Large-scale computing resources at CU Boulder

Jonathon Anderson and Shelley Knuth

rc-help@colorado.edu

www.colorado.edu/rc

Slides:
https://github.com/ResearchComputing/CEASE_Webinar_2021

Outline

- Research Computing
  - Resources
  - User facing services and support
- CRDDS
  - Services and support
- Rocky Mountain Advanced Computing Consortium
  - Leveraging this collaboration
- Procurement of our new computing system

What is Research Computing?

- Provide services for researchers that include:
  - Large scale computing
  - Data storage
  - High speed data transfer
  - Data management support
  - Consulting
  - Training

- We are likely best known for:
  - Summit Supercomputer
  - PetaLibrary storage

What Would I Use Summit For?

- Research Computing is more than just Summit
- What would you use Summit For?
    - Solving large problems that require more:
        - Memory than you have on your personal computer
        - Cores/nodes/power than you have on your personal computer
    - High performance GPU computing
    - High memory jobs
    - Visualization rendering
- Not a place for:
    - Large data storage

RMACC Summit

- 67% CU, 23% CSU, 10% RMACC

- 500+ compute nodes
  - Majority 24 cores / node "Haswell" (shas) nodes

- 13,000+ cpu cores

- Intel Omni-Path Architecture interconnect

- GPFS scratch file system
  - 1.2 PB

RMACC Summit node types

- 10 Nvidia K80 GPU nodes (sgpu)
  - 2 Nvidia Tesla K80 accelerators / node

- 5 "high-memory" nodes (smem)
  - 2 TB of memory / node
  - 48 cores / node

- 20 Intel Xeon "Knights Landing" Phi nodes (sknl)
  - 68 cores / node

Research Computing
UNIVERSITY OF COLORADO **BOULDER**

Be Boulder.

Blanca condo compute cluster

- Aggregate cluster from researcher capital purchase contributions
- Pervasively heterogeneous
- 200+ nodes and almost 3,400 CPU cores
- 30 distinct contributors
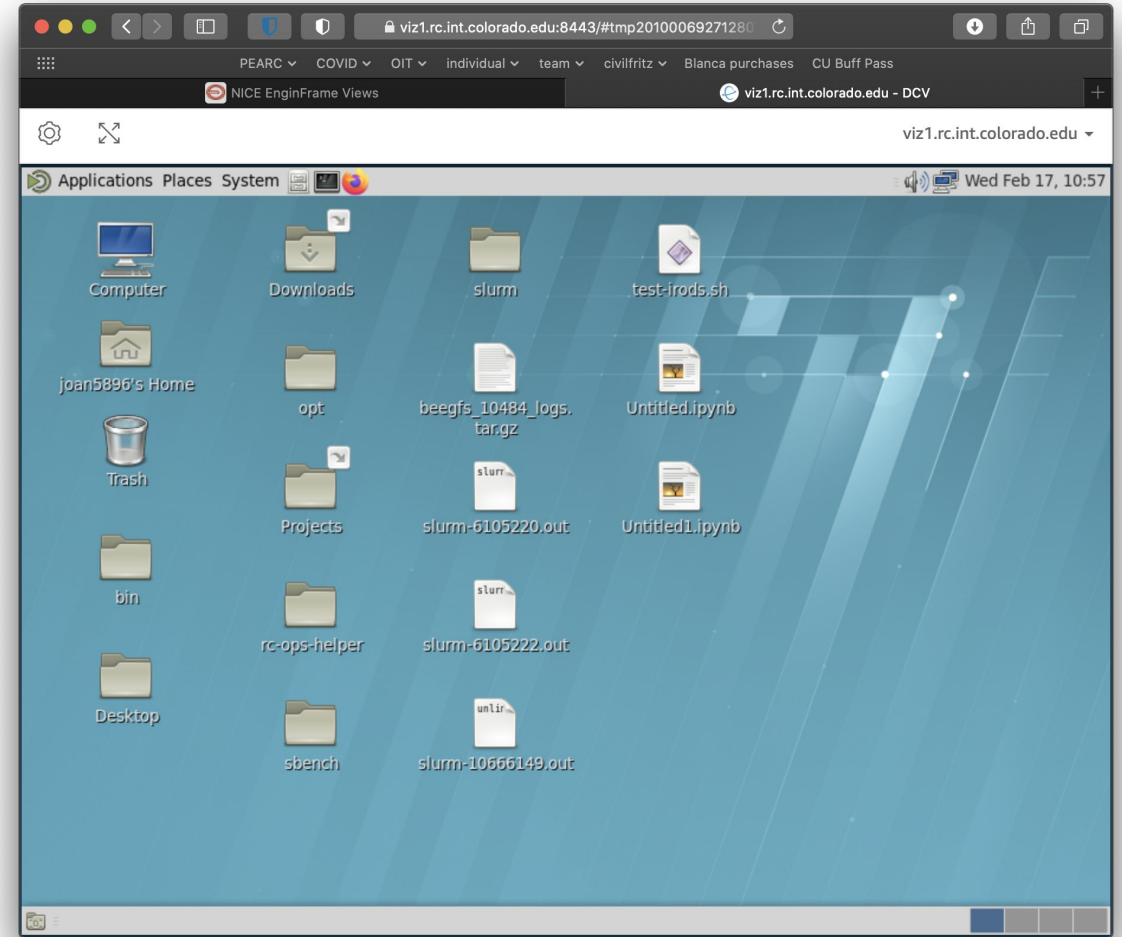- Oldest node 2012; newest node 2021
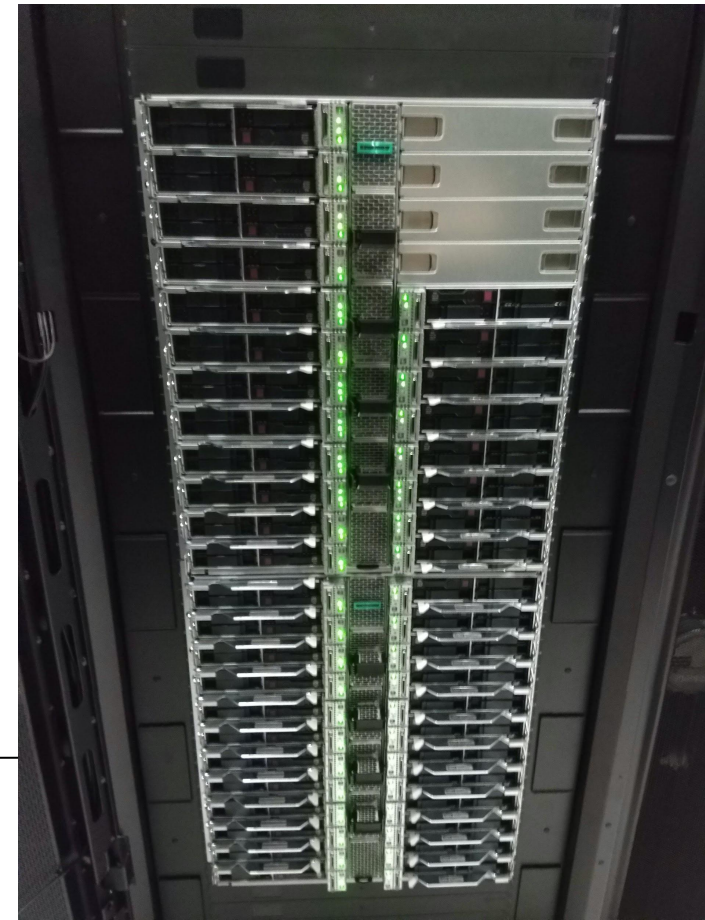
PetaLibrary

- Active storage on ZFS (hard disk)
  - $45/TB/year
  - Accessible from all RC compute nodes
- Archive storage on LTFS (tape)
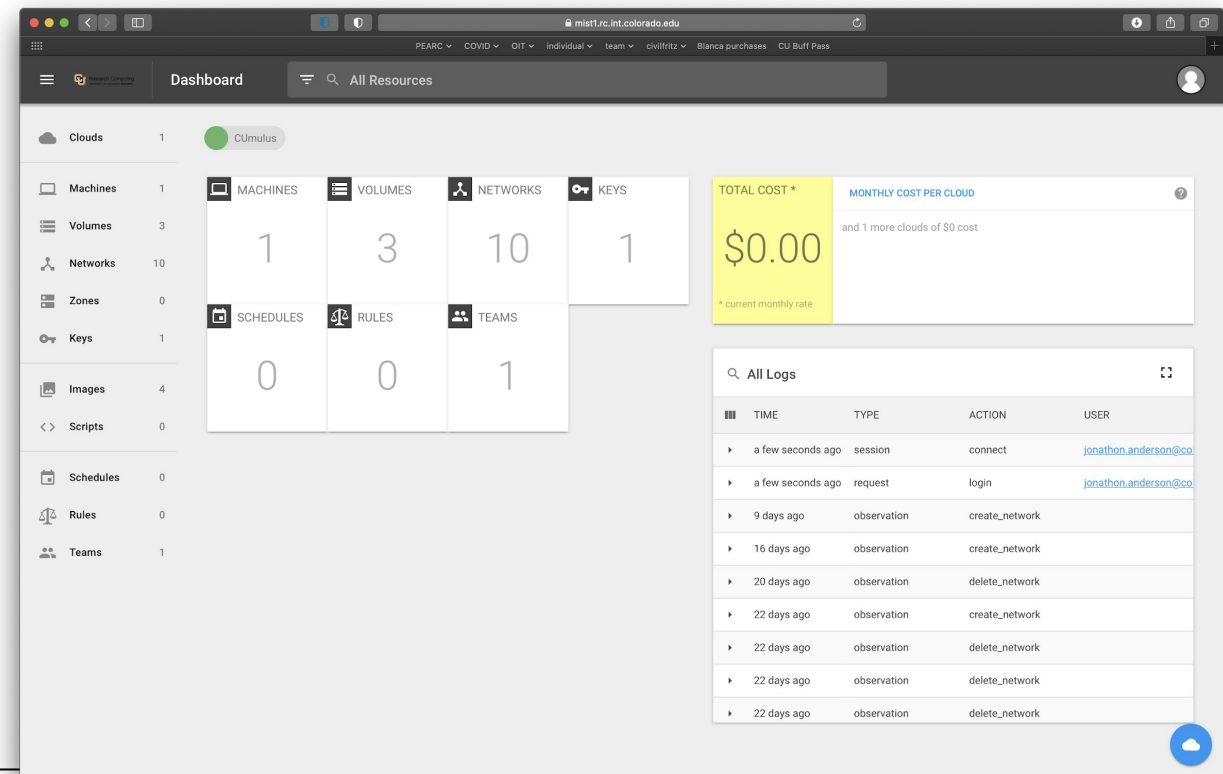 - $20/TB/year
  - iRODS

EnginFrame

- 3D accelerated virtual desktop

- Web-accessible

- Backed by 2 visualization nodes
  - 2 Nvidia Quadro RTX 8000 accelerators / node
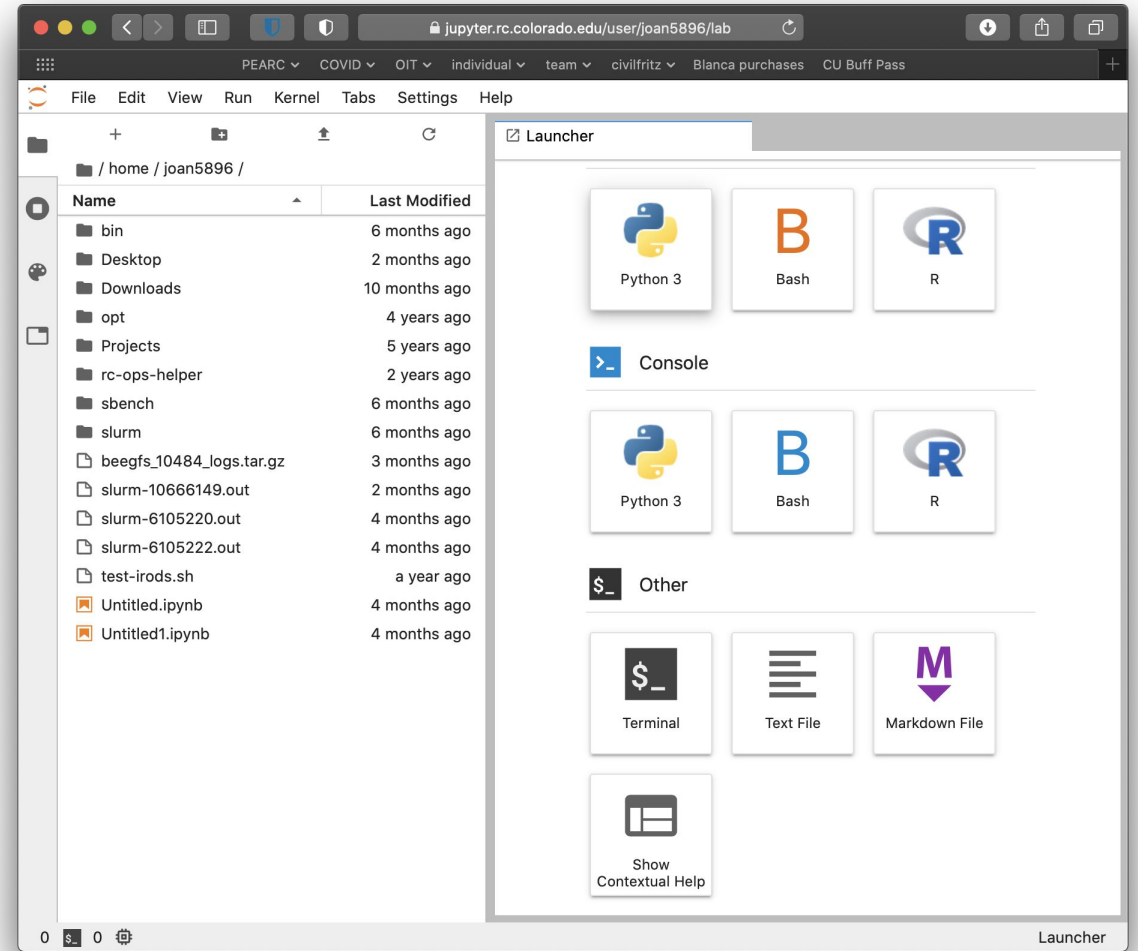 - 48 GiB memory / accelerator

CUmulus Hybrid Cloud

- On-premises OpenStack "IaaS" cloud

- Access to commercial clouds

- mist.io multi-cloud interface

- Hosting for
  - Database servers
  - Science Gateways
  - Compute pipelines

JupyterHub

- Python, R/RStudio, Bash

- Supports user-provided custom kernels

- Able to run on RMACC Summit and Blanca

"CURC-3" next-generation HPC

- Designed for regular expansion

- Estimated compute:
  - 160 CPU compute nodes (7,680 cores)
  - 20 GPU compute nodes
    (60 accelerators, mixture of Nvidia and AMD)

- Anticipated partitioned HDR-100 interconnect

- Anticipated 25Gb Ethernet to storage (including PetaLibrary)

- GPFS scratch
  - 22 GB/s
  - 2 PB

Faculty Support

- Ease access to resources
  - User support services
- CRDDS
  - Review of data management plans for external grants
  - Data storage and publication
    - CU Scholar
    - PetaLibrary
- Facilities statements
- Staff support for grant work

CU Research Computing
UNIVERSITY OF COLORADO **BOULDER**

Be Boulder.

Description of User Support Services

- Trainings
- Consultations
- Ticketing system
- Office hours
- Outreach
- Group configuration
  - Team lead, three full time staff, one 20% staff member
- Approximately 800 active users

Trainings

- 2019-2020 offered nearly 200 trainings
  - "Introduction to Python"
  - "Containerization for Research and Development Applications"
  - "Fundamentals of HPC"
- Over 1000 attendees
- Online option
- Topics chosen by request or by staff

Office Hours/Consulting

- Began in mid-2018
- Collaboration with CRDDS
- Tuesdays 12-1 pm, Thursdays 1-2 pm
  - RC only available on Tuesdays
- 400 consultations per year
- Fully online right now
- Normally in person
- 1:1 consultations also available

Supporting interdisciplinary data-intensive research & education

= CU Center for Research Data and Digital Scholarship (CRDDS)

https://www.colorado.edu/crdds/   crdds@colorado.edu

- Partnership between University Libraries and Research Computing
- Partnerships with other entities that have a similar mission
- Support of faculty & students with data related topics
- Center space (Norlin E206)

Data management

- Curation and DOI registration for 219 data sets in the CU Scholar institutional repository in first year
- DataCite DOI agreement with National Solar Observatory (NSO) and Laboratory for Atmospheric and Space Physics (LASP)
- CoreTrustSeal repository certification in collaboration with National Center for Atmospheric Research (NCAR)
- Data Publishing of large data sets
  - Metadata in CU Scholar linked through Globus to data sets on PetaLibrary
- Support of grant proposals

Data Storage and Publishing

- Meet funding agency or publication demands
- PetaLibrary
    - Large-scale
    - No sensitive data
- CU Scholar
    - Under 10 GB
    - DOI
    - Freely accessible
- Combination
    - PetaLibrary backend
    - Publish larger datasets
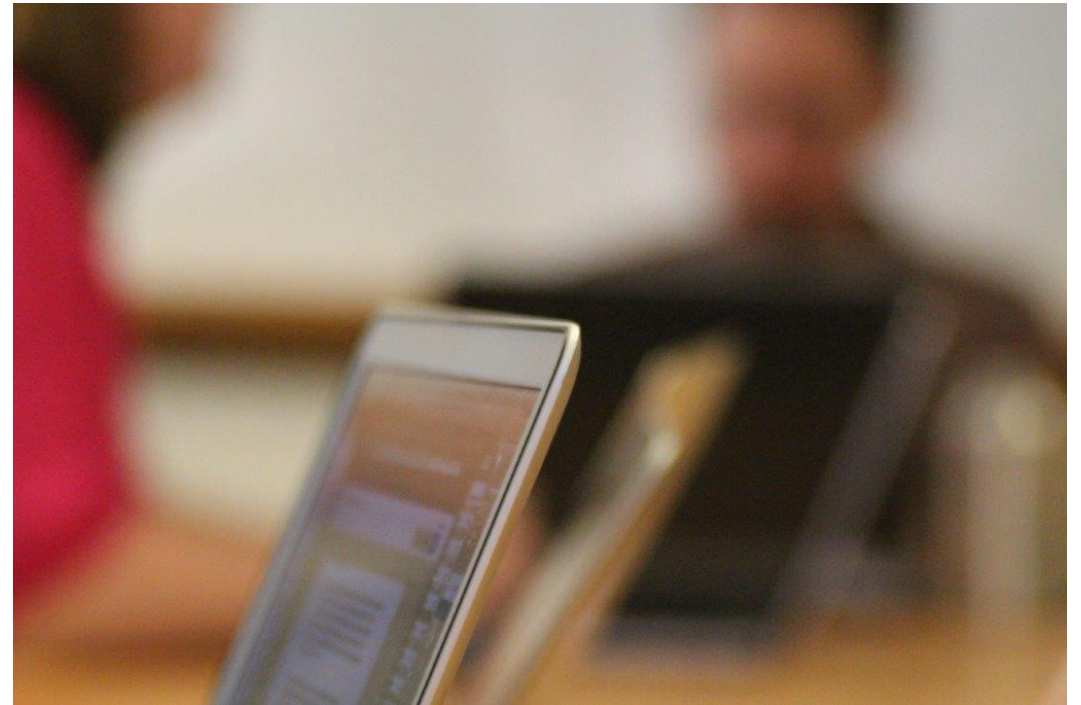
RMACC
the Rocky Mountain Advanced Computing Consortium

- Group of institutions in Rocky Mountain west
- NSF projects
  - MRI grant to procure Summit
  - Cyberteam project
    - Data workflows
    - Access to HIPAA compliant data storage
  - Hybrid cloud
- Share of Summit – 10%
- Share of hybrid cloud - 20%
- 70 users across six institutions
- Ongoing efforts to reach out
- CU provides user support – not dedicated
- Funding agencies love collaborations!

Questions?

Email:
Research Computing rc-help@colorado.edu
CRDDS crdds@colorado.edu

Website: www.colorado.edu/rc

Slides:

https://github.com/ResearchComputing/CEASE_Webinar_2021