# The CI Milieu: Systems, Service Providers, Technologies

**Douglas M. Jennewein, Arizona State University,**

**Campus Champions Leadership Team**

**Virtual Residency**

**June 7, 2021**

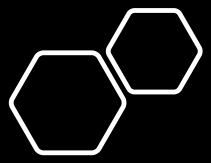
CAMPUS CHAMPIONS

*With support from:*

NSF

XSEDE

Extreme Science and Engineering Discovery Environment



# This talk will (briefly) cover

- NSF/XSEDE Systems: Stampede2, Bridges2, Comet, Expanse, Jetstream(2), Frontera
- Open Science Grid (OSG)
- Science Gateways
- NERSC: Cori and Perlmutter
- ESnet and EPOC
- Internet2
- International Orgs: GEANT, Compute Canada, PRACE
- Tools: Globus, XDMod, Spack, Open OnDemand, ask.ci

# But first…

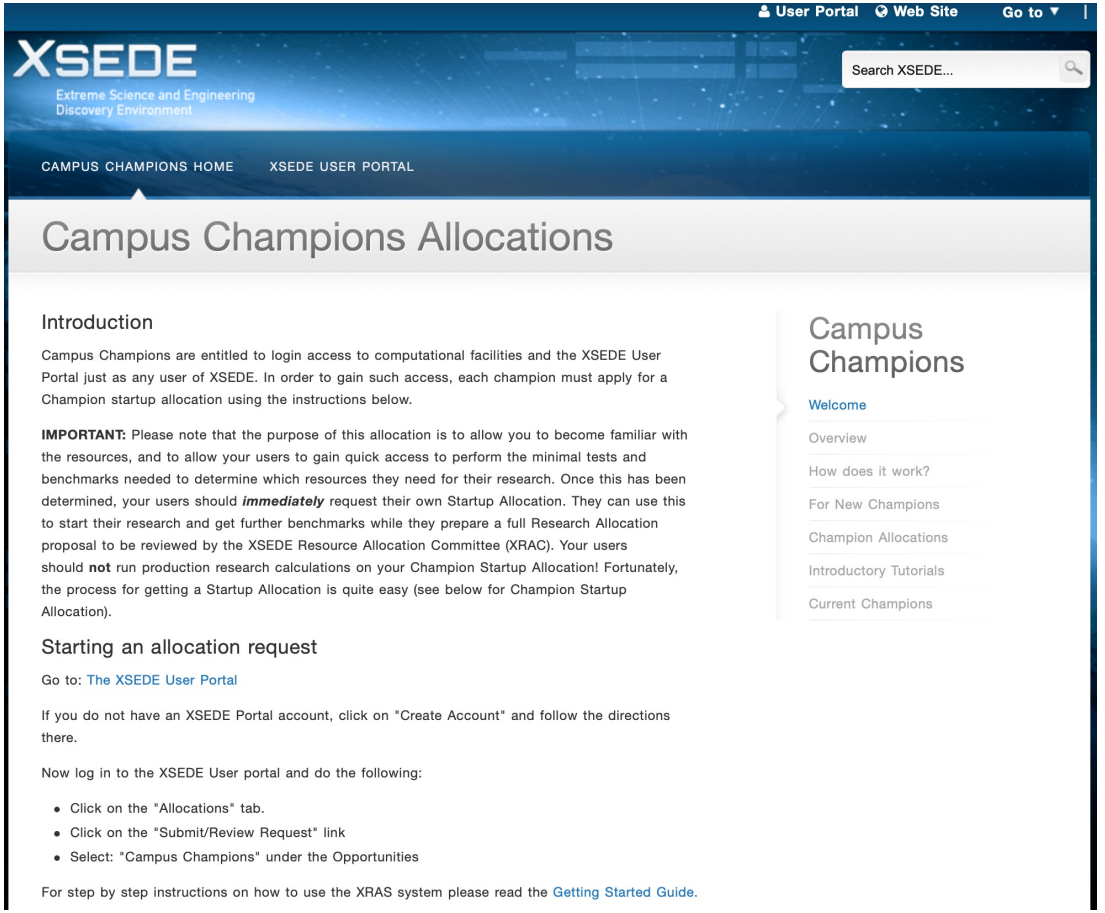
This talk will NOT cover:

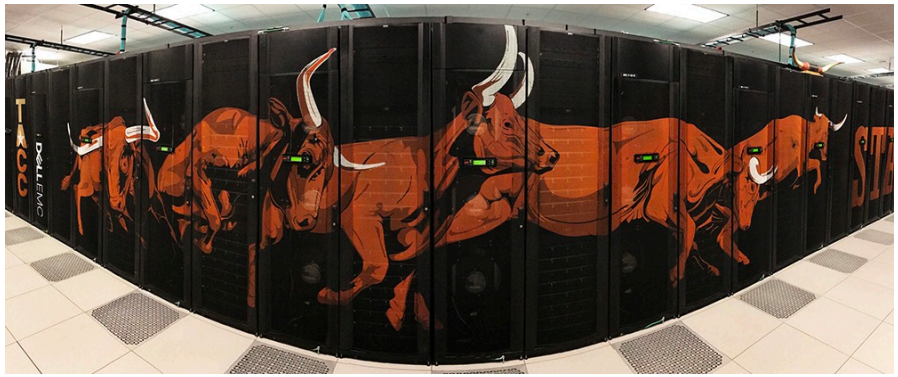
- NSF Comet, Voyager, Ookami, Neocortex, Darwin, Delta, KYRIC, Anvil, CloudLab, Chameleon, and CloudBank
- NIH resources such as BioWulf
- Commercial Cloud
- Regional Networks and The Quilt

Only because of:

1. Time
2. Availability

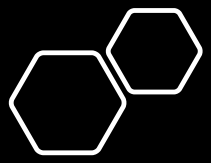
# About XSEDE allocations





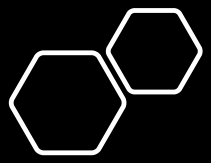

# Stampede2

The Flagship Supercomputer of XSEDE



# Stampede2: Everything's Bigger in Texas

- Brought online in 2017
- Largest XSEDE system (and largest university-based system at the time)
- 18 petaflops of peak performance
- Successfully launched 350,000MPI tasks in a single job

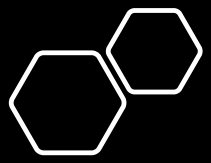


# Stampede2: Everything's Bigger in Texas

4,200 Intel Knights Landing nodes, each with 68 cores, 96GB of DDR RAM, and 16GB of high speed MCDRAM

1,736 Intel Xeon Skylake nodes, each with 48 cores and 192GB of RAM

100 Gb/sec Intel Omni-Path network with a fat tree topology employing six core switches



# Stampede2: Everything's Bigger in Texas

Stampede2 is intended primarily for **parallel applications scalable to tens of thousands of cores**, as well as general purpose and throughput computing.

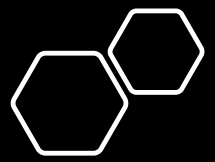
Case in point: Successfully launched 350,000MPI tasks in a single job

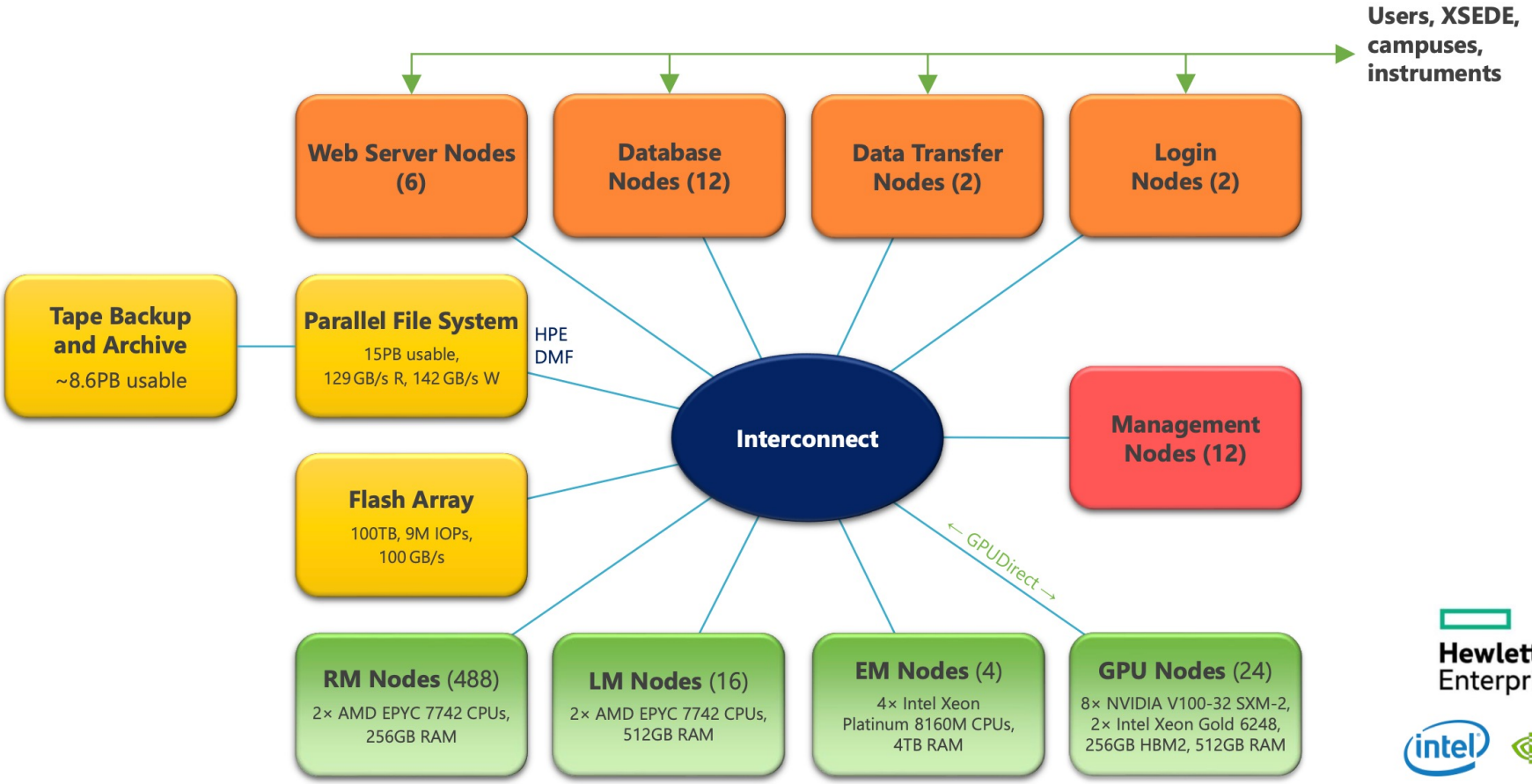

PITTSBURGH SUPERCOMPUTING CENTER

Carnegie Mellon University

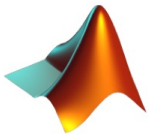
University of Pittsburgh

# Empowering New Research Communities, Bringing Together HPC, AI, and Big Data




Users, XSEDE, campuses, instruments
Web Server Nodes (6)
Database Nodes (12)
Data Transfer Nodes (2)
Login Nodes (2)
Tape Backup and Archive
~8.6PB usable
Parallel File System
15PB usable,
129 GB/s R, 142 GB/s W
HPE DMF
Interconnect
Management Nodes (12)
Flash Array
100TB, 9M IOPs,
100 GB/s
← GPUDirect →
RM Nodes (488)
2× AMD EPYC 7742 CPUs, 256GB RAM
LM Nodes (16)
2× AMD EPYC 7742 CPUs, 512GB RAM
EM Nodes (4)
4× Intel Xeon Platinum 8160M CPUs, 4TB RAM
GPU Nodes (24)
8× NVIDIA V100-32 SXM-2, 2× Intel Xeon Gold 6248, 256GB HBM2, 512GB RAM
Carnegie Mellon University
PSC
PITTSBURGH SUPERCOMPUTING CENTER
University of Pittsburgh
Hewlett Packard Enterprise
intel
NVIDIA
AMD

- Interactivity
- Popular languages and frameworks:
  Python, Anaconda, R, MATLAB, Java, Spark, Hadoop
- AI frameworks: TensorFlow, Caffe2, PyTorch, etc.
- Containers and virtual machines (VMs)
- Databases
- Gateways and distributed (web) services
- Large collection of applications and libraries

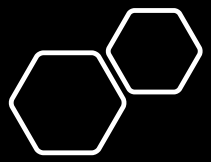








Carnegie Mellon University
PSC
PITTSBURGH SUPERCOMPUTING CENTER
University of Pittsburgh

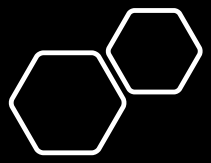


# Memory Intensive Applications

- de novo and metagenome sequence assembly
- graph analytics
- large in-memory databases
- machine learning applications
- large-memory applications written in threaded languages
- bioinformatics
- causal analysis
- machine learning
- graph analytics

# Interactivity

- Interactivity is the feature **most frequently requested** by nontraditional HPC communities.
- Interactivity provides **immediate feedback** for doing exploratory
data analytics and testing hypotheses.
- Bridges2 offers interactivity through a combination of **shared, dedicated, and persistent resources** to maximize availability while accommodating diverse needs.
- Provides languages and environments that **users already know**: High-Productivity Programming



# Shared, dedicated, and persistent

Dedicated database nodes power persistent relational and NoSQL databases

Dedicated web server nodes

OpenStack, KVM, Singularity

Persistent community data collections: e.g. BLAST, ImageNet

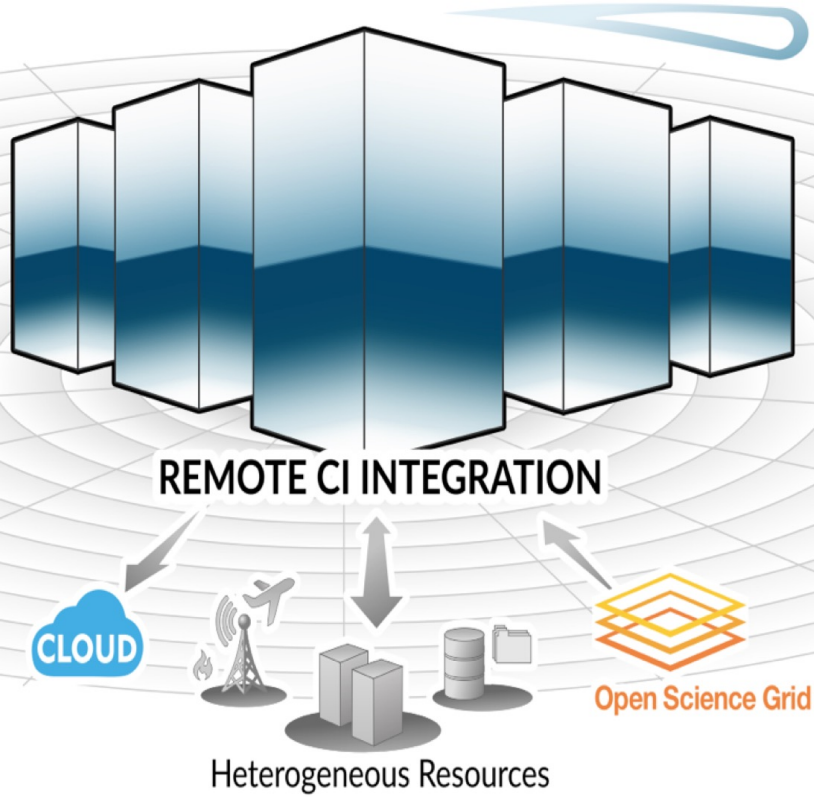

# EXPANSE

# EXPANSE

## COMPUTING WITHOUT BOUNDARIES
5 PETAFLOP/S HPC and DATA RESOURCE

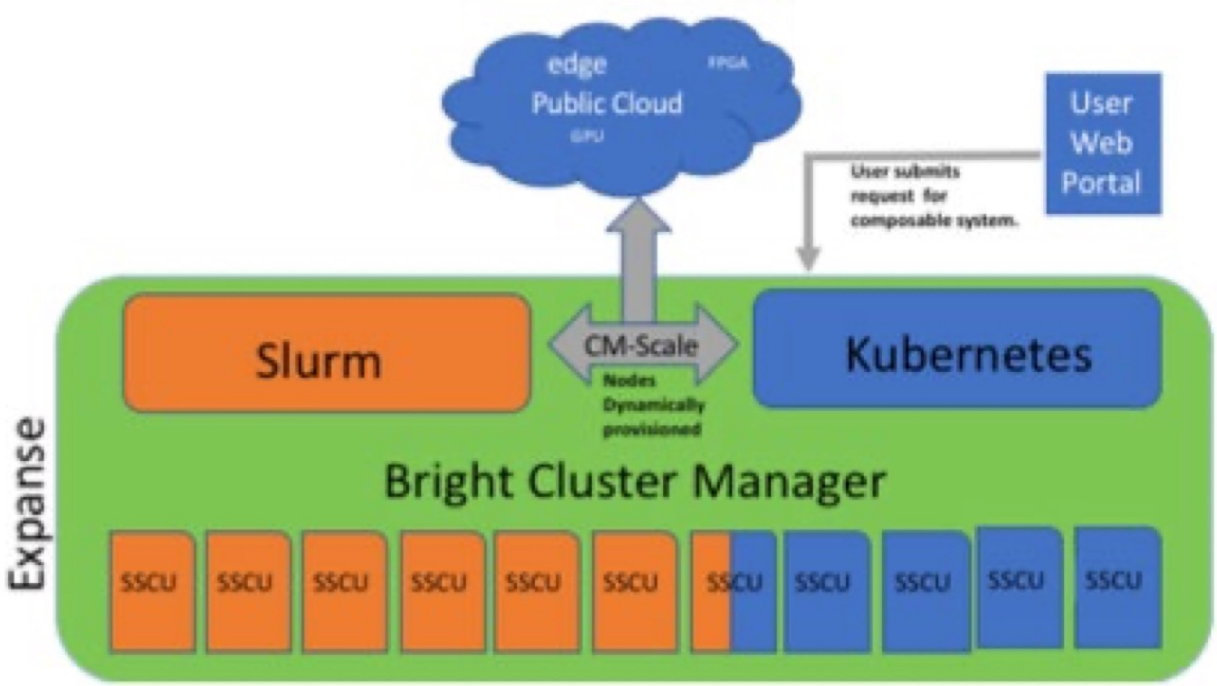
**HPC RESOURCE**
13 Scalable Compute Units
728 Standard Compute Nodes
52 GPU Nodes: 208 GPUs
4 Large Memory Nodes

**DATA CENTRIC ARCHITECTURE**
12PB Perf. Storage: 140GB/s, 200k IOPS
Fast I/O Node-Local NVMe Storage
7PB Ceph Object Storage
High-Performance R&E Networking




**LONG-TAIL SCIENCE**
Multi-Messenger Astronomy
Genomics
Earth Science
Social Science

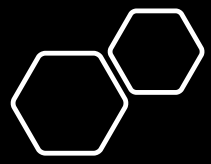
**INNOVATIVE OPERATIONS**
Composable Systems
High-Throughput Computing
Science Gateways
Interactive Computing
Containerized Computing
Cloud Bursting

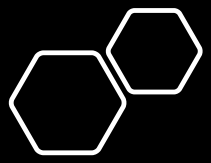

edge
FPGA
Public Cloud
GPU
User
Web
Portal
User submits
request for
composable system.
Slurm
CM-Scale
Nodes
Dynamically
provisioned
Kubernetes
Expanse
Bright Cluster Manager
SSCU
SSCU
SSCU
SSCU
SSCU
SSCU
SSCU
SSCU
SSCU
SSCU
SSCU

COMPOSABLE



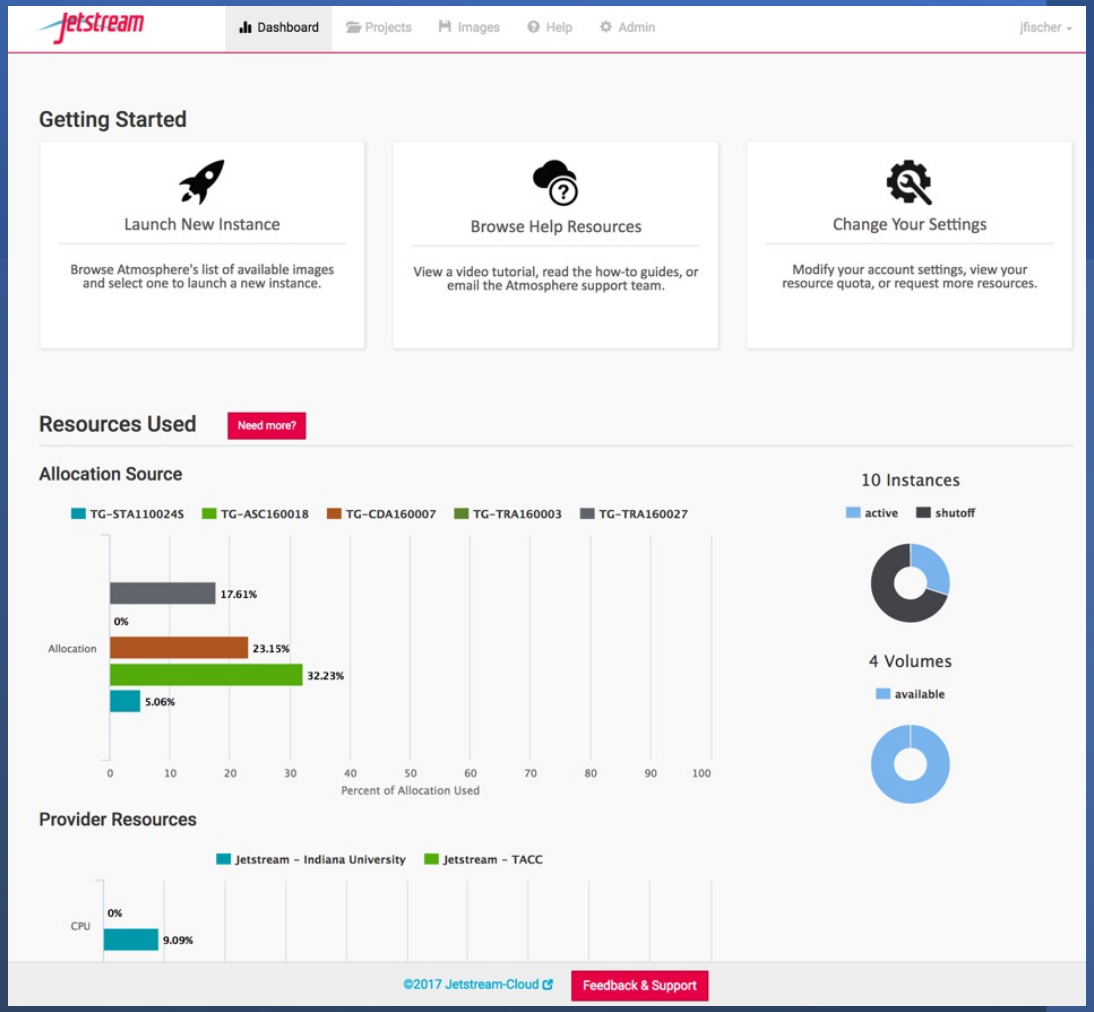
# A NATIONAL SCIENCE AND ENGINEERING CLOUD

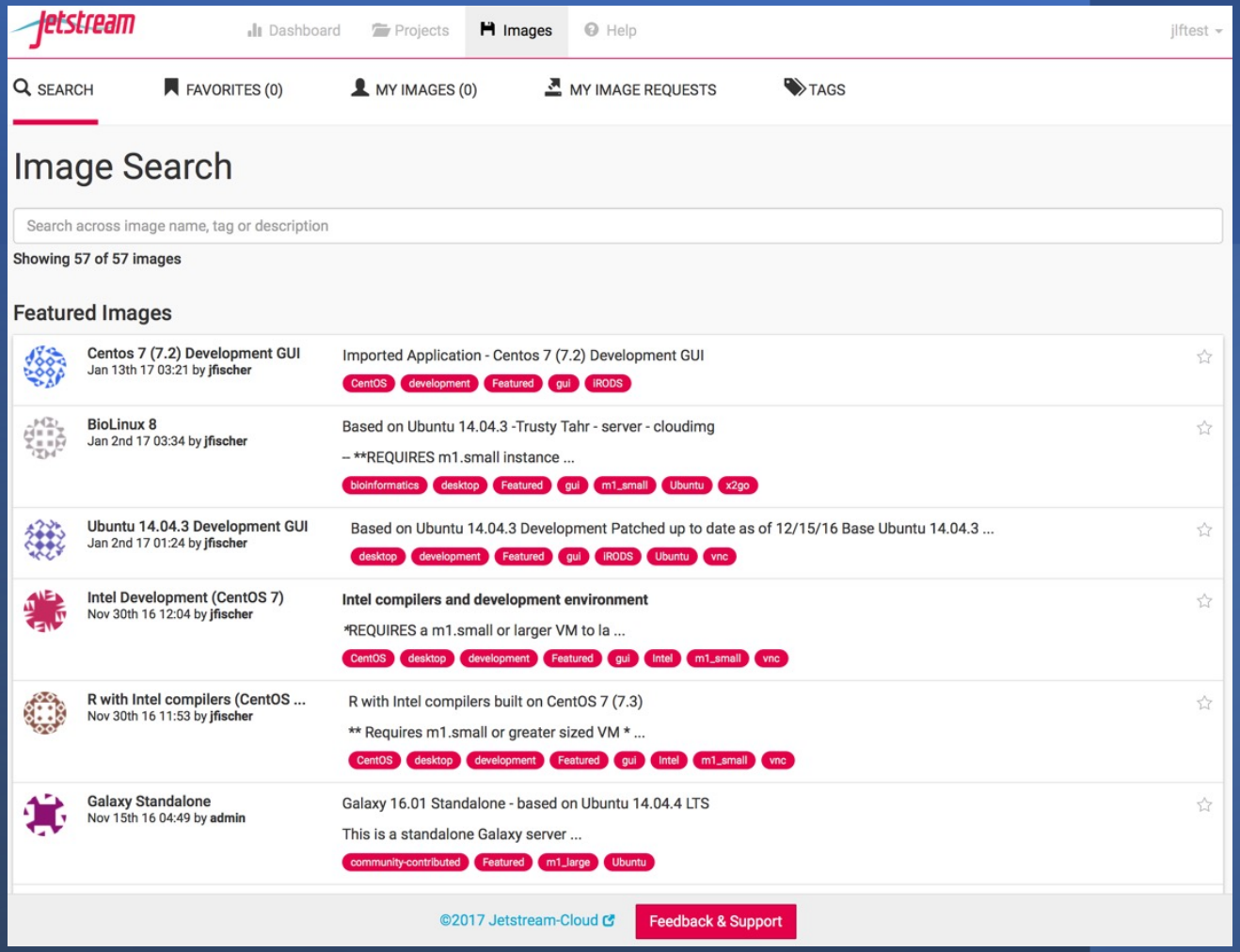


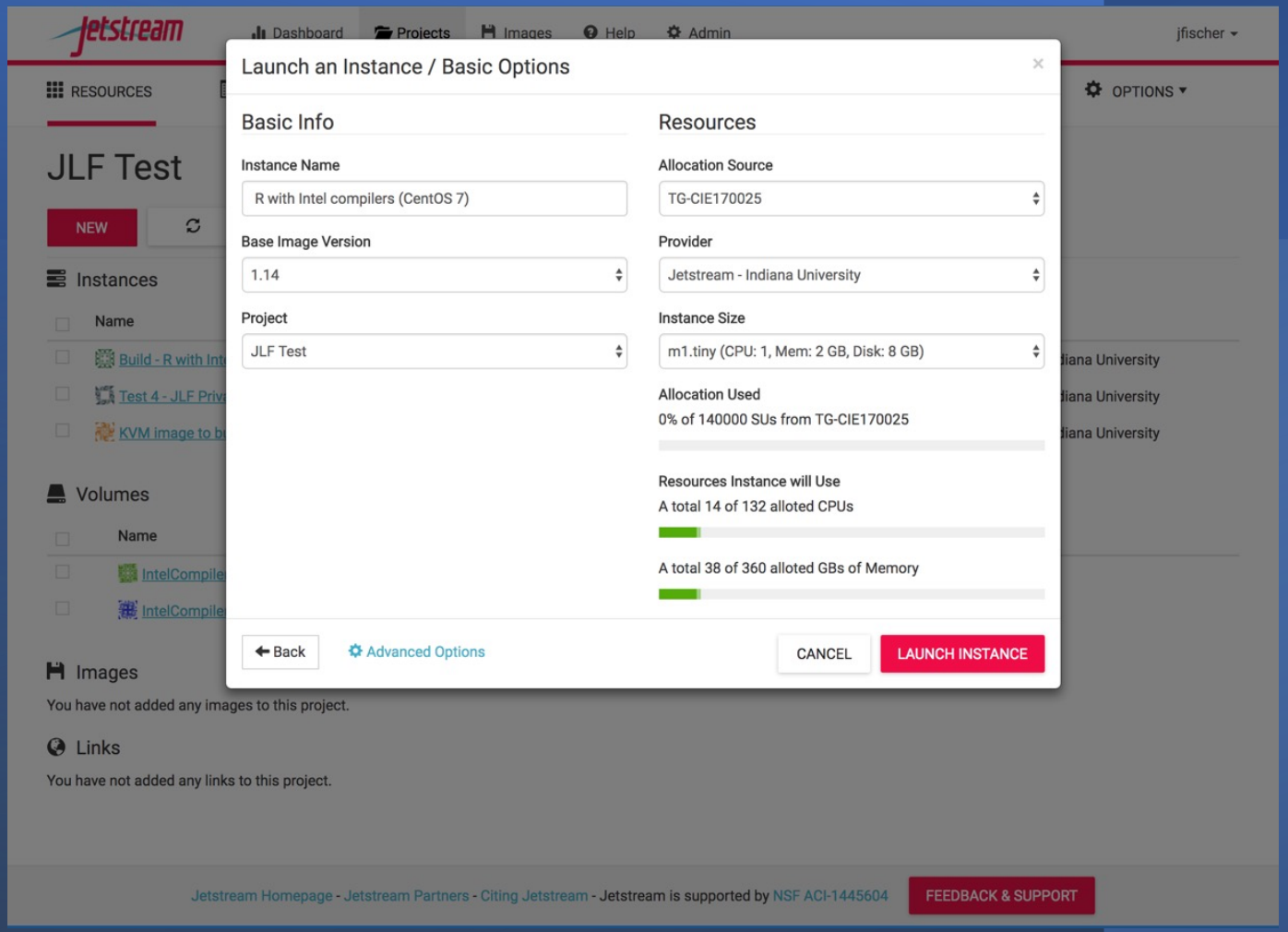


# All cloud all the time

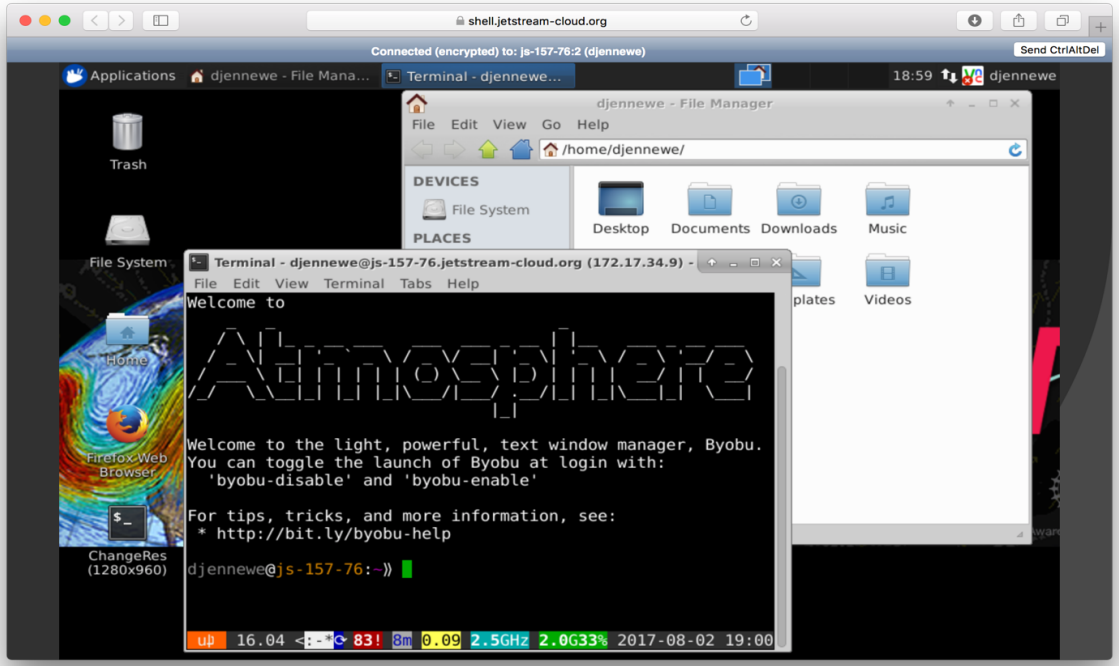
For the researcher needing a handful of cores on demand as well as for software creators and researchers needing to create their own **customized virtual machine** environments. Jetstream is accessible ONLY via web interface.


Jetstream
Dashboard
Projects
Images
Help
Admin
jfischer
Getting Started
Launch New Instance
Browse Atmosphere's list of available images and select one to launch a new instance.
Browse Help Resources
View a video tutorial, read the how-to guides, or email the Atmosphere support team.
Change Your Settings
Modify your account settings, view your resource quota, or request more resources.
Resources Used
Need more?
Allocation Source
TG-STA110024S
TG-ASC160018
TG-CDA160007
TG-TRA160003
TG-TRA160027
Allocation
17.61%
0%
23.15%
32.23%
5.06%
0
10
20
30
40
50
60
70
80
90
100
Percent of Allocation Used
10 Instances
active
shutoff
4 Volumes
available
Provider Resources
Jetstream - Indiana University
Jetstream - TACC
CPU
0%
9.09%
©2017 Jetstream-Cloud
Feedback & Support


jetstream
Dashboard
Projects
Images
Help
jlftest
SEARCH
FAVORITES (0)
MY IMAGES (0)
MY IMAGE REQUESTS
TAGS
Image Search
Search across image name, tag or description
Showing 57 of 57 images
Featured Images
Centos 7 (7.2) Development GUI
Jan 13th 17 03:21 by jfischer
Imported Application - Centos 7 (7.2) Development GUI
CentOS development Featured gui iRODS
BioLinux 8
Jan 2nd 17 03:34 by jfischer
Based on Ubuntu 14.04.3 -Trusty Tahr - server - cloudimg
-- **REQUIRES m1.small instance ...
bioinformatics desktop Featured gui m1_small Ubuntu x2go
Ubuntu 14.04.3 Development GUI
Jan 2nd 17 01:24 by jfischer
Based on Ubuntu 14.04.3 Development Patched up to date as of 12/15/16 Base Ubuntu 14.04.3 ...
desktop development Featured gui iRODS Ubuntu vnc
Intel Development (CentOS 7)
Nov 30th 16 12:04 by jfischer
Intel compilers and development environment
*REQUIRES a m1.small or larger VM to la ...
CentOS desktop development Featured gui Intel m1_small vnc
R with Intel compilers (CentOS ...
Nov 30th 16 11:53 by jfischer
R with Intel compilers built on CentOS 7 (7.3)
** Requires m1.small or greater sized VM * ...
CentOS desktop development Featured gui Intel m1_small vnc
Galaxy Standalone
Nov 15th 16 04:49 by admin
Galaxy 16.01 Standalone - based on Ubuntu 14.04.4 LTS
This is a standalone Galaxy server ...
community-contributed Featured m1_large Ubuntu
©2017 Jetstream-Cloud
Feedback & Support


Jetstream
Dashboard
Projects
Images
Help
Admin
jfischer
RESOURCES
OPTIONS
JLF Test
NEW
Instances
Name
Volumes
Images
You have not added any images to this project.
Links
You have not added any links to this project.
Launch an Instance / Basic Options
Basic Info
Instance Name
R with Intel compilers (CentOS 7)
Base Image Version
1.14
Project
JLF Test
Resources
Allocation Source
TG-CIE170025
Provider
Jetstream - Indiana University
Instance Size
m1.tiny (CPU: 1, Mem: 2 GB, Disk: 8 GB)
Allocation Used
0% of 140000 SUs from TG-CIE170025
Resources Instance will Use
A total 14 of 132 alloted CPUs
A total 38 of 360 alloted GBs of Memory
Back
Advanced Options
CANCEL
LAUNCH INSTANCE
Jetstream Homepage - Jetstream Partners - Citing Jetstream - Jetstream is supported by NSF ACI-1445604
FEEDBACK & SUPPORT

# Real screenshot from real iPad



## Jetstream2 Capabilities

Enhancing IaaS model of Jetstream:

- Improved orchestration support
- Elastic virtual clusters
- Federated JupyterHubs

Commitment to >99% uptime

- Critical for science gateway hosting
- Hybrid-cloud support

Revamped User Interface

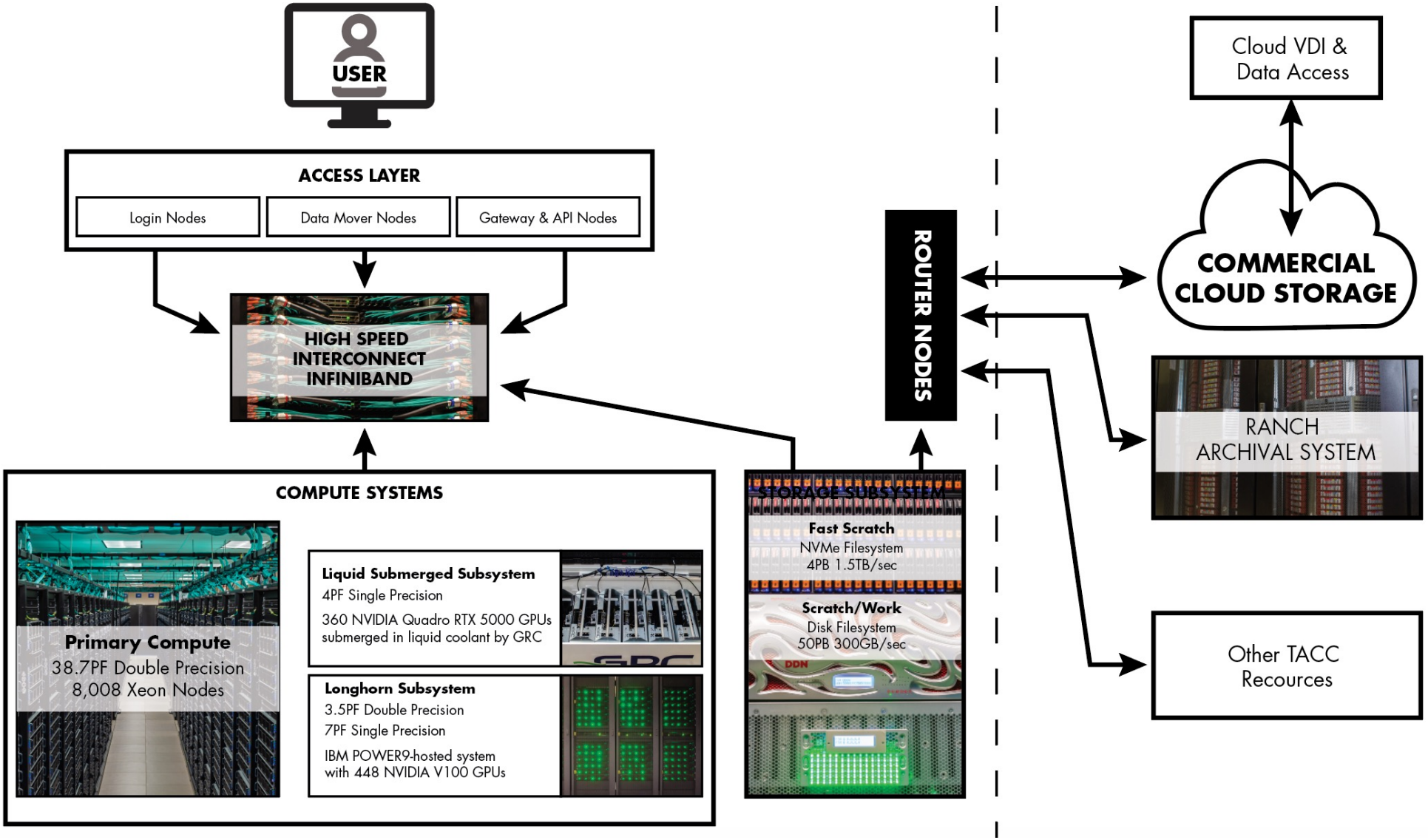
- Unified instance management
- Multi-instance launch

Feb 12, 2019 – Jet stream region called "Jet N6"
NASA/JPL-Caltech/SwRI/MSSS/Kevin M. Gill

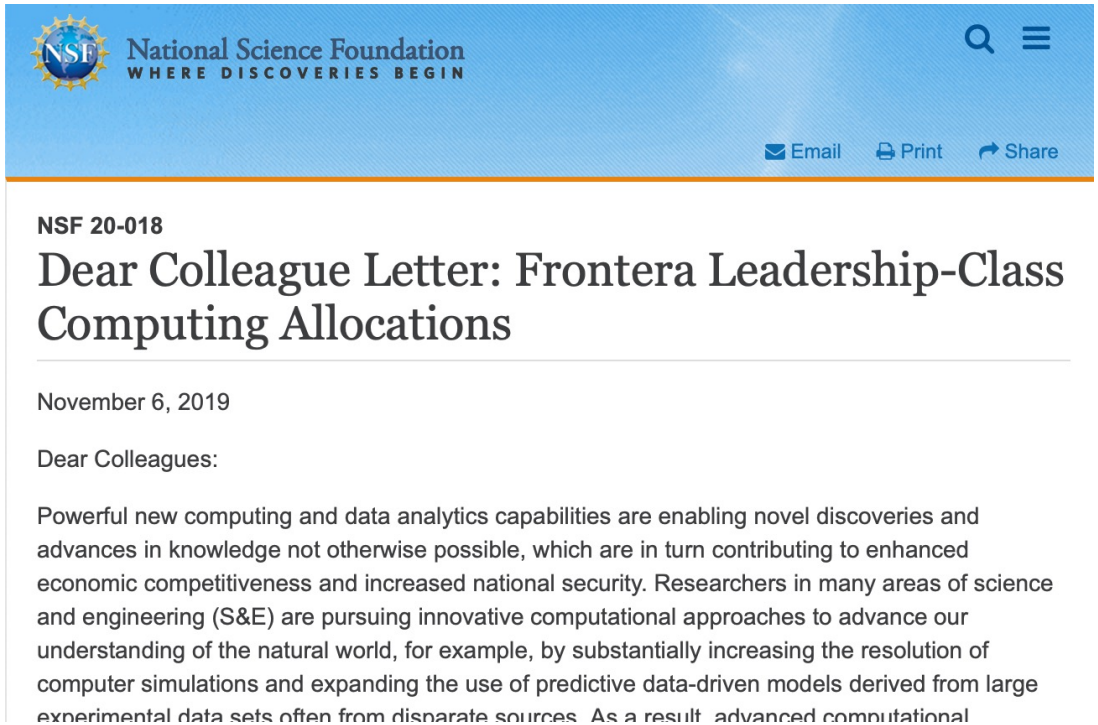
- >57K cores of next-gen AMD EPYC processors
- >360 NVIDIA A100 GPUs will provide vGPUs via NVIDIA's MIG feature
- >18PB of storage (NVMe and disk hybrid)
- 100GbE Mellanox network

# Jetstream2 coming soon!


# FRONTERA – LEADERSHIP CLASS COMPUTING


USER
ACCESS LAYER
Login Nodes
Data Mover Nodes
Gateway & API Nodes
HIGH SPEED INTERCONNECT INFINIBAND
COMPUTE SYSTEMS
Primary Compute
38.7PF Double Precision
8,008 Xeon Nodes
Liquid Submerged Subsystem
4PF Single Precision
360 NVIDIA Quadro RTX 5000 GPUs submerged in liquid coolant by GRC
Longhorn Subsystem
3.5PF Double Precision
7PF Single Precision
IBM POWER9-hosted system with 448 NVIDIA V100 GPUs
STORAGE SUBSYSTEM
Fast Scratch
NVMe Filesystem
4PB 1.5TB/sec
Scratch/Work
Disk Filesystem
50PB 300GB/sec
ROUTER NODES
Cloud VDI & Data Access
COMMERCIAL CLOUD STORAGE
RANCH ARCHIVAL SYSTEM
Other TACC Recources

# ALLOCATIONS ARE DIFFERENT



National Science Foundation
WHERE DISCOVERIES BEGIN

Email Print Share

**NSF 20-018**

## Dear Colleague Letter: Frontera Leadership-Class Computing Allocations

November 6, 2019

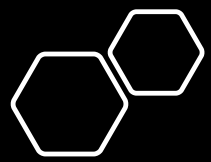
Dear Colleagues:

Powerful new computing and data analytics capabilities are enabling novel discoveries and advances in knowledge not otherwise possible, which are in turn contributing to enhanced economic competitiveness and increased national security. Researchers in many areas of science and engineering (S&E) are pursuing innovative computational approaches to advance our understanding of the natural world, for example, by substantially increasing the resolution of computer simulations and expanding the use of predictive data-driven models derived from large experimental data sets often from disparate sources. As a result, advanced computational

U.S. DEPARTMENT OF
ENERGY

Office of Science


# THE NATIONAL ENERGY RESEARCH SCIENTIFIC COMPUTING CENTER




The National Energy Research Scientific Computing Center (NERSC) is the primary scientific computing facility for the **Office of Science** in the **U.S. Department of Energy**.

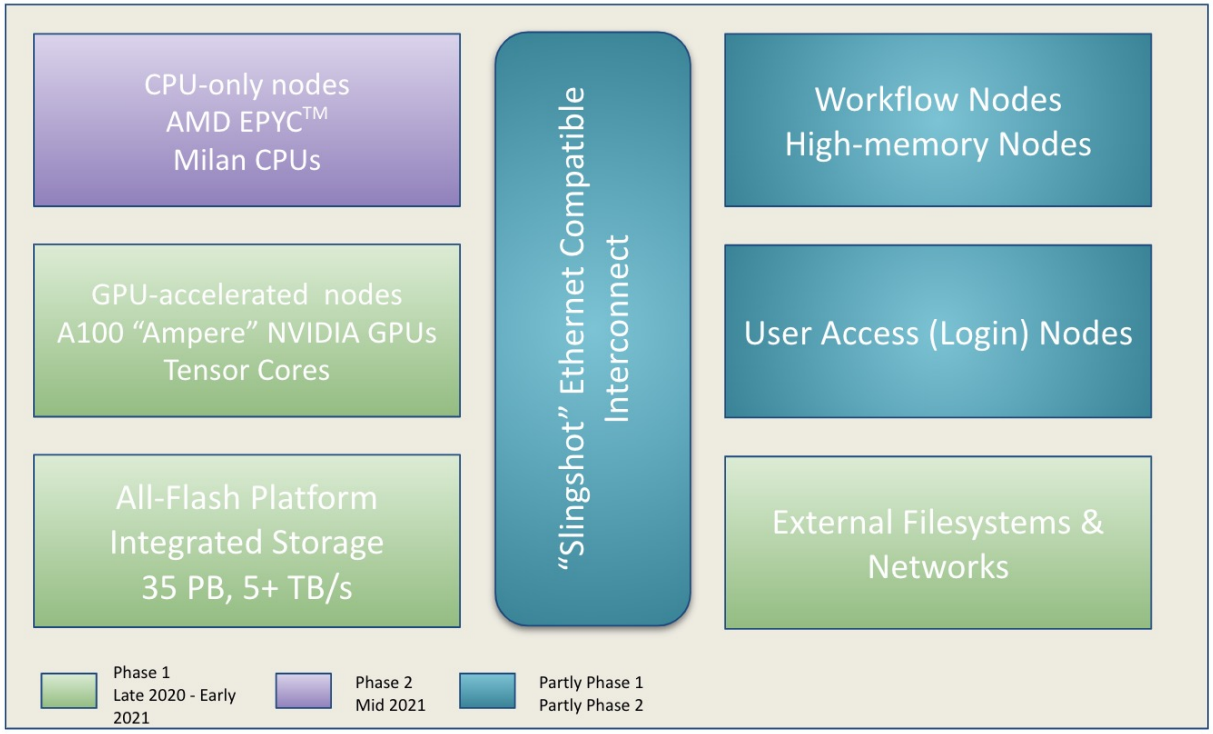
All research projects that are funded by the DOE Office of Science and require high performance computing support are eligible to apply to use NERSC resources. **Projects that are not funded by the DOE Office of Science, but that conduct research that supports the Office of Science mission may also apply.**


## System Overview

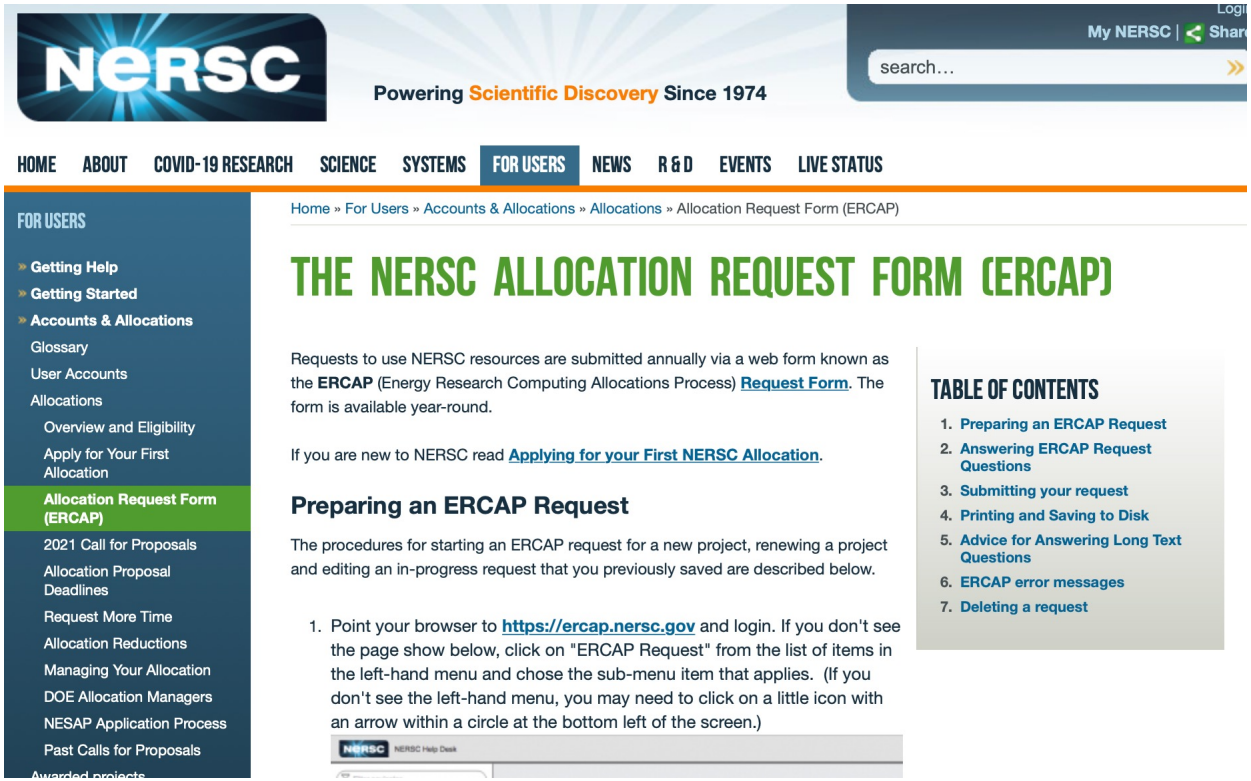
| System Partition | # of cabinets | # of nodes | Aggregate Theoretical Peak | Aggregate Memory |
|---|---|---|---|---|
| Login | - | 20 | - | - |
| Haswell | 14 | 2,388 | 2.81 PFlops | 298.5 TB |
| KNL | 54 | 9,688 | 29.5 PFlops | 1.09 PB |
| Large Memory | - | 20 | 30.7 TFlops | 40 TB |


NERSC
Perlmutter

# COMING SOON


CPU-only nodes
AMD EPYC™
Milan CPUs
GPU-accelerated nodes
A100 "Ampere" NVIDIA GPUs
Tensor Cores
All-Flash Platform
Integrated Storage
35 PB, 5+ TB/s
"Slingshot" Ethernet Compatible
Interconnect
Workflow Nodes
High-memory Nodes
User Access (Login) Nodes
External Filesystems &
Networks
Phase 1
Late 2020 - Early
2021
Phase 2
Mid 2021
Partly Phase 1
Partly Phase 2

# ALLOCATIONS ARE DIFFERENT



# The DOE Energy Sciences Network

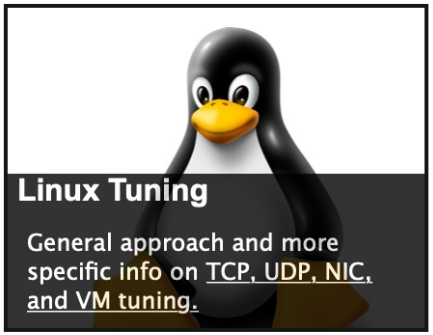
ESnet
ENERGY SCIENCES NETWORK

ESnet provides the high-bandwidth, reliable connections that link scientists at **national laboratories, universities, and other research institutions**. Funded by the DOE Office of Science, ESnet is managed and operated by the Scientific Networking Division at **Lawrence Berkeley National Laboratory**.

# Network Performance Knowledge Base

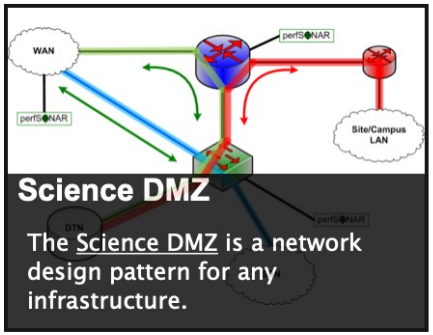
## ESnet Fasterdata Knowledge Base

An Expert Guide for End-to-End Performance Tuning, Tools and Techniques



**Linux Tuning**

General approach and more specific info on TCP, UDP, NIC, and VM tuning.




**Science DMZ**

The Science DMZ is a network design pattern for any infrastructure.

**Network Tuning**

Fix issues with TCP, firewalls, routers, switches, and more.

| Data set size | |
|---|---|
| 10PB | 1,333.33 Tbps |
| 1PB | 133.33 Tbps |
| 100TB | 13.33 Tbps |
| 10TB | 1.33 Tbps |
| 100GB | 133.33 Mbps |

**Network Expectations**

Information about benchmarks for networks and data transfer speeds.

# Engagement and Performance Operations Center (EOPC)

https://epoc.global

Established in 2018 with funding from NSF

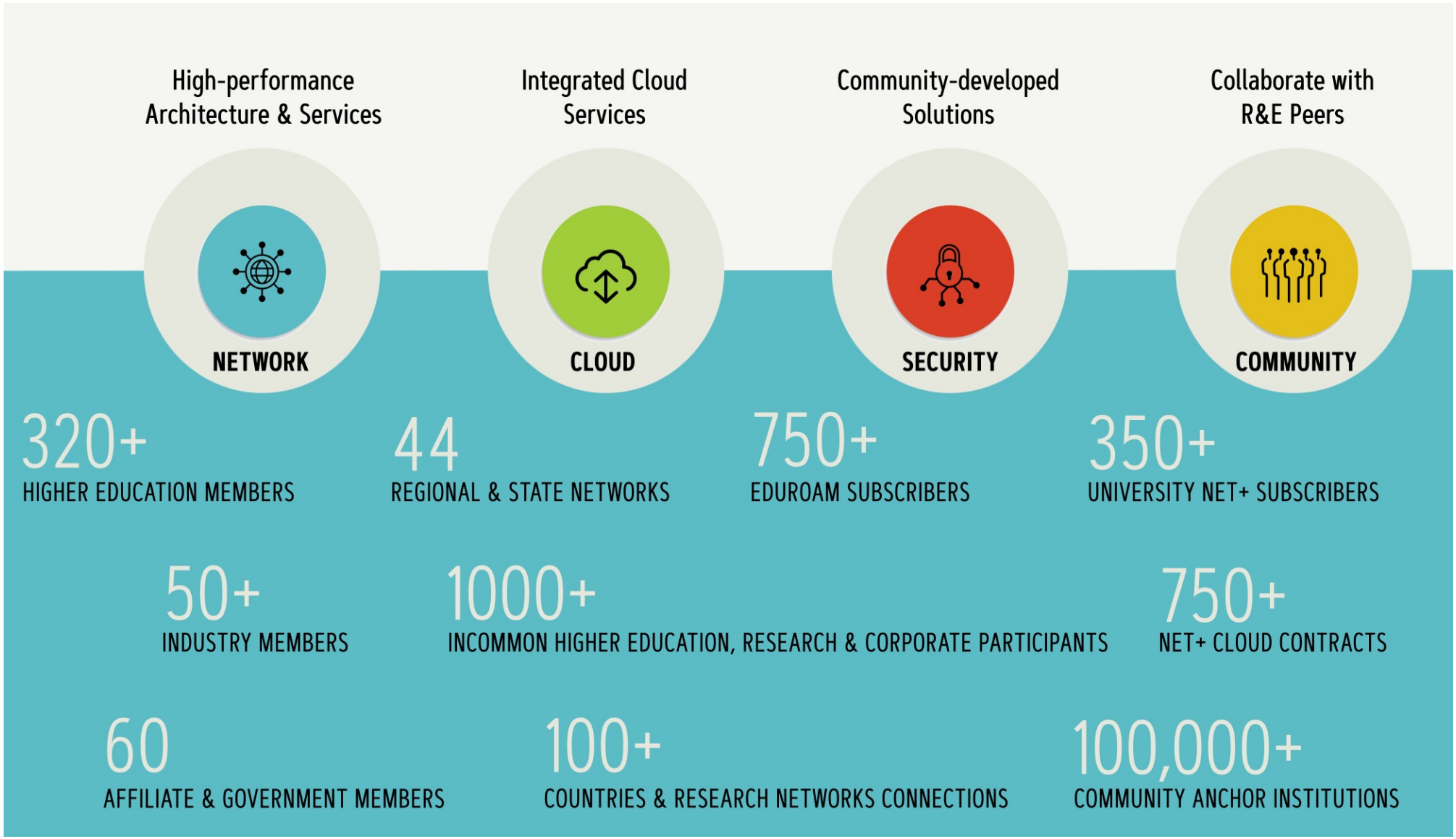
Collaboration of Indiana U and U.S. DOE Energy Sciences Network (ES.net)

# Engagement and Performance Operations Center (EOPC)

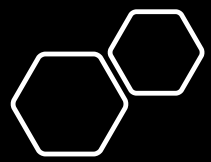
*EPOC provides researchers with a holistic set of tools and services needed to debug performance issues and enable reliable and robust data transfers.*

- Roadside Assistance and consultation via a coordinated Operations Center to resolve network performance problems with end-to-end data transfers reactively
- Application Deep Dive to work more closely with application communities to understand full workflows for diverse research teams in order to evaluate bottlenecks and potential capacity issues
- Network Analysis enabled by the NetSage monitoring suite to proactively discover and resolve performance issues
- Provision of managed data services via support through the IU GlobalNOC and our Regional Network Partners
- Coordinated Training to ensure effective use of network tools and science support


A not-for-profit US computer networking consortium led by members from the **research and education** communities, **industry**, and **government.**


High-performance Architecture & Services
Integrated Cloud Services
Community-developed Solutions
Collaborate with R&E Peers
NETWORK
CLOUD
SECURITY
COMMUNITY
320+
HIGHER EDUCATION MEMBERS
44
REGIONAL & STATE NETWORKS
750+
EDUROAM SUBSCRIBERS
350+
UNIVERSITY NET+ SUBSCRIBERS
50+
INDUSTRY MEMBERS
1000+
INCOMMON HIGHER EDUCATION, RESEARCH & CORPORATE PARTICIPANTS
750+
NET+ CLOUD CONTRACTS
60
AFFILIATE & GOVERNMENT MEMBERS
100+
COUNTRIES & RESEARCH NETWORKS CONNECTIONS
100,000+
COMMUNITY ANCHOR INSTITUTIONS

# Open Science Grid



A National, Distributed Computing Partnership for Data-Intensive Research

# Submit locally, compute globally

OSG Consortium founded in 2004

Created to facilitate data analysis from the Large Hadron Collider

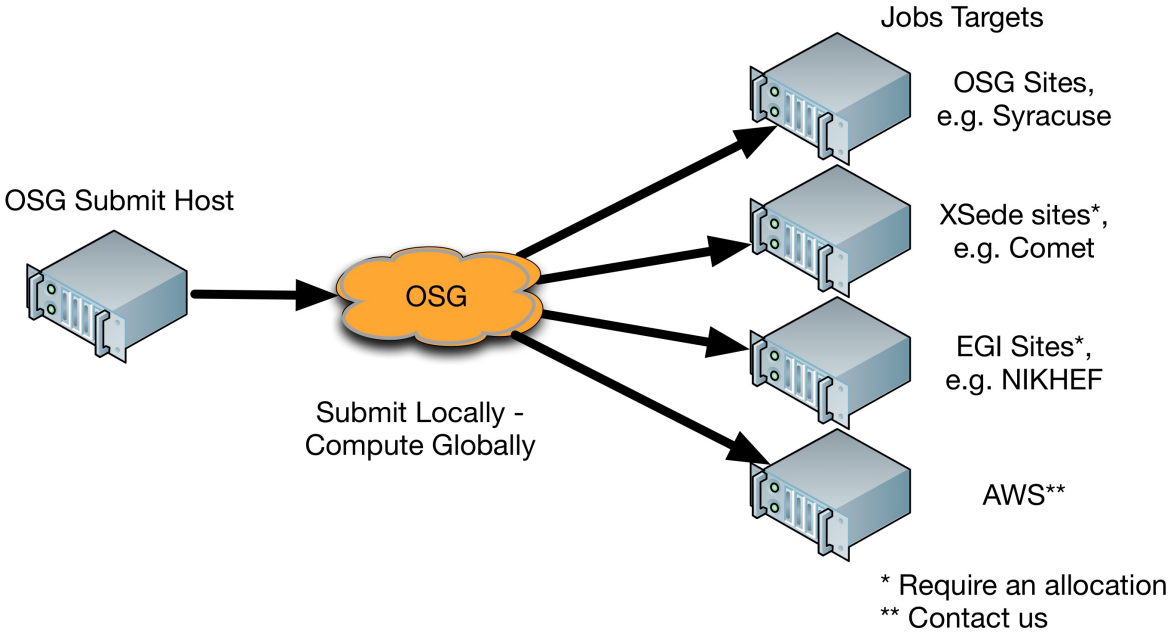
Lends itself well to (High) Throughput Computing

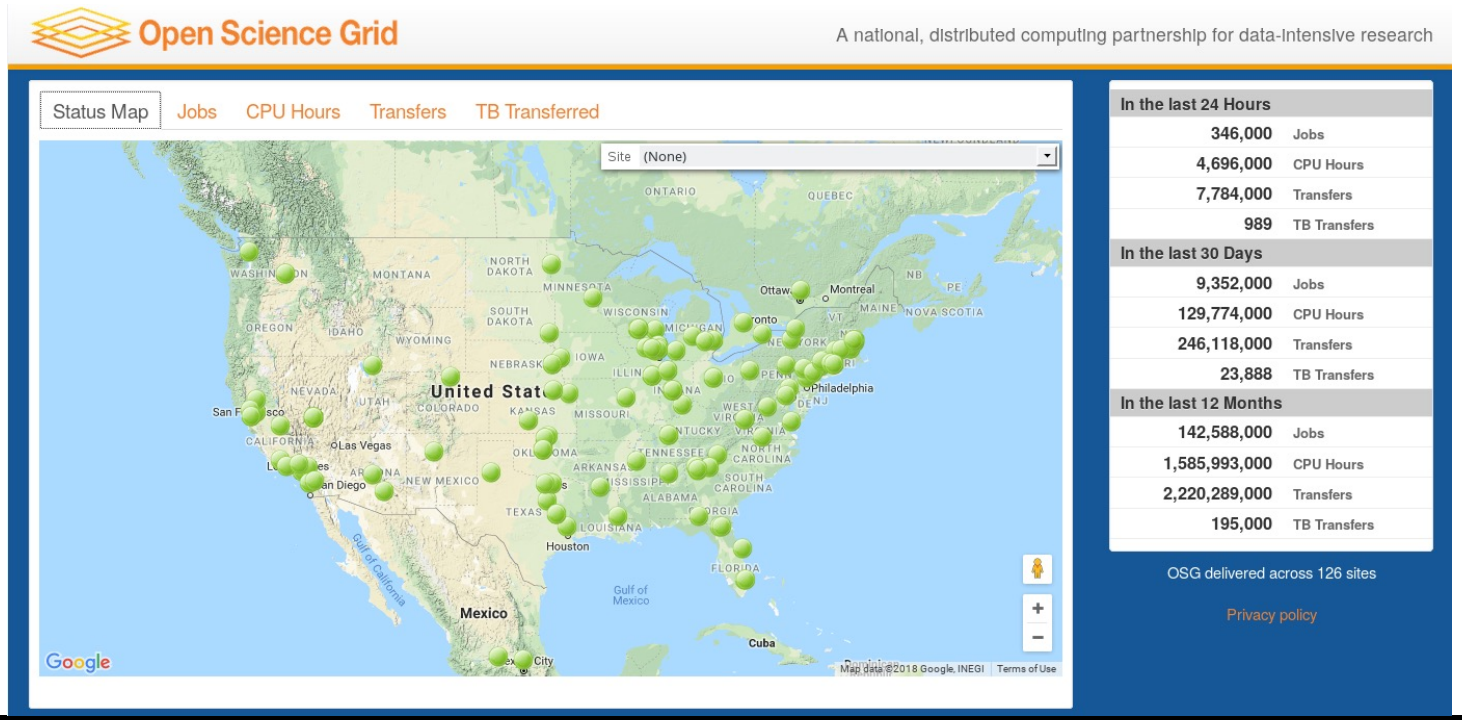
Consists of over 25,000 computers with over 43,000 processors

Funded by DOE as well as NSF

Employs the HTCondor software platform

# How does it work?

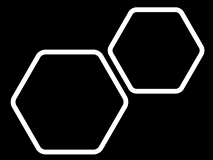


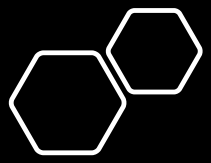



How distributed is it (as of 2018…)?

# SGCI

SCIENCE GATEWAYS COMMUNITY INSTITUTE: CONNECTING PEOPLE AND RESOURCES TO ACCELERATE DISCOVERY BY EMPOWERING THE SCIENCE GATEWAY COMMUNITY





# Gateways

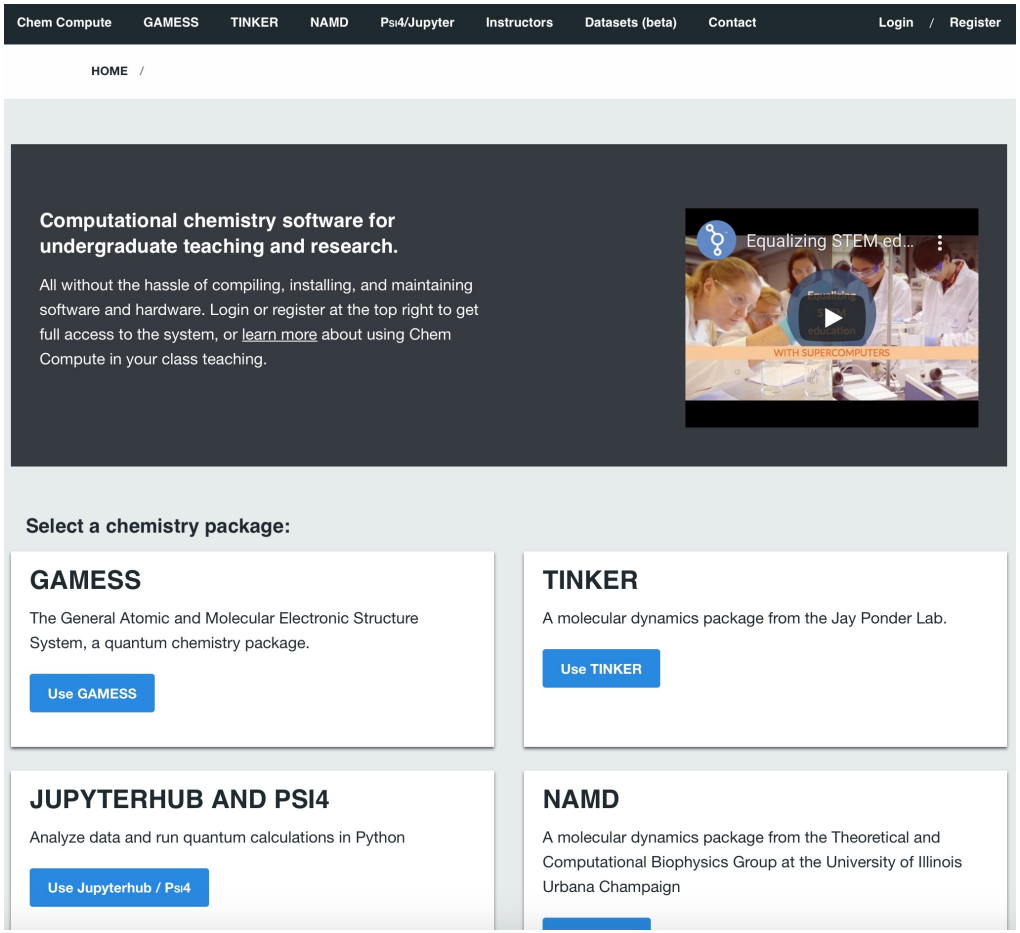
- Science Gateways simplify access to computing resources by **hiding infrastructure complexities**.

- Science Gateways provide **higher level user interface** for XSEDE resources that are tailored to specific scientific communities.

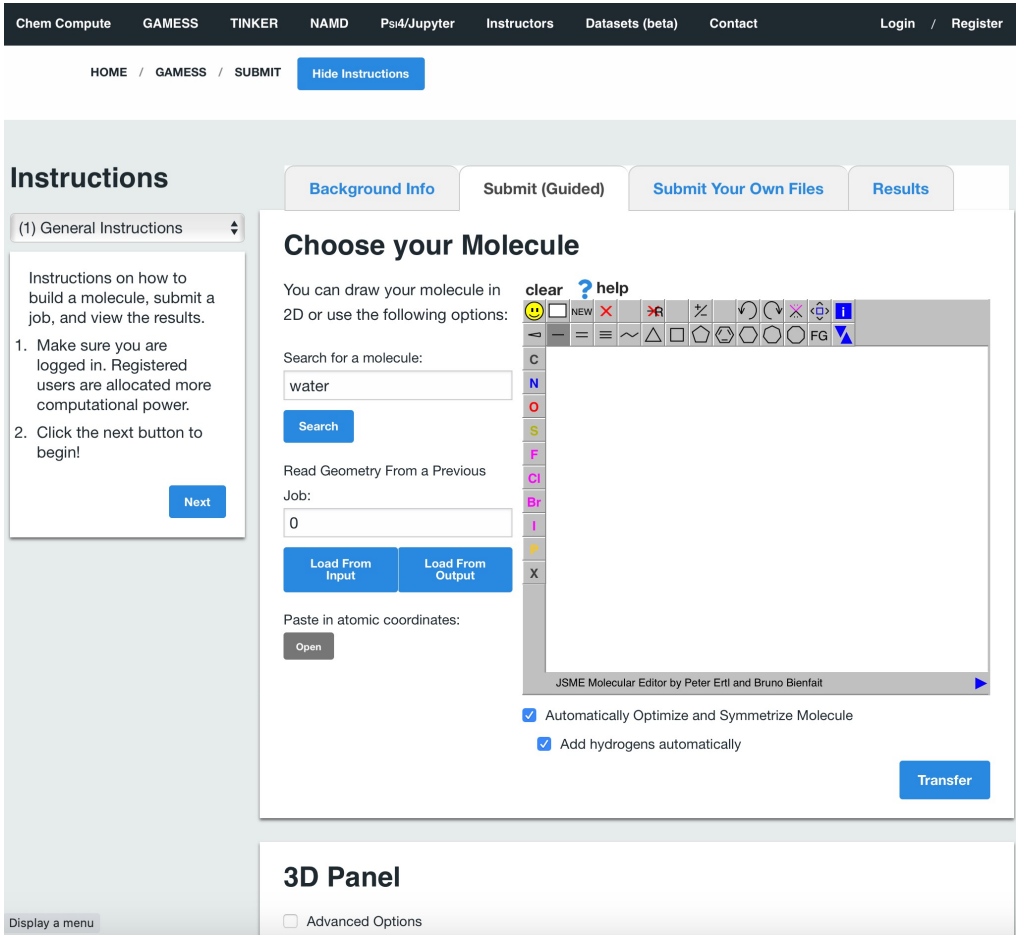
- A Science Gateway is a community-developed set of tools, applications, and data that are **integrated via a portal** or a suite of applications, usually in a graphical user interface, that is further customized to meet the needs of a specific community.

There are a lot of them.

Let’s look at one: Chem Compute



Let’s look at one: Chem Compute



# (Some) International Organizations

- The Partnership for Advanced Computing in Europe (**PRACE**) enables high-impact scientific discovery and engineering research and development across all disciplines by offering world class computing and data management resources and services.

- **G É ANT**: the pan-European data network for the research and education community.

- **Compute Canada** leads the acceleration of research and innovation by deploying state-of-the-art advanced research computing (ARC) systems, storage and software solutions for Canadian researchers and their collaborators in all academic and industrial sectors.

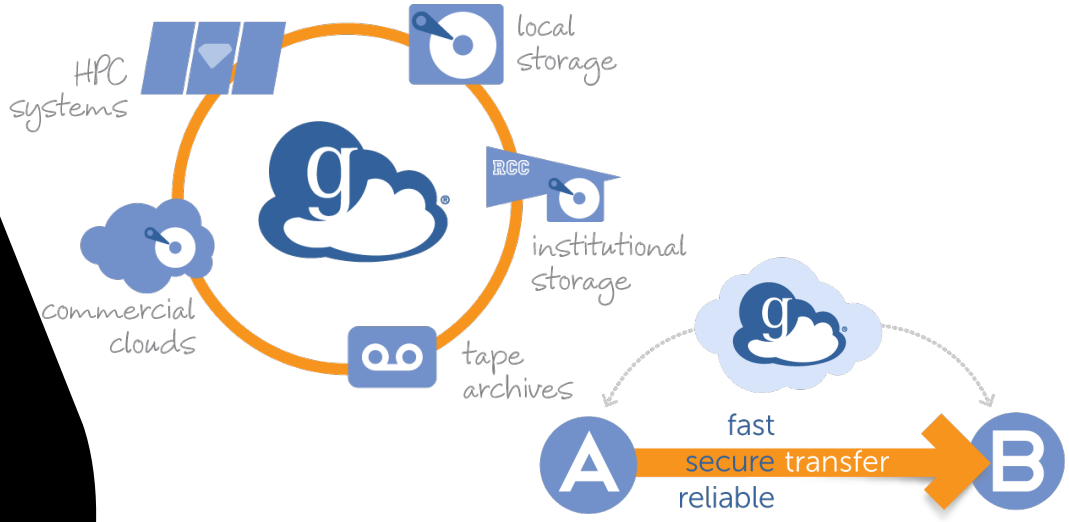
PRACE
PARTNERSHIP FOR ADVANCED COMPUTING IN EUROPE

GÉANT
Networks • Services • People

compute canada | calcul canada

# Tools

- Globus
- XDMoD
- Open OnDemand
- Spack




XDMoD
METRICS ON DEMAND

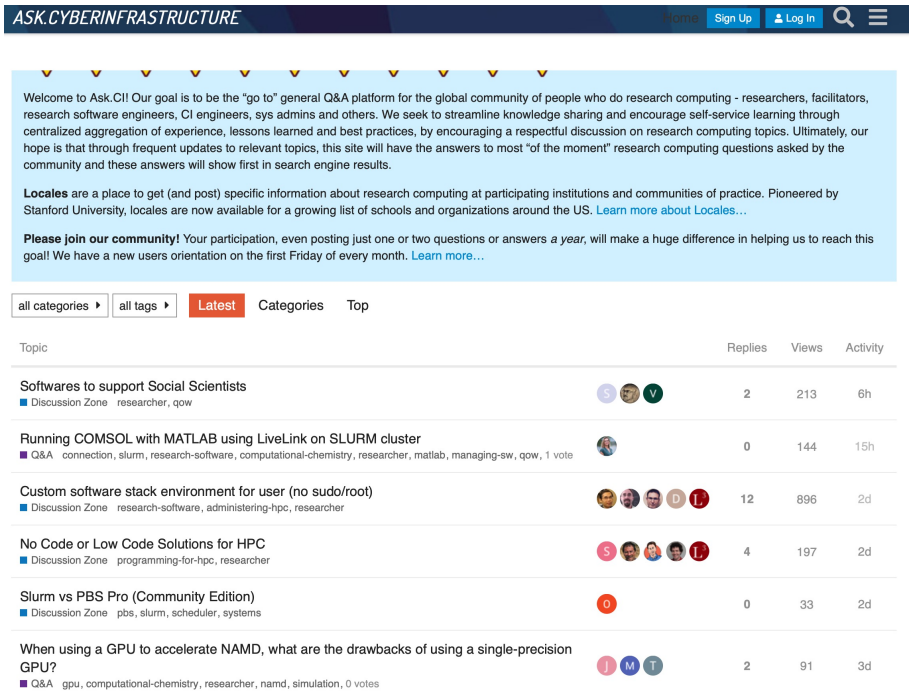
OPEN OnDemand

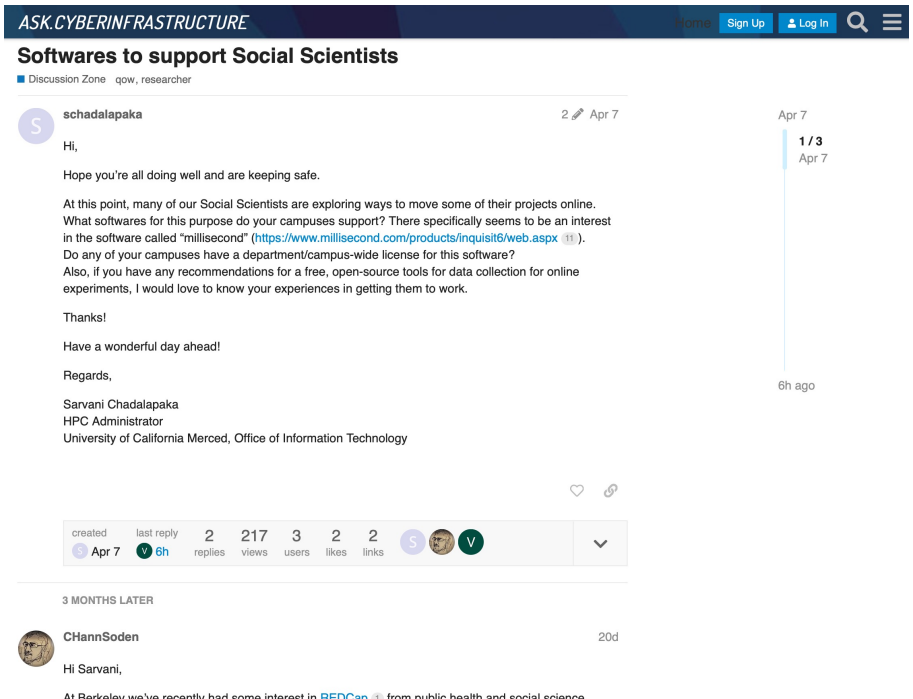
Spack

Ask.Cyberinfrastructure.org
https://ask.ci

Public, searchable, archived Q&A platform for Research Computing

Joint project of Northeast Cyberteam Initiative at MGHPCC and Campus Champions




- Sites like Stack Exchange are dominated by the much larger enterprise IT sector.

- Research Computing questions benefit from **discussion**, not always just one answer.

# Thank you!

**Douglas M. Jennewein,**
**Arizona State University,**
**Campus Champions Leadership Team**

**Douglas.Jennewein@asu.edu**

CAMPUS CHAMPIONS

*With support from:*

NSF

XSEDE
Extreme Science and Engineering Discovery Environment