The Cyberinfrastructure Landscape: Systems, Providers, Technologies

Mike Renfro^{1,2}

¹Tennessee Tech University

²Campus Champions Leadership Team

2024-06-25

Who am I? (Who are any of us, really?)

Back in the day

- ME student at a medium-sized public STEM-ish university who should have studied more instead of helping people do things in computer labs.
- Sysadmin/CAD/FEA co-op student at Oak Ridge National Lab before SGI Irix got its cameo in "Jurassic Park" ("It's a Unix system: I know this!").



Figure 1: Some skinny nerd, 1990

Who am I? (Who are any of us, really?)

Now

- Three ME degrees from the now-R2 university (1995, 1998, 2018)
- Mostly-solo practitioner of all things RCD at the same university (2000–2017, 2017–)
- Perpetually online member of multiple RCD organizations (2018–)
- Member of Campus Champions Leadership Team (2022–)
- Compulsive advice-giver



Figure 2: Same nerd, not remotely skinny, 2023

National Science Foundation Department of Energy

ACCESS Allocations Portal

	allocations.access-cl.org	<u>c</u>)	⊕ Å +
ALLOCATIONS SUPPORT OPERATIONS METRI	CS		♠ Q ⊟ Log
Access			
	ne Get Started Available Resour		Policies & How-To Abo
Ho	ne Get Started Available Resour	rces ACCESS Impact P	-01010501100-10 4001
Need access to computing, data You're in the right place! Read more below, or login to	analysis, or storage resourc	ees?	
Need access to computing, data	analysis, or storage resourc	ees?	to get started?
Need access to computing, data You're in the right place! Read more below, or login to	analysis, or storage resourc	Ready I Ready I R costs you nothing an NSF award	



https://allocations.access-ci.org/

National Science Foundation Department of Energy

ACCESS Project Types

Explore for resource evaluation, grad student projects, small classes/training events, benchmarking, development/porting, other small-scale cases Discover for grants with modest needs, Campus Champions, large classes/training events, NSF graduate fellowships, gateway development Accelerate for experienced users with mid-scale needs, multi-grant programs, collaborative projects, growing gateways Maximize largest-scale research activities



National Science Foundation Department of Energy

ACCESS Project Types

Explore, Discover, Accelerate:

- run for grant duration or 12 months
- can be requested any time
- allow multiple projects
 - Explore 400k credits, only an overview required
 - Discover 1.5M credits, 1-page proposal required
- Accelerate 3M credits, 3-page max proposal required, subject to merit review
- Maximize usually only 1 project allowed, 10-page max proposal required, subject to merit review, requests accepted every 6 months



National Science Foundation Department of Energy

ACCESS Resource Catalog

• • • < >	o o		E allocation	s.access-ci.org/resources	٢	(€ 🖞 +
ALLOCATIONS SUPPORT	OPERATIONS	METRICS				ନ ଦ	≡ Login
🍥 🛪 🛛							
		Home	Get Started	Available Resources	ACCESS Impact	Policies & How-To	About
Resource Catalo				adad	d - Noti de las como de del	T = () + -++ -++ -++ -+	
	e below to read m			anded uses. You'll also find	l a link to its user guide	ə. To filter the list to th	e resource
Click on a resource name	e below to read m	ters at the le	ft. ESOURCES IACS at Stony Br	ook Ookami	l a link to its user guide	a. To filter the list to th	e resource
Click on a resource name types you're most interes FILTERS Allocations	e below to read m	ters at the le	ft. ESOURCES	ook Ookami	t a link to its user guide	b. To filter the list to th	e resource



National Science Foundation

Stampede3 at Texas Advanced Computing Center (TACC)

- 20 Intel Sapphire Rapids nodes each with 4 Intel GPUs, 128 GB HBM
- 560 Intel Sapphire Rapids nodes (no GPUs), 128 GB HBM
- 1060 Intel Skylake nodes, 192 GB RAM
- 224 Intel Ice Lake nodes, 256 GB RAM
- ▶ 10 PB VAST (\$SCRATCH) + 1 TB Lustre (\$WORK)
- 100 Gb Omni-Path networking
- Intended for:
 - parallel applications scalable to 10,000+ cores
 - general purpose computing
 - throughput computing



Delta at National Center for Supercomputing Applications (NCSA)

- All using AMD 7763 CPUs
- 132 nodes with 128 cores and 256 GB RAM
- ▶ 100 nodes with 64 cores, 256 GB RAM, and 4 NVIDIA A40 CPUs
- 100 nodes with 64 cores, 256 GB RAM, and 4 NVIDIA A100 GPUs
- 6 nodes with 128 cores, 2048 GB RAM, and 8 NVIDIA A100 GPUs
- 1 node with 128 cores, 2048 GB RAM, and 8 AMD MI100 GPUs
- ▶ 6 PB Lustre for \$HOME and \$SCRATCH
- 200 Gb HPE/Cray Slingshot networking



National Science Foundation

Bridges-2 at Pittsburgh Supercomputing Center (PSC)

- 488 nodes with 128 AMD 7742 cores and 256 GB RAM
- 16 nodes with 128 AMD 7742 cores and 512 GB RAM
- 4 nodes with 96 Intel Cascade Lake cores and 4096 GB RAM
- 24 nodes with 40 Intel Cascade Lake cores, 512 GB RAM, and 8 NVIDIA V100 GPUs (32 GB)
- 9 nodes with 40 Intel Cascade Lake cores, 192 GB RAM, and 8 NVIDIA V100 GPUs (16 GB)
- 1 node with 48 Intel Skylake cores, 1536 GB RAM, and 16 NVIDIA V100 GPUs (32 GB)
- ▶ 15 PB Lustre for \$PR0JECT
- 200 Gb Infiniband networking



National Science Foundation

ACES at Texas A&M University

- 130 nodes, 11888 cores
- Mostly Intel Sapphire Rapids, some Intel Ice Lake, Intel Cascade Lake, and AMD Rome
- Tons of mostly-composable accelerators:
 - GPUs: NVIDIA H100 and A30, Intel (coming soon)
 - ► FPGAs: Bittware Agilex, Intel D5005
 - Coprocessors: NextSilicon
 - Optane memory modules
- Non-composable accelerators:
 - Graphcore IPUs: GC200, Bow-2000
 - NEC Vector Engine: Type 20B-P
- 2.3 PB Lustre
- 200 Gb Infiniband networking



TECH

National Science Foundation

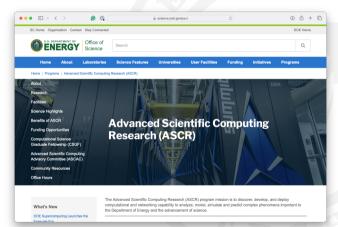
Jetstream2 at Indiana University

- Hybrid-cloud platform for flexible, on-demand, programmable cyberinfrastructure tools
- Interactive virtual machine services
- Infrastructure and orchestration services for research and education
- AMD Milan CPUs (128 per node)
- 360 NVIDIA A100 GPUs
- 512–1024 GB RAM
- 100 Gb Ethernet



National Science Foundation Department of Energy

Advanced Scientific Computing Research (ASCR)





Department of Energy

Accessing ASCR Facilities

- Innovative and Novel Computational Impact on Theory and Experiment (INCITE): multi-year awards for open science using majority of machine at Oak Ridge or Argonne
- ASCR Leadership Computing Challenge (ALCC): 1-year awards for advancing DOE mission or broadening the community capable of using large computing resources at Oak Ridge, Argonne, or NERSC
- Energy Research Computing Allocations Process (ERCAP): 1-year awards for advancing DOE Office of Science and SBIR/STTR mission at NERSC
- Center Reserves: 1-year awards for advancing science and engineering fields at Oak Ridge, Argonne, or NERSC



MATCH Plus MATCH Premier Engagement and Performance Operations Center (EPOC) Science Gateways

Multi-tier Assistance, Training & Computational Help (MATCH) Plus

MATCH Plus:

- takes requests from researchers with a support need,
- identifies a student and mentor that can provide that support,
- connects the researcher to the student and mentor with regular meetings and updates,
- for 5–10 student hours and 2–3 mentor hours per week for 3–6 months,
- at no charge.



MATCH Plus MATCH Premier Engagement and Performance Operations Center (EPOC) Science Gateways

Multi-tier Assistance, Training & Computational Help (MATCH) Premier

MATCH Premier:

- takes requests from already-funded projects,
- identifies an expert consultant and arranges payment,
- ▶ for a 6–12 month period.



MATCH Plus MATCH Premier Engagement and Performance Operations Center (EPOC) Science Gateways

Engagement and Performance Operations Center (EPOC)

EPOC provides researchers with a holistic set of tools and services needed to debug performance issues and enable reliable and robust data transfers. By considering the full end-to-end data movement pipeline, EPOC is uniquely able to support collaborative science, allowing researchers to make the most effective use of shared data, computing, and storage resources to accelerate the discovery process.

– https://epoc.global/



MATCH Plus MATCH Premier Engagement and Performance Operations Center (EPOC) Science Gateways

Science Gateways

. . .

[Science g]ateways are online interfaces that give researchers, educators, and students easy access to shared resources that are otherwise inaccessible or unaffordable for a large segment of the scientific community.

The SGCI was founded to provide services and resources that advance the state of the art in science gateways, that help gateway creators use accepted practices in developing and operating gateways, and that catalyze the formation of a community that may be diverse in discipline but has a common need to advance science through gateways. SGX3



OpenHPC Open XDMoD Open OnDemand Spack

OpenHPC

OpenHPC is a Linux Foundation Collaborative Project whose mission is to provide a reference collection of open-source HPC software components and best practices, lowering barriers to deployment, advancement, and use of modern HPC methods and tools.

OpenHPC components and best practices will enable and accelerate innovation and discoveries by broadening access to state-of-the-art, open-source HPC methods and tools in a consistent environment, supported by a collaborative, worldwide community of HPC users, developers, researchers, administrators, and vendors.

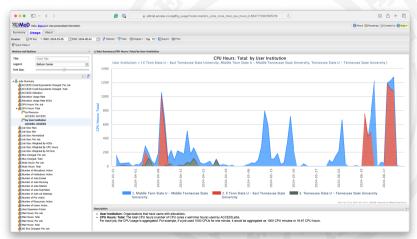
– https://openhpc.community/about-us/





OpenHPC Open XDMoD Open OnDemand Spack

Open XDMoD





OpenHPC Open XDMoD Open OnDemand Spack

Open OnDemand

	• <>		O	J.E.	Irc-ondem	and.lbl.gov C		<u>A</u> 0
BNI	L OnDema	nd Files -	Jobs -	Clusters -	Interactive Apps -	۱ ، ، ، ، ، ، ، ، ، ، ، ، ، ، ، ، ، ، ،	> Develop -	😯 Help 🚽 🚨 🗘 Log 🤇
	OI	PEN				Home / My Interactive Sessions		
					-1			En compilers En OPU En Tutoriais En jugiterituit En contain
		nDe	ema	an			D Jupyter	2-1-numpy-Introduction Last Overkpoint 05/24/201
	-					Interactive Apps	D + H 0	tes tout out table trapes top
	Welc	come to LBN	IL Lawre	ncium S	upercluster.	Servers		
								Plot:
	OnDen	nand provides an	integrated, s	single acces	s point for all of your HPC	🗃 Jupyter Server		Let's import the numpy module.
Activ	ve Jobs	nand provides an	integrated, s	lingle acces	s point for all of your HPC	 Jupyter Server RStudio Server 		Let's import the rampy module. import matplotlib import matplotlib. import many as np
		nano provides an	integrated, s	angle acces	s point for all of your HPC		26 (24)) 26 (25))	<pre>import matplotlib import matplotlib.pyplot as pit import numpy as np # Data for ploting t = np.upmaps(0.2, 2.6, 5.01)</pre>
	ve Jobs		11 User	Account	Time Used Queue Status	RStudio Server	28 (24)) 28 (25))	<pre>import matplotlb import matplotlb.pplot as plt impart rouge as np / data for plotlog t = np.eraspe(0.0, 0.0) = 1 + np.els(1 + np.el + 1)</pre>
	ve Jobs 0 v entries 10 11 29343201 FW.jot		Il User petretto	II Account II It.mp	These Used II Oursus II Status 07:00:18 cm Conversion		36 (24)) 36 (25))	<pre>import matplotlib import matplotlib.pyplot as pit import numpy as np # Data for ploting t = np.upmaps(0.2, 2.6, 5.01)</pre>
	ID II Name 293543241 FW_jot 29054314 dispe		11 User petretto yanianiku	II Account II Ir.mp ac.aome	Time Used Queue Bistus 07.06.14 eff Commer 00.29.67 to Commer	RStudio Server	36 (24)) 36 (25))	impurt metplotlik impurt metplotlik.pyplot as plt impurt mey any d Data for plotling t - sparamge(r, s), s, s, s), s - 1 - sparamge(r, s), s, s), s - 1 - sparamge(r, s), s), (s, ss - s), respectively), as, plot(t, s) (cmasplotlik.lines.line2) at 0x2041c090e6900
	No II Name 0 III Name 293943261 FW_J00 29305414 dispe 29305319 AlEn.7		Il User petretto	II Account II It.mp	These Used II Oursus II Status 07:00:18 cm Conversion	RStudio Server Interactive Apps [Sandbox] Servers	36 (24)) 36 (25))	import metplotlik import metplotlik-pyplot me plt import meny metplotlik # Data for plotlik # or separamped (0,2,0,0,10) % = 1 * (0,2,10,10) % = 1 * (0,2,10,10)\\% = 1
Show 5	KP JODEs 0 entries 80 II 29943261 PW_JM 29054314 dispe 2905531 AIEA.7 2905555 AIEA.7	5	II User petreto yanlanitu riporter	II Account II Ir.mp ac.acme alice	Time Used Open II Status 67:36:56 crt Communic 00:35:57 IC Communic 44:37:67 alice Communic	RStudio Server Interactive Apps [Sandbox]	36 (24)) 36 (25))	import satgletlik import satgletlik pipet as pit import sange as as final as a satgletlik pipet as pit import sange as a satgletlik pipet as a figure spitality spit as a figure spitality as an and a satgletlik import satgletlik likes.LikeB at SatBletlikhedHor
Show 5	XP JODUS 10 entrilion 20042201 PHC.jati 2005201 PHC.jati 20050319 AHEN.7 20050303 AHEN.7 20050203 AHEN.7	5 984.115 984.115	User petrato yantaniu riporter riporter	II Account II Ir.mp ac.aone alce alce	Time Used II Come II Ethin (2:26:54 off Common Common (0:22:37 IoI Common Common (0:22:37 IoI Common Common (0:25:37 IoI Common Common (0:25:37 IoI Common Common (0:25:37 IoI Common Common (0:25:37 IoI Common Common	RStudio Server Interactive Appe [Sandbox] Servers MATLAB	36 (24)) 36 (25))	Appert Matgle(1) Speet Matgle(2) Speet Matgle(2) 1 - Speet Matgle(2)
Show 5	Ke Jowenstee 0 entities 80 II Name 29562301 FW Jail depe 29505319 Alber, 7 Alber, 7 29505310 Alber, 7 Alber, 7 29505310 Alber, 7 Alber, 7 2950414 Alber, 7 Alber, 7 295041210 13,001 28977201	9 984.115 984.115 984.115 986.0002_protonaled.in.sh #	Uter petrato yanianiu yanianiu yanianiu yooter yooter jooter joter kitacke	II Account II Ir.mp allos allos allos mhg nano	Time Used B Canase B States 4/3/26 4 1 Convert 4/3/26 4/6 Convert Convert	RStudio Server Interactive Apps [Sandbox] Servers	36 (24)) 36 (25))	Henry weight and a split of the
Show 5	No Image: Second S	9 984.115 984.115 984.115 9.00032 protocoles in uh	User petrato yoster riporter riporter joster	II Account II Ic.mp acc.some alice alice alice mitig	Time Used II Outer II Default 07.06.51 off Control Control 04.05.07 oil Control Control 44.05.09 dires Control Control 4.05.04.01 dires Control Control 4.05.04.01 dires Control Control 6.05.01 dires Control Control	RStudio Server Interactive Appe [Sandbox] Servers MATLAB	36 (24)) 36 (25))	Never any set of the s



OpenHPC Open XDMoD Open OnDemand **Spack**

Spack

- Spack is a package management tool designed to support multiple versions and configurations of software on a wide variety of platforms and environments.
- It was designed for large supercomputing centers, where many users and application teams share common installations of software on clusters with exotic architectures, using libraries that do not have a standard ABI.
- Spack is non-destructive: installing a new version does not break existing installations, so many configurations can coexist on the same system.
- https://spack.readthedocs.io/

