Yale Center for Research Computing Fall 2019 Newsletter

Message from the Executive Director



On behalf of the Yale Center for Research Computing (YCRC), I am delighted to share with you the activities taking place at Yale in the area of research computing.

I am pleased to welcome the YCRC Steering Committee for the 2019 - 2020 fiscal year, with the full membership list available at <u>https://research.computing.yale.edu/steering-committee</u>. Many thanks to the committee members who served last year, and to this year's new members.

This past year, the YCRC continued to introduce new enhancements to the cyber-infrastructure, provide research services through in-person consulting, and facilitate an interdisciplinary approach to the development and application of advanced computing and data processing technology through a variety of community engagement activities. The newsletter includes details of our new advanced computing capabilities, expert staff members, training, and office hours.

This year, the YCRC will expand the workshops, provide more support for sensitive data research, and enhance the infrastructure. We will be introducing a new storage offering, providing cloud technology support, and adding more accelerator technology to accommodate the growing use of machine learning techniques and other accelerator-dependent applications.

An open invitation is extended to you to visit the YCRC. You are welcome to use our research computing environment, take advantage of our skills-based trainings, have an individual consulting session, or use our powerful visualization workstations which allow for viewing of whole datasets on a single screen.

We look forward to supporting you in the year ahead!

Sincerely,

Kiran D. Keshav Executive Director Yale Center for Research Computing

Fall 2019 Newsletter





High-Performance Computing

One of our most significant jobs at the YCRC is to provide advanced computing facilities that enable a wide range of research in the natural and social sciences, and engineering at Yale. During the 2018-2019 fiscal year, the YCRC operated five high-performance computing clusters (Farnam, Grace, Milgram, Omega, and Ruddle) that, in aggregate, contain thousands of central processing units (CPUs) and roughly 10 petabytes of storage in support of many hundreds of active users. With such a large state-of-the-art facility, change is an essential and common occurrence so that the YCRC can continue to provide Yale's researchers with access to the latest computing technologies.

Refresh of Grace and Farnam

Two of our high-performance computing (HPC) clusters, Grace and Farnam, are due for a refresh. Grace supports research from across the university, including data science, neuroscience, environmental science, high-energy physics, the social sciences, and many other science and engineering fields. Farnam mainly supports research in the life sciences. Some of the oldest nodes on Grace and Farnam will soon be decommissioned, with new nodes introduced. Each of the new nodes has two Intel Xeon Gold 6240 processors for a total of 36 cores per node running at 2.6 GHz.

A portion of these new nodes will be available to the entire community, and some dedicated to specific research groups. The dedicated nodes will support research in molecular modeling, mechanical engineering, energy science, life sciences, computational chemistry, chemical engineering, electrical engineering, economics, and molecular biophysics. When not being used by the research group that purchased them, the dedicated notes will be available for general use.

For the Grace refresh, 163 new nodes are being added, of which 103 are common nodes, and 60 are dedicated nodes. Each of these nodes has 192 GB of memory, except for a few with more memory. Two of the new commons nodes are large-memory nodes, each with 1.5 TB of memory. This refresh also adds a few new nodes with graphics processing units (GPUs). GPUs can be used to accelerate applications in fields such as data science and neuroscience. The commons nodes include ten nodes with four GPUs each.

On Farnam, 32 new nodes are being installed with 192 GB to 768 GB of RAM each. Of these, 19 are commons nodes and 13 are dedicated nodes.

Y CRC

Omega Replacement

The HPC cluster Omega has run tightly coupled parallel jobs, supporting simulations in such fields as cosmology, climate science, nuclear physics, quantum chemistry, and materials science. After eight years of service, Omega will be retired. Initially, we are replacing Omega with a total of 129 high-performance Intel 6136 nodes, each with 24 CPU cores running at 3 GHz, 96 GB of RAM, and the latest high-speed networking, based on InfiniBand HDR100. These nodes will be used for tightly coupled parallel simulations, such as those previously run on Omega. We plan to make 113 of these nodes available in a commons partition, while the remainder will be dedicated nodes supporting research in climate science. We anticipate purchasing additional high-performance nodes in the spring of 2020.

New Faculty Nodes Added to the Clusters

Beyond the major projects above, the YCRC has installed several new compute nodes dedicated to research groups on the HPC clusters. These nodes support research in urban and land change science, molecular modeling, mechanical engineering, and computational chemistry. These new compute nodes include large memory nodes with up to one terabyte of memory, as well as standard nodes with the Intel 6136 CPU. Several nodes contain up to four GPUs. When dedicated nodes are not being used by the research group that purchased them, they are available for general use.

Data Center Tours



Interested in touring the YCRC Data Center?

The YCRC offers guided tours of our Data Center, located at Yale West Campus. Visitors are welcome to get an up-close look at our High-Performance Computing clusters and learn how these machines make advanced computing and data processing a reality for our research affiliates.

Please contact <u>hpc@yale.edu</u> to schedule a tour. We look forward to seeing you!





Yale University is pleased to announce that we are now institutional members of Dryad. Dryad is an open-source, research data curation and publication platform, making data publishing easy for the researcher. The Dryad platform accepts data from any discipline. As an institutional member, Yale researchers can deposit their data free of charge without limitation on the number of datasets deposited.

For additional information, please see Dryad's FAQ page at <u>https://datadryad.org/stash/faq</u> or contact <u>researchdata@yale.edu</u> for questions and recommendations about your specific research needs.

Research Support

The research support team offers office hours every day and consults one-on-one with many YCRC users. During FY19, we met with more than 130 users for a total of 226 hours of consulting.



The figure to the left summarizes the hours we have spent one-onone with students, postdocs, and faculty over the last few years.

We enjoy learning about what the research community is working on and have been energized by all the new faces we saw this year from the Yale School of Management and the Faculty of Arts and Sciences. Be it curing a simple configuration issue, advising on a computational problem, or optimizing code, we always look forward to in-person consultations.

Page 4

Yale Center for Research Computing Copyright @ 2019 Yale University \cdot All rights reserved



Fall 2019 Newsletter



"I'm not sure what we would have done without Ben's (Evans) knowledge, expertise, and willingness to find resolutions to the various hurdles that have come up. He has taken the initiative to contact people at Google, Berkeley, and some entities at Yale to investigate what is needed to ensure the various systems we are developing can speak to each other. Ben is also great at explaining the various aspects of the infrastructure and I have been learning quite a bit from him."

Jessi Cisewski-Kehe, Assistant Professor of Statistics and Data Science

Focused Research Collaboration

The YCRC was involved in a number of focused collaborations with researchers across the University. Ben Evans, Ph.D. continues to work closely with the Dunn Lab in the Department of Ecology and Evolutionary Biology on projects ranging from lab hardware monitoring to phylogenetic analyses. Kaylea Nelson, Ph.D. continues to provide expert support to the Department of Geology & Geophysics, and Rob Bjornson, Ph.D. assists both the Yale Center for Genome Analysis and the Yale Stem Cell Center and their users with computing and sequence analysis.

This summer Dong Wang, Ph.D. from the Department of Mechanical Engineering and Materials Science, and David Huberdeau, Ph.D., from the Department of Psychology joined the YCRC's Research Support Program, bringing their particular knowledge and skill sets to helping YCRC users. Dong is a member of the O'Hern Group and is interested in GPU computing. David is a member of the Turk-Browne Lab and focuses on MRI brain imaging.

Training and Workshops

From July 2018 to June 2019, the YCRC hosted over 40 training events attended by more than 600 community members from over 50 Yale departments. Ten different "bootcamps" are now offered at beginner and intermediate levels, covering various subjects for users of the HPC clusters and other scientific computing environments. Thank you to YCRC staff members Giuseppe Amatulli, Ph.D., Rob Bjornson, Ben Evans, Tom Langford, Ph.D., and Kaylea Nelson for all of their hard work in preparing and presenting these workshops.

Y CRC

In August 2018, the YCRC hosted the Linux Clusters Institute Intermediate Workshop for HPC system administrators from all over the country. Many thanks to YCRC staff member Tyler Trafford for his assistance as a workshop presenter.

We also welcomed a number of vendors including Amazon Web Services, StataCorp, and Wolfram Research, who gave workshops on their software products. In June 2019, IBM presented a deep learning workshop to a packed crowd at the YCRC auditorium. This two-hour session was a prelude to a two-day hands-on workshop coming in December.



IBM Deep Learning workshop held June 13, 2019 in the YCRC Auditorium.

"I loved getting exposed to a wide variety of tools, and I appreciated whenever the instructors described how a tool in a particular language resembled or paralleled another tool in Python, ArcGIS, etc. I also REALLY appreciated the small group size! It was so easy for individuals to ask questions and for all participants to keep up."

- Geo-Computation and Environmental Analysis Workshop participant

Our partnership with the Extreme Science and Engineering Discovery Environment (XSEDE) allowed us to be a satellite host site for their monthly HPC workshops. More information about these events can be found at <u>https://ycrc.yale.edu/xsede</u>.

Please visit <u>https://ycrc.yale.edu/calendar</u> to see the full calendar of workshops and events, and to register.

Training Collaboration

In May 2019, Ben Evans, and Michael Strickler, Ph.D., were key to the successful completion of the "Practical Cryo-EM Workshop" organized by Yong Xiong, Ph.D., Professor of Molecular Biophysics and Biochemistry.

Fall 2019 Newsletter



YCRC

Page 6



"YCRC has been critical in helping run the cryo-EM workshop course, both for having the Farnam pi_cryoem GPU nodes ready (reserved for the workshop) and actually helping the participants. Ben (Evans) gave an overview presentation on Farnam and helped the workshop participants to set up accounts and use the cluster. I also want to add that Michael Strickler was also instrumental in the success of running the workshop. He generated the bootable USB disk for every participant with all the computing environment and software needed, stayed every day in the workshop to help the participants to troubleshoot problems on the fly. A big shoutout to both Ben and Michael!"

Yong Xiong, Professor of Molecular Biophysics and Biochemistry

Outreach Activities

Monthly User Group Meetings

The YCRC User Group had a very successful year, hosting monthly meetings for students and researchers from across campus. Topics have ranged from data management to GPU-based programming, with multiple lightning talks presented by users at each meeting.



YCRC User Group's Data Management meeting held March 6, 2019 in the YCRC Auditorium.

The highlight of the year was a standing-room-only session (50+ attendees) detailing a Guided Tour of Machine Learning, presented by Physics postdoc Chase Shimmin, Ph.D. Nearly half the attendees were first-time visitors to the YCRC, including many from departments outside of the traditional HPC user base.

Y CRC



This academic year, the YCRC User Group will meet on the second Wednesday of each month at 4:00 pm in the YCRC Auditorium at 160 St. Ronan Street. Meeting topics for this fall include Data Science with Python and Pandas, Computational Social Science, and a visit to the YCRC Data Center at West Campus.

New Ask.CI Site

We have just launched a YCRC-specific section on Ask.CI, a question and answer platform for people who do research computing – researchers, system administrators and others. Addressing technical topics of interest to the research computing audience, it has been designed to aggregate answers to a broad spectrum of questions that are commonly asked as researchers utilize advanced computing resources, creating a self-service knowledge base. With the new YCRC section, you will also be able to ask (and answer!) questions specific to our clusters and research computing environments. For additional information, please visit https://ask.cyberinfrastructure.org/c/yale.

Yale Day of Data

The YCRC co-sponsored the sixth annual Yale Day of Data which was held on November 30, 2018, with the theme "Data on Earth." Researchers and data experts from across the university came together to share experiences, challenges, and best practices related to data-intensive research.



The 2019 Yale Day of Data will be held on December 6 with the theme "Data Privacy." Additional information can be found at https://elischolar.library.yale.edu/dayofdata/2019.

YCRC in the Classroom

Members of the YCRC staff work with a number of Yale faculty to impact teaching and learning at Yale. Andrew Sherman, Ph.D. regularly teaches a senior/graduate-level class on HPC and parallel programming in the Department of Computer Science.

Ben Evans assisted Professors John Lafferty and Jessi Cisewski-Kehe with the Department of Statistics and Data Science's YData courses. The core course involved 110 students using a Jupyter hub running on the Google Cloud Platform as the compute environment—a first at Yale.



In addition, Rob Bjornson gave two lectures on Python to Professor Jakub Szefer's EENG 201 Introduction to Computer Engineering class.

Spinup Self-Service



This past year, IT @ Yale launched a new service to provide Yale University faculty, staff, researchers and graduate students with a quick and easy way to deploy cloud resources. Spinup was built with Yale's mission in mind. Spinup resources are behind Yale's enterprise firewall, preventing attack by malicious actors from outside of Yale's network, and are preapproved for moderate and high-risk data such as FERPA and HIPAA.

For additional information or to get started, please see the Yale Spinup website: <u>https://research.cloud.yale.edu</u>.

Regional and National Organizations

Members of YCRC staff are active participants in a number of regional and national computing organizations, keeping the team up to date on the latest technologies. Among these groups are the XSEDE Campus Champions (Kaylea Nelson, Andrew Sherman, Ben Evans), Coalition for Academic Scientific Computation (Andrew Sherman), Campus Research Computing Consortium (Andrew Sherman, Kaylea Nelson), and the NorthEast Big Data Hub (Kiran Keshav, Andrew Sherman).

This year YCRC participated in two new regional initiatives: Eastern Regional Network (ERN), and a quarterly meetup of people in the New York area involved in HPC for biomedical applications. The aim of the ERN effort is to determine whether there may be benefits to Yale researchers from access to distributed HPC facilities hosted at a number of institutions in the region with whom Yale has research collaborations in areas such as cryo-electron microscopy (Cryo-EM), brain imaging, and genomics.

Rob Bjornson organizes the quarterly biomedical HPC meetups in New York, which bring 20-30 participants from most of the academic research institutions in New York and surrounding areas.

Several YCRC staff members participated in the Practice and Experience in Advanced Research Computing (PEARC19) conference held in Chicago this past July. The theme of this year's conference was machine learning and artificial intelligence. Andrew Sherman served on the

Fall 2019 Newsletter

YCRC



PEARC19 executive committee and as the Technical Program Chair. In addition, YCRC staff members Rob Bjornson, Kaylea Nelson, Ben Evans, Tom Langford and Ping Luo attended the conference and contributed to the conference as technical reviewers.

Staff News



YCRC Staff at the Yale Farm, June 6, 2019. Left to right: David Huberdeau, Andrew Sherman, Tom Langford, Dave Logie, Eric Peskin, Alexander Behzad, Paul Gluhosky, Kaylea Nelson, Kiran Keshav, Ben Evans, Erica Gilbert, Rob Bjornson, Michael Strickler, Mark Lisi, Tyler Trafford

Welcome New Staff

In March, we welcomed Erica Gilbert as the new Senior Administrative Assistant at the YCRC. Erica joined us from the Yale School of Medicine and has been with Yale for over 5 years.

Michael Strickler joined the YCRC in May as a Computational Research Support Analyst, reporting to Rob Bjornson, Senior Research Scientist for Biomedical Research Support. Michael divides his time between installation, maintenance, and user support for various life science applications on the Yale clusters; and providing system and application support for the Center for Structural Biology, a shared scientific computing facility in the Bass Center. Most recently, Michael was a Research Specialist with Howard Hughes Medical Institute in the Thomas Steitz Lab at Yale University. Michael has a Ph.D. in biophysics from the University of Rochester.

Christine Costantino joined the YCRC in July as Program Manager reporting to Dave Logie, Director of Projects and Programs. In this role, Christine is responsible for the day-to-day

Fall 2019 Newsletter



management of the YCRC skills-based training program, and the advancement and growth of the program. Christine was most recently Assistant Director of Educational Technology at the Yale Poorvu Center for Teaching and Learning where she supported and trained faculty to use Yale's new learning management system, Canvas. Christine holds a master's degree in business administration from Johns Hopkins University.

In August, we welcomed Ping Luo as a Computational Research Support Analyst on the Arts and Sciences Research Support team reporting to Andrew Sherman, Senior Research Scientist. Ping came to us from the Texas A&M High Performance Research Computing group where she was a Senior Lead IT Consultant who helped faculty members and researchers use HPC resources for their research. Ping holds a master's degree in mathematics from Indiana State University and a master's degree in computer science from Texas A&M University.

Congratulations



YCRC Associate Research Scientist Tom Langford was selected to give the Sambamurti Memorial Lecture for 2019. This is a prize lectureship for young scientists working in particle and nuclear physics.

On July 25, 2019, Tom presented "Fingerprinting Nuclear Reactors with Neutrinos" to a group of summer interns at Brookhaven National Laboratory.

YCRC Research Scientist Giuseppe Amatulli was

invited to present his research on "Global Monitoring of Fresh Water at High Spatial and Temporal Resolutions: Assessing Stream and Lake Hydrological / Physical Features within a Machine Learning Framework" as part of the NASA Jet Propulsion Laboratory's Earth Science Seminar series on August 29, 2019. A recording of his talk is posted on our website at https://ycrc.yale.edu/news/fresh-water.

Summer Interns

YCRC welcomed two college students for 10-week internships from June to August 2019.

Mark Lisi, a rising junior studying computer science and mathematics at Tulane University, joined the Research Support team. Mark tackled every project we threw his way with enthusiasm. During his internship, he developed tools to allow our researchers to better understand our cluster systems and to automatically update our user guides to the most current

Fall 2019 Newsletter



Page 11

configuration of the HPC clusters. Mark also made significant improvements to our web dashboard to more clearly show storage and node utilization.

Allen Wang, a rising junior studying computer engineering at the University Connecticut, joined the Engineering team. During his internship, Allen learned about the day-to-day operation and maintenance of the HPC clusters. He was involved in ticket resolution, hardware repairs, tracking deliveries of new hardware, and the installation of new compute nodes and network switches on Farnam and Grace. Allen very quickly got up to speed on his role and responsibilities and was a very effective member of the team.

We were sad to see the interns leave at the end of the summer but wished them well for the remainder of their studies.

In Memoriam

On July 10, 2019, YCRC High Performance Computing Specialist Stephen Weston lost his battle with ALS. Steve joined Yale as a full-time employee in 2011 and was instrumental in providing technology services to the research community, including co-authorship of the O'Reilly publication *Parallel R: Data Analysis in the Distributed World*. We will miss him dearly. May he rest in peace.



YCRC Staff at the Center, Sept 18, 2019. Front row, left to right: Tyler Trafford, Kaylea Nelson, Ping Luo, Jonathan Aliwalas, Giuseppe Amatulli. Second row, left to right: Erica Gilbert, Christine Costantino, Eric Peskin, David Backeberg, Michael Strickler, Andrew Sherman, Kiran Keshav, Alexander Behzad, Dave Logie. Back row, left to right: Rob Bjornson, Jay Kubeck, Ben Evans, Paul Gluhosky.

Fall 2019 Newsletter

Yale Center for Research Computing Copyright @ 2019 Yale University \cdot All rights reserved



Page 12