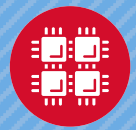
Ohio Supercomputer Center

An OH·TECH Consortium Member

Computing Services to Accelerate Research and Innovation

An introduction to OSC services



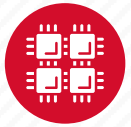
Kate Cahill

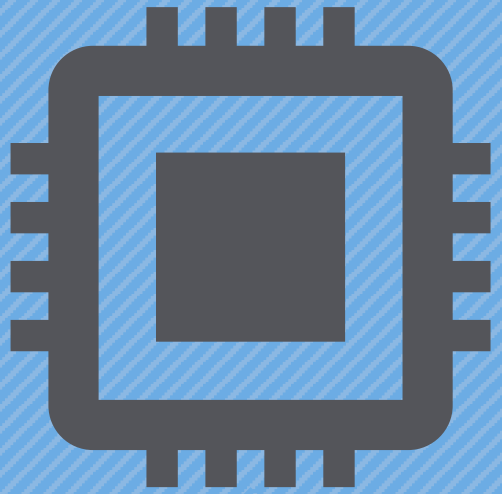
Education & Training Specialist

“OSC is here to empower your research.”

Outline

- What is OSC?
- Overview of Services
  - Hardware Overview
  - Data Storage Systems
  - Software Overview
  - Upcoming services
- OSC Usage and Case Studies
- Research Partnership
- Getting Started at OSC
  - Getting access
  - Learning how to use OSC resources



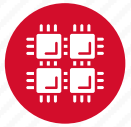


What is the Ohio Supercomputer Center?

"640K ought to be enough for anybody." – Not Bill Gates

About OSC

- Founded in 1987
- Statewide resource for all universities in Ohio
  - high performance computing services
  - computational science expertise
  - “ … propel Ohio's research universities and private industry to the forefront of computational based research.”
- Funded through the Ohio Department of Higher Education
- Reports to the Chancellor
- Located on OSU’s west campus
- Fiscal agent is OSU



The OH-TECH Consortium

Ohio Supercomputer Center provides high performance computing, software, storage and support services for Ohio's scientists, faculty, students, businesses and their research partners.

OARnet connects Ohio's universities, colleges, K-12, health care and state and local governments to its high-speed fiber optic network backbone. OARnet services include co-location, support desk, federated identity and virtualization.

OhioLINK serves nearly 600,000 higher education students and faculty by providing a statewide system for sharing 50 million books and library materials, while aggregating costs among its 90 member institutions.

Service Catalog

Cluster Computing

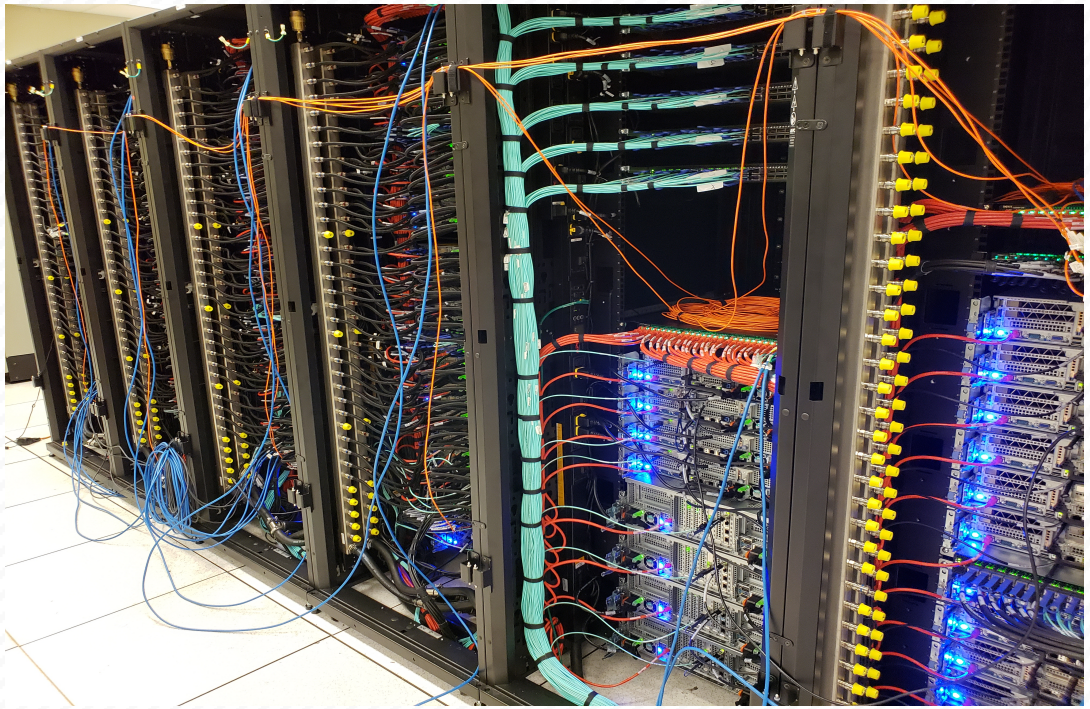
A fully scalable center with mid-range machines to match those found at National Science Foundation centers and other national labs.

Research Data Storage

High-performance, large capacity data storage spaces along with others that are perfect for a wide variety of research data.

Education

High performance computing and networking resources come together to create an exciting and innovative teaching and research environment.



Web Software Development

Our expert web development team helps you create custom web interfaces to simplify the use of powerful HPC resources.

Scientific Software Development

Deep expertise in developing and deploying software that runs efficiently and correctly on large scale cluster computing platforms.

Production Capacity

CY2017

221,400,000+ core-hours consumed

78% average HPC system utilization

4,400,000+ computational jobs

98% up-time

44% average storage system utilization

1.5 PB data stored

2 PB data transferred

79% jobs started within one hour

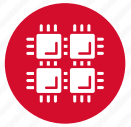
Hardware Overview

“To err is human, but to really foul things up you need a computer.” – Paul Ehrlich

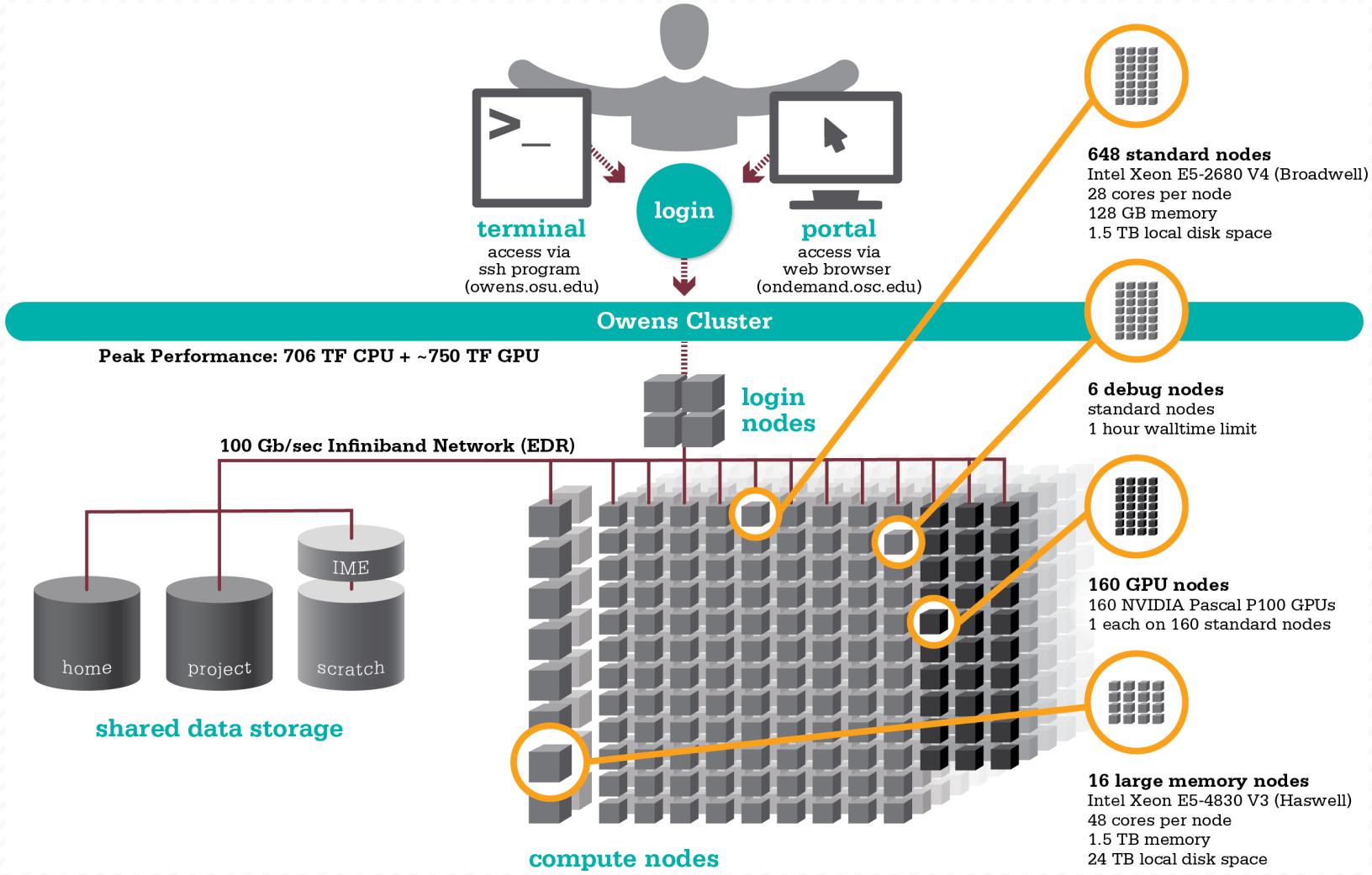
System Status (2019)

SYSTEMS	Ruby	Owens	Pitzer
Date	2014	2016	2018
Cost	$1.5 million	$7 million	$3.35 million
Theoretical Perf.	~144 TF	~1600 TF	~1300 TF
Nodes	240	824	260
CPU Cores	4800	23392	10560
RAM	~15.3 TB	~120 TB	~ 70.6 TB
GPUs	20 NVIDIA Tesla K40	160 NVIDIA Pascal P100	64 NVIDIA Volta V100
	Total compute: ~3,044 TF		

STORAGE	Home	Project	Scratch	Tape Library
Capacity	0.8 PB	3.4 PB	1.1 PB	7+ PB
Current utilization Feb, 18	47%	48%	59%	47%



Owens Cluster Specifications



New HPC Cluster “Pitzer”

- **Named after Russell M. Pitzer**
  - Emeritus professor of chemistry at The Ohio State University
  - Instrumental in founding both OSC and OARnet
- **Goals**
  - Complement existing systems
  - Replace Oakley with a petaflop class system
- **Timeline**
  - System delivery August 15, 2018
  - Full production November 2018
  - Oakley decommissioning Dec 2018

Pitzer Cluster

Characteristics relative to Oakley

- Delivers 8x the processing power (1,300 vs 154 TF)
- Costs 15% less ($4M vs $3.35M)
- Provides 25% more cores (10,560 vs 8,304)
- Has 2X the memory (70.6Tb vs 33.4TB)
- Uses 20% less power

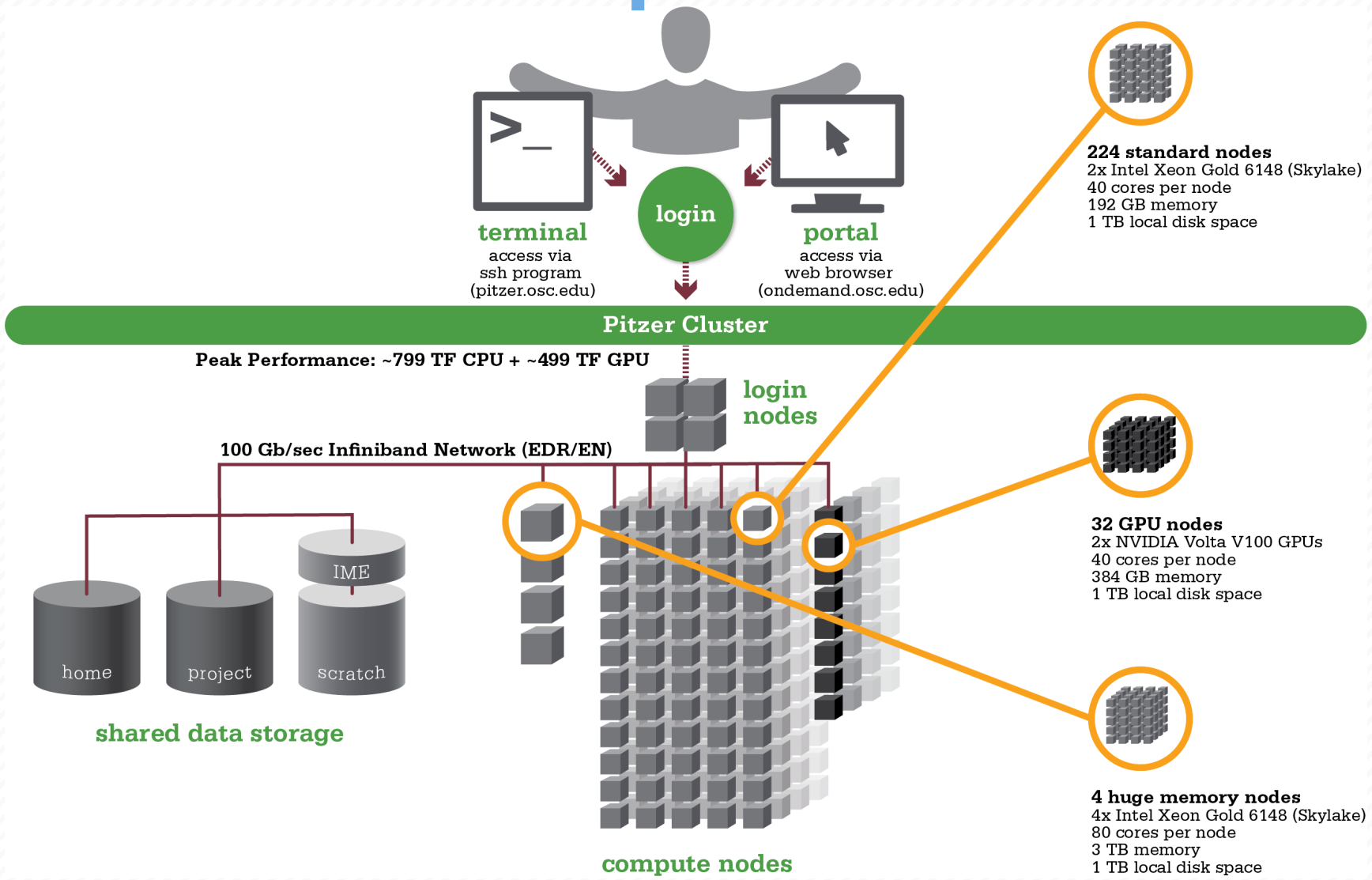
Highlights

- 10,560 processor cores, ~1.3 petaflop peak
- Latest generation: SkyLake processors, 100Gb InfiniBand
- Warm water cooling supports high density, increased performance and efficiency

3 Types of Resources and Example Workload

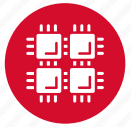
- **Standard compute** (224 nodes) / Modeling and simulation for industry
- **GPU** (32 nodes) / Machine learning, artificial intelligence (AI)
- **Large memory** (4 nodes) / Genomics Sequence Assembly

Pitzer Cluster Specifications

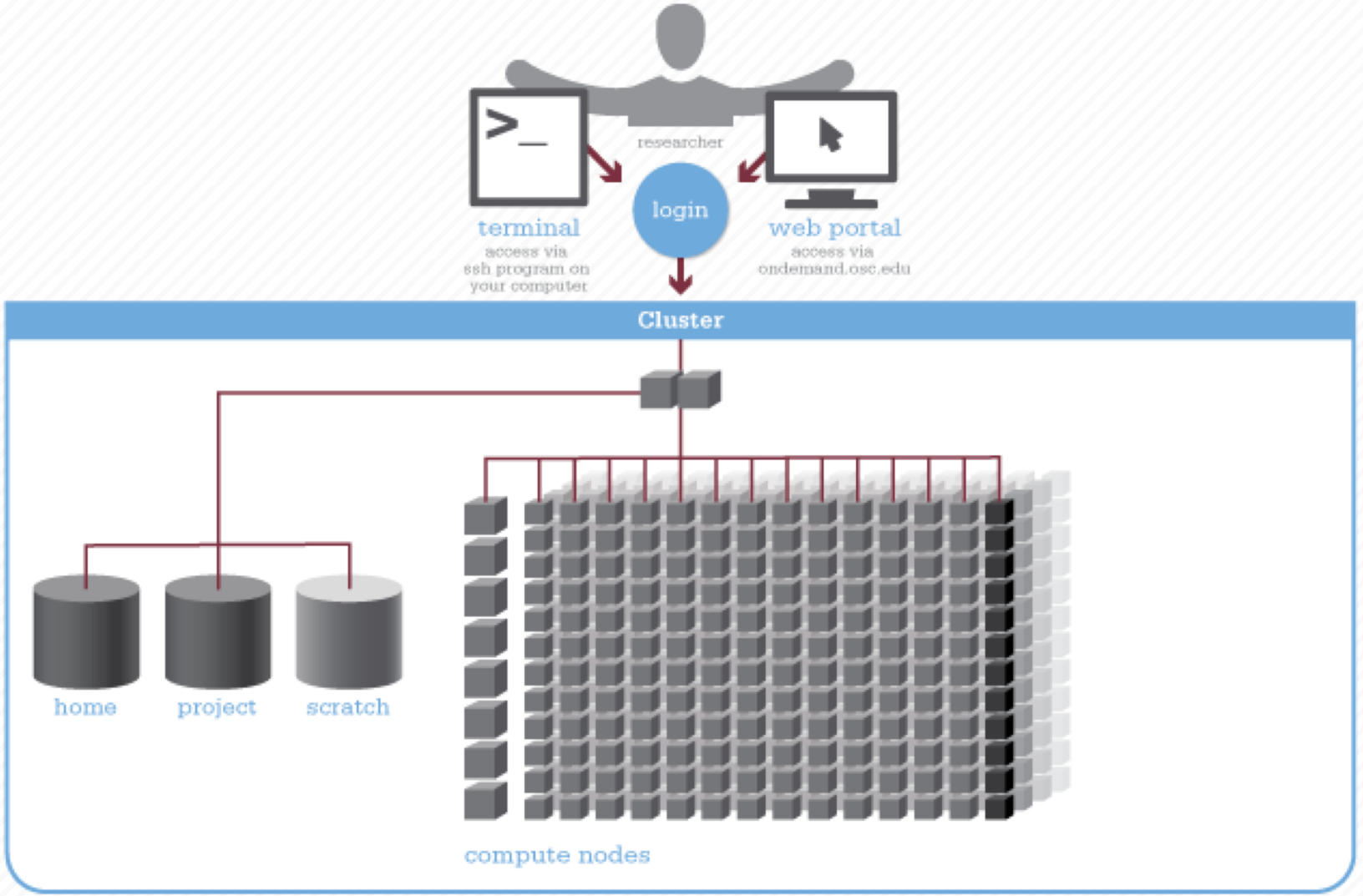


Data Storage Systems

”War is ninety percent information.” – Napoleon Bonaparte

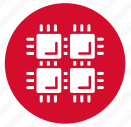


Four different file systems



Filesystem Overview

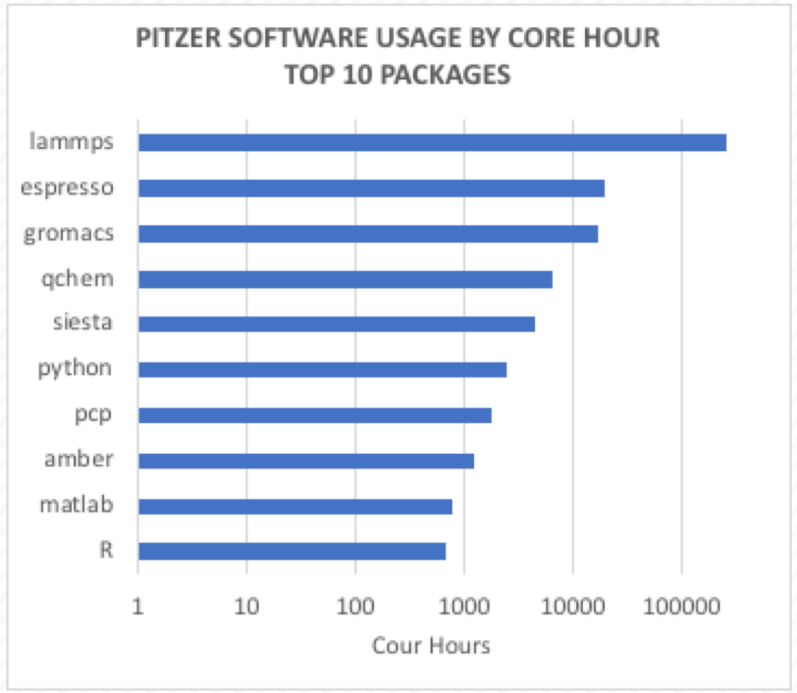
- Home
  - Store files for individual users
  - Backed up daily
- Project
  - Large scale persistent storage
  - Available to Project PIs by request
  - Shared by all users on a project, backed up daily
- Scratch
  - Store large input or output files for HPC simulation or analysis
  - Faster I/O than Home or Project
  - Temporary storage, not backed up
- $TMPDIR
  - Storage on compute nodes, for use during a job
  - Results must be copied to another location
  - All data purged when job quits

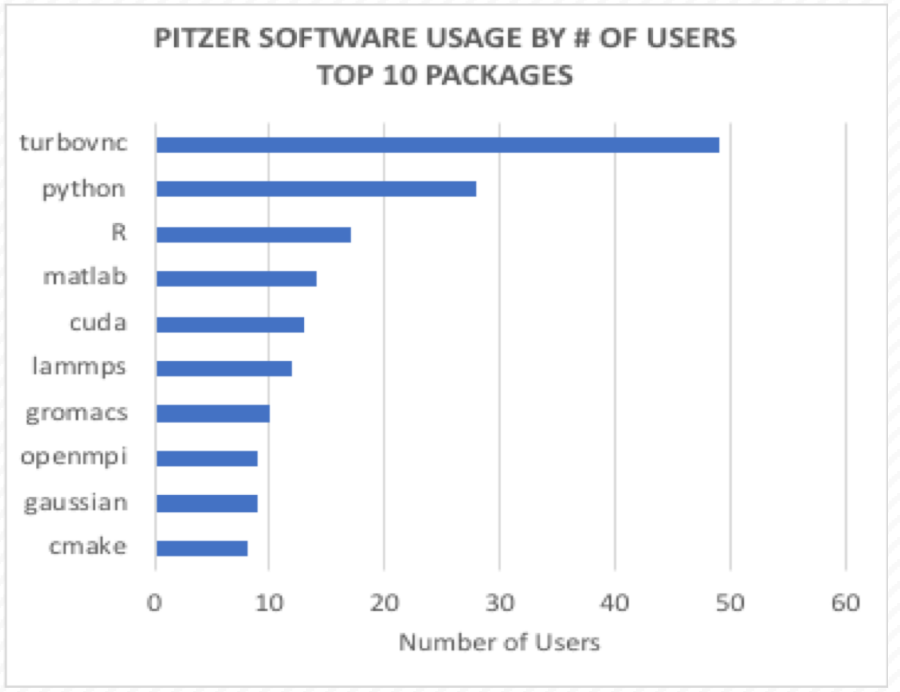


Software Overview

"It works on my machine." – unknown

Software Resources and Use

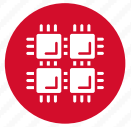




- Software installed and maintained for users
  - ~145 software packages overall, see https://www.osc.edu/resources/available_software
  - Simulations, data analysis tools, numerical libraries, software development tools
  - Support for containers (via singularity) on Pitzer and Owens
- Mix of ISV, free and open source packages
  - 19 purchased software packages currently supported
  - More than $90K spent on license renewals/maintenance in 2018

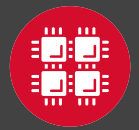
Overview – software development tools

- **Compilers:** Intel, gnu, PGI
- **MPI:** MVAPICH, IntelMPI, OpenMPI
- **OpenMP:** supported by Intel, gnu and PGI compilers
- **GPU:** latest CUDA, OpenACC (PGI compilers)
- **Performance and debug tools:** Arm, Intel and opensource tools
- **Containers:** singularity to run docker containers, with support for GPU



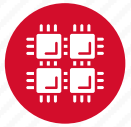
Software Maintained by OSC

- Always first check software page on https://www.osc.edu/resources/available_software
  - Version information for all clusters
  - License information – some software you must request access
  - Usage examples



If OSC doesn't have the software you need?

- List of applications can be found at Software page:
  http://www.osc.edu/supercomputing/software/
- Commercial/Licensed software
  - Fill out a request form (see our FAQ)
  - SUG will consider it
- Open-source software
  - You can install it yourself in your home directory
  - If there's enough demand, we can install it for shared use
- Have your own license?
  - Contact OSC Help



New & Upcoming Services

“In God we trust. All others must bring data.” – W. Edwards Deming

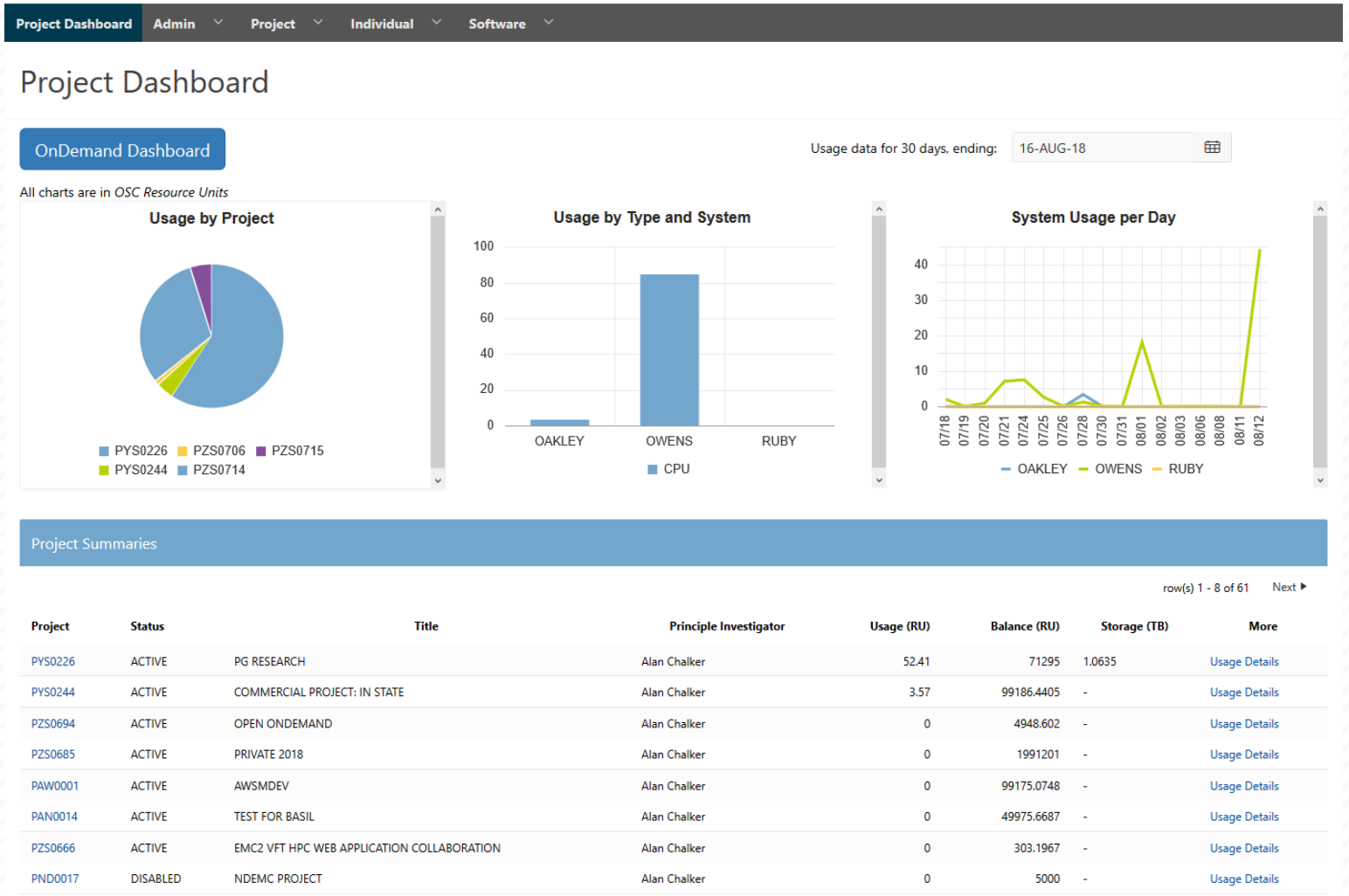
Client Portal Project

Replacement for my.osc.edu

- Friendly user testing started on August 17
- Scheduled go-live for all clients on October 23

New reports.osc.edu

- Provides OSC staff with robust client usage and billing reporting capabilities



Data Focused Projects

Protected Data Environment

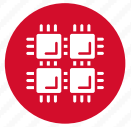
- Unique resource supporting HIPAA, ITAR, or other sensitive data sets
- Initial requirements gathering (OSU Wexner Medical Center, Nationwide Children's Hospital)
  - Provide standard OSC environment with security assurances, assistance with data management plans
  - Collect requirements for different services for new communities

Research Data Archive

- Simplify data management plans for sponsored research
- Provide publishing and other abstraction capabilities
- Additional off-site copy of data for resiliency/availability

Project file system expansion

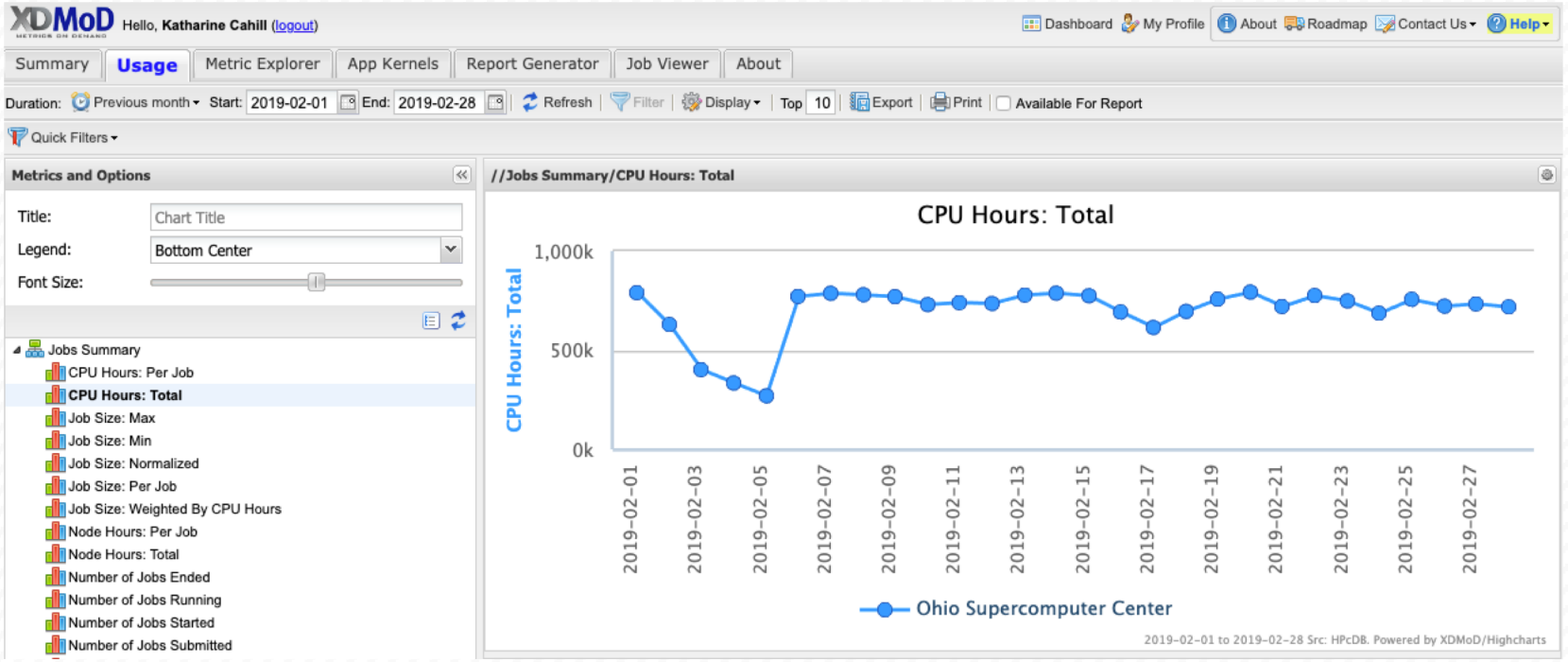
- Increase space for metadata, 2-3B files/directories (1B today)
- Slower tier of storage for infrequently accessed files



xdmod.osc.edu

Report tool for job activity

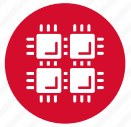
- Information and tutorials: https://www.osc.edu/supercomputing/knowledge-base/xdmod_tool



Other Capital Projects

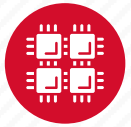
Production infrastructure refresh

- “C20” HPC cluster (replaces the Ruby cluster)
- Storage upgrades
  - Additional fast tier storage
  - Other performance/capacity upgrades
  - Tape media capacity expansion
- Hardware to support protected data environment and research data archive
  - May include “Cloud” resources
- Other EOL hardware



HPC Usage and Case Studies

“The difference between us and a computer is that, the computer is blindingly stupid, but it is capable of being stupid many, many million times a second.” – Douglas Adams



Client Services

CY2017

23 academic institutions

48 companies

2,202 clients

256 awards made

23 training opportunities

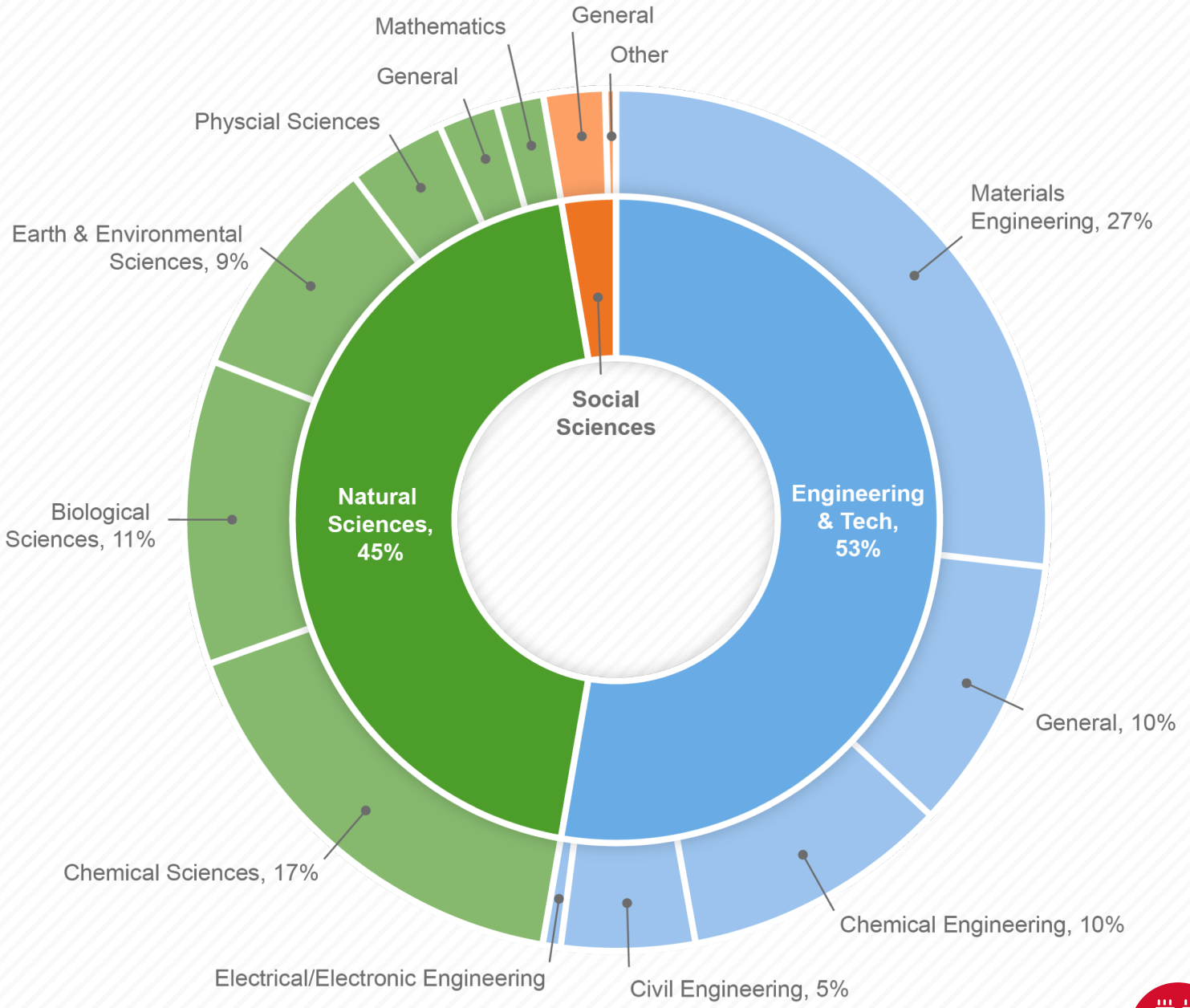
461 trainees

604 projects served

33 courses used OSC

Usage by Field of Science

CY2017



Academic Course Enrollment

CY2017

Department	
Bluffton Mathematics	9
OU Chemical & Biomolecular Engineering	10
OU Chemistry & Biochemistry	2
OSU Materials Science & Engineering	39
OSU Chemical & Biomolecular Engineering	19
OSU Materials Science & Engineering	16
OSU Materials Science & Engineering	6
OSU Computer Science & Engineering	30
OSU Computer Science & Engineering	21
OSU Evolution, Ecology, & Organismal Biology	18
OSU Computer Science & Engineering	3
OSU Evolution, Ecology, & Organismal Biology	12
OSU Biostatistics	3
OSU Chemistry & Biochemistry	28
OSU Computer Science & Engineering	34
OSU Computer Science & Engineering	75
OSU Geography	13

Department	
OSU Computer Science & Engineering	49
OSU Computer Science & Engineering	5
OSU Chemistry	400
OSU Chemistry	14
OSU Chemistry	10
OSU Chemistry	129
OSU Chemistry	105
OSU Chemistry	10
Akron Mechanical Engineering	12
UC Electrical Engineering & Computer Systems	18
UC Physics	23
UC Electrical Engineering & Computer Systems	63
UC Electrical Engineering & Computer Systems	60
UD Electrical & Computer Engineering	13
UD Electrical & Computer Engineering	14
Toledo Chemistry	16

1,279 total students

Why would HPC be necessary for your work?

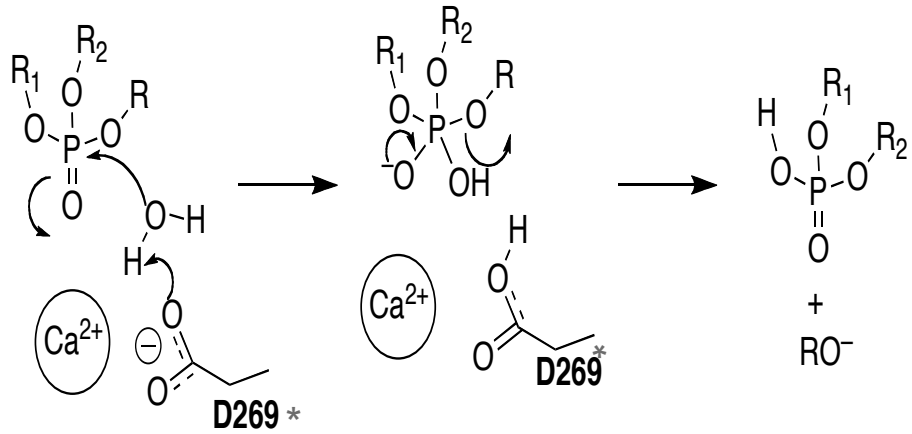
- Your simulations or analyses take too long on your personal computer
- The size of your data is too large to be contained (storage) or accessed (memory) on your computer
- You would like to free up your own system to do other tasks
- You need particular software for your work

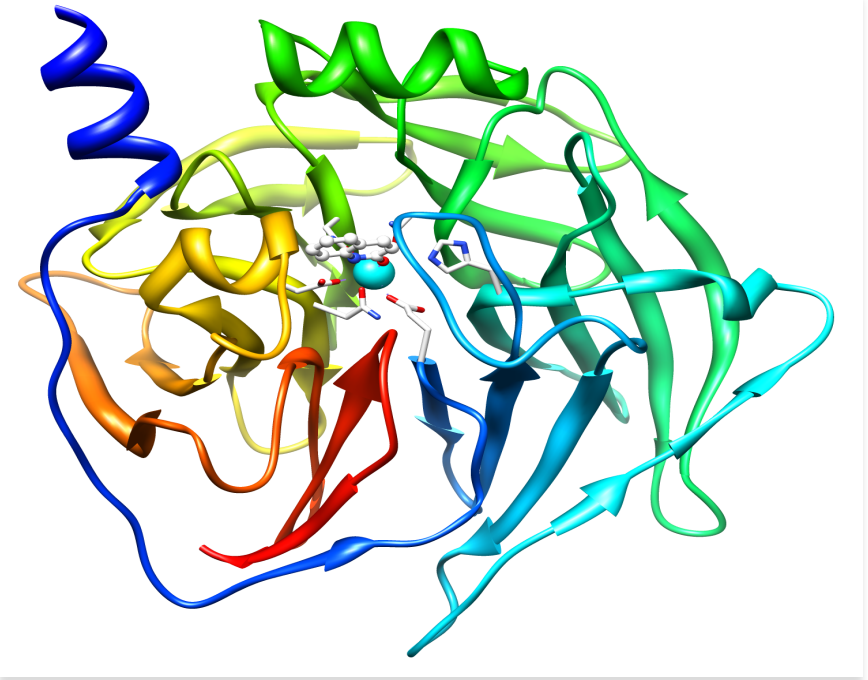
Renewable Energy

Researchers using computational fluid dynamics studies to design new hydropower turbines

PI: Subramania Sritharan, Ph.D., Associate Director for research, Land-Grant Program, Central State University Credit: WKMS

Treating Nerve Agent Exposure





Treating Nerve Agent Exposure

With the power of OSC computing services, a team studies how to capture and destroy organophosphorus nerve agents using modified proteins.

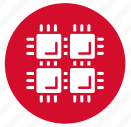
PI: Christopher Hadad, Ohio State University

Research Partnership

"Alone we can do so little; together we can do so much." – Helen Keller

Research Partnership Opportunities

- **Letters of commitment** - sufficient resources needed for your project
- **Boilerplate** text regarding computing and storage **facilities** and **data retention** policy
- **Quotes for specialized services** such as: dedicated computing resources, large amounts of storage, HPC consulting
- **Expertise** in areas such as: scientific software development, web software development and virtual environments by collaborating as co-investigators or senior personnel.
- Opportunities for **Outreach/Broader Impact** activities with OSC's K12 summer educational programs
- **Review of proposals** for research computing infrastructure or research software development



Summer Institute

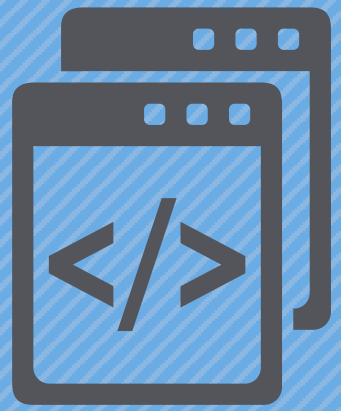
si

- Two-week residential program for high school students
- 16-20 participants each year since 1989
- Students solve complex problems under mentorship of OSC faculty clients

Young Women's Summer Institute

ywsi

- One-week residential program for middle school girls
- 16 participants each year since 2000
- Promotes STEM skills under mentorship of middle school teachers



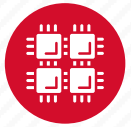
Getting Started at OSC

“If you were plowing a field, which would you rather use? Two strong oxen or 1024 chickens?” - Seymour Cray

Who can get an OSC project?

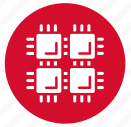
- Academic project
  - Principal investigator (PI) must be a full-time faculty member or research scientist at an Ohio academic institution
  - PI may authorize accounts for students, post-docs, collaborators, etc.
  - Classroom projects are also available
  - No cost** to Ohio academic users
- Commercial projects
  - Commercial organizations may purchase time on OSC systems

** See slide on Fee Structure Transition



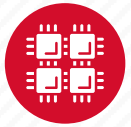
Allocations and Charges

- Charges are in terms of resource units
- Resource units
  - OWENS 1 resource unit (RU) = 10 CPU hours
  - RUBY 1 resource unit (RU) = 20 CPU hours
  - Pitzer 1 resource unit (RU) = 10 CPU Hours
  - CPU hour = walltime x (total # of cores requested)
- Project receives an allocation of RUs
- Jobs are charged to a project



Requesting a New Project-
my.osc.edu

- Startup grant
  - One per PI per lifetime
  - Provide contact info, institution, department
  - 5000 RUs
- Additional allocations for a project
  - Submit a proposal for more RUs
    - Standard: 10,000
    - Major: 30,000
    - Discovery: >30,000
  - Peer-reviewed
  - Grants awarded by Statewide Users Group (SUG)
- Classroom accounts



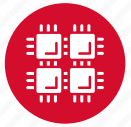
Fee Structure Transition

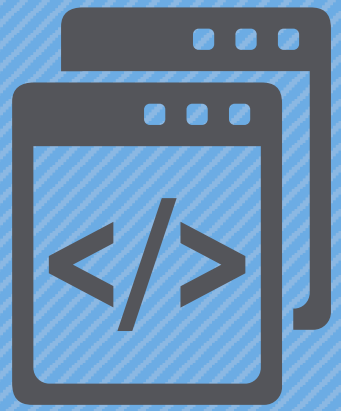
FY19 Plan details

- FY19 rate of $0.075 / RU for cycles > 10K RUs per project; no storage charges
- Projected to provide the $1M needed to cover the OSC budget gap
- FY19 MOUs signed by the 6 biggest universities (OSU, BGSU, OU, UA, CWRU, UC: signing in process)
- OSC will be reaching out to other schools (e.g. Cleveland State) that have low expenses projected ($5K or more for the year).

Implementation

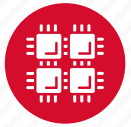
- Initial OSC Pricing Committee meeting March 12th, recent meeting on June 1st
- Implementation issues are being worked out, including an updated allocations process, regular institutional usage reporting, and refund policies
- Discussions are beginning with the OSC pricing committee regarding a revised model for FY20 that will likely include other charges / services





Onboarding new users

“After growing wildly for years, the field of computing appears to be reaching its infancy.” – John Pierce



Beyond Command Line Access

OSC OnDemand ondemand.osc.edu

1: User Interface
- **Web based**
  - Usable from computers, tablets, smartphones
  - Zero installation
- **Single point of entry**
  - User needs three things
    - ondemand.osc.edu
    - OSC Username
    - OSC Password
  - Connected to all resources at OSC

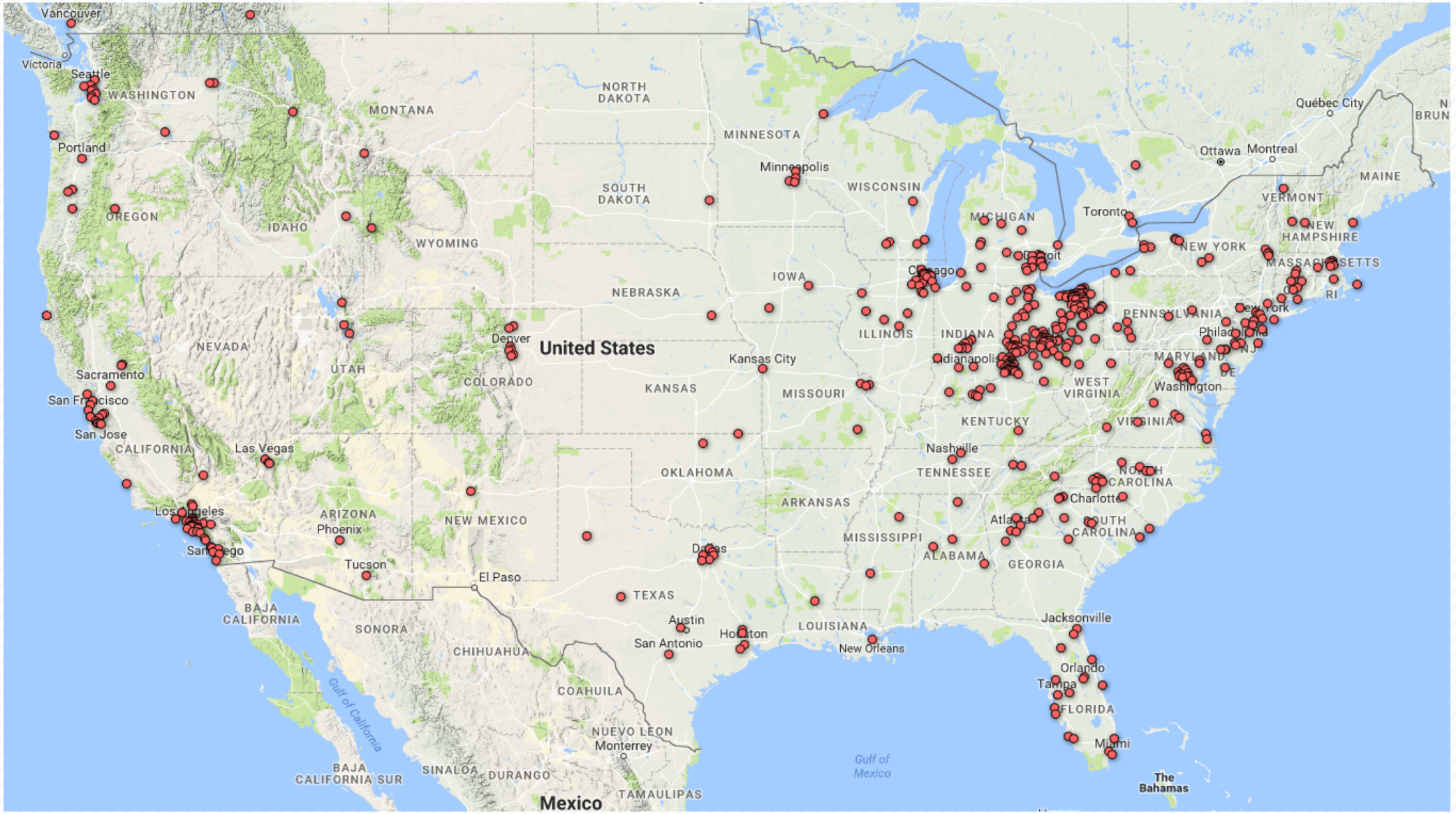
2: Interactive Services
- File Access
- Job Management
- **Visualization** Apps
  - Desktop access
  - Single-click apps (**RStudio, Jupyter, Matlab**, Abaqus, Ansys, Comsol, Paraview)
- Terminal Access

Tutorial available at

osc.edu/ondemand

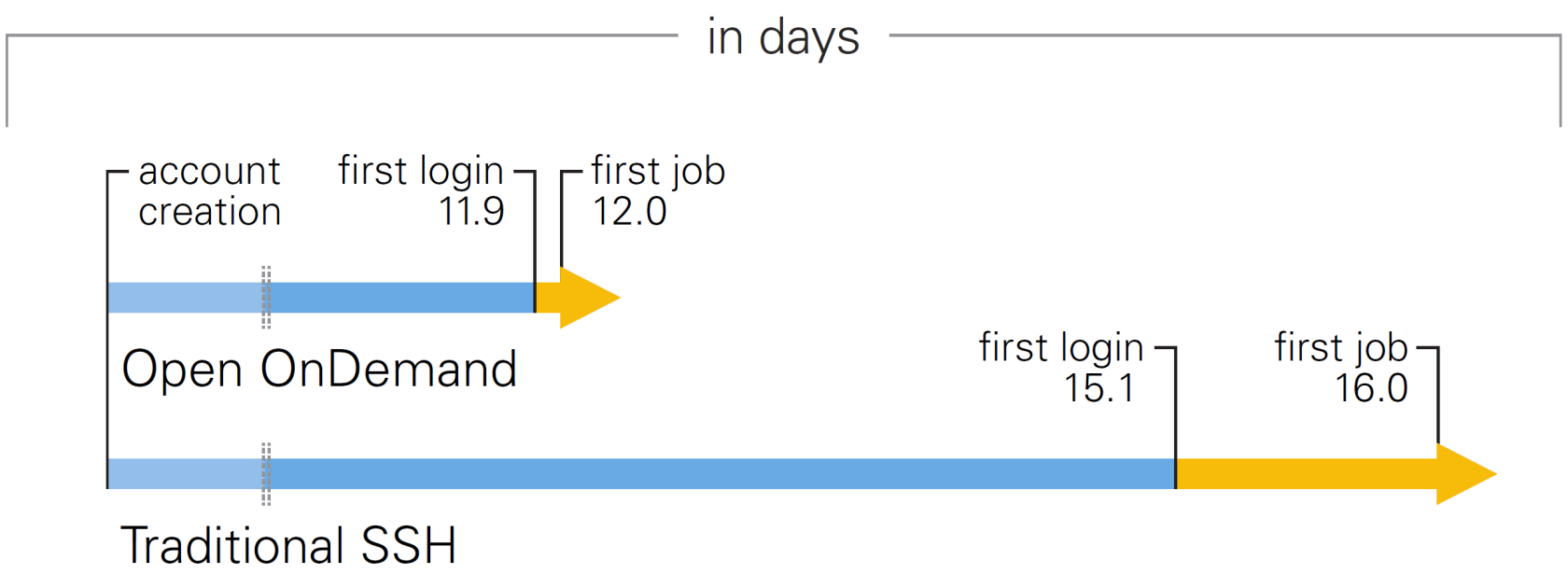
Web-based US Logins

CY2017



Open OnDemand Project

- New OSC OnDemand users start faster than ssh users: first login & job

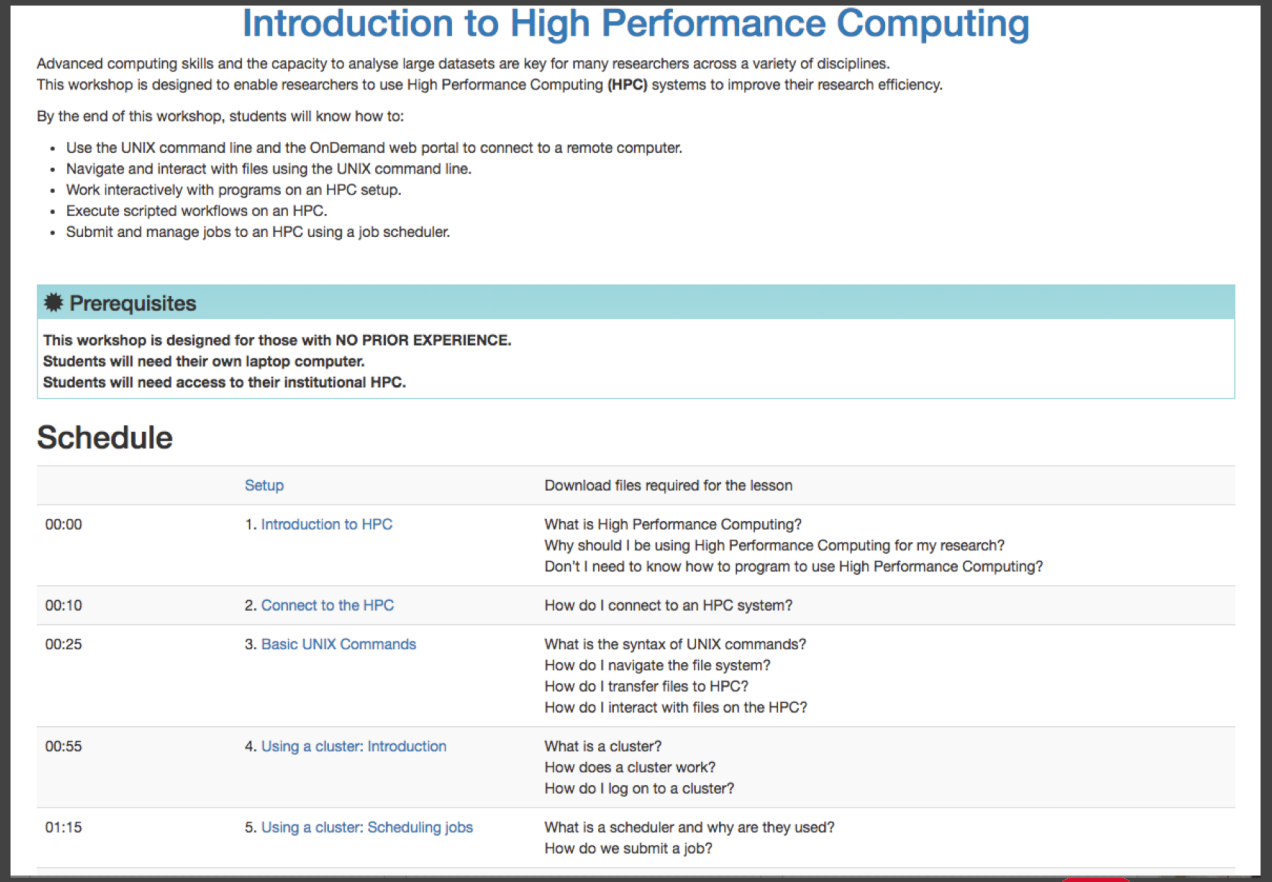


- Open OnDemand currently in use / evaluation at 30+ organizations
- Open OnDemand 2.0: NSF CSSI award, Jan'19 – Dec'22, $3.5M

New Intro to OSC materials

https://khill42.github.io/OSC_IntroHPC/

- Highlights OnDemand for OSC access
- Always available training
- Continually being updated and improved
- Based on HPC/Data carpentry materials



Introduction to High Performance Computing

Advanced computing skills and the capacity to analyse large datasets are key for many researchers across a variety of disciplines.
This workshop is designed to enable researchers to use High Performance Computing **(HPC)** systems to improve their research efficiency.

By the end of this workshop, students will know how to:

- Use the UNIX command line and the OnDemand web portal to connect to a remote computer.
- Navigate and interact with files using the UNIX command line.
- Work interactively with programs on an HPC setup.
- Execute scripted workflows on an HPC.
- Submit and manage jobs to an HPC using a job scheduler.

Prerequisites

This workshop is designed for those with NO PRIOR EXPERIENCE.
Students will need their own laptop computer.
Students will need access to their institutional HPC.

Schedule

	Setup	Download files required for the lesson
00:00	1. Introduction to HPC	What is High Performance Computing? Why should I be using High Performance Computing for my research? Don't I need to know how to program to use High Performance Computing?
00:10	2. Connect to the HPC	How do I connect to an HPC system?
00:25	3. Basic UNIX Commands	What is the syntax of UNIX commands? How do I navigate the file system? How do I transfer files to HPC? How do I interact with files on the HPC?
00:55	4. Using a cluster: Introduction	What is a cluster? How does a cluster work? How do I log on to a cluster?
01:15	5. Using a cluster: Scheduling jobs	What is a scheduler and why are they used? How do we submit a job?

Statewide Users Group (SUG)

- The Statewide Users Group (SUG) is made up of OSC users
  - Provides program and policy advice to OSC
  - Meets twice a year
- Standing committees
  - Allocations
  - Software and Activities
  - Hardware and Operations
- Get involved!
  - Next meeting will be in October

Get your questions answered

FAQs: https://www.osc.edu/resources/getting_started/supercomputing_faq

HOW TOs: https://www.osc.edu/resources/getting_started/howto

New User Guide:
https://www.osc.edu/resources/getting_started/new_user_resource_guide

Updated presentations: https://www.osc.edu/~kcahill/NewUser

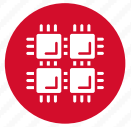
Office Hours and contacting OSC support staff:

go.osu.edu/rc-osc Tuesdays 1-3 p.m. or Weekdays 4-5 p.m. at OSU Pomerene Hall

Email help: oschelp@osc.edu

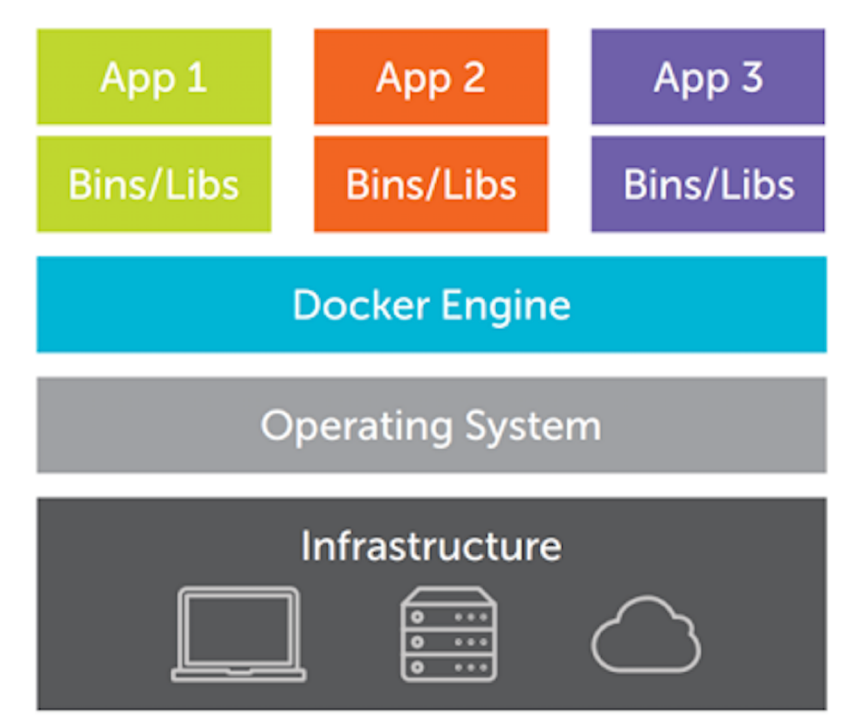
System updates

- Read Message of the Day on login
- Follow @HPCNotices on Twitter



Using Containers at OSC

Containers for Research Computing



- A container image is a lightweight, stand-alone, executable package of a piece of software that includes everything needed to run it: code, runtime, system tools, system libraries, settings.
- Effective at distributing applications and their dependencies

- Docker is common container engine in use today; commonly used to containerize enterprise applications on cloud services

  https://www.docker.com

- Singularity is an emerging leader in research computing (HPC) containers, and has a different approach to security model from Docker

  https://www.sylabs.io

Advantages of Containers

Performance

- Near-native application performance

Freedom

- Bring your own software environment

Reproducibility

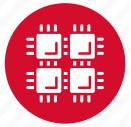
- Package complex software applications into easy to manage, verifiable software units

Compatibility

- Built on open standards available in all major Linux distributions

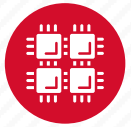
Portability

- Build once, run (almost) anywhere



Containers at OSC

- Both Docker and Singularity are available on all clusters at OSC
- OSC Resources for Getting Started:
  - Singularity software page
  - HOWTO: Use Docker and Singularity Containers at OSC
- Non-OSC references:
  - https://www.sdsc.edu/assets/docs/events/introduction-to-singularity.pdf
  - Singularity user guide: OSC currently using 3.1



OH·TECH

Ohio Technology Consortium

A Division of the **Ohio Department of Higher Education**

info@osc.edu

osc.edu

twitter.com/osc

oh-tech.org/blog

facebook.com/ohiosupercomputercenter

linkedin.com/company/ohio-supercomputer-center