The Cyberinfrastructure Landscape: Systems, Providers, Technologies

Mike Renfro^{1,2}

¹Tennessee Tech University

²Campus Champions Leadership Team

2025-06-23

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–), CaRCC Emerging Centers Steering Committee (2024–)
- Compulsive advice-giver



Figure 2: Same nerd, not remotely skinny, 2023

Computational and/or Storage Resources Services/Consultancies

Tools and Software

National Science Foundation Department of Energy Joint NSF, other federal agencies and non-governmental partners

ACCESS Allocations Portal

ALLOCATIONS SUPPORT OPERATIONS METRIC	28		♠ Q ≡	Log
Access				
Hor	ne Get Started Available Resources	ACCESS Impact	Policies & How-To	Abou
Need access to computing, data You're in the right place! Read more below, or login to What is an allocation?	analysis, or storage resources? get started. Which resources?	Rea	dy to got <mark>startod?</mark>	

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



National Science Foundation Department of Energy Joint NSF, other federal agencies and non-governmental partners

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 Joint NSF, other federal agencies and non-governmental partners

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



Computational and/or Storage Resources Services/Consultancies

Tools and Software

National Science Foundation Department of Energy Joint NSF, other federal agencies and non-governmental partners

ACCESS Resource Catalog





National Science Foundation Department of Energy Joint NSF, other federal agencies and non-governmental partners

New: DeltaAI at National Center for Supercomputing Applications (NCSA)

- 152 NVIDIA Grace-Hopper nodes, with 288 cores and 4 NVIDIA H100 GPUs (96 GB) each
- Superchip architecture
- ► 6 PB Lustre for \$HOME and \$SCRATCH
- 200 Gb HPE/Cray Slingshot networking



 Computational and/or Storage Resources
 National Science Foundation

 Services/Consultancies
 Department of Energy

 Tools and Software
 Joint NSF, other federal agencies and non-governmental partner

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
- Same storage and networking as DeltaAl



Computational and/or Storage Resources National Science Foundation Services/Consultancies Department of Energy Tools and Software Joint NSF, other federal agencies and non-governmental partners

Stampede3 at Texas Advanced Computing Center (TACC)

- New: 24 Sapphire Rapids nodes with 1024 GB RAM, and 4 NVIDIA H100 (96 GB) GPUs
- 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



 Resources
 National Science Foundation

 sultancies
 Department of Energy

 J Software
 Joint NSF, other federal agencies and non-governmental partne

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
- New: 10 nodes with 104 Intel Sapphire Rapids cores, 2048 GB RAM, and 8 NVIDIA H100 GPUs (80 GB)
- 31 nodes with 40 Intel Cascade Lake cores, 102–512 GB RAM, and 8 NVIDIA V100 GPUs (16–32 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 Department of Energy Joint NSF, other federal agencies and non-governmental partners

Derecho at National Center for Atmospheric Research (NCAR)

- All using AMD 7763 CPUs
- 2488 nodes with 128 cores and 256 GB RAM
- 82 nodes with 64 cores, 512 GB RAM, and 4 NVIDIA A100 GPUs (40 GB)
- 200 Gb HPE/Cray Slingshot networking



National Science Foundation Department of Energy Joint NSF, other federal agencies and non-governmental partners

Anvil at Purdue University

- Supporting CPU, GPU simulation and large memory simulation
- "Composable Subsystem" offers Kubernetes support for science gateways and other workloads
- All using dual AMD 7763 CPUs (128 cores)
- 1000 CPU nodes with 256 GB RAM, 32 with 1024 GB RAM
- 16 GPU nodes with 512 GB RAM and 4 NVIDIA A100 GPUs



National Science Foundation Department of Energy Joint NSF, other federal agencies and non-governmental partners

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

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



National Science Foundation Department of Energy Joint NSF, other federal agencies and non-governmental partners

Advanced Scientific Computing Research (ASCR)





Mike Renfro The Cyberinfrastructure Landscape: Systems, Providers, Technologies 16 / 26

National Science Foundation Department of Energy Joint NSF, other federal agencies and non-governmental partners

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



National Artificial Intelligence Research Resource (NAIRR) Pilot

Early concept for a shared national research infrastructure connecting US researchers to:

- computational and AI,
- 🕨 data,
- software,
- training, and
- educational

resources. Many of the federally-funded resources are available through ACCESS or other routes, but NAIRR also facilitates access to commercial resources including AI models, inference services, and software as a service offerings.



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

		X	0		linc-ond	emand.lbl.gov 🔿			4 Ø
BNL O	Demand	Files -	Jobs - (Clusters -	Interactive Apps -	a 4	/> Develop -	🕑 Help - 🚨	G+ Log O
	OPE	N				Home / My Interactive Sessions			
					-1		En sec En en	B Compilers ES OPU ES Tutorians E	3 julytetub ES containe
	D	De	m	an	a	Interactive Apps	File Edit	View Insert Cell Kernel	Widgets Help Markdown V B
	Welcom	ie to LBN	L Lawren	icium S	upercluster.	Servers			
	OnDemand	provides an	integrated, si		e point for all of your LIDC	Therease Conver		Plot:	
		promoto an	integrated, ai	ngle acces	s point for all of your HPC	Jupyter Server		Let's import the sumpy module.	
ctive .	Jobs	protition art	integrated, si	ngie acces	s point for all of your HPC	 Supprer Server RStudio Server 	3n (24)	Let's import the numpy module. import matplotlib import matplotlib.pyplot as import numpy as mp	plt
	Jobs entries	, pro 11000 011	ntogratod, an	ngie acces	s point for all of your HPC	 Supyter Server RStudio Server 	In (24)	Let's import the numpy module. i import matplotlik import matplotlik import matplotlik.pplot as import numpy as no i f data for plotling t = np.exampt(0.0, 2.0, 0.0)	րեւ
tow 50 v	Jobs entries		I User I	Account II	Time Used II: Ourse II: Status	G RStudio Server Interactive Apps [Sandbox]	3n (24) 3n (25)	Let's import the inamy: module. i import matplotlib import matplotlib.pyplot as import mompy as np i d data for plotting t = np.arange(0,0,2.0, 0.0) s = 1 + np.sin(2 + np	pit }
xow 50 - 4	Jobs entries II Name 201 PW_job 1014 dispc	,	II User I patrito yanianiu	I Account II	Time Used Ourse Status 07.06.16 cm 002257 10 Common	RStudio Server Interactive Apps [Sandbox]	In (24)	Let's import the issumpy module. i Support morpholic lib Support morpholic pupplet as Support morpholic pupplet t = pupplet pupplet figs, as = pit.excepter() as.piel(t, a) figs.compared by library at	pix }
Active .	II Name 2011 FW, Job 814 disp: 2019 AlEx.7904.115	,	II User I petreto yanianiu riporter	Account II k_mp ac_acre aice	The Used Game Subar 520-04 of Subar 6220-14 of Subar 6220-14 Compared 6220-14 Compared 6220	Supprer server Retudio Server Interactive Apps [Sandbox] Servers	In (24) In (25) Out(25)	Let's import the issuey: module. i deport matpletlib deport matpletlib deport maps i / dear for ploting to - mp.eracego(-0, 2.0, 0.0) n - 1 + sp.stn(1 + sp.p.k + 1 fig. ax = ploting ax.plot(x, s) i (emephetlib-lises.time2 at 200	pit
xow 50 0 0	Image: Name Name 201 FW, Job 814 disp: 019 AlEx.7904.115 255 AlEx.7904.115		II User I petreto yanlantiu riporter riporter	Account II K.mp acc.acm aice aice	Spoint for all of your HPC Time Used II Oaxe II Setue 620.04 ett Connel Connel 620.05 ett Connel Connel 620.07 ette Connel Connel 630.06 ette Connel Connel	Supprer Server Studio Server Interactive Apps [Sandbox] Servers MATLAB	2n (24) 2n (25) Out(25)	Let's import the issuegy module.	plt } 0x2b81cb93e4500)
xow 80 - 4 xow 80 - 4 > 290483 > 29058 > 29058 > 29058 > 29058	Image: Name III Name 2011 PR/300 1814 dispe 2019 AEEx7064.115 2020 AEEx7064.115		II User I putetto yantaritu riporter riporter	Account II K.mp ac.acme alice alice alice	Tere Used II Outer III Status 620.04 off Execute Constat Constat 620.04 off Execute Constat Constat 643.07 des Execute Constat Constat 443.070 des Execute Constat Constat 430.040 des Execute Constat Constat	Charactive Apps [Sandbox] Bervers MatLAB	36 (24) 36 (25) 001(25)	Lethingoithe sumpy module.	11q (00000000000000000000000000000000000
Active 100 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	Image Name 11 Name 1281 FW_Job 1814 dispc 1919 AREn.7964.115 2020 AREn.7964.115 2020 AREn.7964.115 2020 AREn.7964.115	protocated in sh	II User I potetto yacitatiu igooter igooter igooter jotarg	Account II K.mp accome alice alice mtg	Three lated Occurs II Status 0738.64 0738.64 01 Control 0738.65 071 Control Control 0738.65 071 Control Control 0738.65 071 Control Control 0738.65 071 Control Control 0738.65 070 Control Control 0738.65 070 Control Control 0738.65 070 Control Control 0738.65 070 Control Control	Opymr server Totractive Apps (Sandbox) Server MatLAB Sonk Jupyter Server	36 (24) 36 (35) 06(25)	Lat's report the 'singly' module.	pit
Active . 100% 60 0 2005 20	Image: Name Name 201 PW/Jeb 201 PW/Jeb 201 Adfo.7964.115 202 Adfo.7964.115 203 Adfo.7964.115 204 Adfo.7964.115 205 Adfo.7964.115 206 Adfo.7964.115 207 Adfo.7964.115 208 Adfo.7964.115	protocated in sh	II User I potetto yaritettu igonter igonter jotarg kifuska	Account II K.mp accome alice alice mng nano	These blood O state If Status (2020)11 (2011)1 (2011)1 (2011)1 (2020)13 (2111)1 (2011)1 (2011)1 (2020)13 (2011)1 (2011)1 (2011)1 (2020)14 (2011)1 (2011)1 (2011)1 (2020)15 (2011)1 (2011)1 (2011)1 (2020)15 (2011)1 (2011)1 (2011)1 (2020)15 (2011)1 (2011)1 (2011)1	Charactive Apps [Sandbox] Interactive Apps [Sandbox] Bervers MATLAB Sandbox] Matlandbox Matland	36 (24) 36 (35) 06(25)	Lativeporthe support nodes.	pit) 0422441c093444500;
Active . 10 10 10 10 10 10 10 10 10 10 10 10 10	Image: Name Name 201 FW, Joh 814 disp: 203 ARE, 7964, 115 204 ARE, 7964, 115 203 ARE, 7964, 115 204 ARE, 7964, 115 207 Start, 81	protonabed in sh	II User I petetta yaniaciu ripoter ripoter jalang kyhanke sahaliken	Account k_mp ac_some alice alice ming nano nano	The bad One Itela 0'2014 0'2014 0'2014 0'2014 0'1 0'2014 0'2014 0'1 0'2014 0'2014 0'1 0'2014 0'2014 0'1 0'2014 0'2014 0'1 0'2014 0'2014 0'1 0'2014 0'2014 0'1 0'2014 0'2014 0'1 0'2014 0'2014 0'14 0'2014	Charactive Apps [Sandbox] Interactive Apps [Sandbox] Bervers MATLAB Server MATLAB MATLAB Server MATLAB Server MATLAB Server MATLAB Server MATLAB MATLAB Server MATLAB MATLAB MATLAB Server MATLAB Server MATLAB Server MATLAB Server MATLAB Server MATLAB Server MATLAB MATLAB Server MATLA	30, [24] 30, [25]	Let's report the 'suppy' module.	214 (CONSERVINGENCE)



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/

