendations for developing BPC outreach strategies to engage underrepresented populations at all levels of computing from K-12 to careers. Attendees should bring questions about BPC that pertain to their research, academic community, or professional context. The panelists will address questions about their projects, as well as questions about attendee projects that pertain to incorporating BPC. Attendees will leave with a sense of the depth of BPC networks and resources, as well as insight into the next phases of BPC in K-12, higher education, and career pathways. To address the need for increased diversity in CS, the National Science Foundation's Directorate for Computer Information Science and Engineering (CISE) launched the Broadening Participation in Computing Pilot effort in 2017, via a Dear Colleague Letter (Kurose, 2017). This effort built on years of federal support for BPC, including the 2016 announcement of CS for All (The White House, Office of the Press Secretary, 2016), the CISE Strategic Plan for Broadening Participation (NSF, 2012), the initiation of the BPC Alliances in 2005 (Chubin, Johnson, 2011). As more and more researchers and practitioners recognize the value of diversity in CS, from K-12 and higher education, to advanced computing and industry, researchers and practitioners are hungry for more information. From department plans and outreach efforts, to hiring strategies, student recruitment, and funding requests, BPC is everywhere. Considering how to proceed can be overwhelming, however there are a myriad of in-the-box solutions, and national networks, like the NSF-funded Broadening Participation in Computing Alliances, which are available to support anyone seeking to incorporate BPC into their work.

8:00am –
9:30am

P Panel: Edge-to-Edge Distributed CI for Distributed Research

Brella

Speakers: Scotty Strachan, Barr von Oehsen

We are pleased to welcome a panel of diverse scientists and engineers to present and converse on the challenges facing modern data-driven science workflows. This discussion will prove informative and useful to all aspects of the cyberinfrastructure community: computing, software, networks, funding agencies, researchers, and community organizers.

Organizers:

Barr von Oehsen (Rutgers University; Advanced Research Computing)
Scotty Strachan (University of Nevada, Reno; Cyberinfrastructure)

Panelists:

Jayne Bormann (AlertWildfire camera network)
Ivan Seskar (COSMOS wireless testbed)
Robert Casey (IRIS seismology data services)
David Coyle (USGS Next-gen Water Observing System)
Kobus Van der Merwe (POWDER wireless testbed)

Session abstract:

Technology evolution in academic science and engineering research has arguably created a new, disruptive workflow that is challenging traditional cyberinfrastructure models designed to solve "elephant flow" and "HPC" problems. These new "Edge-to-Edge" (E2E) workflows typically begin with distributed Internet of Things (IoT) data acquisition, traverse sometimes vast heterogeneous networks, and ultimately result in a spectrum of data products delivered in near-real-time to diverse end stakeholders. E2E is a key workflow type in emerging convergence research, as new observational data with rapid delivery become useful across multiple disciplines, and IoT devices are increasingly affordable for the individual researcher. Dominant domains of E2E workflows include environmental science (public hazards, ecosystem health, and resources management) and smart cities (technology testbeds, social services, and infrastructure planning). A recent "Dear Colleague" RFI letter from NSF (20-015) recognized the need for data-focused, ecosystem-scale CI services and capabilities that transcend disciplines and strive to make data available for seamless interoperability. Traditional academic CI at both the national and institutional scale is highly biased towards single lumps of data and scalable batch computing. E2E workflows represent a "ground zero" for this problem, as IoT data integration and automation is highly complex. The CI community needs to initiate a purposeful pivot in focus, expertise, and practices development to meet these impending needs and the resulting deluge of data flows and varieties. Our panel highlights some mature

E2E projects and workflows, and poses challenging questions to both the domain science and CI provider audiences.

8:00am – 9:30am	P	PEARC20 Panel Submission: Towards a National Cyberinfrastructure Ecosystem from Campus to National Facilities	Brella
		<i>Speakers: Donald F. McMullen, J. Eric Coulter, Richard Knepper, Resa Reynolds</i> This panel will bring together cyberinfrastructure professionals from colleges and universities to discuss their approaches to meeting researchers' needs using both on-campus and external CI resources. Existing programs such as the XSEDE Campus Champions, the Linux Cluster Institute, and HPC University offer opportunities for personal development, and organizations like the Coalition of Campus Research Computing Centers (CaRCC) provide fora for staff and leadership of campus research computing centers. For smaller institutions, however, or those that have difficulties in providing research CI, there is little in the way of dedicated on-site expertise to give a hand up or overcome immediate gaps in knowledge blocking progress toward a strong campus CI support environment for researchers. The XSEDE CI Resource Integration (CRI) group provides services that enable the development of campus CI that is harmonized with XSEDE facilities and offerings. CRI also provides in-person and on-site targeted support for campus research CI staff to fill knowledge or skills gaps and to help research computing programs solve immediate, "show stopping" problems. In this panel, we will discuss lessons learned from the trenches about how to enable continuous growth and quality improvement in campus research computing capabilities, and what additional common issues and problems still need to be addressed to build campus CI that is properly integrated with regional and national facilities and services. These individual experiences from campuses are valuable for the broader community considering how to optimize choices about local investments in equipment and expertise while leveraging national resources like XSEDE, the Open Science Grid, and cloud services. Discussion will be guided by the XSEDE CRI team, and will seek to foster coordination among resource providers, exposing common issues in developing a stronger U.S. research CI ecosystem.	
8:00am – 12:00pm	T	Fundamentals of Accelerated Data Science with RAPIDS	Brella
		<i>Speakers: Zahra Ronaghi</i> The open source RAPIDS project allows data scientists to GPU-accelerate end-to-end data science workflows and analytics applications, and leverage multiple GPUs for larger than memory datasets. Learn how to GPU-accelerate your data science applications by: Using cuDF (GPU-enabled Pandas-like dataframes) and Dask-cuDF to ingest and manipulate large datasets directly on the GPU in preparation for a variety of machine learning algorithms Utilizing GPU-accelerated machine learning algorithms (cuML) such as K-means and logistic regression, as well as cuGraph for graph analytics Understanding key differences between CPU-driven and GPU-driven data science, including API specifics and best practices for refactoring, learning techniques and approaches for end-to-end data science on GPUs Upon completion, you'll be able to refactor existing CPU-only data science workloads to run much faster on GPUs and perform a wide variety of end-to-end data science tasks using large datasets	
8:00am – 12:00pm	T	How to build your own Deep Neural Network Framework	Brella
		<i>Speakers: Kwai Wong, Stanimire Tomov, Daniel Nichols, Rocco Febbo, Xianfeng Ma</i> There are many capable Deep Neural Network software frameworks available freely. Most of them typically provide a python user interface which lets a user quickly train and test his/her dataset without bothering how DNN actually works. Using these DNN frameworks is clearly advantageous to a majority of users, allowing them spending more time on their own projects without worrying the integrity of their solution because these DNN frameworks are well known and accepted as standard in the data science community. However, in some cases, when a researcher may in fact want to find out why and how a method of one of these open source DNN frameworks works or fails, particularly with respect to his/her problems, it will be a difficult task for the researcher because he/she lacks the experience to understand how in general the DNN algorithm is implemented in the backend compute engine which is in general written in C++. As such this tutorial intends to prepare the users of such needs, broadening their knowledge in DNN via a step by step illustration starting from the mathematical formulation, numerical algebra interpretation, to programming implementation.	
8:00am – 12:00pm	T	TUTORIAL FULL - Easy Deployment of Edge Services for the ScienceDMZ with SLATE	Brella
		<i>Speakers: Joe Breen, Robert Gardner, Lincoln Bryant, Mitchell Steinman, Shawn Mckee, Muhammed Akhdhor, Jason Stidd, Chris Weaver</i> Connecting campus HPC resources to national cyberinfrastructure or multi-institution research platforms can be labor intensive and present numerous challenges. Overcoming them is precisely what's needed to meet the increasing demands on cyberinfrastructure from NSF Large Facilities and other data producing instruments , and the analysis and processing requirements of their scientific collaborations. The SLATE2 platform harnesses	

Kubernetes and Helm to facilitate deployment and operation of edge services (such as Globus, PerfSONAR, HTCondor CE, StashCache, Jupyter, etc.) to reduce the complexity and operational load of bringing campus compute and storage resources to shared national cyberinfrastructure, such as the Open Science Grid. SLATE runs on-premise and allows administrators at sites, or operators from trusted external organizations, to securely deploy service-applications from a curated catalog. Attendees will learn SLATE platform basics, how to easily share a campus resource with the Open Science Grid, and deploying a selection of storage connectors (e.g. Globus, MinIO). The tutorial will be of special interest to NSF Campus Cyberinfrastructure (CC*) awardees and Science Gateway developers with an interest in connecting to, and building, multi-institution research platforms.

8:00am –
12:00pm

T TUTORIAL FULL - Introduction to Docker, Singularity, and Kubernetes

Brella

Speakers: Ritu Arora, Charlie Dey, Ian Wang, Greg Zynda

A container is a portable software unit that has one or more applications and all their dependencies bundled together in a single package. Containers make software and data distribution easy, and can thus save time in software installation and maintenance. They are also useful in ensuring the reliability and reproducibility of the applications by future-proofing them against the changes in hardware and system software. In this beginner-level, half-day tutorial (3 hours duration), the participants will get an overview of containers and their relevance in cloud computing and high performance computing. Demonstrations and a hands-on session on the use of Docker, Singularity, and Kubernetes will be included. The participants will be required to create an account on DockerHub to participate in the hands-on session and would need to bring their laptops.

8:00am –
12:00pm

K High Throughput Science

Brella

Speakers: Igor Sfiligoi, Frank Würthwein

Science has long relied on computing for major breakthroughs. The type of computing resources and infrastructure services needed by a science application can however vary significantly; while most sciences require significant compute resources that can only be obtained by spreading the compute over many compute units, the needed throughput and latency requirements of the communication between such compute units can vary from negligible to near-continuous. The Distributed High Throughput Computing (dHTC) paradigm aims to serve applications that are tolerant of communication latency, since such resources both can be acquired at significantly lower cost and allow for a distributed funding model, and thus allow for more total compute resources to be available to such applications. This allows scientists to fully exploit resources offered by both distributed research cyberinfrastructures, such as the Open Science Grid (OSG) and Pacific Research Platform (PRP/TNRP), and public Cloud providers, for example Amazon Web Services, Microsoft Azure and Google Cloud Platform, in addition to more traditional HPC systems with low latency interconnects. The aim of this workshop is to engage with scientific communities that could make effective use of dHTC, both in the form of application developers and supporting institutions. It will provide an overview of the dHTC concept, the most commonly used tools involved, as well as testimonials from existing dHTC scientific user communities and their supporting organizations. Active attendee participation is highly encouraged.

8:00am –
5:00pm

K Leveraging a Research Computing and Data Capabilities Model for Strategic Decision-Making

Brella

Speakers: Patrick Schmitz, Dana Brunson, Gail Krovitz, Thomas Cheatham, Claire Mizumoto

Research computing and data (computing, data, and related infrastructure and services) is changing at an accelerating rate, while the range of fields and disciplines depending on the cyberinfrastructure is expanding and becoming increasingly diverse. A new "Capabilities Model" has been developed that identifies the range of relevant approaches to fully supporting research computing and data. This model allows institutions to assess their current capabilities, and provides input into strategic decision making. This Capabilities Model is designed to be useful to a diverse mix of stakeholders on the front lines of cyberinfrastructure, including campus research computing and data practitioners, along with the principal investigators and research team members (faculty, staff, etc.) with whom they work, and campus leadership. This workshop will introduce participants to the framework and background of the model, and how it can be applied to a range of different institutions. Participants will work together to fill out a copy of the questionnaire and explore specifics of applying the questionnaire for their respective institutions. Finally, participants will provide feedback to improve the model, explore future plans to make the model available as an online service, and discuss how the resulting aggregated data could be used in strategic decision making by institutions.

8:00am –
5:00pm

K Research Software Engineers Community Workshop

Brella

Speakers: Christina Maimone, Ian Cosden, Charles Ferenbaugh, Daniel Katz, Chris Hill, Kenton McHenry, Kaspar Mueller

Research software engineers (RSEs) have existed for decades, but the development of common job titles and career paths, the creation of professional networks to connect RSEs, and the emergence of RSE groups at

universities have gained momentum in the past few years. This workshop will build on that momentum and bring RSEs together to 1) grow the RSE community by establishing and strengthening professional networks of current RSEs and RSE leaders, and 2) generate materials that promote and support current and future RSEs. In support of growing and strengthening the network of RSEs, the workshop will include presentations from RSEs about their work and their groups, and about the development of the RSE community. The workshop will also bring RSEs together in small groups for discussions and to create draft materials to share with the community. Workshop participants will be asked to propose and prioritize topics for breakout group work, but possible topics include: - Writing profiles of career paths of RSEs - Documenting RSE-approved best practices, techniques, and technologies - Developing a plan for outreach to RSEs around the country - Creating an outline of an RSE skills training workshop - Reviewing and editing RSE job description/career progression models - Setting up, funding, sustaining, and managing RSE groups - Writing profiles of RSE projects and their impact - RSE project management - Document RSE in academia versus industry Each breakout group will produce a draft blog post, charter for a working group, or other document. Groups will assign a member to be responsible for finishing and publishing the materials after the workshop or coordinating follow-up meetings for extended work. Materials that are created can be hosted and publicized on multiple forums including the US-RSE website (us-rse.org).

8:00am – 5:00pm	<p>K What Does it Mean to be a Campus Champion? Brella</p> <p><i>Speakers: Shelley Knuth, Jay Alameda, Hussein Al-Azzawi, Torey Battelle, Marisa Brazil, Dana Brunson, Douglas Jennewein, Bj Lougee, Julie Ma, Timothy Middelkoop, Henry Neeman</i></p> <p>The importance of utilizing research computing and data infrastructure for scientific discovery and scholarly achievement has grown as research questions become more complicated and datasets get larger. Many higher education institutions have built infrastructure utilized by researchers and supported by local research computing staff. Some of these research computing groups are large, with several infrastructure and support staff, while others may be supported by only one or two staff members. For both of these groups, the Campus Champions program has provided an opportunity for knowledge exchange, professional development, and growth (Brazil 2019). Over the past twelve years, the Campus Champions program has grown to over 630 Champions at over 300 research institutions. A significant number of Champions attend the PEARC conference; in 2019, over 20% of PEARC attendees were Champions. With on average 50 new Champions joining the program per year, a cohesive and all-encompassing on-boarding program must be in place to ensure that the Champions understand how to assist members at their institutions as well as their peers in the community. For several years, a formal on-boarding process has served as a basic introduction, with members of the community relying on the web or their colleagues to determine where appropriate resources and information might be located. A more extensive process that includes mentoring and exposure to resources is desired. Given the number of Campus Champions who attend PEARC, and its vision to foster "the creation of a dynamic and connected community of advanced research computing professionals who promote leading practices and the frontiers of research, scholarship, teaching, and industry application" [PEARC 2019], hosting a workshop dedicated to the on-boarding of Campus Champions would have maximum impact.</p>
10:00am – 10:30am	<p>R Break and Networking Brella</p>
12:00pm – 1:00pm	<p>L Lunch Break and Networking Brella</p>
1:00pm – 5:00pm	<p>T Best Practices for High Performance Research Computing in the Cloud Brella</p> <p><i>Speakers: Tara Madhyastha, Ankit Malhotra, Pierre-Yves Aquilanti, Kevin Jorissen, Lee Pang</i></p> <p>Cloud computing lets researchers match computational infrastructure to the needs of specific research applications, execute without queue wait times, scale to meet workload demands, and share and reproduce scientific analyses. However, there are many ways to run high performance computing (HPC) applications in the cloud, and selecting the best and most cost-effective way can be daunting. This tutorial describes the specific benefits of cloud computing to research workloads, defines cloud-based computational methods such as auto-scaling and serverless computing, and teaches best practices for computationally intensive research workloads in the cloud. Attendees will apply what they have learned in a hands-on tutorial on Amazon Web Services (AWS). We will use specific scientific examples from genomics and weather forecasting to illustrate the appropriate use of both serverless batch computation and tightly-coupled clusters running applications with the Message Passing Interface (MPI) library. We will describe how to parallelize workflows and ensure reproducibility through workflow management systems and containers, illustrated with the open-source software NextFlow. We will also cover practical aspects of cloud computing for research, such as budgeting and minimizing cost, leveraging ongoing technological advances, sharing data and research infrastructure, and adapting to changing research methods and standards.</p>

1:00pm – 5:00pm	T Leveraging Tapis For Portable, Reproducible High Performance Computing In the Cloud <i>Speakers: Julia Looney, Charlie Dey, Anagha Jamthe, Joe Stubbs, Steve Black</i>	Brella
<p>This tutorial will focus on providing attendees exposure to state-of-the-art techniques for portable, reproducible research computing, enabling them to easily transport analyses from cloud to HPC resources. We will introduce open source technologies such as Jupyter, Docker and Singularity, and how to utilize these tools within NSF-funded cyberinfrastructure platform, Tapis API. The approaches introduced not only increase application portability and reproducibility but also reduce or eliminate the need for investigators to maintain physical infrastructure so that more time can be spent on analysis. For the tutorial, attendees will have access to allocations on XSEDE JetStream and one or more HPC resources such as TACC's Stampede2 or Frontera.</p>		
1:00pm – 5:00pm	T Modern Tools for Supercomputers <i>Speakers: Si Liu, Robert McLay, Albert Lu, Kent Milfeld</i>	Brella
<p>Powerful supercomputers have played an important role in the computational research community. However, the increasing complexity on modern systems may defer or hinder their work. A large amount of precious time and effort has been spent unnecessarily managing the user environment, reproducing standard workflow, handling large scale I/O work, profiling and monitoring users' jobs, understanding and resolving unnecessary system issues, etc. To help supercomputer users focus on their scientific and technical work and to minimize the workload for the consulting team, we designed and developed a series of powerful tools for supercomputer users. These tools are portable and effective on almost all supercomputers and are now serving thousands of supercomputer users of TACC, XSEDE, and other institutions every day. In this tutorial, we will present and practice with supercomputer tools specifically designed for complex user environment (LMod, Sanity Tool), tools for workflow management (ibrun, launcher, launcher-GPU), tools for job monitoring and profiling (Remora, TACC-Stat, core_usage, amask, etc.), and several other convenient tools. Attendees will learn how these tools are designed and used in their daily work. Detailed hands-on exercises are prepared beforehand and will be executed mainly on the Stampede2 and Frontera supercomputers at the Texas Advanced Computing Center (TACC).</p>		
1:00pm – 5:00pm	T TUTORIAL FULL - Programming Modern Multicore Processors <i>Speakers: Amit Ruhela, Lars Koesterke, John Cazes, Chris Ramos, Joe Garcia, Virginia Trueheart</i>	Brella
<p>Modern processors from Intel (Xeon), AMD (EPYC), ARM (ThunderX2), and IBM (Power9) rely on complex hardware features to achieve high floating-point performance. Because the base frequencies for current multi-core designs hover somewhere between 2-3 GHz, researchers can no longer rely on frequency scaling for increased performance. Instead, developers learn to program for parallel hardware found in modern cluster environments. To achieve good performance, developers leverage features of advanced processors and write optimized code aware of vectorization, memory access patterns, and caches, etc. Developers also understand how to balance multi-process (MPI) and multi-threaded programming (OpenMP). This tutorial will cover serial and thread-parallel optimization including introductory and intermediate concepts of vectorization and multi-threaded programming principles. We will address profiling techniques and tools and give a brief overview of modern HPC architectures. The tutorial will include hands-on exercises. We will also demonstrate the use of profiling tools on TACC systems. This tutorial is designed for intermediate programmers, familiar with OpenMP and MPI, who wish to learn how to program for performance on modern architectures.</p>		
3:00pm – 3:30pm	R Break and Networking	Brella