

# BEST PRACTICES FOR BUILDING AN HPC/CI TRAINING PROGRAM FOR THE NEXT GENERATION NSF WORKFORCE

June 23, 2023

EXPANSE  
COMPUTING WITHOUT BOUNDARIES

- **Mary Thomas**
- **SDSC HPC Training Lead**



# ***Best Practices For Building an HPC/CI Training Program for the Next Generation NSF Workforce***

The San Diego Supercomputer Center's HPC/CI Training program is designed to train the next generation of the NSF workforce community. Our program focuses on training our user communities in emerging technologies and relevant topics, including advanced cyberinfrastructure, high-performance computing, data management, reproducible computing, and scientific applications such as ML and deep learning. Our training team consists of SDSC/UCSD faculty and staff, collaborators, and vendors. We offer various types of training events, including webinars, tutorials, workshops, and institutes, which range from one hour to several days and cater to different communities, including K-12 students, undergraduates, graduates, and working professionals. All events and materials are archived for data provenance and converted into online, accessible training materials. We ensure that all training materials follow the FAIR (Findable, Accessible, Interoperable, and Reusable) practices and are freely available online through GitHub and our interactive video web pages. It is important to note that our training team collaborates with the UCSD EDI offices to ensure that our outreach follows established practices and meets diversity goals. Since the launch of this program, we have successfully hosted more than 110 events and made an impact on over 13,000 participants. In this presentation, we will describe our training program approach and development, highlight its current impact, and discuss our future plans

# Outline

- Introduction
- Defining HPC/CI
- Training Program Design
- Current SDSC HPC/CI Training Program
- Synergistic Activities Broaden Impact
- HPC Students Program
- Broadening Impact
- Conclusions & Future Plans



# The San Diego Supercomputer Center

*A leader in high-performance and data-intensive computing and cyberinfrastructure*

<https://www.sdsc.edu/>



SDSC provides resources, services and expertise to the local, regional, and national research community, including industry and academia. It supports hundreds of multidisciplinary programs spanning a wide variety of domains.

<https://timeline.sdsc.edu/>



# SDSC by the Numbers

**250++  
Employees**

**~3,000  
Training & Event  
Participants/year**

**~10,000  
Active Unix  
Accounts  
on HPC systems**

**1M++  
users on  
Science gateways**

**1M++  
Students took  
our Big Data courses**

**4 HPC Systems  
~200,000 x86 cores  
~1,500 GPUs**

**AI/ML Supercomputing  
Habana/Intel hardware  
SDSC Expertise**

**Universal Scale Storage  
Open for business from  
200TB to 10's of PB**

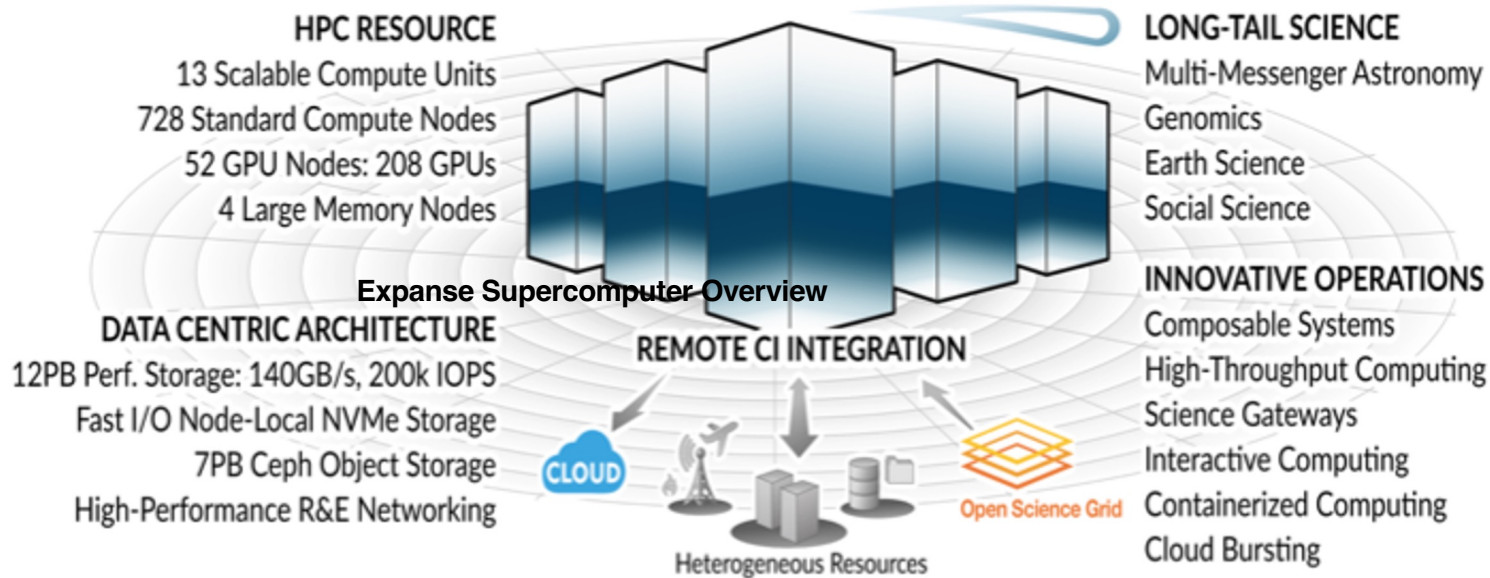
**Globally Federated  
Cyberinfrastructure  
100++ institutions  
on 5 continents**

**We design, deploy, and operate end-to-end solutions for our partners  
from academia, government, industry & non-profits**

# Advanced Computing Systems:

Needed for large scale HPC applications

EXPANSE COMPUTING WITHOUT BOUNDARIES  
5 PETAFL0P/S HPC and DATA RESOURCE



For more details see the Expanse user guide @ [https://www.sdsc.edu/support/user\\_guides/expanse.html](https://www.sdsc.edu/support/user_guides/expanse.html)  
and the "Introduction to Expanse" webinar @ [https://www.sdsc.edu/event\\_items/202006\\_Introduction\\_to\\_Expanse.html](https://www.sdsc.edu/event_items/202006_Introduction_to_Expanse.html)

SDSC SAN DIEGO  
SUPERCOMPUTER CENTER

UC San Diego

SDSC SAN DIEGO  
SUPERCOMPUTER CENTER

UC San Diego



# Expanse





# Outline

- Introduction
- **Defining HPC/CI**
- Training Program Design
- Current SDSC HPC/CI Training Program
- Synergistic Activities Broaden Impact
- HPC Students Program
- Broadening Impact
- Conclusions & Future Plans





# Early parallel computers based on commodity hardware

Stone Soupercomputer (2003):  
Cheapest cost/flop=\$0, ~20 Mflops

<https://web.archive.org/web/20031121211117/http://stonesoup.esd.ornl.gov>

**SDSC** SAN DIEGO SUPERCOMPUTER CENTER  
OF THE UNIVERSITY OF CALIFORNIA SAN DIEGO  
COMPUTING WITHOUT BOUNDARIES

ABOUT SDSC SERVICES SUPPORT RESEARCH & DEVELOPMENT EDUCATION & TRAINING NEWS & EVENTS

## College through Career

SDSC offers hands-on training in computational thinking, high-performance computing and big data exploration to students and researchers at nearly every level of sophistication. SDSC "Training" for research professionals focuses on building their skills to use high-performance computing, data-intensive computing, and data analytics within their own research disciplines – from neuroscience and geophysics to the humanities, arts and social sciences.

### TRAINING PROGRAMS

- Advanced Computing**  
Training classes and workshops introduce new and current users to high-performance computing, data and visualization resources available at SDSC, and provide the skills necessary to use them.
- CIML Summer Institute**  
SDSC's **Cyberinfrastructure-Enabled Machine Learning (CIML)** project teaches researchers and students the best practices for effectively running machine learning and data science applications on advanced cyberinfrastructure and high-performance computers.
- Convergence Research (CORE) Institute**  
This training program provides graduate students and early and mid-career researchers and professionals with a foundational experience to position them for impact throughout their careers on the most challenging societal issues of our time.
- HPC & Data Science Summer Institute**  
This week-long summer workshop focuses on a broad
- Internet Data Analysis**  
The Center for Applied Internet Data Analysis (CAIDA)  
workshops cover various

**SDSC** SAN DIEGO SUPERCOMPUTER CENTER  
OF THE UNIVERSITY OF CALIFORNIA SAN DIEGO  
COMPUTING WITHOUT BOUNDARIES

UC San Diego  
Directory Status Business

ABOUT SDSC SERVICES SUPPORT RESEARCH & DEVELOPMENT EDUCATION & TRAINING NEWS & EVENTS

HOME > EDUCATION & TRAINING > ADVANCED COMPUTING TRAINING

## Advanced Computing Training

Training classes and workshops introduce new and current users to the high-performance computing, data, and visualization resources available at SDSC and provide the programming skills necessary to use SDSC resources effectively and efficiently. Workshops offer experienced users more in-depth instruction, including hands-on assistance with their own codes and collaborative discussions with other users and parallel computing experts.

### Upcoming

- JUNE 27-29 2023** **2023 Cyberinfrastructure-Enabled Machine Learning Summer Institute**  
The CIML Summer Institute will involve introducing ML researchers, developers and educators to the techniques and methods needed to migrate their ML applications from smaller, locally run resources, such as laptops and workstations, to large-scale HPC systems, such as the SDSC Expanse supercomputer. (Application deadline is Friday, April 7.)  
*Location: 15th Floor at the Village Room A & B, UC San Diego Campus and remote*
- JULY 26 2023** **Scientific Computing with Kubernetes**  
In this webinar, we provide recipes for transitioning scientific workloads that currently run on traditional batch systems to Kubernetes systems. Kubernetes is batch-like in nature, but there are some differences that science users should be aware of. We will also briefly describe capabilities that are not found in traditional batch systems that can improve the effectiveness of scientific computing.  
*This event will be held remotely.*
- AUG 7-11 2023** **2023 HPC and Data Science Summer Institute**  
The SDSC Summer Institute is a week-long workshop hosted by the San Diego Supercomputer Center (SDSC) at the University of California, San Diego, focusing on a broad spectrum of

### Training Links

- Training Catalog
- Training for Expense Users
- All SDSC Events
- Resource Documentation (User Guides)
- Expanse Home
- NRP Home
- TSCC Home
- HPC Students

**Training Coordinator**  
training-info@sdsc.edu

**Industry Partners**  
ipp@sdsc.edu

### Subscribe

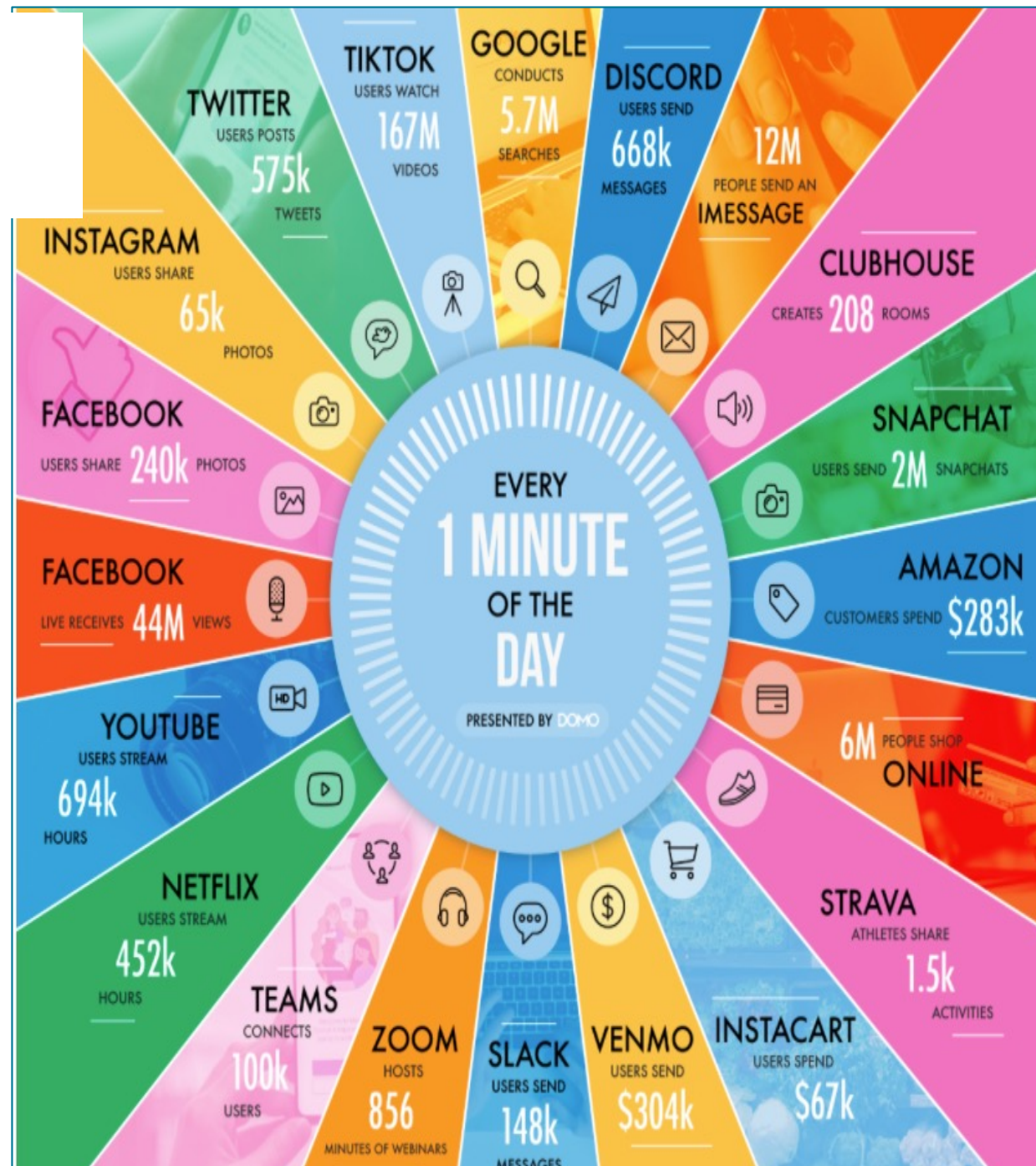
Get the latest SDSC news and events via email:  
**SIGN UP TODAY!**

# HPC/CI Is Everywhere: You use it daily!

- The “Internet”
- Reaches 65% of the world (5.6 billion)
- 93% access via mobile technologies
- Data consumed globally in 2021: 79 zettabytes\* (2021), projected to be 180 zettabytes by 2025

\*Zettabyte=1,000,000,000,000,000,000 [10<sup>21</sup>] bytes

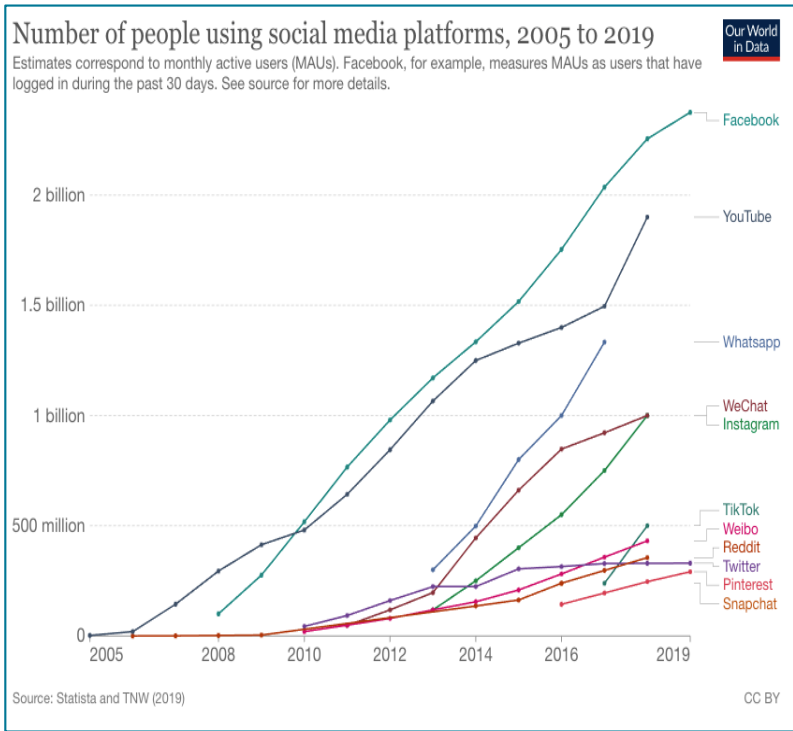
Source: <https://www.visualcapitalist.com/from-amazon-to-zoom-what-happens-in-an-internet-minute-in-2021/>



Src: <https://www.washingtonpost.com/graphics/national/security-of-the-internet/bgp/>



# HPC/CI Is Everywhere: billions of users performing billions of transactions sending billions of data packets

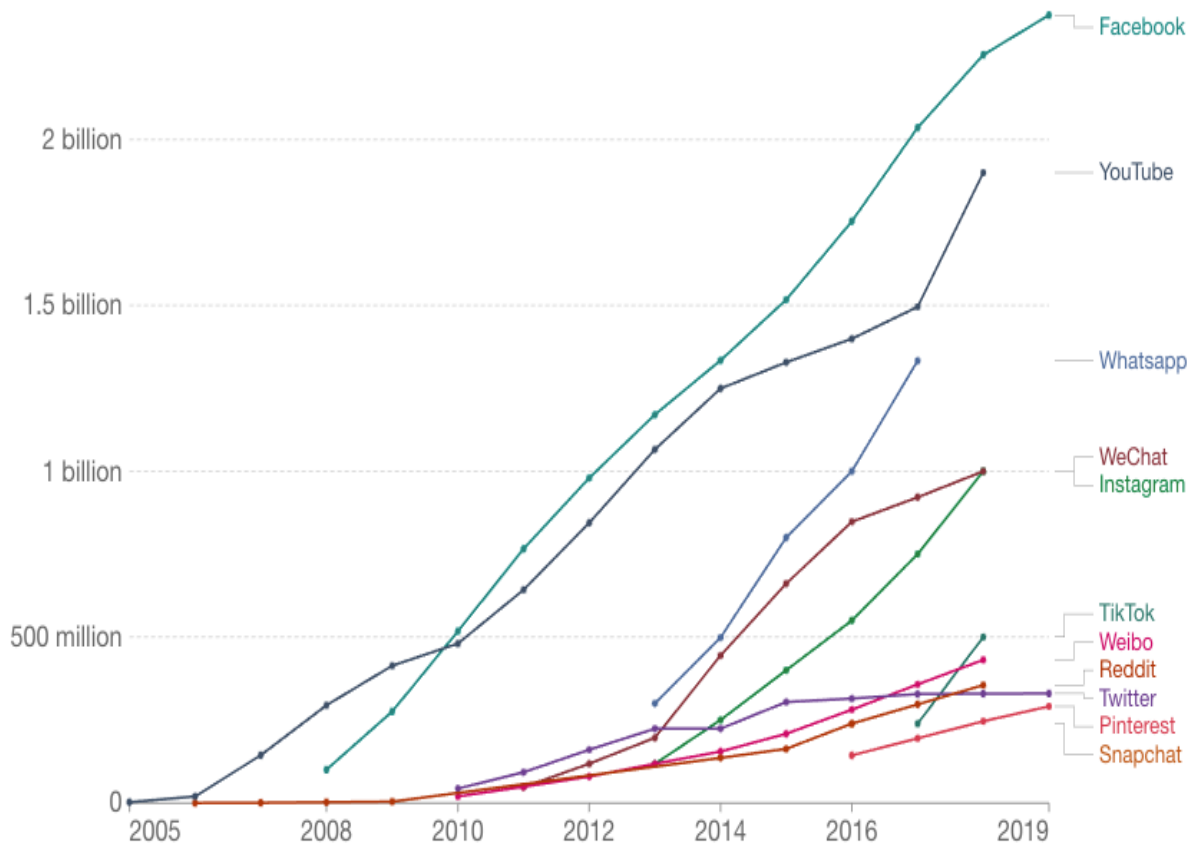


Source: [https://ourworldindata.org/exports/users-by-social-media-platform\\_v17\\_850x600.svg](https://ourworldindata.org/exports/users-by-social-media-platform_v17_850x600.svg)

Source <https://www.visualcapitalist.com/from-amazon-to-zoom-what-happens-in-an-internet-minute-in-2021/>

## Number of people using social media platforms, 2005 to 2019

Estimates correspond to monthly active users (MAUs). Facebook, for example, measures MAUs as users that have logged in during the past 30 days. See source for more details.



Source: Statista and TNW (2019)

CC BY

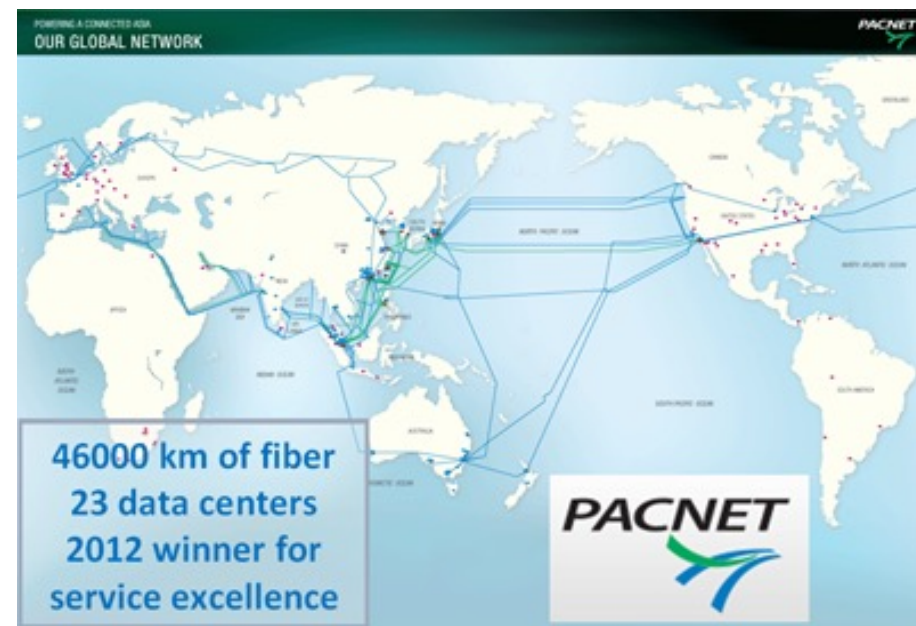
# HPC/CI Is Everywhere: Where do we house all this data?

- Locate it where it is easy/cheap to keep cool (e.g. Iceland)
- Current Idea: build skyscraper sized data storage facilities
- Need to keep it “localized” for efficient access



# HPC/CI Is Everywhere: How do we send all this data?

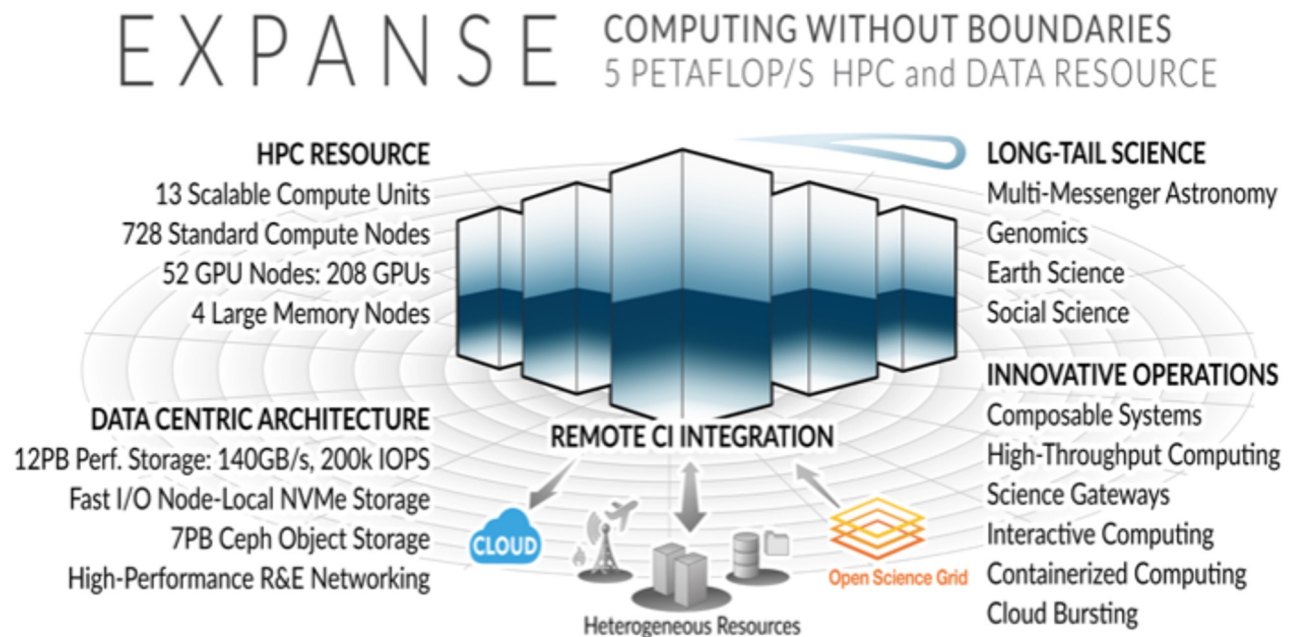
- Gigabit+ networks





# What is High Performance Computing (HPC) and Cyberinfrastructure (CI)?

- *HPC: Aggregating computing power*
- *CI: Connecting resources with distribute emerging technologies*
- *Deliver much higher performance than desktop computer or workstation*
- *Solve large problems in science, engineering, or business.*



For more details see the Expance user guide @ [https://www.sdsc.edu/support/user\\_guides/expance.html](https://www.sdsc.edu/support/user_guides/expance.html)  
and the "Introduction to Expance" webinar @ [https://www.sdsc.edu/event\\_items/202006\\_Introduction\\_to\\_Expance.html](https://www.sdsc.edu/event_items/202006_Introduction_to_Expance.html)

**SDSC** SAN DIEGO  
SUPERCOMPUTER CENTER

UC San Diego

# What is a Supercomputer Cluster?

A very large set of compute nodes ( $10^5 - 10^6+$ ) that work together...



← SDSC Expanse: 13+ racks; 728 AMD EPYC nodes,  $\sim 10^5$  cores; 52 GPU nodes; > 12 PB  
→ 5.2 petaflops.  
( $10^{15}$  floating-point operations per second.)

ORNL Frontier: 100+ racks; 9000 AMD EPYC nodes,  $> 10^6$  cores; > 32 PB  
→ 1.5 exaflops  
( $10^{18}$  floating-point operations per second)





# Outline

- Introduction
- Defining HPC/CI
- **Training Program Design**
- Current SDSC HPC/CI Training Program
- Synergistic Activities Broaden Impact
- HPC Students Program
- Broadening Impact
- Conclusions & Future Plans



# Training Program: The Challenge

- 2018 search of “what HPC training is available” nationwide?
  - In-depth search yielded a large number of events at several large HPC centers
  - But very few events led to training materials (videos, presentations, etc.)
- **Serious lack of training material provenance**
- **Decision to reorganize HPC training events, and ensure that training material available persistently.**



# Training Program Redesign: Clarify nature of training modalities

- **Webinars:** somewhat passive video presentations
  - Record & archive; large groups (zoom limits)
- **Tutorials & workshops:**
  - A little more interactive, can be a few hours to a few days.
  - Record & archive; limited to large groups
- **Annual Summer institutes** (multiday, multiple topics)
  - Highly interactive, small groups
  - Record & archive
- **HPC User Training:** 12-14 weeks of 2 hours sessions
  - Highly interactive, small groups
  - Record & archive
- **Self-guided training:** base on webinars, tutorials, etc.

# Training Program Redesign: Organization of Training Materials

- Archive all training materials and repurpose/reuse
- Define keywords/tags to describe/search events:
  - Title, abstract, date, presenter, related training materials, related training pages, GitHub repos, pdfs, YouTube, ....
- Integrate training terms into the SDSC Web content management system (Cascade)
- Create dynamic event pages for upcoming events
- Develop training catalog that lists past events
- Pull event data from content management system (Cascade) and use to build other pages
- **Develop Tools** as needed for accessing materials



# Training Program Redesign: Identify/Expand Training Communities

- Users of SDSC HPC/CI systems and projects:
  - Systems: Expanse, Voyager (AI), TSCC (condo)
  - National programs: XSEDE/ACCESS
- HPC Students Program:
  - Undergrad, grad;
  - focus on UCSD, but keep open for all schools
  - Research Experience for High School (REHS)
- Identify ways to improve EDI of our programs
- **NSF WFD**: faculty, researchers, postdocs, gov, industry
  - Leveraging funding from CyberTraining awards

# Outline

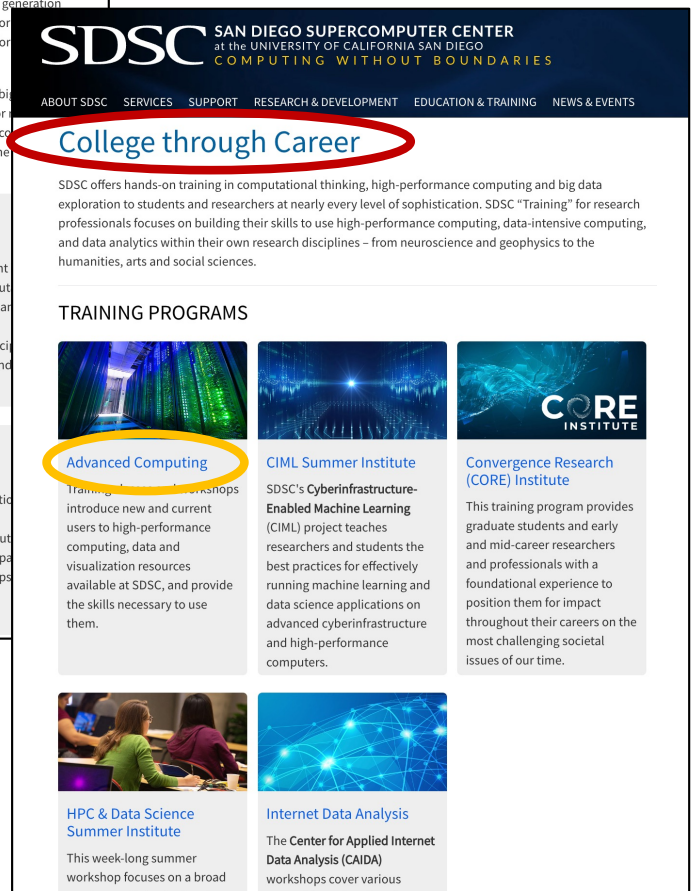
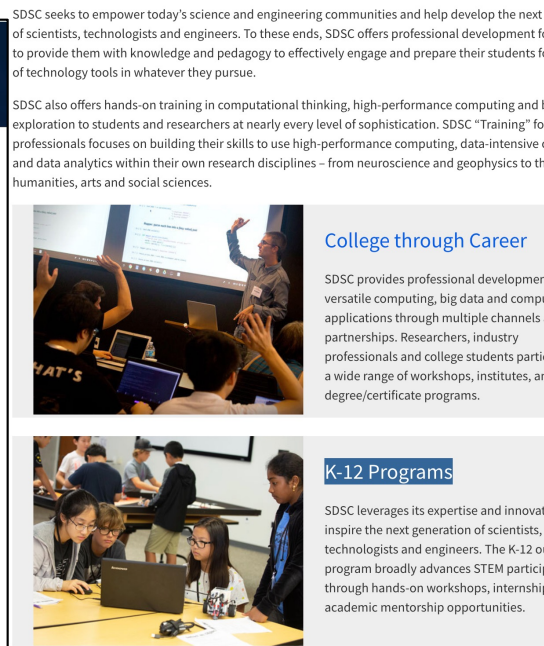
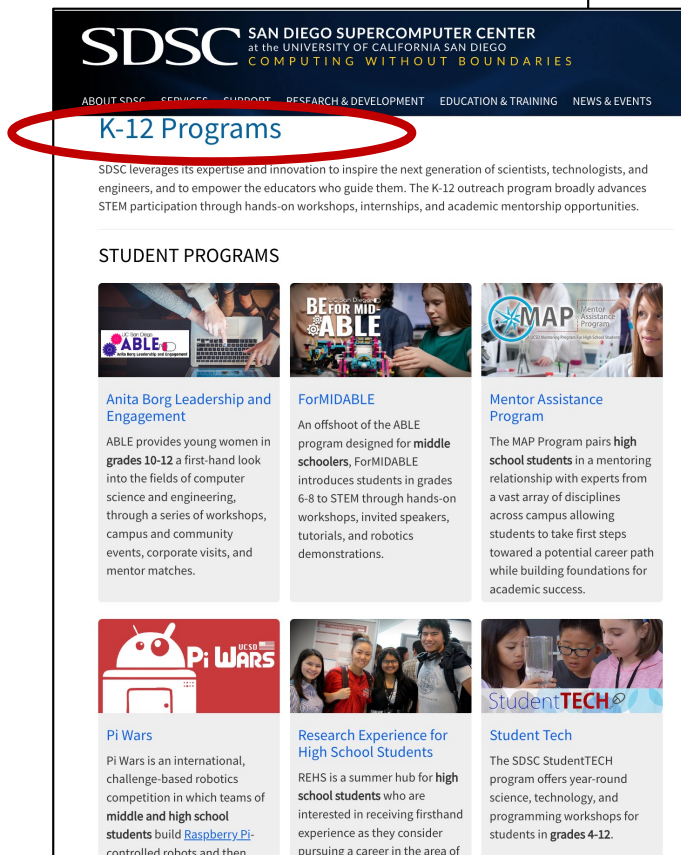
- Introduction
- Defining HPC/CI
- Training Program Design
- **Current SDSC HPC/CI Training Program**
- Synergistic Activities Broaden Impact
- HPC Students Program
- Broadening Impact
- Conclustions & Future Plans



# Refactor SDSC Home Pages to Reflect Diverse Training Programs

**K-12 (REHS)**  
**Coordinated by Ange Mason**  
**~12-15 years**

**Multiple programs from**  
**Coursera to HPC Training**





# Advanced Computing Training Program

- All events are in the CMS system
  - Event data provenance
- Dynamically build pages using CMS database:
  - Upcoming events
  - Training event catalog
  - Interactive video pages
- New tools to support training
- Moving towards searchable content
  - Metadata

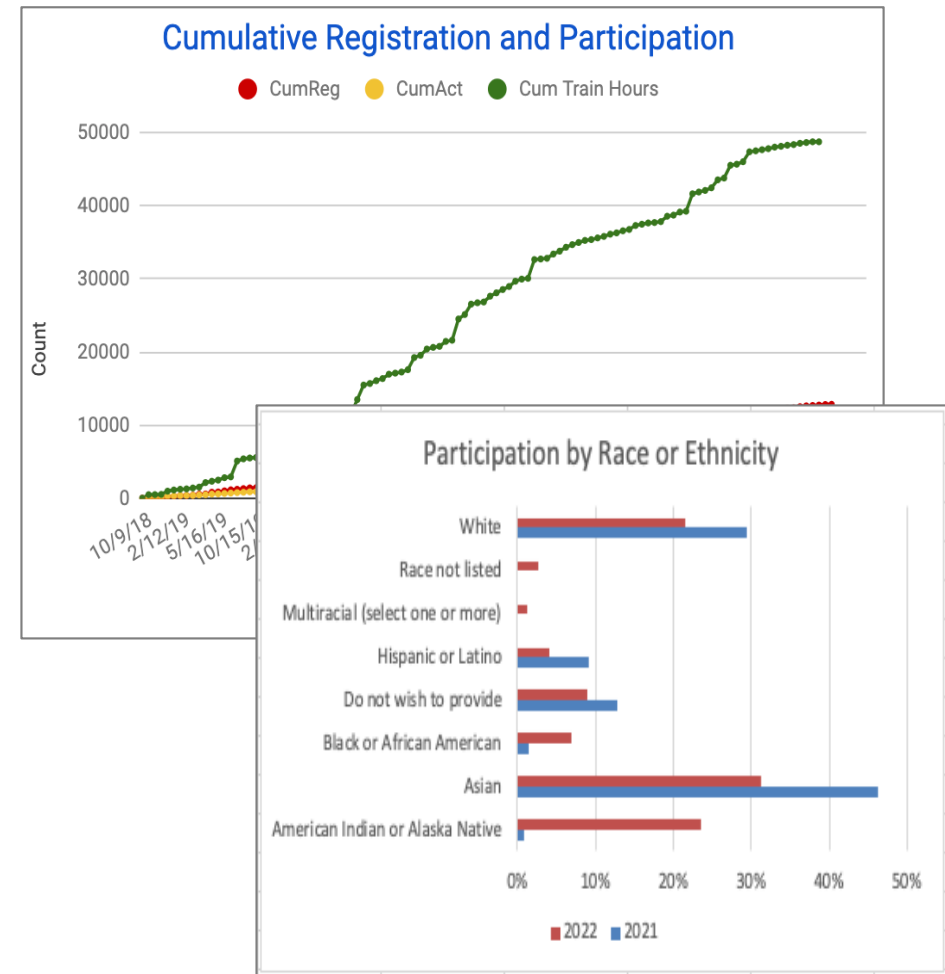


The screenshot shows the SDSC website's 'Advanced Computing Training' page. The header includes the SDSC logo, 'SAN DIEGO SUPERCOMPUTER CENTER at the UNIVERSITY OF CALIFORNIA SAN DIEGO', and the tagline 'COMPUTING WITHOUT BOUNDARIES'. The UC San Diego logo is in the top right. A navigation menu contains 'ABOUT SDSC', 'SERVICES', 'SUPPORT', 'RESEARCH & DEVELOPMENT', 'EDUCATION & TRAINING', and 'NEWS & EVENTS'. A search bar is also present. The main content area features the title 'Advanced Computing Training' and a paragraph describing training classes and workshops. Below this is an 'Upcoming' section with three entries: '2023 Cyberinfrastructure-Enabled Machine Learning Summer Institute' (June 27-29, 2023), 'Scientific Computing with Kubernetes' (July 20, 2023), and '2023 HPC and Data Science Summer Institute' (August 7-11, 2023). A 'Training Links' sidebar on the right lists 'Training Catalog', 'Training for Expense Users', 'All SDSC Events', 'Resource Documentation (User Guides)', 'Expense Home', 'NRP Home', 'TSCC Home', and 'HPC Students'. At the bottom of the sidebar are 'Training Coordinator' (training-info@sdsdc.edu), 'Industry Partners' (ipp@sdsdc.edu), and a 'Subscribe' section with the text 'Get the latest SDSC news and events via email: SIGN UP TODAY!'.

[https://www.sdsdc.edu/education\\_and\\_training/training\\_hpc.html](https://www.sdsdc.edu/education_and_training/training_hpc.html)

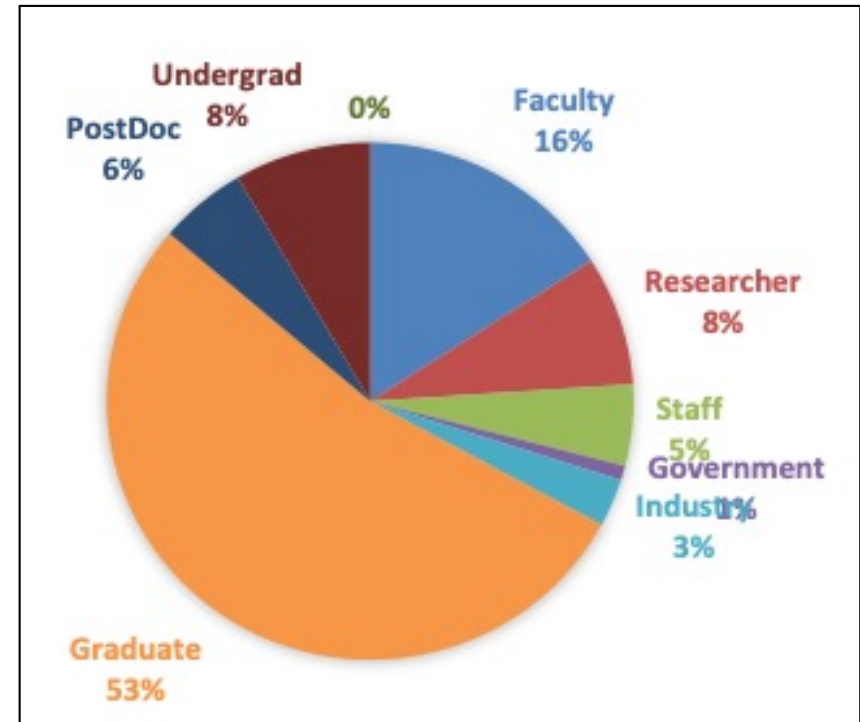
# HPC Training pages provide information for future and past events

- Training event data stored in web site database
  - Used to list upcoming events
  - Customize for HPC systems
  - Training repos accessed > 1000+ / cloned over 500+ times during reporting period
- Event information archived and *dynamically* available.
- Training Catalog
  - 49 archived events
  - Links to associated training materials
  - Interactive search via keywords/Tags: year, topics, type, system, level of difficulty, etc.



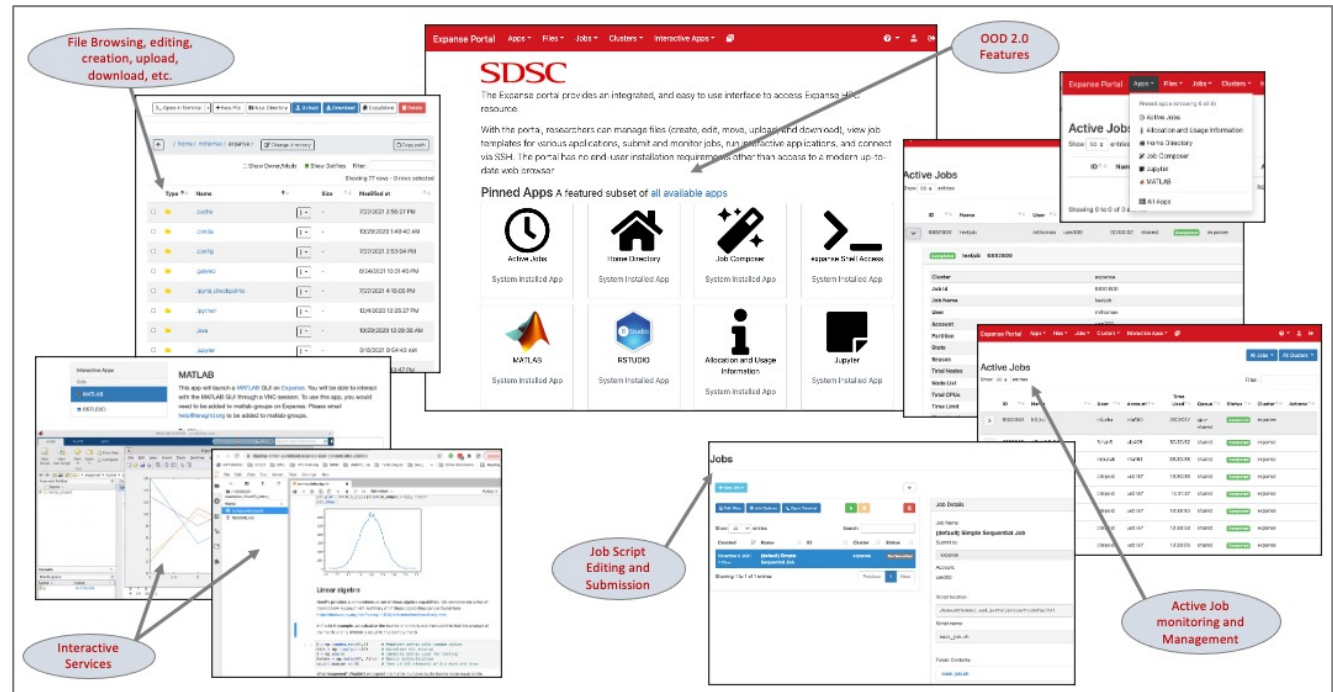
# Interactive videos continue to be a rich source of training content

- Hosting 60+ current and archived events
  - Very popular: 4105 unique visitors; Up from 724 in 2021
  - Every training event with recordings is converted
  - During reporting period, converted 37+ Events, > 180 hours of recordings
- Features:
  - Synchronized, searchable table of contents and transcripts; keyword queries
  - Embedded annotations
  - Links to associated tutorials & GitHub repos
- Pages created dynamically:
  - Can now search by year and title keywords
  - Working on developing metadata tag system





# Expansive User Portal



- Features: file browser & management; Pinned Apps; Job templates, submission, management
- Interactive applications: Notebooks, Jupyter Lab, Matlab, Rstudio
- Authentication: ACCESS (Advanced Cyberinfrastructure Coordination Ecosystem: Services & Support) credentials
- Used for most training events
- Based on Open On Demand

# Satellite Reverse Proxy Service: Galyleo Client

- Continues to be used daily on Expanse, TSCC, and Comet
- Users launch secure (HTTPS) Jupyter Notebooks:
  - Simple command line client (galyleo)
  - Also use behind Expanse portal job submission
  - Supports system and custom containers, and Conda environments



## CIP Cyberinfrastructure Professional FELLOWS PROGRAM

**CyberTraining: Training and Developing a Research Computing and Data (RCD) CI Professionals Community**  
*A joint program between UC San Diego, San Diego State and CSU San Bernardino partners to develop a CIP workforce.*

### Program Overview

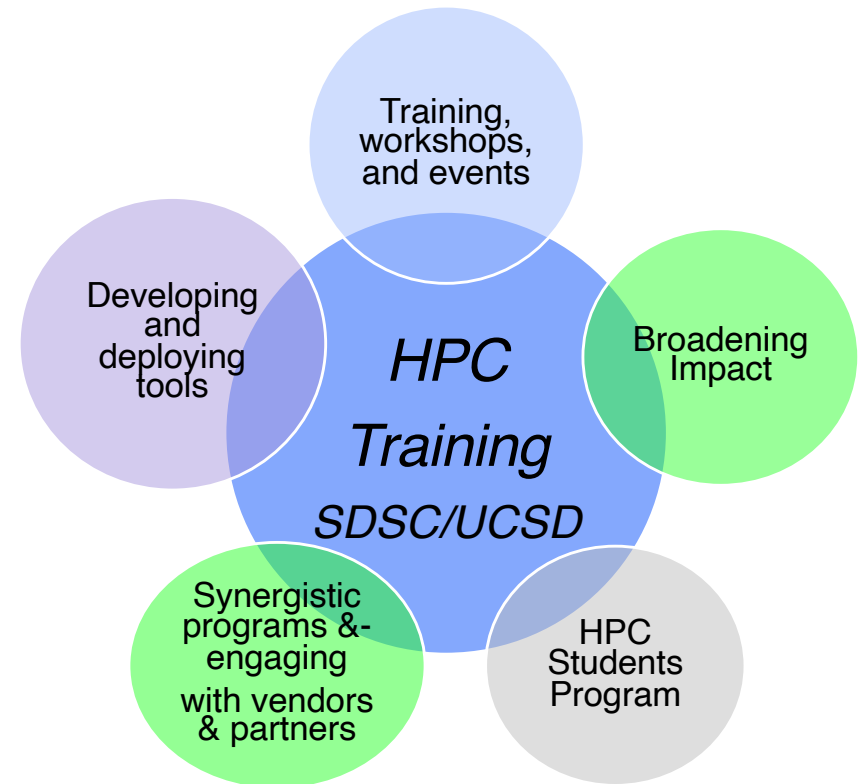
A critical role within scientific research teams is the cyberinfrastructure (CI) professional who has interdisciplinary skills that bridge the science, engineering, and computing specialties. Readily finding such personnel, with the necessary experience needed to navigate CI ecosystems has become a bottleneck for many projects. The Cyberinfrastructure Professional (CIP) Fellows Program is designed to address this need by training and mentoring a team of interdisciplinary Researcher-Facing (CARCC) CI Professionals with individualized training program in advanced computing and workforce development topics, embedding them into institutional departments, teaching them how to support and facilitate research projects, fostering a long-term CIP community, and defining sustainable career paths specifically for these kinds of professionals.

Eligible candidates include those who have domain science and/or computing backgrounds and want to advance their skills in CI, HPC, data science, and have an interest in facilitating scientific research teams who need to use CI resources and services. Potential candidates include: CI research support staff, software engineers, and system administrators, data curators, computational research scientists and engineers, and CI facilitators (people who work directly with researchers to help them to make effective use of Cyberinfrastructure (CI)). The creation of these new CIP-Fellow positions supports the goal of creating long-term career development paths and opening doors to new opportunities for CI Professionals.

**The CIP-Fellows program is seeking candidates who have the appropriate science and/or computing background and the motivation to work with research teams. [see below]**

# Building WFD community through synergistic and connected activities

- **Hosting training, workshops, and events:**
  - Workshops on preparing for Expanse (CPU/GPU): Data science, & machine learning Institutes; Neurosciences Gateway; HSI-STEM workshops
- **Deploying tools** to make it easier to learn about and use HPC/CI
- **Continuing engagements with partners:**
  - AMD User Forum; OpenACC Hack-a-thon; HPC@MSI to support minority serving institutions; HPC Student sponsorships
- **HPC Students**
  - Campus impact: CSE collaboration; Supercomputing Club part IEEE chapter;
  - SC22 Student Cluster Competition team winning HPL benchmark
- **Broadening impact:**
  - Awarded two CyberTraining Grants; contributing to other grants & projects
  - Survey & DEI metrics; meet Campus and NSF standards



Targeting the NSF Workforce Development (WFD) Community



# HPC/CI Training Program: Topics

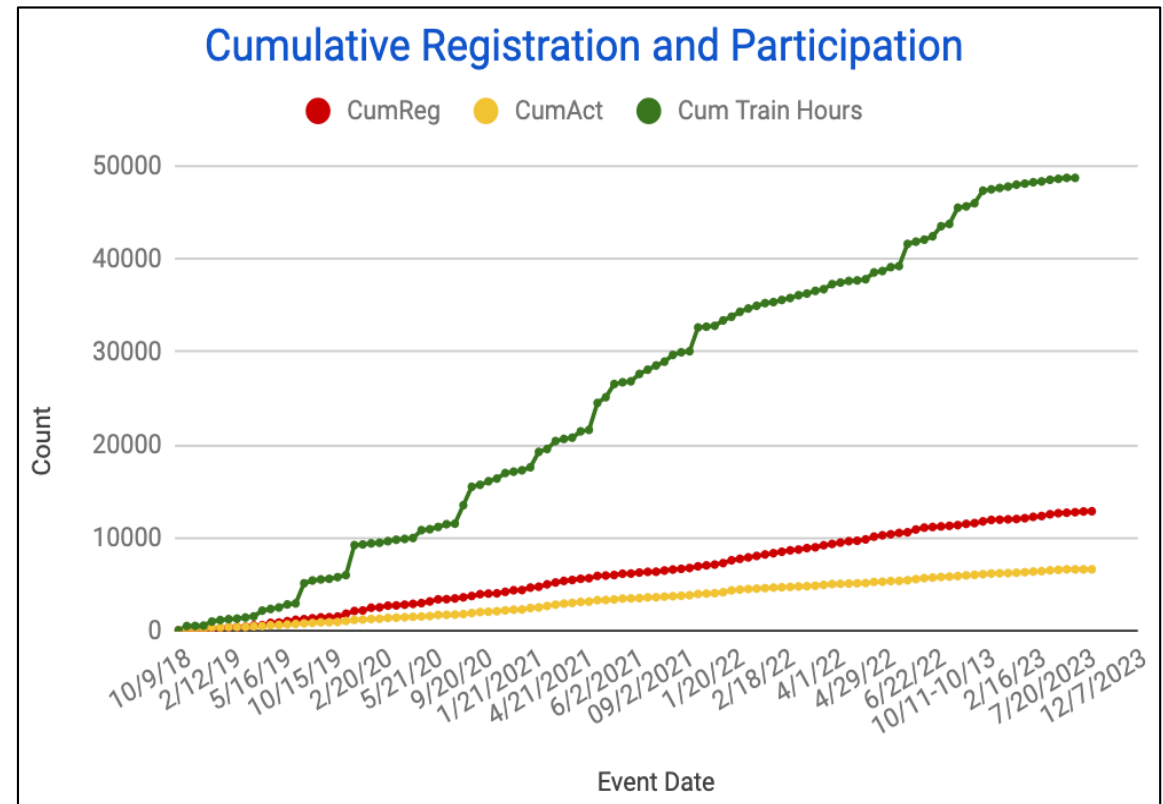
- Parallel Programming Concepts HPC architectures (CPU, GPU), software, Linux, admin skills.
- Running HPC applications in the areas of bioinformatics, numerical methods, password security, and other applications.
- Cloud Computing
- Visualization and analysis of big data sets.
- Learn the basics of parallel programming, including MPI using C, Fortran, and possibly other languages.
- Performance characterization of the cluster
- Current Trends: Anything AI/ML; Singularity; Kubernetes; Jupyter Notebooks; cloud computing; accelerators

# Training activities focus on productive use of SDSC HPC Systems & new concepts

| Date     | Topic  | Type  |             |
|----------|--|---|-------------|
| 12/14/21 | Expance Webinar - Running Jupyter Notebooks on Expance   | W   |             |
| 01/14/22 | [Week 1] Advanced HPC/CI User Training: Session 1 - Parallel Programming Concepts                  | T   |             |
| 1/20/22  | Expance Webinar - Parallel Computing Concepts  | W   |             |
| 1/21/22  | [Week 2] HPC/CI Training Series: Parallel Computing Concepts & HPC overview & Expance Architecture | T   |             |
| 1/28/22  | [Week 3] HPC/CI Training Series: Job Submission - Queues and ba                                    |   |             |
| 2/4/22   | [Week 4] HPC/CI Training Series:   |   |             |
| 2/11/22  | [Week 5] HPC/CI Training Series:   |   |             |
| 2/17/22  | Expance Webinar - Accessing an   |   |             |
| 2/18/22  | [Week 6] HPC/CI Training Series:   |   |             |
| 2/25/22  | [Week 7] HPC/CI Training Series:   |   |             |
| 3/2/22   | Voyager Training session   |   |             |
| 3/4/22   | [Week 8] HPC/CI Training Series:   |   |             |
| 3/3/22   | TSCC 101   |   |             |
| 3/17/22  | Expance Webinar - Singularity -  |   |             |
| 4/1/22   | [Week 9] Advanced HPC/CI User  |   |             |
| 4/8/22   | [Week 10] HPC/CI Training Series:  |   |             |
| 4/15/22  | [Week 11] HPC/CI Training Series:  |   |             |
| 4/21/22  | Expance Webinar - Enduring Sec   |   |             |
| 04/22/22 | [Week 12] HPC/CI Training Series:  |   |             |
| 04/27/22 | Parallel and GPU Computing with  |   |             |
|          | <b>Date</b>  | <b>Topic</b>  | <b>Type</b> |
|          | 4/29/22  | [Week 13] HPC/CI Training Series: GPU Architectures, CUDA GPU Computing with CUDA Python (TBD to confirm with Andy on this topic) | T           |
|          | 5/4/22   | Kubernetes Tutorial   | T           |
|          | 5/6/2022   | [Week 14] HPC/CI Training Series: GPU Profiling, monitoring: comms, NSight  | T           |
|          | 5/3/22   | NVIDIA GPU Hackathon - Prep Day   | T           |
|          | 05/10 - 05/12  | NVIDIA GPU Hackathon (5 days)   | T           |
|          | 05/19/22   | Expance Webinar - Intro to Neural Networks and Deep Learning on Expance   | W           |
|          | 6/16/22  | Expance Webinar - Reproducibility   | W           |
|          | 6/22/22  | CIML Prep Day   | T           |
|          | 06/27-06/29  | CIML 2022 Summer Institute (#applicants)  | T           |
|          | 7/27/22  | SDSC HPC/DS Prep Day  | T           |
|          | 08/01 - 08/05  | SDSC HPC/DS Summer Institute (#applicants)  | T           |
|          | 09/22/22   | Expance Webinar - Parallel Concepts   | W           |
|          | 10/13/22   | TSCC Interactive Computing Training   | t           |
|          | 10/11-10/13  | AMD HPC User Forum Fall Meeting - 2022  | O           |
|          | 10/20/22   | Expance Webinar - Kubernetes  | W           |
|          | 10/21/22   | Jupyter Notebook workshop for HSI STEM faculties and their collaborators  | T           |
|          | 11/8/22  | Voyager Training - Part-II: Voyager Habana arch + Porting TensorFlow & PyTorch  | T           |
|          | 11/15/22   | AMD HPC User Forum Meeting @ SC22   | O           |

# Training activities focus on productive use of SDSC Systems and new concepts

- Participation Summary:
  - 10/2018 – 12/2023
  - 110 scheduled events
  - 207 training days
  - Cum Registration ~ 12,907
  - Cum Actual ~ 6,664
  - Participation: ~54%
    - Mostly virtual
- Training Effort/Contact hours:
  - ~ 822 training hours
  - ~ 47,340 cum participant hours





# Tracking impact: outreach and diversity efforts

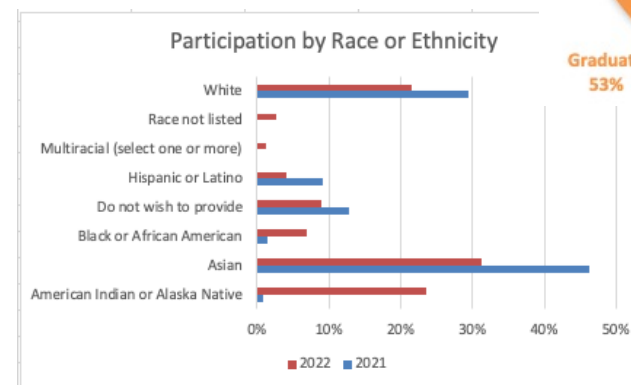
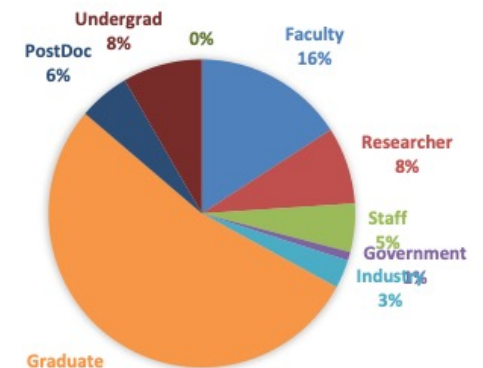
- Current efforts to expand/enhance/ensure/ diversity in our activities:
  - Work with UCSD EDI for input/guidance
  - Track diversity for event registrations; reach out to more organizations
  - Track all event registrations & participation
  - Reach out to more organizations & institutions
  - Survey participants before, during, after
- Expanse MSI program: allocation awards
- Focussed events: Summer Institutes, NSG workshop; SCC teams
- University of California orgs: designated as HSIs

# Training Program Outreach Goal: Impacting Underrepresented STEM Communities

- Focussed events: Summer Institutes, NSG workshop; Student Cluster Competition team
- Training Metric Successes:
  - Participation: Somewhat reflects UCSD; some improvement in some cases
  - Graduates@50%
  - Improved participation by gender
- CIML'23: 67% MSI Appl's

| Gender                 | Count | 2022  | 2021  |
|------------------------|-------|-------|-------|
| Man                    | 234   | 67.8% | 68.6% |
| Woman                  | 94    | 27.2% | 20.1% |
| Do not wish to provide | 14    | 4.1%  | 4.3%  |
| Nonbinary              | 3     | 0.9%  |       |

Participation by Gender: some improvement



# Outline

- Introduction
- Defining HPC/CI
- Training Program Design
- Current SDSC HPC/CI Training Program
- **Synergistic Activities Broaden Impact**
- HPC Students Program
- Broadening Impact
- Conclusions & Future Plans



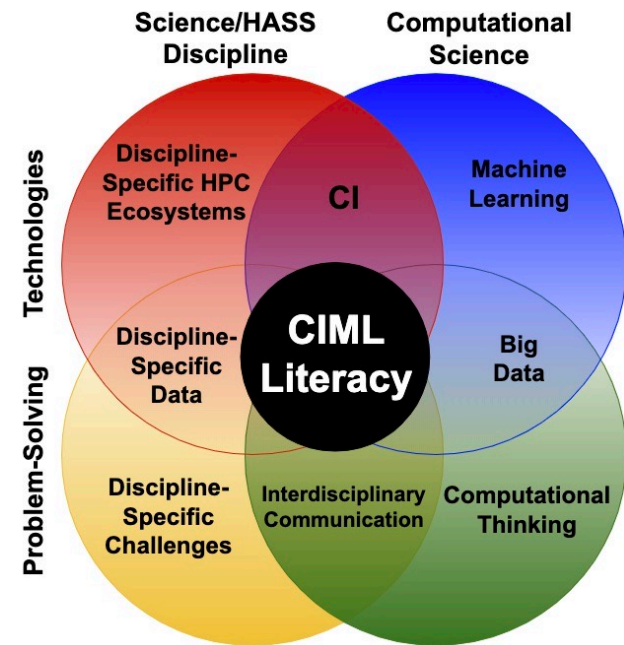
# Synergistic programs broaden impact

- **Broadening impact:**
  - Awarded two CyberTraining Grants; contributing to other grants
  - Monitor metrics, meet Campus and NSF standards
- **Continuing engagement between SDSC and partners:**
  - AMD User Forum; GPU Hack-a-thon; HPC@MSI to support minority serving institutions; HPC Student sponsorships
- **Deploying tools** to make it easier to learn about and use Expanse
- **HPC Students**
  - Campus impact: CSE collaboration; Supercomputing Club part IEEE chapter; SC23 Student Cluster Competition team winning HPL benchmark
- **Partners include:**
  - Dell, AMD, NVIDIA, Supermicro, Liquid, OpenACC, HPE
  - CloudBank, Univ. of Washington, NRP Project, CalIT2
  - Engage Partners in Training:



# Cyberinfrastructure-Enabled Machine Learning Project

- CIML Summer Institute
  - introduces machine learning (ML) researchers, developers and educators to HPC and the techniques needed to run their models at scale
- Annual ML Summer Institutes:
  - 2021, 2022: virtual
  - 2023, in person
- NSF Award #2017767:  
“CyberTraining: Implementation: Small: Developing a Best Practices Training Program in Cyberinfrastructure-Enabled Machine Learning Research”



Thomas, M. P., Götz, A. W., KANDES, M. C., & Sinkovits, R. S. (2023). Developing a Best Practices Training Program in Cyberinfrastructure-Enabled Machine Learning Research (Accepted). *PEARC 2023 Conference Series - Practice and Experience in Advanced Research Computing*.  
<https://doi.org/https://doi.org/10.1145/3569951.3597543>

# Cyberinfrastructure Professional (CIP) Fellows Program (CIP)

## Program will Train (CI) professionals to

- Have interdisciplinary skills that bridge the science, engineering, and computing specialties;
- Support and facilitate research projects, fostering a long-term CIP community
- Identify sustainable CIP career paths
- Program includes: advanced computing, workforce development topics,
- CIPs will be embedded into institutional departments
- NSF, award number 2230127, CyberTraining: Training and Developing a Research Computing and Data (RCD) CI Professionals Community



## CIP Cyberinfrastructure Professional FELLOWS PROGRAM

CyberTraining: Training and Developing a Research Computing and Data (RCD) CI Professionals Community  
*A joint program between UC San Diego, San Diego State and CSU San Bernardino partners to develop a CIP workforce.*

### Program Overview

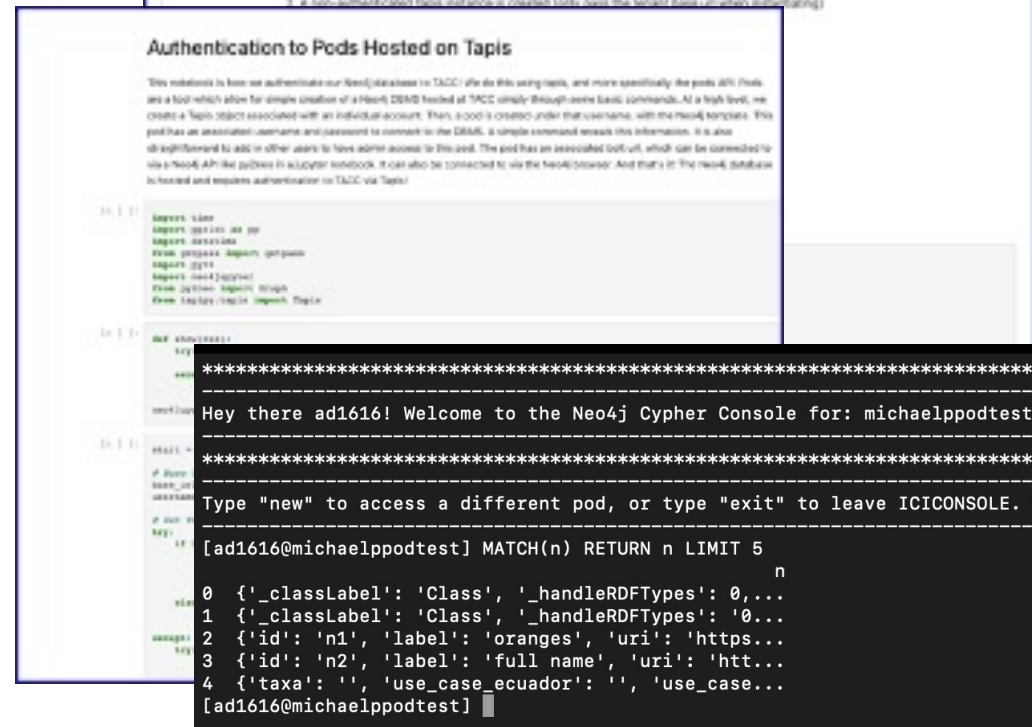
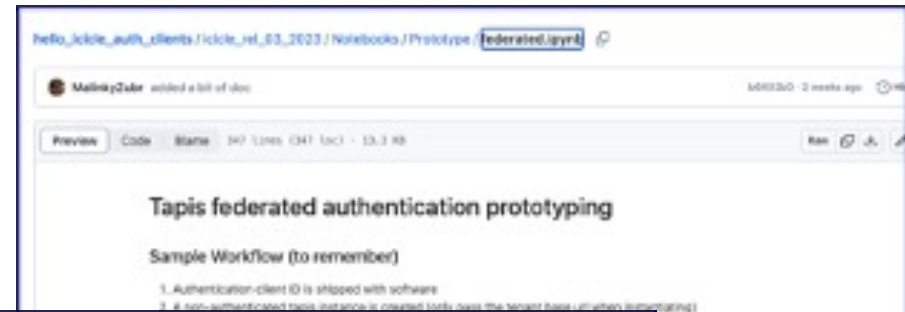
A critical role within scientific research teams is the cyberinfrastructure (CI) professional who has interdisciplinary skills that bridge the science, engineering, and computing specialties. Readily finding such personnel, with the necessary experience needed to navigate CI ecosystems has become a bottleneck for many projects. The Cyberinfrastructure Professional (CIP) Fellows Program is designed to address this need by training and mentoring a team of interdisciplinary Researcher-Facing (CARCC) CI Professionals with individualized training program in advanced computing and workforce development topics, embedding them into institutional departments, teaching them how to support and facilitate research projects, fostering a long-term CIP community, and defining sustainable career paths specifically for these kinds of professionals.

Eligible candidates include those who have domain science and/or computing backgrounds and want to advance their skills in CI, HPC, data science, and have an interest in facilitating scientific research teams who need to use CI resources and services. Potential candidates include: CI research support staff, software engineers, and system administrators, data curators, computational research scientists and engineers, and CI facilitators (people who work directly with researchers to help them to make effective use of Cyberinfrastructure (CI)). The creation of these new CIP-Fellow positions supports the goal of creating long-term career development paths and opening doors to new opportunities for CI Professionals.

The CIP-Fellows program is seeking candidates who have the appropriate science and/or computing background and the motivation to work with research teams. [see below]

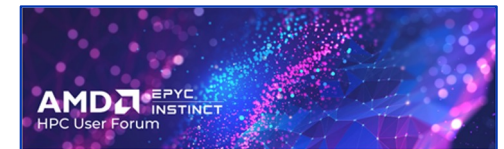
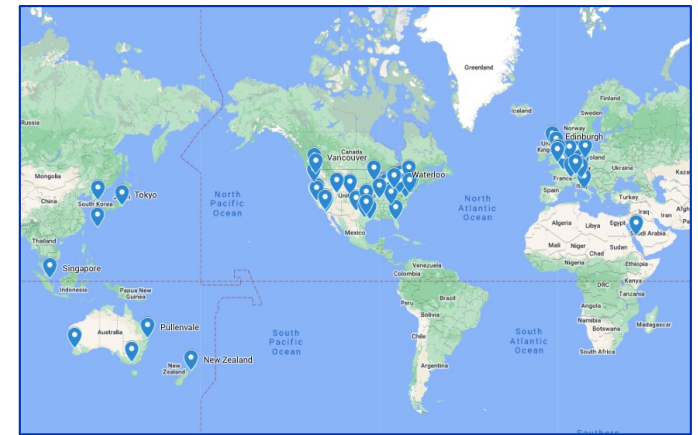
# ICICLE REHS Project: Creation of Remote Knowledge Graphs

- Students worked on the Intelligent Cyberinfrastructure with Computational Learning in the Environment (ICICLE) project
- Created command line and Jupyter Notebooks to create/populate/modify knowledge graphs
- Software contributions to the production SW components catalog



# AMD HPC User Forum: an Active and International Community

- AMD Support/Sponsorship continues
- 3 years old and growing strong:
  - 75+ Institutions; 17 countries; 125 members
  - 9+ events (over 1000 regs/750 attendees).
  - Strong technical meetings: e.g. Fall'22 Program: AMD Instinct (Day 1); Adaptive & Embedded Compute + AI (Day2); EPYC + Q&A (Day3)
- Organization: SDSC continues to lead
  - Executive Team: President: Mary Thomas (SDSC); Vice President: Melyssa Fratkin (TACC); Chief Communications Officer: Susan Rathbun (SDSC)
  - AMD Liason: Guy Ludden (C. Fronczak)
  - Programming & Events: Bob Sinkovits (SDSC), Lev Gorenstein (Purdue)
  - Special Interest Groups (SIG): Mahi Tatineni, All Sills (Texas Tech)
  - Planning for 2023: meetings, syncs, new SIGs





# The HPC student program

**A program for educating and training the *next generation* of HPC professionals and researchers**

- Motivate students to pursue HPC careers
- Serve as a bridge between students and HPC-related research and researchers at SDSC.
- Work with UCSD undergraduates, graduates; expand to other groups as time and budget permits.
- Advances UCSD EDI goals

# HPC Students continues to connect SDSC activities to students at UCSD and elsewhere

- Sponsor UCSD Supercomputing Club: merged with IEEE campus chapter
  - 2023 kickoff meeting: >65 attendees; planning 5+ significant projects
- Annual HPC User Training 2022 (14-weeks) focused on UCSD students; open to research/EDU communities:
  - 90+ Registrants; 50+ Expanse accounts; 19 active.
- Research Experience for High School Students (REHS) brings students to SDSC to work on HPC projects.
  - REHS'22 contributed software to ICICLE Project



# HPC Students: Participation at Technical Meetings

- Student Cluster Competition:
  - Competition team of 6 undergrads.
  - Appx 15 undergrads + a few graduate students are part of “home” team → extends impact
  - Undergrads: independent study credit through CSE department
  - Team composition: 33% from underrepresented communities
  - Appx 15 mentors from SDSC, UCSD, and partners
  - SCC competition history:
    - SCC23: planning
    - SCC22: in person; AMD EPYC+INSTINCT; placed 3rd in US and won the HPL benchmark;
    - SCC21: virtual competition, placed 4th overall;
- Student Volunteers at SC (& PEARC): SDSC sponsors students at these meetings.
- 2023: Supercomputing club hosted its own student cluster competition.



# HPC Student Activities

- SDSC Internships (paid, CCR)
  - HPC Training (CCR)
  - Supercomputing Club
  - Club Projects:
    - Raspberry PI Cluster build:
      - SDSC Supplies hardware, location, network, etc.
    - Supercomputing Club students build-out/admin
- @Supercomputing:**
- Student cluster competition (SCC) teams
    - Selected for: SCC20 & 21
    - SCC22: planning
  - Student volunteers:
    - 1 selected for SC21
    - 4 selected for SC19
  - All expenses paid by SDSC and sponsors!

[https://www.sdsc.edu/education\\_and\\_training/hpc\\_students.html](https://www.sdsc.edu/education_and_training/hpc_students.html)



# K-12 Education Programs

- ABLE
- ForMIDABLE speakers
- MAP Symposium changes
- REHS
- Summer Workshops.
- K-8 summer camps?
- PI-Wars

# Outline

- Introduction
- Defining HPC/CI
- Training Program Design
- Current SDSC HPC/CI Training Program
- Synergistic Activities Broaden Impact
- HPC Students Program
- Broadening Impact
- **Conclusions & Future Plans**



# Conclusions & Future Plans

- Training program changes have resulted in an active program with room to grow and adapt to emerging technologies
- 2023 HPC Training:
  - 12 Webinars scheduled for 2023 + Training events: SDSC DSI'23, CIML'23, GPU Bootcamp (July'23), others
  - Add/develop tutorials, bootcamps, workshops as requested
  - Refactoring to reduce staff time, train CIP-Fellows program and SCC23 effort
- Training material catalogue webpage:
  - [https://www.sdsc.edu/education\\_and\\_training/training\\_catalog.html](https://www.sdsc.edu/education_and_training/training_catalog.html)
  - Prototype searchable custom pages using metadata/taxonomy
  - HPC Training gateway
- HPC Students:
  - Club's 2023 kickoff meeting scheduled for January 19, 2023, 5-7pm
  - SCC23 preparation and planning

# SDSC HPC Training Team

## Faculty & Staff:

- Mary Thomas [mpthomas@ucsd.edu](mailto:mpthomas@ucsd.edu)
- Bryan Chin [b5chin@ucsd.edu](mailto:b5chin@ucsd.edu)
- Marty Kandes, [mkandes@sdsc.edu](mailto:mkandes@sdsc.edu)
- Christopher Irving, [cirving@ucsd.edu](mailto:cirving@ucsd.edu)
- Andy Goetz, [agoetz@sdsc.edu](mailto:agoetz@sdsc.edu)
- Tom Hutton, [hutton@sdsc.edu](mailto:hutton@sdsc.edu)
- Mai Nguyen, [mhnguyen@ucsd.edu](mailto:mhnguyen@ucsd.edu)
- Paul Rodriguez, [prodriguez@sdsc.edu](mailto:prodriguez@sdsc.edu)
- Peter Rose, [pwrose@ucsd.edu](mailto:pwrose@ucsd.edu)
- Darshan Sarojini, [dsarojini@eng.ucsd.edu](mailto:dsarojini@eng.ucsd.edu)
- Sakai, Scott, [ssakai@ucsd.edu](mailto:ssakai@ucsd.edu)
- Bob Sinkovits, [sinkovit@sdsc.edu](mailto:sinkovit@sdsc.edu)
- Subha Sivagnanam, [sivagnan@sdsc.edu](mailto:sivagnan@sdsc.edu)
- Mahi Tatineni, [mahidhar@sdsc.edu](mailto:mahidhar@sdsc.edu)
- Nicole Wolter, [nickel@sdsc.edu](mailto:nickel@sdsc.edu)
- Chen, Yuwu, [ychen64@sdsc.edu](mailto:ychen64@sdsc.edu)

## ER Team

- Csori, Csilla, [ccsori@sdsc.edu](mailto:ccsori@sdsc.edu)
- Dwyer, Michael, [mdwyer@sdsc.edu](mailto:mdwyer@sdsc.edu)
- Susan Rathbun, [susan@sdsc.edu](mailto:susan@sdsc.edu)
- Ben Tolo, [btolo@sdsc.edu](mailto:btolo@sdsc.edu)
- Cindy Wong, [cwong@sdsc.edu](mailto:cwong@sdsc.edu)

## Sponsors/Vendors

- AMD
- DDN
- CloudBank
- LIQID
- NVIDIA
- Supermicro
- Intel
- Oracle



# Thank You

# Key Links

- SDSC HPC Training Program:
  - [https://www.sdsc.edu/education\\_and\\_training/training\\_hpc.html](https://www.sdsc.edu/education_and_training/training_hpc.html)
- SDSC HPC Students Program:
  - [https://www.sdsc.edu/education\\_and\\_training/hpc\\_students.html](https://www.sdsc.edu/education_and_training/hpc_students.html)
- SDSC's HPC/CI Training Series
  - [https://www.sdsc.edu/event\\_items/202201\\_HPC-CI-Training-Series.html](https://www.sdsc.edu/event_items/202201_HPC-CI-Training-Series.html)
- SDSC HPC Students Site:
  - <https://hpc-students.sdsc.edu>
  - SCC activities: <https://hpc-students.sdsc.edu/scc>