Research Engagement Speaker

NRP Engagement Webinar

Typically on the fourth Monday of the month at 1 ET - 12 CT - 11 MT - 10 PT

Zoom: see your email for coordinates!

Internet2 Research Engagement YouTube Playlist

To hear about future activities, please join the NRP engagement email list **via this link**.

Feel free to share this with anyone that may be interested. Calls will typically be on the fourth Monday of the month.

If you would like to present or have a suggestion for a session, please write to **Dana Brunson**.

PAST CALLS:

Monday, August 24, 2020 at 1 ET - 12 CT - 11 MT - 10 PT

Title: The Pacific Research Platform, National Research Platform, Global Research Platform and Nautilus

Presenters: John Graham and Dima Mishin, UC San Diego

Dmitry Mishin: Applications Developer at University of California San Diego

- Dmitry graduated with a Master's from Moscow Institute of Radio Engineering, Electronics and Automation, speciality computing machines, complexes, systems and networks. Dmitry also has a PhD in Geophysics from Earth Physics Institute, Russian Academy of Sciences. He is currently working as Applications Developer in University of California San Diego, San Diego Supercomputer Center on Comet Supercomputer system enhancements, and Calit2 on supporting and expanding the Nautilus – global Kubernetes cluster. The development focuses on supporting the large-scale computations, data visualization, IoT, data storage solutions. His research interests are HPC systems, data storage and access, microservice architectures, performance measurement and analysis.

John Graham: Senior Development Engineer at University of California San Diego

- John Graham is a Senior Development Engineer at QI, the UC San Diego division of the California Institute for Telecommunications and Information Technology (Calit2). Primary role is the lead engineer for the Pacific Research Platform. Other projects include GPU JupyterHub deployments for deep leaning and pattern recognition pipelines. GrowBotics is an ongoing project to create autonomous roof-top farming robots with a goal of providing food for the buildings occupants. John is a charter member of the Open Source Geospatial Foundation, and previously collaborated with other members on the "Open GeoData Pipeline" for disaster response-related satellite data and aerial photography coming from the U.S. Navy and the National Oceanic and Atmospheric Administration (NOAA). He is an highly experienced systems engineer and data visualization expert. John's past efforts include many years working with San Diego State University, Visualization Center. In the early 90's, previous to working with universities in San Diego, John ran a successful internet startup, a streaming media technology and service called Broadware. It was sold to Cisco and the technology is still part of their product line. John also spent several years in the nano-technology field building and operating Scanning Tunneling Microscopes at the University of Akron, Atomic Force Microscopes for Angstrom Technologies and Arizona State University and finally Low Temperature and Electrochemical Scanning Probe Microscopes for Park Scientific Instruments.

Recording: https://www.youtube.com/playlist?list=PLLIsQFBBoJG_jXSxkoGtDzcFjcrkT2y-2

Slides: https://docs.google.com/presentation/d/1Vrbf_-a9VZ1GSck9Z850BlfrlaxksvdB7oDNHca_4EA/edit#slide=id.g90f5ee2b76_0_22

Abstract:

 The PRP/NRP/GRP now represents a partnership among more than 50 institutions. Most host Flash I/O Network Appliances (FIONAs), which are rack-mounted PCs. FIONAs are advanced Science DMZ Data Transfer Nodes, optimized for 10-100Gbps data transfer and sharing. Many sport two to eight GPU add-in boards.

165 of these 10-100G connected FIONAs have been joined into a "hyper-converged cluster" called *Nautilus*, which uses Google's open-source Kubernetes to orchestrate software containers across the distributed system. Kubernetes is a now widely adopted way to manage containerized software. In fact, more than two thirds of the Fortune 500 companies have adopted it, and it is available within all the major commercial cloud providers. Nautilus currently has 550 GPUs, 6000 CPU cores, and more than 2PB of disk, all distributed among campus Science DMZs.

John Graham will talk about how you can participate and explore Nautilus, add your own node to Nautilus to get the benefits of informal Potluck Supercomputing (PLSC), and how to build a Kubernetes cluster that will federate with Nautilus, making it an persistent experiment in community-based infrastructure. He will explain how RocketChat has helped create a community of users who need to share big data and compute on it

Dima Mishin will then discuss Nautilus' use of Ansible to automate sysadmin tasks, Ceph storage pools, advanced measuring and monitoring, as well as Nautilus' coming federation with the new NSF SDSC Expanse supercomputer and its data-centric architecture. He will finish up with their new work with InMon and Reservoir Labs.

Monday, June 22, 2020 at 1 ET - 12 CT - 11 MT - 10 PT

The Eastern Regional Network – Simplifying Multi-Campus Research Collaborations - Barr von Oehsen

Abstract:

The Eastern Regional Network (ERN) was formed to address challenges related to simplifying multi-campus collaborations and partnerships in the Northeast that advance the frontiers of research, pedagogy, and innovation. The ERN is first and foremost a network of people interested in pursuing this vision, and who manage and use the campus and regional research computing, data, storage and network resources that can make it happen. We have chosen to make this a regional effort for two principal reasons: (1) face to face interactions are relatively straightforward and inexpensive to arrange; (2) the characteristics of our region are unique – for example, the Northeast contains eight different state university systems in a geographic area whose size is comparable to that of California. During this presentation I will give a brief overview of the ERN. It starts with motivation, mission and vision, and follows with a summary of the steps that we are taking to realize that vision, including accomplishments to date and ambitions for the future.

Bio:

Dr. James Barr von Oehsen is the Associate Vice President of the Rutgers University Office of Advanced Research Computing (OARC). Dr. von Oehsen is responsible for providing strategic leadership in advancing Rutgers University's research and scholarly achievements through next generation computing, networking, cloud services, and data science infrastructure. Prior to joining Rutgers, he was employed by Clemson University Computing and Information Technology (CCIT) as the Executive Director of the Cyberinfrastructure Technology Integration (CITI) group. Dr. von Oehsen has extensive experience working with diverse campus research communities throughout the nation as well as within the US industry sector. His interests are in advanced research computing, data science, machine/deep learning, cybersecurity, mathematical modeling, commercial and campus cloud solutions, and hardware architectures. In 2018 he received the NJEdge Technology Innovation Award for his work involving the convergence of software defined networking, research computing, tiered storage, commercial cloud, and federation of services. He is also a founding member of the Eastern Regional Network, a consortium of universities, regional network providers, and research centers with a vision to simplify multi-campus collaborations and partnerships that advance the frontiers of research, pedagogy, and innovation.

Monday, April 27, 2020 at 1 ET - 12 CT - 11 MT - 10 PT

Dan Schmiedt, "Life of a Packet"

Abstract:

Early in his career as a network engineer, while interviewing at a large networking company, the interview team lead told Dan, "You have an impressive-looking resume. But we don't really care about that." He pointed at a whiteboard. "Tell us exactly how a packet is built, travels across a network, and arrives at a destination host, and don't leave anything out."

Dan nervously stood up and started drawing and telling the story of the life of a packet. Heads nodded and he got the job, although he didn't stay long before he decided to go back to work at Clemson.

Today, any time someone asks Dan how a network works, he stands in front of a whiteboard and tells the story of the life of a packet.

Bio: Dan Schmiedt is the Director of Network Services and Telecommunications at Clemson University, where he has worked (with a few meaningful interruptions) since he was a student employee there in the early 1990's.

Inter-campus collaboration in the era of big data enabled by the Pacific Research Platform with UCSC, WUSTL, and UCSF - David Parks

Abstract:

UCSC is utilizing the PRP in collaboration with both the Washington University in St. Louis and UCSF. These collaborations involve cross-campus collaboration on datasets that are well into the multi-terabyte sizes. The PRP provides an enabling platform for these collaborations. The Henglab, lead by Keith Hengen, out of the University of Washington in St. Louis is producing terabyte and even petabytes scale data from novel longitudinal studies of neural activity in live mouse subjects. The PRP has enabled collaboration between UCSC and WUSTL where one campus provides data and the other provide the algorithms and compute, and both meet seamlessly on the PRP. UCSF is involved in recording organoid cell cultures of human glial neural cells, producing individual recordings that can reach into terabyte-scale, the PRP is enabling standard tools and analysis of data shared between UCSC and UCSF. At UCSC we are scaling up live organoid cell culture experimentation such that hundreds or even thousands of simultaneous experiments can be undertaken in parallel, generating massive datasets and the PRP enables scalable processing and streaming solutions. In this talk, we will introduce the science being performed at the scale of big data and how the PRP is enabling both the collaboration and the science in novel ways.

Bio:

David Parks is a graduate student researcher at UCSC pursuing his Ph.D. in Bimolecular Engineering with a focus on deep learning technologies. He works under Professor Haussler in the Braingeeers lab, a multi-disciplinary lab scaling up cell culture experimentation and bringing it into the open-source ecosystem. David has over a decade of experience in Silicon Valley working in enterprise software and deploying big data systems on platforms such as Hadoop.

Monday, February 24, 2020 at 1 ET - 12 CT - 11 MT - 10 PT

IceCube Computing Grid - Benedikt Riedel

Slides

Recording

Abstract: We present how the IceCube computing efforts have evolved over the last 15 years from mostly home-grown solutions to a globally distributed computing infrastructure. It will highlight advantages and disadvantages of such an approach for an experiment with broad science goals ranging from astrophysics to particle physics to geophysics, and what we see in our future as we engage in more external collaborations in computing.

Bio: Benedikt Riedel is the Global Computing Coordinator for the IceCube Neutrino Observatory and Computing Manager for the Wisconsin IceCube Particle Astrophysics Center. Previously he worked on the Open Science Grid at University of Chicago. He received a Ph.D. in 2014 from University of Wisconsin-Madison working on supernova neutrino signals in the the IceCube Neutrino Observatory.

Monday, January 27, 2020 at 1 ET - 12 CT - 11 MT - 10 PT

Running a 380PFLOP32s GPU burst for Multi-Messenger Astrophysics with IceCube across all available GPUs in the Cloud

Igor Sfiligoi and Frank Würthwein

Slides

Recording

The IceCube Neutrino Observatory is the National Science Foundations (NSF)'s premier facility to detect neutrinos with energies above approximately 10 GeV and a pillar for NSF's Multi-Messenger Astrophysics (MMA) program, one of NSF's 10 Big Ideas. The detector is located at the geographic South Pole and is designed to detect interactions of neutrinos of astrophysical origin by instrumenting over a gigaton of polar ice with 5160 optical sensors. The sensors are buried between 1450 and 2450 meters below the surface of the South Pole ice sheet. To understand the impact of ice properties on the incoming neutrino detection, and origin, photon propagation simulations on GPUs are used. We report on a few hour GPU burst across Amazon Web Services, Microsoft Azure, and Google Cloud Platform that harvested all available for sale GPUs across the three cloud providers the weekend before SC19, reaching over 51k GPUs total and 380 PFLOP32s. GPU types span the full range of generations from the NVIDIA GRID K520 to the most modern NVIDIA T4 and V100. We report the scale and science performance achieved across all the various GPU types, as well as the science motivation to do so.

Igor Sfiligoi is Lead Scientific Software Developer and Researcher at UCSD/SDSC. He has been active in distributed computing for over 20 years. He has started in real-time systems, moved to local clusters, worked with leadership HPC systems, but spent most of his career in computing spanning continents. For about 10 years, he has been working on one such world-wide system, called glideinWMS, which he brought from the design table to being de-facto standard for many scientific communities. He has recently moved his attention in supporting users on top of Kubernetes clusters and Cloud resources. He has a M.S. in Computer Science equivalent from Universita degli studi di Udine, Italy. He has presented at many workshops and conferences over the years, with several published papers.

Frank Würthwein is the Executive Director of the Open Science Grid, a national cyberinfrastructure to advance the sharing of resources, software, and knowledge, and a physics professor at UC San Diego. He received his Ph.D. from Cornell in 1995. After holding appointments at Caltech and MIT, he joined the UC San Diego faculty in 2003. His research focuses on experimental particle physics and distributed high-throughput computing. His primary physics interests lie in searching for new phenomena at the high energy frontier with the CMS detector at the Large Hadron Collider. His topics of interest include, but are not limited to, the search for dark matter, supersymmetry, and electroweak symmetry breaking. As an experimentalist, he is interested in instrumentation and data analysis. In the last few years, this meant developing, deploying, and now operating a worldwide distributed computing system for high-throughput computing with large data volumes. In 2010, "large" data volumes are measured in Petabytes. By 2025, they are expected to grow to Exabytes.

Monday, October 28, 2019 at 1 ET - 12 CT - 11 MT - 10 PT

Running Genomics Workflows on the Pacific Research Platform's Nautilus Kubernetes Cluster

Alex Feltus, Ph.D.

Professor|Genetics & Biochemistry @ Clemson University; Co-Founder|Praxis AI

Slides

Recording

Abstract: Our core biological research mission is to discover causal alleles underlying complex trait expression in plants and animals. Active projects include (A) discovery of genetic subsystems driving legume-microbe symbiosis that can be engineered into other plants so they can make their own fertilizer, (B) elucidation of gene expression pattern shifts between normal and disordered brain tissue for better diagnosis of intellectual disability, and (C) detection of tumor specific alterations in kidney and other tumors of relevance to precision medicine. Our scientific instrument is the high performance/throughput computer where we run bioinformatic, machine learning, and network biology workflows on tens to thousands of terabytes of in-house and open source deep DNA sequencing datasets. In recent years, we have wrapped applications in containerized NextFlow workflows and now run data intensive experiments on Kubernetes (K8s) clusters including the PRP Nautilus cluster (we have added a node at Clemson) and the Google Cloud Platform. In this webinar we will (A) present results from a large tumor biomarker screen generated with the Nautilus cluster, (B) describe broadly useful open source genomics workflows (GEMmaker, KINC, and Gene Oracle) with Nautilus-specific usage documentation, (C) outline a grassroots strategy to add nodes to the Nautilus cluster and train people how to use that super-awesome system, and (D) discuss a business model where one can build an elastic K8s cluster for a small virtual organization that can be dynamically linked to larger national compute fabrics platforms via aggregation or federation.

Dr. F. Alex Feltus received a B.Sc. in Biochemistry from Auburn University in 1992, served two years in the Peace Corps, and then completed advanced training in biomedical sciences at Vanderbilt and Emory. Since 2002, he has performed research in bioinformatics, high-performance computing, cyberinfrastructure, network biology, genome assembly, systems genetics, paleogenomics, and bioenergy feedstock genetics. Currently, Feltus is an Professor in Clemson University's Dept. of Genetics & Biochemistry, CEO of Allele Systems LLC, Core Faculty in the CU-MUSC Biomedical Data Science and Informatics (BDSI) program, member of the Center for Human Genetics, and serves on the Internet2 Board of Trustees as well as various "Advance Research Computing" engagement workgroups. Feltus has published numerous scientific articles in peer-reviewed journals, teaches undergrad and PhD students in bioinformatics, biochemistry, and genetics. At present, he is funded by multiple NSF grants and is engaged in tethering together extremely smart people from diverse technical backgrounds in an effort to propel genomics research from the Excel-scale towards the Exascale.