Bring Open Science Pool Capacity to Researchers

Showmic Islam May 22, 2024

OSPool Capacity, May 22, 2024

Who are we?

The OSG Consortium provides computational capacity and services for researchers, faculty, staff, and students at academic, government, and non-profit institutes.

What does this mean for you?

We provide a place (OSPool) to run computations and help you with doing so!



Community Opportunities

Throughput Computing 2024 (July 8-12)

Annual high-throughput computing community conference

https://osg-htc.org/events/throughput-computing-2024/



OSG School 2024 (August 5 - 9)

Weeklong summer school

https://osg-htc.org/school-2024/



OSPool: the Quick Facts



We provide computing capacity for:

- RNA/DNA sequencing, protein binding analyses
- Image analyses
- Machine learning/ artificial intelligence
- Weather, economic, and other systems modeling
- Statistical computations

- Astronomy and black hole research
- Chemical reaction chemistry
- Library sciences, academic classes, seminars, other educational events
- Computer science, architecture analyses ... And SO MUCH MORE!

Who can use your computing resources?

Researchers, faculty, students, and staff affiliated with a **US-academic, government, or non-profit institute** *and* **their collaborators**.



Who can use your computing resources?

Researchers, faculty, students, and staff affiliated with a **US-academic, government, or non-profit institute** *and* **their collaborators**.

What does it cost?

All resources and services are funded by the US government and thus are provided **free-of-charge** (no allocation necessary!).



Who can use your computing resources?

Researchers, faculty, students, and staff affiliated with a **US-academic, government, or non-profit institute** *and* **their collaborators.**

What does it cost?

All resources and services are funded by the US government and thus are provided **free-of-charge** (no allocation necessary!).

What can I use your resources for?

Research, collaboration, data sharing, teaching/lectures, homework assignments, capstone projects, open-source technology development, etc.

Who can use your computing resources?

Researchers, faculty, students, and staff affiliated with a **US-academic, government, or non-profit institute** *and* **their collaborators.**

What does it cost?

All resources and services are funded by the US government and thus are provided **free-of-charge** (no allocation necessary!).

What can I use your resources for?

Research, collaboration, data sharing, teaching/lectures, homework assignments, capstone projects, open-source technology development, etc.

Is there help for me or my students using your resources?

Yes! The Facilitation team provides support, including workflow logistics (software install, data management) and skill building (computing concepts, workflow strategizing, reproducibility).

Is this you?



Researcher

- New data sources
- New computational methods
- Scaling up current work



Instructor

- Adding computing examples to domain curriculum
- Teaching core stats and CS skills



Leader

- Connect with other institutions around computing
- See examples of research at work

Take Advantage of the OSPool Community

Take Advantage of the OSPool Community But, what is the OSPool ... ?

Building Locally

Often, if a campus wants to provide more computing to its researchers, it will invest in a local computing cluster.



Doing More Together

But what if institutions pooled their "extra" resources together in a virtual computing cluster? (and provided access to researchers without a local cluster?)



Open Science Pool (OSPool)

This is exactly the idea behind the OSPool.



Running work





Access Our Services

Request an Account

OSG Portal

No Proposals. No Allocations.

 \equiv

Harness the Capacity of the OSPool

If you are a researcher affiliated with a US Academic Institution, the capacity of the <u>OSPool</u> is available for you to harness, just sign up!

About Yourself

Full Name*

First Last

Institutional Affiliation*

Email (Please use the email address related to your institutional affiliation)*

username@domain.ext

Briefly describe your research or research-related role*

https://portal.osg-htc.org/application



Orientation Meeting

Discuss science and computation details, identify goals and motivation.

Communicate about HTC approaches / OSPool capacity (<u>New User Slides</u>)

Provide specifics for getting started (<u>Roadmap</u>)

End with ongoing support options and account creation.



Open an internet browser and enter: https://notebook.ospool.osg-htc.org

C Jupyterhub	
	Sign in with CILogon

Log into an OSPool Access Point

Login using any of the available authentication options. Some choices:

- ORCID
- GitHub
- Google (Gmail)

CÎ CI	Logon			
	Consent to Attribute Release	~		
	<u>Jupyter for OSG</u> requests access to the following information. If you do not approve this request, do not proceed. Your CILogon user identifier Your email address Your username and affiliation from your identity provider 			
	Select an Identity Provider			
	Log On			

Getting Connected

💭 jupyterhub				
	ClLogon			
Sign in with ClLogo		Consent to Attribute Release		
	Jupyter for OSG requests access to the follo	wing information. If you do not approve		Server Options
	Your CILogon user identifier Your email address Your username and affiliation from you	Ir identity provider	0	Basic Includes basic command-line tools. Includes the HTCondor command-line utilities and Python bindings.
		Select an Identity Provide	0	Data Science
		ORCID - @		Includes libraries for data analysis from the Julia, Python, and R communities. Includes the HTCondor command-line utilities and Python bindings.
		Remember this selection Log On	•	Explore OSPool Notebooks
	By s	electing "Log On", you agree to <u>CILogon's privac</u>		Explore OSPool notebooks by interacting with two built-in tutorials: one for analyzing many independent files (tutorial-fastqc) and one for running many random predictive simulations (tutorial-R).
	For questions ab	out this site, please see the FAQs or send email		
				Start

OSPool Notebooks (Guest Version)

()	File Edit View Run Kernel Ta	bs Settings Help							
	+ 🗈 🛨 C	☑ Launcher ×		°°					
■ ● *	+ • Filter files by name Q • / Name ▲ Last Modified • work a month ago	Image: Console Image: Cons		40°	Tempo Acces	or	Submit jobs Tary Point	Tempo Test P	rary
	Simple 👥 0 🕵 0 🤠		Launch	ner					

OSPool Notebooks (Full Account Version)

Image: Console Image	💭 File Edit View Run Kernel Ta	abs Settings Help	
Image: Start Hold Hold Image: Start Hold Hold Hold Hold Hold Hold Hold Hold	🖿 🕂 🗈 🛨 C	Z Launcher X	φ.
Nome Last Modified Work a moth ago Pyton 3 Bah Pyton 3 Bah Pyton 3 Bah Pyton 3 Bah OSPool Access Point	Filter files by name Q ► /	Notebook	
 work a month age Python 3 Console Console Console B Bah Python 3 B Console Console Consol	E Name A Last Modified		
	work a month ago	Python 3 Console Python 3 Python 3 Source Source Source Terminal Image: Console Image:	Submit jobs OSPool OSPool Access Point
Simple 💽 0 🕼 Launcher	Simple 0 🛐 0 🗇	Launcher	r

Launch "Explore OSPool Notebooks" Notebook

- 1. Click the "Explore OSPool Notebooks" box
- 2. Click orange "Start" button

Server Options

• Basic

Includes basic command-line tools. Includes the HTCondor command-line utilities and Python bindings.

O Data Science

Includes libraries for data analysis from the Julia, Python, and R communities. Includes the HTCondor command-line utilities and Python bindings.

Explore OSPool Notebooks

Explore OSPool notebooks by interacting with two built-in tutorials: one for analyzing many independent files (tutorial-fastqc) and one for running many random predictive simulations (tutorial-R).

Use Case 1: Analyzing Many Files

Example Research Areas

Bioinformatics, MRI Image Analyses, Audio Analyses, Evolutionary Sciences, Neuroscience, Plant Science, & Protein Docking.







Queue Multiple Jobs

Syntax	List of Values	Variable Name
queue N	Integers: 0 through N-1	\$(ProcID)
queue Var matching pattern*	List of values that match the wildcard pattern.	
queue Var in (item1 item2)	List of values within parentheses.	\$(<i>Var</i>) If no variable name is
queue <i>Var</i> from <i>list.txt</i>	List of values from <i>list.txt</i> where each value is on its own line.	provided, default is \$(Item)

Real-World Research Impact: Docking



Rousselene Larson, Utah State University

Research: Identification of plant-based natural compounds that interact with multiple myeloma; drug development

Importance: Multiple myeloma currently no known cure and can appear spontaneously.

Number of OSPool jobs in the last year: **1.7 million** Core Hours: **2.25 million** 40,453,470 total docking runs in one month

Contributing Facility	Core Hours
University of California San Diego	578 K
Syracuse University	446 K
Lancium	264 K
University of Connecticut	151 K
Fermi National Accelerator Laboratory	129 K
University of Wisconsin	129 K
University of Washington	89.2 K
Great Plains Network	79.8 K
University of Notre Dame	72.1 K

Use Case 2: Random Simulations

Example Research Areas

Monte Carlo Simulations, Statistics, Weather Forecasting, Vehicle Safety Research, Risk Assessment, Physics, & Economics.



Real-World Research: Monte Carlo Simulations



Image Credit: Connor Natzke

Connor Natzke, Colorado School of Mines/ TRIUMF

Research: Simulating strong nuclear force, the force that binds the atomic nucleus together

Importance: Improved understanding of one of the four fundamental forces in nature

Number of OSPool jobs in the last year: **312 k** Core Hours: **223 k**

Contributing Facility	Core Hours
University of Chicago	67.2 K
University of Wisconsin	45.8 K
University of California San Diego	20.8 K
Clemson University	12.6 K
University of Michigan	8.91 K
New Mexico State University	5.21 K
Kansas State University	4.93 K
University of Notre Dame	4.71 K
Florida State University	4.71 K

Use Case 3: Different Parameters

Example Research Areas

Machine Learning, Astronomy, Psychology, Economics, Nuclear

Physics, Sensitivity Analyses, & Reaction Chemistry.







Real-World Research: Parameters



Aashish Tripathee, University of Michigan

Research: Search for continuous gravitational waves using LIGO data

Importance: Understand nuclear matter & neutron stars; test for departures from Einstein's theory of relativity.

Number of OSPool jobs in the last year: **603 k** Core Hours: **14.5 million**

Contributing Facility	Core Hours
University of Wisconsin	2.52 Mil
Syracuse University	2.19 Mil
University of California San Diego	1.97 Mil
Lancium	1.64 Mil
Great Plains Network	1.00 Mil
Fermi National Accelerator Laboratory	779 К
University of Chicago	593 K
Villanova University	584 K
American Museum of Natural History	386 K

Overall, what runs well on the OSPool?

Many "Small" Jobs

	Ideal Jobs! (up to 10,000 cores across Jobs, per user!)	Still Very Advantageous!	Less-so, but maybe
Cores (GPUs)	1 (1; non-specific type)	<8 (1; specific GPU type)	>8 (or MPI) (multiple)
Walltime	<10 hrs* *or checkpointable	<20 hrs* *or checkpointable	>20 hrs
RAM	<few gb<="" th=""><th><10s GB</th><th>>10s GB</th></few>	<10s GB	>10s GB
Input	<500 MB	<10 GB	>10 GB
Output	<1 GB	<10 GB	>10 GB
Software	'portable' (pre-compiled binaries, transferable, containerizable, etc.)	most other than \rightarrow	Licensed software; non-Linux

High Throughput Computing (HTC)

One of our favorite HTC analogies: baking the world's largest/longest cake



In computational terms: solving a big problem (the world's longest cake) by executing many small, self-contained tasks (individual cakes) and joining them.

Photos: Arun Sankar via https://www.theguardian.com/world/2020/jan/16/indian-bakers-rise-to-task-of-making-worlds-longest-cake

Workflow Tools

Tools are available to streamline your complex workflow

• DAGMan

(https://portal.osg-htc.org/docum entation/htc_workloads/automat ed_workflows/dagman-workflow s/)

 Pegasus (<u>https://pegasus.isi.edu/</u>)

Matthew Dorsey

North Carolina State University Design new types of magnetic materials from

colloidal building-blocks.



Number of OSPool jobs in last year: 6.5 million

Previous Teaching & Education Events







Distributed Computing at the African School of Fundamental and Applied Physics

Gqeberha, South Africa

African Centers of Excellence in Bioinformatics Global Consortium Meeting

Kampala, Uganda

Great Plains Network Annual Meeting

Kansas City, Missouri

Support instructors and teaching assistants in using OSPool resources for lectures, homework assignments, and workshops

 Use Case 1: Biotechnology professor wanted students to have hands-on experience analyzing genetic data



Utilized OSPool Resources University of Wisconsin system, Colorado School of Mines, Seattle University, and more!

Teaching resources available:

Support instructors and teaching assistants in using OSPool resources for lectures, homework assignments, and workshops

- Use Case 1: Biotechnology professor wanted students to have hands-on experience analyzing genetic data
- Use Case 2: Physics professor wanted to teach a cross-campus course without having to deal with obstacles required to get students local computing accounts



Utilized OSPool Resources University of Wisconsin system, Colorado School of Mines, Seattle University, and more!

Teaching resources available:

Support instructors and teaching assistants in using OSPool resources for lectures, homework assignments, and workshops

- Use Case 1: Biotechnology professor wanted students to have hands-on experience analyzing genetic data
- Use Case 2: Physics professor wanted to teach a cross-campus course without having to deal with obstacles required to get students local computing accounts
- Use Case 3: Statistics professor wanted students to experience working with real data using full size algorithms



Utilized OSPool Resources University of Wisconsin system, Colorado School of Mines, Seattle University, and more!

41

Teaching resources available:

Support instructors and teaching assistants in using OSPool resources for lectures, homework assignments, and workshops

- Use Case 1: Biotechnology professor wanted students to have hands-on experience analyzing genetic data
- Use Case 2: Physics professor wanted to teach a cross-campus course without having to deal with obstacles required to get students local computing accounts
- Use Case 3: Statistics professor wanted students to experience working with real data using full size algorithms
- Use Case 4: Informatics professor hosted a GPU seminar to introduce students to new concepts and hands on experience in machine learning and GPUs



Utilized OSPool ResourcesUniversity of Wisconsin
system,Colorado School of Mines,
Seattle University,
and more!

Teaching resources available:

Expand Your Computational Skills with the OSPool

Twice-a-week Office Hours

Drop-in virtual help sessions are provided twice-a-week, every week.

Research Computing Facilitators are Here to Help!





Christina Koch *Mathematics*

Mats Rynge Computer Science



Showmic Islam Engineering



Rachel Lombardi Biology



Andrew Owen Chemistry

Additional Learning Resources

<u>Synchronous</u>

- Monthly trainings
- OSG User School
- Throughput Week



<u>Asynchronous</u>

- Documentation
- Tutorials
- OSPool Notebooks
- Recordings of events and trainings
 - 160 videos uploaded on OSGaffiliated YouTube channels

Leverage OSPool Services



Aren't sure how HTC can transform your science? Schedule a meeting with a Facilitator at support@osg-htc.org



HTC23 Conference Group Photo

Acknowledgement

This material is based upon work supported by the National Science Foundation under Cooperative Agreement OAC-2030508 as part of the PATh Project. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.