SDSC recently announced the launch of HPC Share, a data sharing resource that will enable users of the Center’s high-performance computing resources to easily transfer, share, and discuss their data within their research teams and beyond. HPC Share is available to all users of SDSC’s Comet supercomputer, with a potential expansion to SDSC’s Triton Shared Computing Cluster (TSCC) and its Expanse supercomputer, the latter slated to enter production later this year.

Continued on page 3
The new year is off to a strong start on several fronts at SDSC as we move into the ‘reality’ phase of several new projects awarded to us by the National Science Foundation (NSF). While we currently have some 40 active awards totaling more than $100 million, two significant ones come to mind: Our new Expanse supercomputer and CloudBank.

Earlier this month we issued purchase orders for Expanse, which at this time is still on for an October launch. Prior to its arrival we are working with UC San Diego Facilities on upgrades to the data center required to accommodate the direct liquid cooling system that will be part of Expanse’s advanced design. This new system is being designed to scale past Expanse so we do it once, like we did with our machine room power density infrastructure.

CloudBank is a five year, NSF-funded collaboration to create a suite of managed services with the goal of simplifying public cloud access for computer science research and education via an extensive outreach and training program. Allocation of cloud resources will be given to projects that belong to NSF programs, and which specify the use of CloudBank in the NSF solicitations. Allocations and an initial set of education and training materials will be available when CloudBank begins formal operations in August.

COVID-19 Preparations, Access to Comet
The other ‘reality’ phase of the first quarter was of course preparing for a potentially significant spread of COVID-19, which is already causing a sizeable number of conferences either to be canceled or morphed into ‘e-conferences’ with remote participation. SDSC already delivers many of its training and workshops virtually, so we look forward to making these programs even more accessible by focusing on remote delivery and support. The most capable resources supported by the NSF’s Office of Advanced Cyberinfrastructure, including Comet, are being made available to support a new COVID-19 HPC Consortium co-launched by IBM and the Department of Energy.

SDSC now has remote work arrangements in place for most operations. SDSC’s co-location operations will continue to be staffed 24/7, and we have identified staff who need an emergency designation from campus that permits them to respond in person if needed.

Please stay healthy during this time by taking the needed precautions and practices!

Michael L. Norman
SDSC Director
SDSC Launches Comprehensive Data Sharing Resource
[Continued from page 1]

“HPC (High Performance Computing) users face a range of hurdles to share their data,” said SDSC Visualization Group Leader Amit Chourasia, also SeedMeLab’s Principal Investigator (PI). “Their collaborators may not have adequate context of the computed data, or may not be able to find, access, or fetch the data from HPC system. They may also have to review data in a fast-paced environment. These hurdles inevitably create a bottleneck at best, or in the worst case can cripple the scientific discovery process.”

“HPC Share bridges a major gap in our current cyberinfrastructure by offering a turnkey system that eliminates barriers among collaborating researchers so that they can quickly access and review scientific results with context,” said SDSC Director Michael Norman. “Users will benefit from reduced complexity, ubiquitous accessibility, and more importantly, rapid knowledge exchange.”

HPC Share is powered by SDSC’s open-source SeedMeLab software that was developed with support from National Science Foundation (NSF). Its built-in web services, coupled with an API extension, make it a versatile platform to create branded data repositories for small research groups to large communities or integrate with existing research data flow while also serving as an important stepping stone for researchers to realize FAIR data management in practice.

In addition to Chourasia, the SeedMeLab project includes SDSC Director Michael Norman as co-PI and David Nadeau as a technical architect with SDSC. Master’s and undergraduate interns at UC San Diego’s Computer Science and Engineering Department, as well as regional high school students, assisted in the project with prototype extensions and comprehensive quality assurance of the software.

SDSC/WIPAC GPU CloudBurst Team Receives 2020 CENIC Innovations Award

The California Research and Education Network and the Pacific Research Platform (CENIC) has awarded researchers at SDSC and the Wisconsin IceCube Particle Astrophysics Center (WIPAC) at the University of Wisconsin–Madison its 2020 ‘Innovations in Networking Award for Experimental Application’ for their bold experiment last November that marshalled all globally available-for-sale GPUs across Amazon Web Services, Microsoft Azure, and the Google Cloud Platform.

About 51,500 GPU processors were used for the experiment, which demonstrated that IceCube can effectively utilize a large number of GPUs in a single pool. The experiment was funded by the National Science Foundation to prepare for the exascale computing era.

Act II
The same research team – coordinated by Frank Würthwein, SDSC lead for high-throughput computing; Igor Sfiligoi, SDSC’s lead scientific software developer for high-throughput computing; Benedikt Riedel, global computing coordinator for the IceCube Neutrino Observatory and computing manager at WIPAC; and David Schultz, a production software manager with IceCube – followed up with a second experiment in early February using a portion of remaining funding from the modest NSF EAGER grant.

“This time we showed that the cloudburst run can actually be sustained during an entire workday instead of just one or two hours, and have moreover measured the cost of using only the two most cost-effective cloud instances for each cloud provider,” said Sfiligoi. “We have some funds remaining, so we will have at least one more Cloud run. We’re now figuring out what to do – we don’t like to repeat ourselves!”

We’re looking forward to Act III!
Michael Zentner joined SDSC in mid-2019 following nine years with Purdue University, where he was an Entrepreneur in Residence at the Purdue Foundry and a senior research scientist. For 18 years prior to that he was a senior team member/co-founder of several startup companies. Upon joining SDSC, Mike was named PI of the Science Gateways Community Institute (SGCI) founded by SDSC Associate Director Nancy Wilkins-Diehr in 2016 under a five-year, $15 million National Science Foundation grant. Mike continues as director of HUBzero®, an open-source science gateway platform. He’s also a co-PI of the Network for Computational Nanotechnology Cyber Platform project, which operates the nanoHUB.org science gateway. Mike earned a Ph.D. in Chemical Engineering from Purdue University and dual MBAs in international business from Purdue University’s Krannert School of Management and the TIAS School for Business and Society Business in Tilburg, Netherlands.

Q: Can you describe your role as Director of ‘Sustainable Scientific Software’ to a broader, less technical audience?
I would describe it as setting up and executing business models that work with software products, which is different than a research model. A sustainability model is more about turning things into regular, repeatable processes and activities that have a business model behind them. This allows us to think about how we capitalize and build on the original investment in the creation and discovery phase that takes place in a research program.

Q: What attracted you to SDSC and UC San Diego?
The key reason was my role in the Science Gateways Community Institute program and its upcoming need for a sustainability model. There’s a lot of good synergy between SGCI and HUBzero and nanoHUB, so it made a lot of sense. What’s worked really well with the SGCI was its initial design and operation. Coming to
SDSC, I knew I was moving into something that Nancy really set up well. At this stage we've pretty well succeeded in doing most of the things that we originally proposed, and are proving that people want those services and are willing to pay for them – in fact wanting to pay for more effort than what we could have given them with our grant funding. We've done approximately 100 consulting projects now for science gateways, and we've identified over 500 gateways in our catalog with millions of users in the aggregate. That's hardly a comprehensive list, so there is still a lot of opportunity for the SGCI out there, and a lot of opportunity to think about how we become self-sustaining.

Q: What's on your radar for 2020 and into next year?
The biggest thing is to write the NSF renewal proposal for SGCI, which includes designing a financial and sustainability model as well as a picture of how we productize the knowledge we've gained during our first five years. This is an opportunity to apply my entrepreneurial experience to the SGCI as part of SDSC's Sustainable Scientific Software effort. I'm also looking at bringing some help on board to assist in selling some of these products, as well as those related to HUBzero, nanoHUB, and products from SDSC that are at a point of market readiness. It's really a full-on sales effort to our current academic audience but also exploring how what we do at SDSC may be attractive to small-to-midsize companies that have cyberinfrastructure needs but not personnel budgets like larger companies.

Q: That sounds like looking for someone with both an academic and business mindset, no?
The best bad thing that ever happened to me was that I was put in the role of sales for three years (selling collaboration software before joining Purdue). I got beat up a lot but I also learned a lot. Nowhere did you get it drilled into you more about how you have to provide value than in that kind of job! It's not very common to find somebody in an academic environment who wants to do the creation side as well as the building out and scaling side of the work. Also, if we find people with more of a business mentality, at what stage do we have to get things before that person would be interested in working on it? It needs to be at a stage where it is not just created, but where others not involved in its creation will speak to its value.

Q: As a senior researcher and entrepreneur what is something I didn't ask you that you would want your SDSC colleagues to know about?
Nobody ever asks me what went wrong in the world of entrepreneurship – and 90% of the time things do go wrong, or at least not as planned, so you learn from things that didn't work. A lot of problems become obvious to the entrepreneur who experienced them but are not obvious at all to others. I recently gave a presentation at the International Workshop on Science Gateways on things that will guarantee that you won't be sustainable, a kind of recipe of things that will hurt if you don't pay attention. Even though I couldn't provide a guaranteed path for success, I did have some folks tell me they would save the list as points of reference for what not to do.

Check out SDSC’s Research Videos, Social Media Feeds!
Did you know that SDSC has a YouTube channel? Just click on the YouTube icon at the bottom of the SDSC homepage under the ‘Get Connected’ section or go to...

www.youtube.com/SanDiegoSupercomputerCenter

twitter.com/SDSC_UCSD
facebook.com/SanDiegoSupercomputerCenter
youtube.com/SanDiegoSupercomputerCenter

STAY CONNECTED
REHS Students Win First Place at December AI Conference

2020 REHS Program Starts June 22

For more than a decade, SDSC researchers have mentored an array of Research Experience for High School Students (REHS) participants including David Huang, Richard Chen, and Alvin You, who recently won first place in the precision medicine and drug design poster session at the December 2019 Artificial Intelligence in Medicine Conference.

During their time at SDSC, Huang, Chen, and You worked with SDSC researchers Igor Tsigelny and Valentina Kouznetsova, focusing their studies on deep learning – a type of machine learning that uses multiple layers of computer processing units. The students used deep learning models to examine inhibitors for conditions ranging from diabetic cataracts to Alzheimer’s disease. Their study tested approximately 250,000 potential drug candidates with 296 predicted inhibitors and resulted in a 96 percent accuracy rate. The findings showed several novel solutions for treating diabetic cataracts.

The REHS program, part of SDSC’s student outreach program, was developed by SDSC Education manager Ange Mason to help increase awareness of computational science and related fields of research to students in the San Diego region. Entering its 11th year in 2020, students are mentored by computational research scientists and other SDSC staff to gain exposure to career options and work-readiness skills through hands-on experience.

Protein Data Bank Archive Adds New Coronavirus Protease Structure

The Protein Data Bank archive, which contains more than 160,000 3D structures for proteins, DNA, and RNA, recently released a new Coronavirus protease structure following the recent coronavirus outbreak, an ongoing viral epidemic that earlier this year spread from China to populations all across the world.

The structure is a high-resolution crystal structure of 2019-nCoV coronavirus 3CL hydrolase (Mpro) as determined by Zihe Rao and Haitao Yang’s research team at ShanghaiTech University.

“Open access to PDB data ensures that rapid access to rigorously validated and expertly curated 3D structure information contributes to research in fundamental biology, biomedicine, bioenergy, and biotechnology,” said Stephen K. Burley, director of the RCSB Protein Data Bank.

The PDB archive is jointly managed by the Worldwide Protein Data Bank partnership, involving data centers in the United States, Europe, and Asia. U.S. operations are led by the RCSB Protein Data Bank at Rutgers, SDSC, and UC San Francisco. PDB data provide a starting point for structure-guided drug discovery.
The spread of the virus causing the COVID-19 outbreak affects the whole world. Virus outbreak such as COVID-19 demonstrate the need for an optimization of data management and data reuse from which we all can benefit.

Access to the immensely valuable data of past and current epidemics is not always equally accessible for different affected populations and countries. Under the urgent need to harness machine-learning and future AI approaches to discover meaningful patterns in epidemic outbreaks, we need to ensure that all data is FAIR (Findable, Accessible, Interoperable, Reusable but in this sense also meaning Federated, AI-Ready).

Study Finds Close Evolutionary Proximity Between Microbial Domains in the ‘Tree of Life’

A comprehensive analysis of 10,575 genomes as part of a multinational study led by UC San Diego researchers reveals close evolutionary proximity between the microbial domains at the base of the tree of life, the branching pattern of evolution described by Charles Darwin more than 160 years ago.

The study, published in Nature Communications in December, found much closer evolutionary proximity between the Archaea and Bacteria microbial domains than have most previous studies. This new result arises from the use of a comprehensive set of 381 marker genes versus a couple of dozen core genes such as ribosomal proteins typically used in previous studies, according to Qiyun Zhu, a postdoctoral scholar in the UC San Diego School of Medicine’s Department of Pediatrics and lead author of the paper.

SDSC Distinguished Scientist Wayne Pfeiffer made more than 2,000 runs on the standard compute nodes of Comet to generate the gene trees, while Uyen Mai, a PhD student in the Mirarab Lab at UC San Diego and co-first author of the paper, combined these trees using ASTRAL on Comet’s GPU nodes.

GO FAIR Foundation Launches Virus Outbreak Data Network (VODAN)

The spread of the virus causing the COVID-19 outbreak affects the whole world. Virus outbreak such as COVID-19 demonstrate the need for an optimization of data management and data reuse from which we all can benefit.

Access to the immensely valuable data of past and current epidemics is not always equally accessible for different affected populations and countries. Under the urgent need to harness machine-learning and future AI approaches to discover meaningful patterns in epidemic outbreaks, we need to ensure that all data is FAIR (Findable, Accessible, Interoperable, Reusable but in this sense also meaning Federated, AI-Ready).
As this column is written, the U.S. is in the throes of an uncertain response to a new virus-caused disease, COVID-19, whose properties are not well understood and for which no vaccine or treatment currently exists. In concert with guidance from the CDC and University of California, we have canceled our Technology Forum breakfast series for the foreseeable future and will greatly miss the opportunity to interact with current and prospective tech forum members. However, SDSC carries on in its mission of operating supercomputers to benefit science and society. And, as highlighted in recent news reports, supercomputing will be a ‘mission critical’ function in combating the COVID-19 disease caused by the SARS-CoV-2 virus.

In a recent news conference, the White House highlighted public-private partnerships as a key element of its game plan, joining and marshalling the efforts of government and private industry to respond to the pandemic.

The graphic provides an excellent illustration of the many benefits of public-private partnerships, and many of those apply to leveraging such partnerships in the current situation. In keeping with the spirit of this approach, SDSC staff and researchers stand ready to partner with industry to tackle this global challenge, whether it involves facilitating access to our supercomputer and storage systems, collaborating with our computational scientists, navigating other computational resources and collaborations available through the National Science Foundation’s XSEDE program, or all of the above.

If you have a current need or want to explore options, please contact us. We look forward to overcoming this challenge together and resuming our normal program of activities.

Ron Hawkins
SDSC Director of Industry Relations