SDSC Receives Top Awards at SC18 Conference

SDSC received two key HPCwire awards for 2018, recognizing the use of its Comet supercomputer in assisting scientists in finding the first evidence of a source of high-energy cosmic neutrinos, and for Comet’s role in a recent autism study led by a team of researchers at the UC San Diego’s School of Medicine.

[Continued on Page 6]
SDSC once again had a very successful showing at this year’s Supercomputing Conference held last month in Dallas. We had a heavy schedule of presentations throughout the event, including sessions by many SDSCers as well as our partners in academia and industry. Many thanks to our External Relations and Industry Relations groups for their preparation, with special kudos to Susan Rathbun for coordinating what is a multi-tiered undertaking for us each year.

Regarding the headline of this issue, the ongoing use of Comet among a diverse and ever-expanding research community was once again recognized by HPCwire at SC18 with two top awards: Editors’ Choice for the Top High-Performance Computing (HPC)- Enabled Scientific Achievement was for Comet’s role in assisting an international team of scientists at the National Science Foundation (NSF)-funded IceCube Neutrino Observatory in finding the first evidence of a source of high-energy cosmic neutrinos; Editors’ Choice for Best Use of High-Performance Computing in the Life Sciences recognized Comet’s use in recent findings related to autism spectrum disorder (ASD) as described in a study published in the April 20, 2018 online edition Science. SDSC has been an HPCwire awards winner for seven years now.

Comet also was recently extended for a sixth year of service under a supplemental National Science Foundation grant valued at almost $2.4 million, meaning it will continue as a researcher resource through March of 2021. Since entering production in mid-2015, Comet has served more than 1,700 unique PIs at some 350 institutions, and is one of the most widely used systems throughout the NSF’s eXtreme Science and Engineering Discovery Environment (XSEDE).

I’d also like to congratulate Ilkay Altintas, our chief data science officer, for being named a Fellow of the Halicioğlu Data Science Institute (HDSI). Ilkay is working with UC San Diego faculty, industry partners, and students at all levels in her new role. She was our very first ‘Pi Person of the Year’ in 2014 when we inaugurated this award, which recognizes SDSC researchers who, like the π symbol, have one ‘leg’ in a science domain and the other in cyberinfrastructure technologies. You can read more about Ilkay’s new appointment on page 4 in this issue.

Last but certainly not least, SDSC’s new Annual Report is now available online as a pdf at www.sdsc.edu/pub/index.html. I encourage you to browse through this document, which really showcases the diversity of our center both in people and the highly innovative projects they pursue. If you are in need of some hard copies of this new report, please let our External Relations group know.

Happy holidays, and may 2019 bring us all joy & prosperity!

Michael L. Norman
SDSC Director
**SDSC Awarded NSF Grant for Data Reproducibility Research**

SDSC has been awarded a three-year National Science Foundation (NSF) grant worth more than $818,000 to design and develop cyberinfrastructure that allows researchers to efficiently share information about their scientific data and securely verify its authenticity while preserving provenance and lineage information. The grant focuses on implementing a web-based cyberinfrastructure platform called Open Science Chain. "Data sharing is an essential element of scientific research and associated publications, and facilitating the future reuse of that data in a secure and verifiable manner is critical," said Subhashini Sivagnanam (top left), principal investigator for the grant and a principal scientific computing specialist with SDSC’s Data-Enabled Scientific Computing division. Viswanath Nandigam (lower left), associate director for SDSC’s Advanced Cyberinfrastructure Development Lab, is co-principal investigator for the project.

https://goo.gl/pvNk1V

**SDSC’s CIPRES Awarded Two Federal Grants to Support Biological Research Innovation**

The CIPRES science gateway, which supports major discoveries about evolutionary relationships among our planet's living creatures, was recently awarded grants from the National Science Foundation (NSF) and National Institutes of Health (NIH) providing more than $2.8 million to sustain and enhance the widely used resource. Short for CyberInfrastructure for Phylogenetic RESearch, CIPRES provides biologists access to sophisticated software tools that run on supercomputers including Comet. "Understanding the evolutionary history of living organisms is a central goal of nearly every discipline in biology," said SDSC Bioinformatics Researcher Mark Miller, also the gateway's principal investigator. "CIPRES's wide adoption by the biological community underscores the importance of this resource to the future of biological research."

https://goo.gl/xuytSJ

**SDSC Opens New BlockLAB Research Laboratory**

SDSC’s Center for Large Scale Data Systems (CLDS) has formally opened a new blockchain research laboratory with the objectives of exploring the principal technologies and business use cases in blockchains, distributed ledgers, digital transactions, and smart contracts. "There is much interest and experimentation underway in evaluating blockchain technologies and their applications in industry and government," said James Short (left), BlockLAB’s director and SDSC’s lead scientist at CLDS. "One of our primary goals is to work closely with industry partners to provide foundational knowledge to help science-based and industrial companies evaluate the potential benefits and risks of applying these new technologies to critical, large-scale transaction and data-intensive business processes." The new lab is founded in partnership with technology firms AEEC, Collibra, Decision Sciences, Dell Technologies, IBM, and Intel.

https://goo.gl/6gjmv1
Q: Can you provide some details about your new appointment at HDSI?
The appointment is for my role in the stewardship of HDSI cyberinfrastructure resources and services. That involves working with multiple organizations on campus and elsewhere to enable the use of computing and data services for education and research at HDSI.

Q: In addition to your research projects including WoRDS and WIFIRE, HDSI Fellows also have a significant academic role in data science areas. Can you elaborate a bit about your experience there?
On the education side, I am a lecturer in UC San Diego's Department of Computer Science and Engineering (CSE), and co-director of the MAS program in Data Science and Engineering, where I teach several courses. I have also taught several MOOCs (Massive Open Online Courses) on big data and data science, reaching out to hundreds of thousands of learners. Beginning this year, HDSI is offering an undergraduate program in Data Science and Engineering, which I hope to be involved with and transfer my experience with data science teaching and research to the activities within the program, such as the undergraduate student grants and projects.

Q: How does this appointment dovetail with other data-enabled science initiatives at SDSC?
This fellowship comes at a time of significant investments by SDSC for the Data Science Hub, where experts at SDSC and other areas of campus can form multidisciplinary data science teams to help provide solutions to regional, national, and global challenges such as smart cities, precision medicine, advanced manufacturing, and data center automation. Within the DSH, we have developed a flexible data gateway for students, faculty, and researchers to get access to data and create data solutions for their needs, which in turn will provide an operational platform for HDSI. We look forward to working with HDSI by not only developing new applications for data-driven research and analytics, but by actively training and thereby establishing a modern data science workforce that can help drive innovation.
This year’s Research Experience for High School Students (REHS) summer internships included a unique project: finding out what universities REHS alumni from the previous eight years of the SDSC program attended, and how their majors aligned with the internships. The study, conducted by students Mihir Gupta from Scripps High School and Andrew Liu from Del Norte High School in the San Diego area, was based on responses from 154 students from a combined 59 high schools from all over greater San Diego, and in some cases even outside the area. The survey found that 70% of REHS alumni who attended college attended a top 20 global university for their undergraduate education as ranked by U.S. News & World Report. Other REHS survey highlights:

- Responding REHS alumni attended 38 different schools for their undergraduate education, with UC Berkeley, UC San Diego, and UCLA being the top three schools, respectively, and accounting for more than 50% of responding alumni.
- Just under 66% of alumni respondents who attended college selected Computer Science & Engineering as their declared undergraduate major, making it the most popular choice well ahead of the next two, Biology (12.8%) and Mathematics & Physics (6.7%).
- 100% of responding students found at least one aspect of the internship valuable, with Research Experience listed as the most popular (97 of the 154 respondents), followed by Mentorship (84/154), Exposure to Computer Science (62/154), and Peer Collaboration (40/154).

“I am very proud of what this data suggests,” said Ange Mason, SDSC’s education manager and founder of the REHS program. “While we realize that students who apply for this internship may be predisposed to computer science and engineering or a related field, it’s gratifying to see how many of them pursue CSE in their university years and into their early careers.”

SuAVE, pandas, Python
The students used SuAVE (Survey Analysis via Visual Exploration), an online tool for exploratory survey data analysis developed by SDSC Researcher Ilya Zaslavsky. SuAVE is used to teach research methods to undergraduate students by helping them to visualize data using techniques from image analytics, faceted search, and online map navigation. Also used was pandas, an open-source BSD-licensed library providing data structures and analysis tools for the Python programming language.

“We also learned many things outside of our project,” said Gupta and Liu. “We got the chance to interview many researchers and ask them about their careers and jobs. Overall, REHS gave us real-life experience in the workplace as well as research experience and mentorship, all of which we consider to be invaluable tools that will definitely help us in our educational and vocational futures.”

More information about SDSC’s REHS program can be found at http://education.sdsc.edu/studenttech/?page_id=657
SDSC was recognized by the online publication with the following honors:

- **Editors' Choice for the Top High-Performance Computing (HPC)-Enabled Scientific Achievement**, for Comet’s role in assisting an international team of scientists at the National Science Foundation (NSF)-funded IceCube Neutrino Observatory in finding the first evidence of a source of high-energy cosmic neutrinos. Comet was also used in the verification and analysis of last year’s landmark discovery of gravitational and light waves generated by the collision of two neutron stars eons ago. That discovery, made by researchers at the NSF’s Laser Interferometer Gravitational Wave Observatory (LIGO), earned three researchers the 2017 Nobel Prize in Physics for their detection of gravitational waves in the universe as hypothesized by Albert Einstein in 1915.

- **Editors’ Choice for Best Use of High-Performance Computing in the Life Sciences**, recognizing Comet’s use in new findings related to autism spectrum disorder (ASD). The new study was published in the online edition of the April 20, 2018 issue of the journal Science.

This year’s awards were presented by HPCwire at the 2018 International Conference for High Performance Computing, Networking, Storage and Analysis (SC18), in Dallas, TX.

SDSC shared this award with several other organizations, including the NSF’s IceCube Neutrino Observatory and its eXtreme Science and Engineering Discovery Environment (XSEDE), as well as Stanford University, the Pittsburgh Supercomputing Center, and Globus, a research data management service.

“As an astrophysicist myself, this award is especially exciting for us,” said SDSC Director Michael Norman. “Comet’s contribution to the discovery of gravity waves occurred in 2015, but since that initial discovery LIGO has detected a total of five binary black hole collisions and one binary neutron star collision. In every case, Comet was part of the cyberinfrastructure that helped detect and verify those fantastic findings. Now Comet is contributing to another spectacular discovery, this time from IceCube. This illustrates the power and potential of an integrated national cyberinfrastructure.”
LIFE SCIENCES AWARD

HPCwire's 'Best Use of High-Performance Computing in the Life Sciences' award for 2018 is for recent findings related to ASD. While researchers established earlier that de novo mutations contribute to approximately one-third of cases of ASD, they identified a culprit that differs from previously known genetic causes of autism.

"For ten years we've known that the genetic causes of autism consist partly of de novo mutations in the protein sequences of genes. However, gene sequences represent only 2 percent of the genome," said Jonathan Sebat, a professor of psychiatry, cellular and molecular medicine and pediatrics at UC San Diego School of Medicine, and head of the Beyster Center for Molecular Genomics of Neuropsychiatric Diseases.

To investigate the other 98 percent of the genome in ASD, Sebat and his colleagues analyzed the complete genomes of 9,274 subjects from 2,600 families. One thousand were sequenced at Human Longevity Inc. (HLI) and at Illumina Inc., and DNA sequences were analyzed using Comet with the assistance of SDSC Distinguished Scientist Wayne Pfeiffer and SDSC Bioinformatics Programmer Analyst Madhusudan Gujral, who is also a co-author of the paper.

"Life sciences is one of our three key strategic priorities at SDSC, so we are gratified for this continuing recognition by HPCwire," said Norman. "Awards such as these underscore the significant role that high-performance systems are playing in helping to advance discovery in many areas of life sciences."

SDSC Activity at SC18

VIDEOS

Gateway to Data-Driven Discovery: An Overview of SDSC Partnerships and Projects by SDSC Director Mike Norman - https://youtu.be/dCO-av00lh0

Applications of Deep Learning at SDSC Mai H. Nguyen, Lead for Data Analytics at SDSC https://youtu.be/dCO-av00lh0

More SDSC videos from SC18 at https://goo.gl/h3BgHs

SDSC HPC Resources Featured in Latest XSEDE Science Highlights Brochure

The sixth edition of the annual eXtreme Science and Engineering Discovery Environment (XSEDE) brochure is now online! The latest edition features SDSC's Comet supercomputer, as well as Trestles and Gordon, among the national HPC resources used to advance scientific discovery. Download and share the Highlights Book via the URL below.

http://hdl.handle.net/2142/102126
Supercomputing Conference Showcases Vibrant Innovation Ecosystem

For organizations involved in High-Performance Computing, the ACM/IEEE Supercomputing (SC) conference held every November is an annual milestone for new product and technology announcements as well as bringing together the HPC community to reconnect and discuss new and existing projects and collaborations in a dynamic atmosphere.

The 2018 conference, held in Dallas, Texas, was no exception, with record attendance of over 13,000. For SDSC and our industry partners, it was an opportunity to showcase projects undertaken in 2018, such as testing a new distributed flash filesystem from partner Weka.IO and implementing real-time, in situ visualization of earthquake simulations on partner GigaIO’s recently announced high-performance interconnect fabric.

The SDSC display was a hub for technology instruction, including presentations by DDN (storage solutions for modern workloads), NVIDIA (machine learning on GPUs), and Sylabs (application containerization using Singularity). Meetings and discussions were held with technology companies big and small to hear about their newest technologies and discuss potential projects and collaborations in the coming year. At the end of the week, everyone returned to homebase tired but full of promise for the coming year.

Through its newly revamped Industry Partners Program, Advanced Technology Lab, upcoming National Science Foundation solicitations, and other programs, SDSC sees multiple opportunities for industry engagement on new supercomputers and other projects in 2019. For more information on how your company can participate, please contact us or visit industry.sdsc.edu.

Ron Hawkins
SDSC Director of Industry Relations