Protected Environment at CHPC

Anita Orendt
Center for High Performance Computing
anita.orendt@utah.edu

Overview

- Background on the protected environment (PE)
- New PE Resources
- How to get a PE account
- Description of PE resources
- How to access PE resources
- People, Policy/Procedures & Tools
- Q & A

What is the CHPC Protected Environment (PE)?

• Developed in 2009 to strengthen the privacy and security protections for health information in scientific research

• Work closely with Security and Privacy office for consultations, security risk and compliance assessments, reviews, mitigation plans, and policy & regulation enforcement

• In 2017/18 -- deployed an updated PE with the assistance of a NIH Shared Instrumentation Grant awarded April 2017.

• Current PE is more reliable and secure, have expanded capabilities, and is scalable in a condominium fashion (similar to the general environment).

See: https://www.chpc.utah.edu/resources/ProtectedEnvironment.php

What is the CHPC Protected Environment (PE)? (2)

•The Protected Environment (PE) provides computing and data infrastructure, services, and facilitators (people) to support research and researchers at the University of Utah involving restricted or sensitive data –including

- •compliance for HIPAA (NIST 800-66 security framework)
- •DBGaP (database of Genotypes and Phenotypes, https://www.ncbi.nlm.nih.gov/gap)
- •RSICC (Radiation Safety Information Computational Center) export control data
- •work on proprietary data.

•At present, the PE is **NOT** certified for FISMA moderate (Federal Information Security Modernization Act, NIST 800-53, https://www.cisa.gov/federal-information-security-modernization-act) or CMMC CUI (confidential unclassified info, 110 requirements aligned with the NIST 800-171 security framework, https://www.acq.osd.mil/cmmc/about-us.html) -- though CHPC is currently working on the deployment of a CMMC 2.0 environment

Why do we have it?

- Researchers need a safe place to compute and work with restricted and sensitive data
- Restricted data can be stolen from insecure places
  - insecure systems, laptops/phones and tablets/removable drives
- Required by law in order to comply with regulations such as HIPAA. PHI security breaches are serious. e.g., fines, potential lawsuits, loss of reputation/credibility/funding.

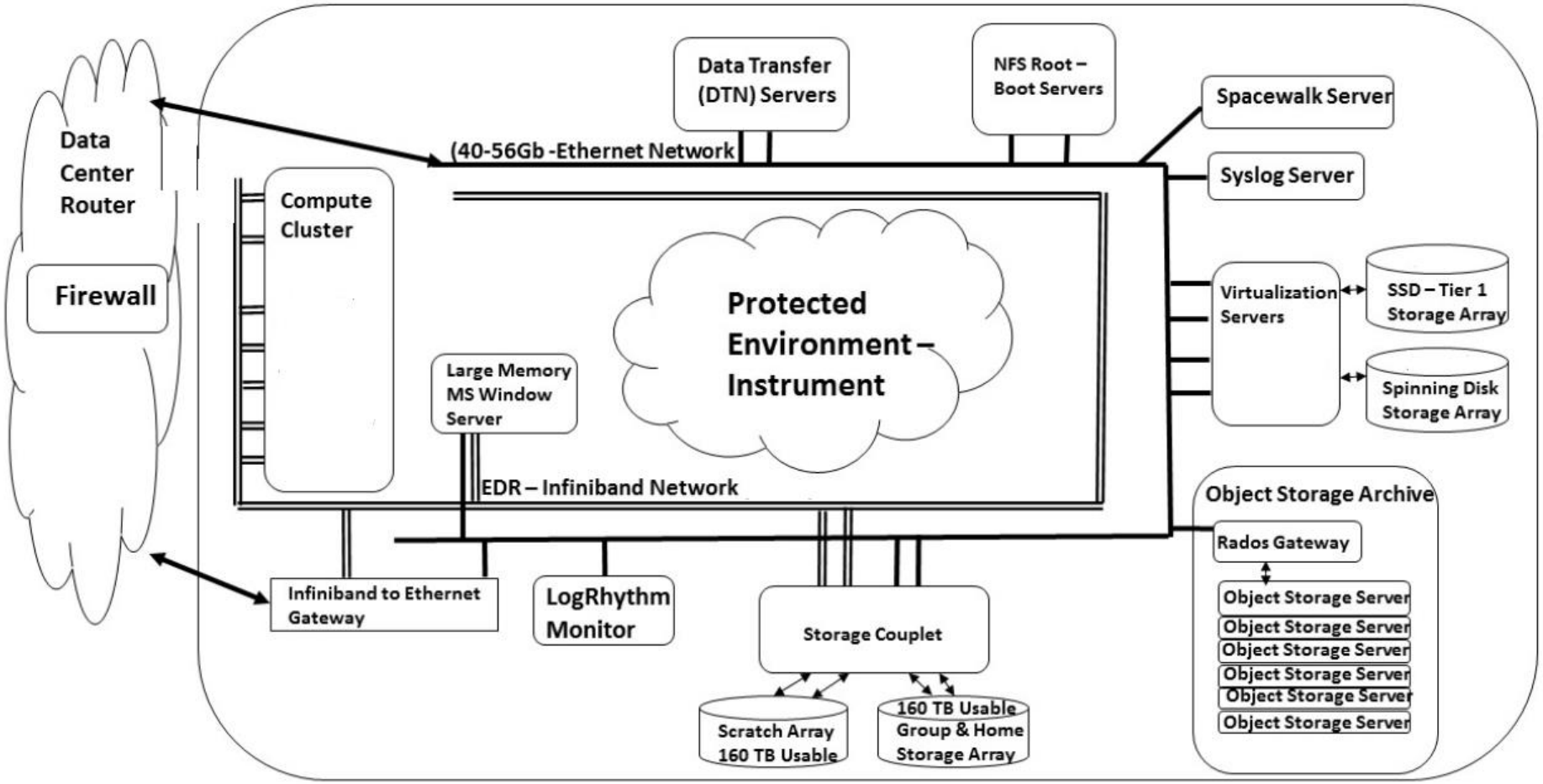
Safeguarding data is important for you as well as your institution

Security is a team effort...

Information security efforts will only be successful when all stakeholders understand the risks and take steps to avoid them.

PE Resources

- HPC Cluster – Redwood with Bristlecone interactive nodes
- Home Directories – Mammoth
- Project Space – Mammoth
- Scratch Space – /scratch/general/pe-nfs1
- Archive Storage – Elm
- VM farm – Prismatic1 and Prismatic 2
- Windows Server – Narwhal


Data Center Router
Firewall
Data Transfer (DTN) Servers
NFS Root – Boot Servers
Spacewalk Server
(40-56Gb -Ethernet Network
Syslog Server
Compute Cluster
Protected Environment – Instrument
Large Memory MS Window Server
Virtualization Servers
SSD – Tier 1 Storage Array
Spinning Disk Storage Array
EDR – Infiniband Network
Object Storage Archive
Rados Gateway
Object Storage Server
Object Storage Server
Object Storage Server
Object Storage Server
Object Storage Server
Object Storage Server
Infiniband to Ethernet Gateway
LogRhythm Monitor
Storage Couplet
Scratch Array 160 TB Usable
160 TB Usable Group & Home Storage Array

Description of Resources

https://www.chpc.utah.edu/documentation/guides/redwood.php

- 2 general login nodes and 2 bristlecone nodes
  - General login - 32 cores, 192 GB memory
  - Bristlecone – 28 cores, 64 GB memory
- 8 general GPU compute nodes
  - 2 with 32 cores, 4 GTX1080Ti, 192 GB memory
  - 5 with 16 cores, 8 GTX1080Ti, 64 GB memory
  - 1 with 16 cores, 5 GTX1080Ti, 64 GB memory
- 18 general CPU compute nodes
  - 4 with 32 cores, 192 GB of memory (Intel skylake)
  - 9 with 28 cores, 128 GB memory (Intel broadwell)
  - 2 with 64 cores, 512 GB memory (AMD Rome)
  - 2 with 64 cores, 512 GB memory (AMD Milan)
  - 1 with 32 cores, 1 TB memory (Intel Xeon X7560 processor)
- Owner nodes (both interactive/login and compute, some with GPUs) – with guest access for all PE users to owner nodes left idle

Description of Resources (2)

- Home/Project Storage – Mammoth
  - Home 50 GB/user home – backed up
  - Project space – 250 GB free; $150/TB for more space – NOT backed up unless group purchases Elm space required for the backup (change Summer 2022)
- Ceph Object Storage – Elm
  - Archive space – used internally for backup of home and project spaces; groups can purchase at $150/TB
- Windows Server – Narwhal
  - https://www.chpc.utah.edu/documentation/guides/narwhal.php
  - 24 CPU cores @3GHz, 512GB RAM, 1TB SSD local space
  - SAS with text miner, AMOS, SPSS, R, STATA, Mathematica, Matlab, and Microsoft Office 2010
  - Can mount PE home and project space (mammoth)

Description of Resources (3)

- Data Transfer Nodes
  - pe-dtn03.chpc.utah.edu and pe-dtn04.chpc.utah.edu
  - DTNs schedulable (12 cores/96 GB memory on each accessible via slurm)
    - redwood-dtn partition and dtn account
  - Also set up in Globus as University of Utah - CHPC PE Clustered Endpoint

Description of Resources (4)

- VM farm (set up with fail over for high availability)
  - Original VM farm has usable 72 cores, 1150 GB RAM, 25 TB SSD, 16 TB SED spinning
  - New VM farm – currently has 32 usable cores, 512 GB RAM, 20 TB SED SSD Storage; usable cores and memory can be expanded.
  - VMs can mount project spaces
- Sizing in incremental blocks (2 core, 4 GB RAM, 50 GB storage)
- Have costing model for VMs with block plus basic installation at hardware cost (next slide)
- Customization billed at $75/hour

VM Pricing

Blocks	Cores	RAM (GB)	Storage (GB)	Price
1	2	4	50	$425
2	2	8	100	$615
4	4	16	200	$990
8	8	32	400	$1745
16	8	64	800	$3250

- Prices are for a 5 year time period
- Additional VM local storage can be purchased, 100GB increments, $1100/TB
- https://www.chpc.utah.edu/resources/virtualmachines.php#pvf

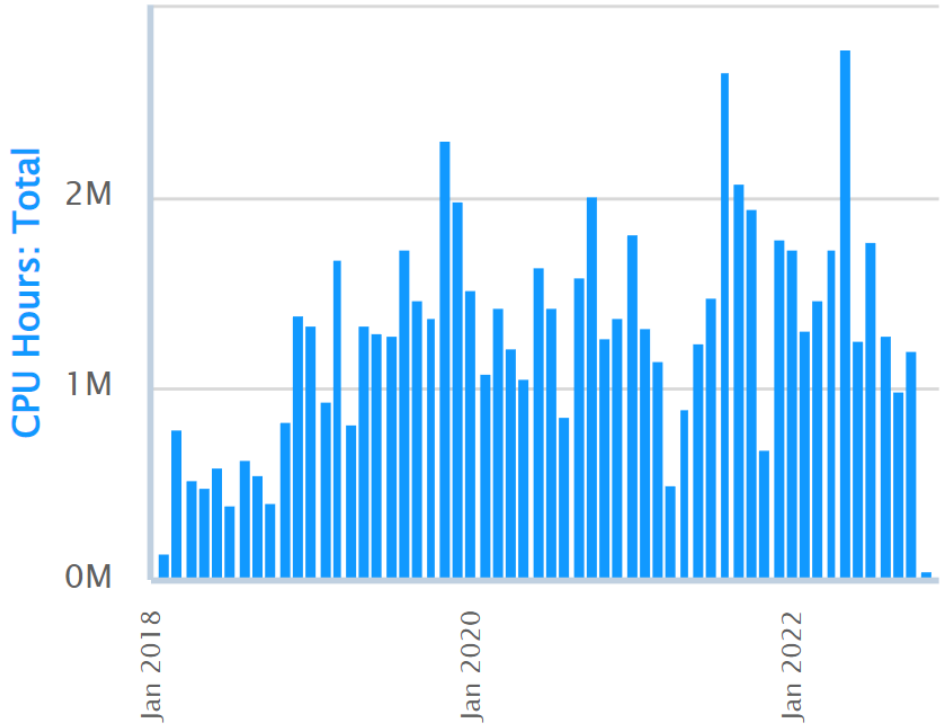
HPC Cluster Allocation Process

- Use of allocation started July 1, 2018
- Quick allocation request -- 1 quarter only, 30,000 wallclock core hours
- Normal Allocation request -- submitted at most quarterly for up to 4 quarters at a time
- Normal allocation requests are accepted 4 times per year, according to the following schedule
  - December 1st for allocation beginning January 1st
  - March 1st for allocations beginning April 1st
  - June 1st for allocations beginning July 1st
  - September 1st for allocations beginning October 1st

Usage Stats

- HPC usage since refresh
- 2022 Needs Assessments
  - 33 YTD
- Total Active Projects
  - > 170 projects
- VMs
  - 110 user requested VMs

Activity		Jobs	CPU Time (h)	
Users:	PIs:	Total:	Total:	Avg (Per Job)
172	58	5,205,959	73,887,992.3	14.19



Some of the Systems Controls in Place

- Standard baseline build list
- Inventory assets & hardware POC
- Qualys scans, Center for Internet Security (CIS) scans, Nessus, nmap, security onion, traffic trending with cacti
- Central Syslog, logwatch reports, network flow reports
- The physical hardware in datacenter with controlled room access; hosts are racked in a locked cabinet and have locked server bezels
- Thorough Documentation!
- Needs assessment, training, MFA/VPN access, IRB certification

Requires constant review of technical & physical security controls

Security Features

- Firewall (palo alto)
  - Classifies all traffic, including encrypted traffic, based on application, application function, user and content. You can create comprehensive, precise security policies, resulting in safe enablement of applications.
  - Innovative features reduce manual tasks and enhance your security posture, for example, by disseminating protections from previously unknown threats globally in near-real time, correlating a series of related threat events to indicate a likely attack on your network
  - Threat prevention feature WildFire identifies unknown malware, zero-day exploits, and advanced persistent threats (APTs) through static and dynamic analysis in a scalable, virtual environment, and automatically disseminates updated protections globally in near-real time

Security Features (2)

- SIEM – Security Information Event Management (LogRhythm)

  Due to growing need for a comprehensive log and event management; provides predefined reports to easily document evidence of compliance

  – enable better security of networks and optimize information technology operations with sophisticated log correlation and analytics.
  – automate collection, organization, analysis, archival, and recovery of log data that enables enterprises to comply with log data retention regulations.
  – ensure compliance with mandates for HIPAA and other government regulations and to protect patient confidentiality and safety.

Access Controls

- Login Access
  - General linux login nodes via ssh: redwood.chpc.utah.edu (round robin of redwood1 and redwood2)
  - Other general interactive nodes via ssh: bristlecone1.chpc.utah.edu, bristlecone2.chpc.utah.edu
  - Windows: narwhal.chpc.utah.edu – connect via RDP
  - Access to all requires DUO 2 factor authentication
  - From non-UofU IP address must first use University VPN
  - Data access – based on IRB number/project name
  - We verify users' right to access the specified data (check IRB)
  - Use unix FACLs (File Access Control Lists)

Getting Started in the PE

https://www.chpc.utah.edu/resources/ProtectedEnvironment.php

- Step 1: Determining if your project fits in REDCAP
- Step 2: Needs Assessment
- Step 3: Requesting access to a PE resource

PE Needs Assessment

https://www.chpc.utah.edu/role/user/needs_assessment_form.php

- Complete Form – one assessment needed for each project
  - Information about PI
  - Project funding information
  - IRB information
  - Project Computing Requirements
    - What do you intend to do in this environment?
    - Do you need a virtual machine?
    - Do you need a database?
    - What services will the software/hardware provide?
    - To whom will these services be provided?
    - Who needs to have remote terminal (ssh/rdp) access?
    - Will there be any information sharing with third parties?
    - Brief description of the research with this project
    - How many people will use this system?
    - Estimate of how many people and records are anticipated to be stored

Requesting PE Access

- Get a CHPC general environment account
- Get a CHPC PE account
  - Note for each PE project, you will need to complete the CHPC PE account application
- Do CHPC's HIPAA training (will get a invite to a Canvas course)
- Set up DUO two factor authentication
- If the resources you need already exist – you are ready to go
- If you need a new VM – work with CHPC to get VM provisioned
  - CHPC will need info on OS, number of cores, amount of memory, disk space and any additional software needs

Acknowledging use of PE

- Sample acknowledgements at https://www.chpc.utah.edu/about/acknowledge.php

  "*The support and resources from the Center for High Performance Computing at the University of Utah are gratefully acknowledged. The computational resources used were partially funded by the NIH Shared Instrumentation Grant 1S10OD021644-01A1.*"

- Link publication to the S10 grant via your "My NCBI", see https://www.ncbi.nlm.nih.gov/books/NBK3842 for information

Resources

These slides can be found at:
https://www.chpc.utah.edu/presentations/ProtectedEnvironmentatCHPC.php

CHPC-PE main page:
https://www.chpc.utah.edu/resources/ProtectedEnvironment.php

CHPC-PE FAQs:
https://www.chpc.utah.edu/documentation/pefaq.php

HHS HIPAA FAQ: http://www.hhs.gov/ocr/privacy/hipaa/faq/index.html

Getting Help

- CHPC website and wiki
  - www.chpc.utah.edu
    - Getting started guide, cluster usage guides, software manual pages, CHPC policies
- Service Now Ticketing System
  - Email: helpdesk@chpc.utah.edu
- Help Desk: 405 INSCC, 581-6440 (9-5 M-F)
- We use chpc-hpc-users@lists.utah.edu for sending messages to users; chpc-hipaa-users@lists.utah.edu for PE specific messages