High-Performance Computing (Resources and Capabilities)

https://github.com/wsphd/csun-hpc/

"High-Performance Computing (Resources and Capabilities)
California State University, Northirdge (CSUN)

Friday, April 11, 2025
CSUN Faculty Retreat - Odyssey Restaurant

Zack Hillbruner, *Information Technology*, zack.hillbruner@csun.edu

Wayne Smith, Ph.D., *Department of Management*, ws@csun.edu

Introduction/Background/Motivation

Introduction/Background/Motivation

- Some n i N needs are ≤ contemporary desktop/laptop and software
  - But double-check new methodologies and growth (and by extension, movement) of data
- Some n j N needs are > contemporary desktop/laptop and software
  - Essentially, "compute-intensive, data-intensive, or network-intensive"
  - Use primarily FOSS (Linux, Open Source, etc.) to complement COTS (Windows, SPSS, etc.)
- Private, "on-premises" servers
  - Usually purchased by an individual faculty member or Dept. (often with a grant or project)
  - Usually located in the on-campus CSUN MDF
  - CSUN IT usually racks and networks the system; Users manage the system and applications
- Public Cloud (AWS, GCP, MS-Azure, OCI, Digital Ocean, etc.)
  - Use "free-tier" (still need to provide a credit card)
  - Purchase credits with a credit card
  - Scholars can ask for resources for substantive reseearch
- Or?
  - NSF-funded, multi-year, inter-institution, STEAM/SocialSTEM, R3s/CCC's too
  - CSUN IT Technology Resources for Research

General Advanced Computing/Data Management

General Advanced Computing/Data Management

- There are plenty of (non-HPC) advanced computing issues too (research and instruction).
- Ecosystem Transition: Compute
    - COTS languages (e.g., SPSS, Stata, MPlus, Matlab) -> FOSS languages (e.g., R, Python, Julia)
    - COTS spreadsheets (e.g., Excel) -> FOSS spreadsheets(e.g., LibreOffice).
    - Operating Systems (e.g., Windows/MacOs) *plus* Linux, Excel -> LibreOffice, etc.
    - Beyond replication -> Reproducibility (not just 'A' journals)
- Ecosystem Transition: Data
    - "Big Data"
    - research results can include output data (and perhaps even source data) too
- Ecosystem Transition: Network
    - "Big Data"
    - research results can include output data (and perhaps even source data) too
- Example: Technology Trends
    - Campus Labs *plus* Home Labs, Open Science, Open Research, Open Data, Open anything...
- I'm happy to discuss these issues too but it's not the focus on this material.

Jetstream2

Jetstream2

- Managed by Indiana University
- 100's of GiB of RAM, 10's of PB of disk, 10's of GPUs, fast networks
    - Best for new learners, data science projects (R, Python, Julia, etc.), large simulations, gateway to other systems, including several supercomputers around the country
- Need an "ACCESS ID"
    - Like an ORCID ID but for Research Computing
    - Have CV or Resume for upload (don't worry, your request will be approved)
- Be willing to learn:
    - How to ask (nicely and well, for more (incrementally) resources, and read a simple dashboard
    - the Command line and Linux
    - Webshell
    - SSH for logging int (and some learning curve for generating SSH keys and passphrases)
    - SCP for file transfer (after the SSH process is done)
- (Live demo)
- Start here:
    - [Jetstream ACCESS page]https://jetstream-cloud.org/get-started/index.html

Nautilus/NRP

Nautilus/NRP

- Managed by University of California, San Diego
- 100's of GiB of RAM, 10's of PB of disk, 100's of GPUs/FPGAs/TPUs/DPUs, very fast networks
    - Best for leading-edge science and engineering, especially w/ funded labs and staff
- Be willing to learn:
    - Must be comfortable with the Command Line, Open Source, and Linux
    - Kubernetes (open source client-server), you use the "kubectl" binary
    - You control just about everyting with ASCII ".yaml" files
- (Static demo)
- Start here:
    - Send Wayne an email - ws@csun.edu

Additional Resources (at no charge)

Additional Resources

- Sometimes, researchers just need an unmanaged or managed (by students, supervised by faculty) resource to host public-facing files and applications
    - Oregon State University Open Source Lab (OSL)
- Recently, CSUN was added to the Cloudbank/2i2C JupyterHub resource (this complements CSUN Apporto and SDSU/CSUSB TIDE)
    - Cal-ICOR JupyterHub Pilot
- Increasingly, Libretexts is moving beyond "texts" and becoming a complete LMS solution, including a JupyterHub resource
    - General System - LibreTexts
    - Specific Application - JupyterHub
- Some researchers want to experiment with real Quantum resources
    - D-Wave LEAP Quantum Launchpad/D-Wave Learn Program (D-Wave)
- Many researchers require an AI system that *is* open, transparent, and reproducible (built *top-down*)
    - NSF National Artificial Intelligence Research Resource Pilot (NAIIR)
- Some researchers desire an AI system that *is* open, transparent, and reproducible (built *bottom-up*)
    - Non-Profit Personal AI Lab (Kwaai)

Conferences/Fellowships

Conferences/Fellowships

- There are plenty of zero-cost and low-cost U.S. domestic events for learning about HPC resources at the *Application*-level.

Name	Venue	Cost	Timeframe
Practice & Experience in Advanced Research Computing (PEARC)	varies	mid $	late July
Science Gateways (SGX3)	varies	$0 (NSF)	varies
Confab (DOE)	varies	low $	early April
Institute for Mathematical and Statistical Innovation (IMSI)	varies	$0 (NSF)	varies

- There are plenty of zero-cost and low-cost U.S. domestic events for learning about HPC resources at the *Infrastructure*-level.

Name	Venue	Cost	Timeframe
Research Computing at Smaller Institutions (RCSI)	Swarthmore, PA	$0 (NSF)	early June
National Research Platform (NRP)	UCSD, CA	$600	late January
Supercomputing (SC)	St. Louis, Denver, Atlanta	low $	mid November
Corporation for Networking and Research (CENIC)	varies	low $	late March
Southern California Linux Expo (SCaLE)	Pasadena, CA	low $	early March

- And the list of *International* events for learning about HPR resources is growing quickly.

Name	Venue	Cost	Timeframe
CINI HPC Summer School (CINI)	Naples, Italy	N/A	mid June

- The following are some of the Fellowships available:
  - ICICLE: Intelligent CI with Computational Learning in the Environment (ICICLE)

National Workshops

Natonal Workshops

- There are plenty of *in-person* events for learning about HPR resources.

Name	Venue	Cost	Timeframe
Open Science Grig (OSG)	U of Wisconsin-Madion, WI	$0 (NSF)	late June
HPC and Data Science Summer Institute (SDSC)	UCSD, CA	$350	late July - early August

Name	Venue	Cost	Timeframe
NERSC International HPC Summer School (NERSC)	varies	$0 (DOE)	early July
KNIT (FABRIC)	varies	$0 (NSF)	mid March

- There are plenty of *virtual* events for learning about HPR resources.

Name	Venue	Cost	Timeframe
OU Supercomputing Center for Education & Research (OU)	virtual, synchronous	$0 (NSF)	late June
HPC Pathways (NCSA)	virtual, asynchronous	$0	on-going
Cornell Roadmaps	virtual	$0	asynchronous, on-going
HPC Carpentry	in-person and virtual, synchronous	$0	varies
(comprehensive, searchable list of resources	N/A	$0	varies

Upskilling - Professional Associations/Societies

Upskilling - Professional Associations/Societies

- Faculty - These HPC resources should be of use to *Faculty* over time.
  - R OpenSci (ROpenSci)
  - PyOpenSci (pyOpenSci)
  - JuliaCon (annual Summer conference abstracts, proceedings)
  - Framework for Open and Reproducible Research Training (FORRT)
  - Open Accelerated Computing (OpenACC) (C/C++ optimizations for research, annual Summer conference)
  - (and check your discipline's pre-conference workshops and related conference themes for HPC events)
- Staff - These HPC resources should be of use to *Staff* over time.
  - US Research Software Engineering Association (US-RSE)
  - Campus Research Computing Consortium (CaRCC)
  - Campus Champions
  - OpenOnDemand
  - Internet2 Research Engagement
  - EduCause Research Computing and Data Community Group
- Administration - These HPC resources should be of use to *Administration* over time.
  - Coalition for Academic Scientific Computing (CASC)
- Sundry - These HPC resources should be of use to various individuals over time.
  - ES NET (DOE)
  - The Quilt
  - Fabric

- Sundry - These open source resources should be of use to various individuals over time.
    - UC Open Source Program Offices)
    - Professional Development for Instructors Interested in Student Participation in Humanitarian Free and Open Source Software (POSSE)