

Where to Compute? Campus & National Resources

Research Computing & Data (RCD)
Community Meeting
June 13, 2023



Meeting Basics



Google Doc notes page in chat



Meeting will be recorded



Use the Zoom chat for questions, ideas, and suggestions



Or Raise Hand
(Reactions > Raise Hand)



Live Transcript (auto-generated closed captions) is available



Meeting content (recording, notes, discussion, slides, chat log) will posted on meeting page





Agenda

- Meeting Basics
- Welcome to the UCSD RCD Community!
- Where to Compute?
 - What to consider before picking a resource
 - Categorizing resources to align with needs
 - Campus and national resources
- Upcoming Meetings & Call to Action



WELCOME TO THE UCSD RCD COMMUNITY!

It's about us!



Research Computing and Data (RCD), also known as Cyberinfrastructure (CI), refers to foundational research support areas. This community includes researchers and their collaboration partners, those RCD professionals who include data scientists, systems engineers, software developers, research librarians, data wranglers, research facilitators, and more.

Meetings



Our bi-monthly meetings are a chance to come together to discuss issues important to the UCSD research community, learn about resources, software, tools, and services relevant to our work.

<https://ucsd-rcd.github.io/meetings/>

Mailing List



Our mailing list is a way to stay updated on the UCSD RCD Community, ask questions, and contribute your knowledge. We'd also love to hear your ideas for this community.

Google Groups [ucsd-rcd](https://groups.google.com/a/ucsd.edu/g/ucsd-rcd-l)
<https://groups.google.com/a/ucsd.edu/g/ucsd-rcd-l>

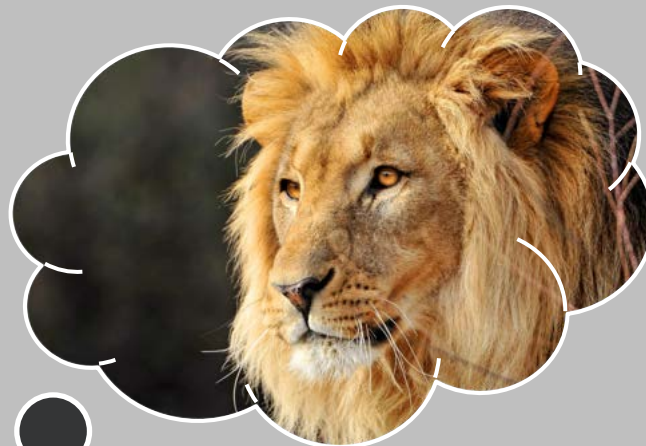
Website & GitHub Repos



Our website and GitHub repos are places to share resources, experiences, and examples in a way that's widely accessible.



<https://ucsd-rcd.github.io/>
<https://github.com/ucsd-rcd>



**Where to compute
when your laptop is
no longer enough?**

Things to Consider



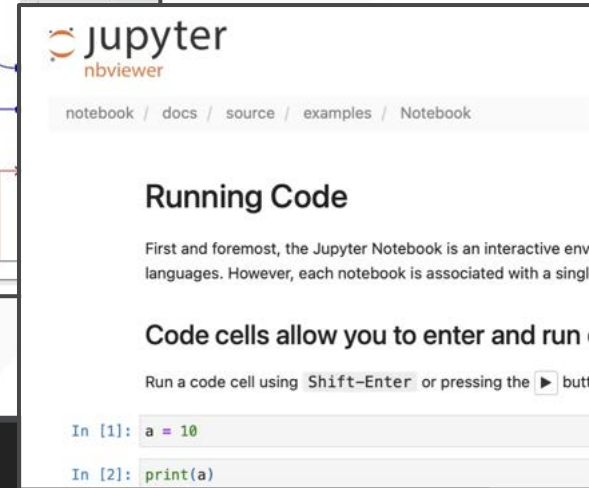
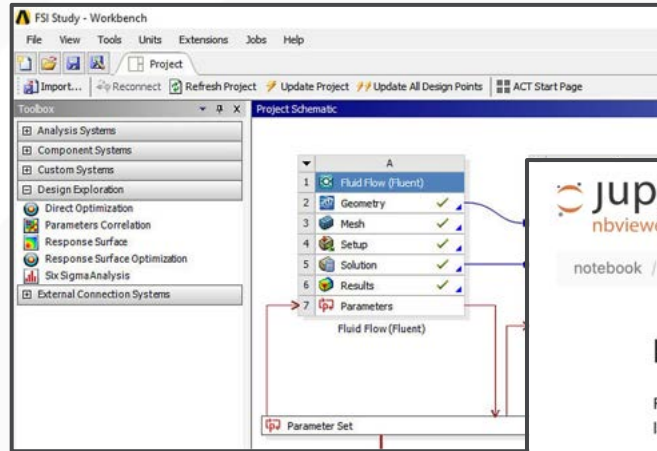
Consideration 1: Current Resource and Experience

- What's the project/research?
- Individual or group driven?
- Student or faculty?
- Level of programming and/or system administration experience
- What is your current compute resource?
- How long to run job on current resource?
- Who needs access to resource?



Consideration 2: The Environment

- Interaction may determine resource
 - I.e., GUI or command line
 - Batch or interactive
- Parallel or sequential (series)?
- Software needs - licensed, custom, open source?
- Hardware requirements: CPU, RAM, GPUs
- Bursty or continuous workloads

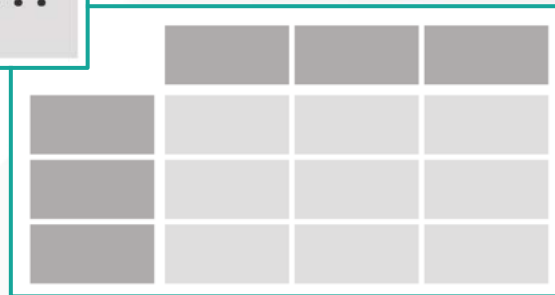
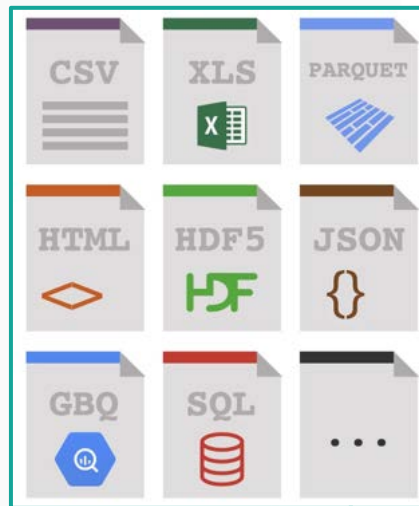


```
#!/bin/bash

echo '#ifndef _SFGEN_BUILDINFO_H
#define _SFGEN_BUILDINFO_H
' > buildinfo.h
a=`git show --oneline --no-patch`
echo "#define SFGEN_LATEST_COMMIT \"$a\\n\"" >> buildinfo.h
git status -s *.c *.cpp *.h Makefile Make.mach.* > t
...
```


Consideration 3: The Data and Storage

- Volume & structure of data
 - Total size
 - Number of files
 - Structured data?
- Source
- Location
- Duration
- Access
- Collaboration
- Restrictions (NDAs, compliance, DUAs, etc.) may determine resources
- May require modifying current data pipelines



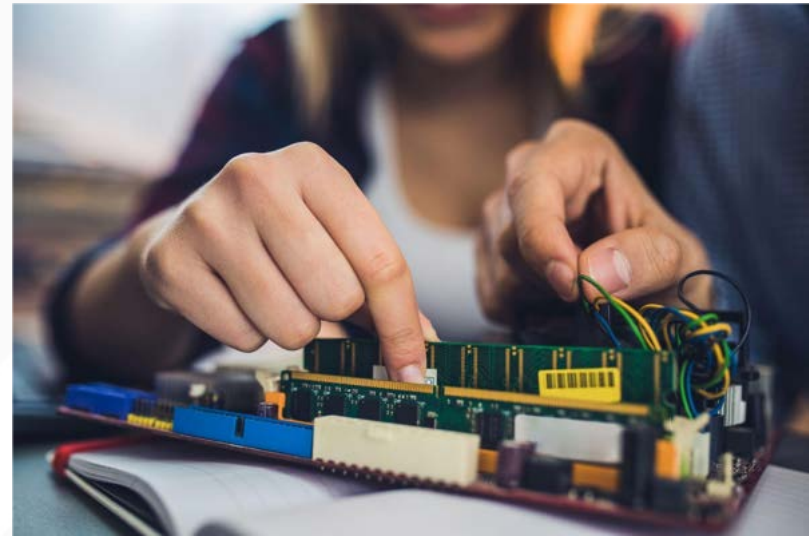
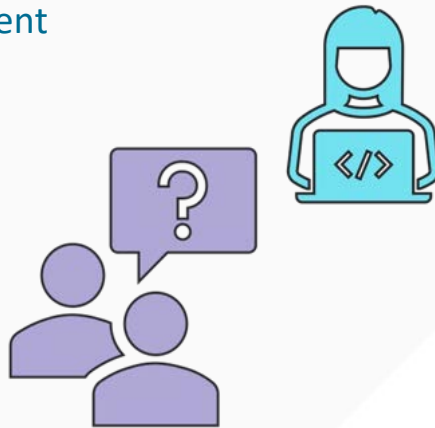
Consideration 4: The Project

- Timeframe
- Collaborations
- Funding
- Science domain
 - Some domains have specialized resources
 - Datasets may already be available
 - E.g., bioinformatics



Consideration 5: The Support

- What current and potential support is provided to the researcher or research group?
- What can you or your research group do effectively and efficiently?
- Different resources and solutions will require more or less effort in various areas
 - Systems administration
 - Programming optimization
 - Software usage/development
 - Cost management
 - Architecture
 - Sustainability
 - Training



Consideration 6: Cost

- Cost: time or money
- The more you need, the more time or money
- Resources are either allocated or purchased
- Allocated: Resources are granted based on proposals
- Purchased: Hardware, cloud services, TSCC compute nodes (condo)
- Consider ongoing costs, especially for purchased resources
- Cost tradeoff between control and capacity
 - Fully “owned” system cost more to operate
 - Hardware may require colocation (renting data center space)
 - Hardware or cloud computing requires system administration



Consideration 5: Cost Examples

- Very rough comparison
- Trying to provide an idea of the time or money needed for different resources

AWS EC2 On-Demand (us-west-2)

Type	vCPU	RAM	GPU	\$/month
t2.nano	~1	512MB	0	\$4.24
c6i.2xlarge	8	16GB	0	\$250
p4d.24xlarge	96	1.1TB	8@A100	\$24k

TSCC Nodes

- 36 cores, 256 GB RAM: \$11k
- 4 NVIDIA A100, 64 cores, 256GB RAM: \$45k
- Setup \$2k, operation \$500 - \$1.1k/year

ACCESS (NSF HPC) Allocations (~Annual)

Allocation	Proposal	CPU hours	GPU hours
Explore	Abstract	400k	7.6k
Discover	1 page	1.5M	28.5k
Accelerate	3 pages	3M	57k
Maximize	10 pages	Unlimited	Unlimited

https://www.sdsc.edu/services/hpc/tsc/condo_details.html

<https://allocations.access-ci.org/prepare-requests-overview#comparison-table>

<https://aws.amazon.com/ec2/pricing/on-demand/>

Resources

Things you can use to crunch numbers, explore data, and develop a dysfunctional relationship with programming

Campus

National

Campus Resources

Buying Hardware & SDSC Data Center (Colocation)

A dedicated workstation or server(s)

- Consider operating needs
 - Location
 - Networking
 - Power
 - System administration
- SDSC Data Center provides colocation
- Some departments also have spaces



The screenshot shows the SDSC website header with the logo and tagline "COMPUTING WITHOUT BOUNDARIES". The navigation menu includes "ABOUT SDSC", "SERVICES", "SUPPORT", "RESEARCH & DEVELOPMENT", "EDUCATION & TRAINING", and "NEWS & EVENTS". The breadcrumb trail is "HOME > SERVICES > CYBERINFRASTRUCTURE SERVICES > COLOCATION". The main heading is "Colocation". Below it is an image of a server rack aisle. The text describes the colocation services, mentioning 13 Megawatts of power, multi-10-gigabit network connectivity, and 24/7 operations staff. It also lists standard services provided, such as ISO-Base seismic protection, aisle containment, and 2x30A power distribution.

SDSC SAN DIEGO SUPERCOMPUTER CENTER
at the UNIVERSITY OF CALIFORNIA SAN DIEGO
COMPUTING WITHOUT BOUNDARIES

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

HOME > SERVICES > CYBERINFRASTRUCTURE SERVICES > COLOCATION

Colocation



For researchers who choose to purchase their own IT equipment, SDSC offers rack colocation services. Our 19,000-square foot climate-controlled and secure datacenter is fully equipped with 13 Megawatts of power, multi- 10-gigabit network connectivity, and a 24/7 operations staff. The service was designed in partnership with representatives from several UC campuses and the UC Office of the President, and is available to UC researchers through their campus representatives.

Standard colocation includes the following services:

- > Standard rack provided with ISO-Base seismic protection, aisle containment, and 2x30A power distribution.
- > 10+ Gb/s networking fabric connectivity both throughout SDSC aggregation fabric and into CENIC.
- > 24/7 operations staff providing facility oversight and emergency "remote hands" hardware assistance.

Additional services available on an as-needed basis.

More Information

services@sdsc.edu

<https://www.sdsc.edu/services/ci/colocation.html>

UCSD Research Cluster

- Research IT managed Kubernetes cluster
- Command line access via login node
- Launch scripts for various containers
- Jupyter Notebooks via datahub.ucsd.edu
- Good next step after laptop

UCSD Research Cluster: User Guide



Created by amoxley
Last updated: Apr 28, 2023 by Kimberly Thomas · 11 min read

UCSD Research Cluster: User Guide



Overview

UC San Diego's Research GPU cluster, a service of Research IT Services, provides researchers in all disciplines and divisions access to hardware nodes located at SDSC. Funding for the cluster was provided by Research IT Services.

To report problems, or to request assistance, please email Research Computing & Data Support: rcd-support@ucsd.edu.



Jobs on the cluster are executed in the form of Docker "containers" which are essentially lightweight virtual machines, e GPU hardware, and each well isolated from other users' processes.



The [Kubernetes container management/orchestration system](#) routes users' containers onto compute nodes, monitors p limits/quotas as appropriate.

Please be considerate and terminate idle containers: while containers share system RAM and CPU resources under the GPU cards are assigned to users on an exclusive basis. When attached to a container they become unusable by others.

More Information

rcd-support@ucsd.edu

<https://ucsdcollab.atlassian.net/wiki/spaces/RESUP/pages/87949427/UCSD+Research+Cluster>

Triton Shared Computing Cluster (TSCC)

Campus HPC resource with 2 cost models

- Condo (a system purchase model)
- Hotel (a pay-as-you-go model)

Condo participant run “at risk” preemptible jobs to glean cycles.

AKA “the grad student queue”.



CPU Nodes

CPU model	CPU quantity	Memory ¹	Estimated Cost
Intel Xeon Gold 6354 (Ice Lake) 18-core 3GHz	2	256 GB (7GB/core)	\$11,000
Intel Xeon Platinum 8358 (Ice Lake) 32-core 2.6GHz	2	1 TB (15GB/core)	\$20,000

GPU Nodes²

GPU model	GPU memory	GPU quantity	CPU model	CPU quantity	Memory ¹	Estimated Cost
NVIDIA Ampere A100	40GB	4 (with NVlink)	Intel Xeon Platinum 8358 32-core 2.6GHz	2	256 GB	\$45,000

More Information

tsc-support@ucsd.edu

<https://www.sdsc.edu/services/hpc/tsc/index.html>

Commercial Cloud

UC currently has negotiated agreements with these commercial cloud providers:

- Amazon Web Services (AWS)
- Google Cloud Platform (GCP)
- Microsoft Azure

Indirect Costs Exemption: Cloud computing services provided by the above vendors (plus SDSC) are exempt from the application of indirect costs (IDC) when used for extramurally funded research.

More Information

UCSD Blink Page

<https://blink.ucsd.edu/technology/cloud/index.html>

August 2022 Meeting

<https://ucsd-rcd.github.io/meetings/events/2022-08-09-Commercial-Cloud-Research-Activity.html>

Cloud Services

Last Updated: July 1, 2022 1:59:14 PM PDT

Give [feedback](#)

Learn about cloud services that have agreements in place with the University of California campuses and how to get started with a Cloud Services Account.

Cloud Services

Cloud Services provides UC San Diego faculty, staff, and other affiliated researchers access to computing resources from approved vendors that have agreements in place with the University of California or UC San Diego.



National Resources

Science Gateways

Science gateways are web platforms that simplify the participation in science by easing the connection to otherwise difficult to use resources (like HPC or instruments), sharing data and computational codes, collaborating, and generating large international audiences for broader impact.

UCSD RCD Community Meetings



Calendar and materials for the UCSD
Research Computing & Data Community
Meetings

Building & Using Science Gateways

October 11, 2022, 2:00 p.m. PDT

Description

Science gateways are web platforms that simplify the participation in science by easing the connection to otherwise difficult to use resources (like HPC or instruments), sharing data and computational codes, collaborating, and generating large international audiences for broader impact. UCSD has been a leader in promoting the use and creation of science gateways to make research capabilities available outside of individual labs or campuses.

This meeting will provide a background on science gateways, including how to find existing gateways and what tools and platforms are out there if you want to host a gateway. Contributors to two successful UCSD science gateways, the CIPRES (phylogenetics) and the Qiita (omics), will describe their portals and can answer questions about the realities of developing and operating a gateway.

More Information

SGX3 NSF Center of Excellence

<https://sciencegateways.org>

October 2022 Meeting

<https://ucsd-rcd.github.io/meetings/events/2022-10-11-ScienceGateways.html>

ACCESS (NSF HPC)

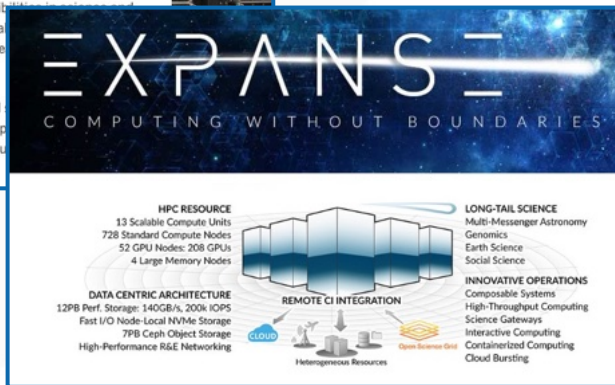
- Multiple NSF funded HPC, data, support, and computing resources.
- Allocated via proposals at different levels.
- Examples:
 - SDSC Expanse
 - TACC Frontera

FRONTERA

The Fastest Academic Supercomputer in the World.

In 2018, the National Science Foundation (NSF) awarded a \$60 million grant to the Texas Advanced Computing Center (TACC) to deploy a new petascale computing system, Frontera. Frontera opens up new possibilities in scientific engineering by providing computational capability that makes it possible for investigators to tackle much larger and more complex research problems across a wide spectrum of domains.

Deployed in June 2019, Frontera is the 19th most powerful supercomputer in the world, and the fastest supercomputer on a university campus. User access began in June 2019, and the system entered full production in September 2019.



ACCESS | Advancing Innovation

Welcome to ACCESS.

Advanced Cyberinfrastructure Coordination
Ecosystem: Services & Support

ABOUT

More Information

<https://access-ci.org>

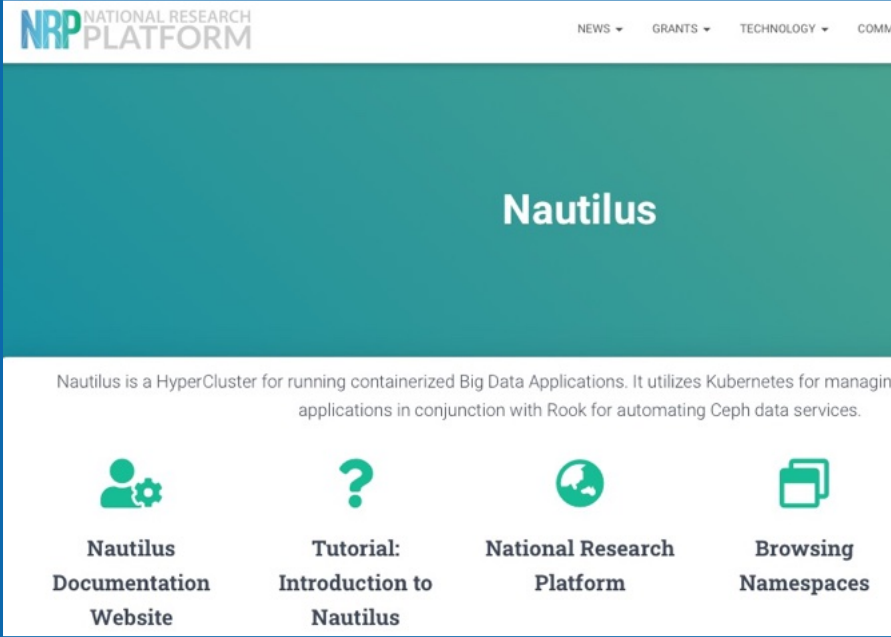
<https://www.sdsc.edu/services/hpc/expanse/>

<https://tacc.utexas.edu/systems/frontera/>

National Research Platform (NRP) Nautilus

Nautilus is a HyperCluster for running containerized Big Data Applications. It utilizes Kubernetes for managing and scaling containerized applications in conjunction with Rook for automating Ceph data services.

- Researchers contribute nodes to the cluster
- Helps small campuses become part of a larger system
- Groups contribute application containers







NRP NATIONAL RESEARCH PLATFORM

NEWS ▾ GRANTS ▾ TECHNOLOGY ▾ COMMUNITY ▾

Nautilus

Nautilus is a HyperCluster for running containerized Big Data Applications. It utilizes Kubernetes for managing applications in conjunction with Rook for automating Ceph data services.

-  **Nautilus Documentation Website**
-  **Tutorial: Introduction to Nautilus**
-  **National Research Platform**
-  **Browsing Namespaces**

More Information

<https://nationalresearchplatform.org/nautilus/>

Open Science Grid OSPool

- Leverages excess capacity of the Open Science Grid
- Great for large numbers of small jobs
 - Single core, a few GB of RAM
 - A few hours max
 - < 1 GB of data in and out
- But! Larger jobs are possible
- Lots of expertise on this campus



More Information

<https://portal.osg-htc.org/documentation/>

Domain-Specific Example, Terra.bio

The screenshot shows the Terra.bio website homepage. At the top, there is a navigation bar with the Terra logo, menu items (ABOUT, SCIENCE, RESOURCES, NIH DMSP, COMMUNITY, BLOG), a LOG IN button, and a search bar. Below the navigation bar, the main content area features a 'Focus on your science' section with a sub-headline 'Access data, run analysis tools, and collaborate in Terra: a scalable platform for biomedical research.' and a 'GET STARTED' button. To the right of this text is a large graphic with the Terra logo and the text 'Focus on your science' next to a play button icon. Below this is a green banner with three statistics: '42,000 TERRA USERS', '2.38 MILLION STUDY PARTICIPANTS', and '90 MILLION SINGLE CELLS'. At the bottom, there is a section titled 'Terra supports researchers in many biomedical disciplines' with four sub-sections: 'Cancer Genomics', 'Single-Cell Transcriptomics', 'Medical and Population Genetics', and 'Infectious Diseases', each accompanied by a representative image of cells or viruses.

Terra ABOUT ▾ SCIENCE ▾ RESOURCES ▾ NIH DMSP COMMUNITY BLOG LOG IN Search Terra.bio

Focus on your science

Access data, run analysis tools, and collaborate in Terra: a scalable platform for biomedical research.

GET STARTED

42,000 TERRA USERS

2.38 MILLION STUDY PARTICIPANTS

90 MILLION SINGLE CELLS

Terra supports researchers in many biomedical disciplines

- Cancer Genomics
- Single-Cell Transcriptomics
- Medical and Population Genetics
- Infectious Diseases

Shared platform for bioinformatics.

- Hosts popular datasets
- Researchers can use their campus cloud accounts

Talk to your collaborators and find out if similar resources exist for your field.

More Information

<https://terra.bio/>

Upcoming Topics

August 8

Ideas? How to build the RCD+ Community at UCSD?

COMMUNITY DISCUSSION

Questions, ideas, suggestions!

UCSD RCD Community?

Meeting Topics?

Needs?



Mailing List Reminder

Google Groups [ucsd-rcd](https://groups.google.com/a/ucsd.edu/g/ucsd-rcd-l)

<https://groups.google.com/a/ucsd.edu/g/ucsd-rcd-l>