Protected Environment at CHPC

Anita Orendt, anita.orendt@utah.edu
Wayne Bradford, wayne.bradford@utah.edu
Center for High Performance Computing

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

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.
- New PE is more reliable and secure, have expanded capabilities, and is scalable in a condominium fashion (similar to the general environment).
- Looking for users with NIH funding who can make use of new PE

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

Why do we have it?

- Researchers need a safe place to compute and work with restricted 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 and your institution

Other Uses of PE

- While need for HIPAA compliance is the most common reason to use the PE, there are other uses, including:
  - ITAR (International Traffic in Arms Regulations) compliance
  - FDA part 11 compliance
  - Any other sensitive or restricted data and/or application
- These each come with their own regulations and requirements

NIH dbGaP

- https://www.ncbi.nlm.nih.gov/gap
- See Security Procedure section
- For “controlled-access human genomic and phenotypic data”
- Do not contain direct identifiers, but the data are sensitive and must be protected

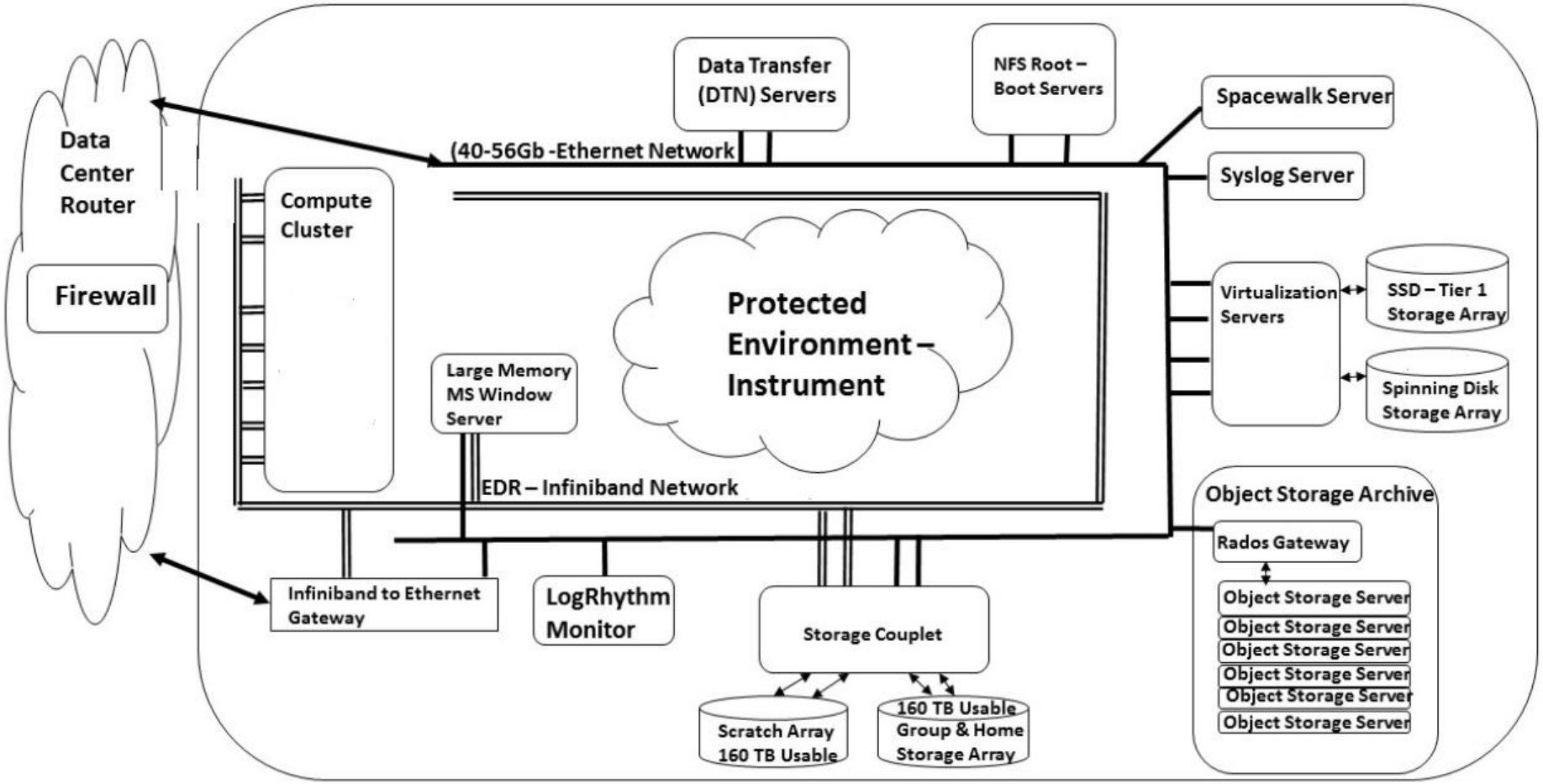
CHPC is promoting move of all human genomic work to PE

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
- Home Directories – Mammoth, 50 GB per user
- Project Space – Mammoth, about 600 TB total space
- Scratch Space – Mammoth, 160 TB
- Archive Storage – Elm, about 1.5 PB usable capacity
- VM farm – Prismatic
  - Both SSD storage (25 TB) and SED spinning disk storage (16 TB)
- 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
Virtualization Servers
SSD – Tier 1 Storage Array
Spinning Disk Storage Array
Large Memory MS Window Server
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

- HPC Cluster – Redwood
  - https://www.chpc.utah.edu/documentation/guides/redwood.php
  - 2 general login nodes (XeonSP, 32 cores, 192GB memory)
  - 2 general GPU compute nodes (32 cores, 4 GTX1080Ti, 192GB)
  - 15 general CPU compute nodes (436 total cores)
    - 4 XeonSP (skylake) nodes with 32 cores, 192GB of memory
    - 11 Broadwell nodes with 28 cores, 128GB memory
  - Owner nodes (both interactive/login and compute)
    - currently 156 compute nodes with over 4600 cores
  - Mellanox EDR Infiniband interconnect (broadwell nodes connect at FDR)
  - 160 TB scratch server
  - Slurm batch system

Description of Resources (2)

- Storage - Mammoth
  - 50 GB/user home – backed up
  - Project space – 250 GB free; $150/TB for more space; up to 5 TB for NIH funded projects
  - 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)

- VM farm (have 4 servers with fail over for availability)
  - Have usable 72 cores, 1150 GB RAM, 25 TB SSD, 16 TB SED spinning
- Sizing in incremental blocks (2 core, 4 GB RAM, 50 GB storage)
  - Storage SSD by default unless encryption needed
  - Can support 275 blocks
  - Can get additional space if needed
  - Can also mount project space
- 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	$630
2	2	8	100	$1025
4	4	16	200	$1810
8	8	32	400	$3380
16	8	64	800	$6525

- NIH funded projects free for lifetime of grant
- Prices are for a 5 year time period
- https://www.chpc.utah.edu/resources/virtualmachines.php#pvf

HPC Cluster Allocation Process

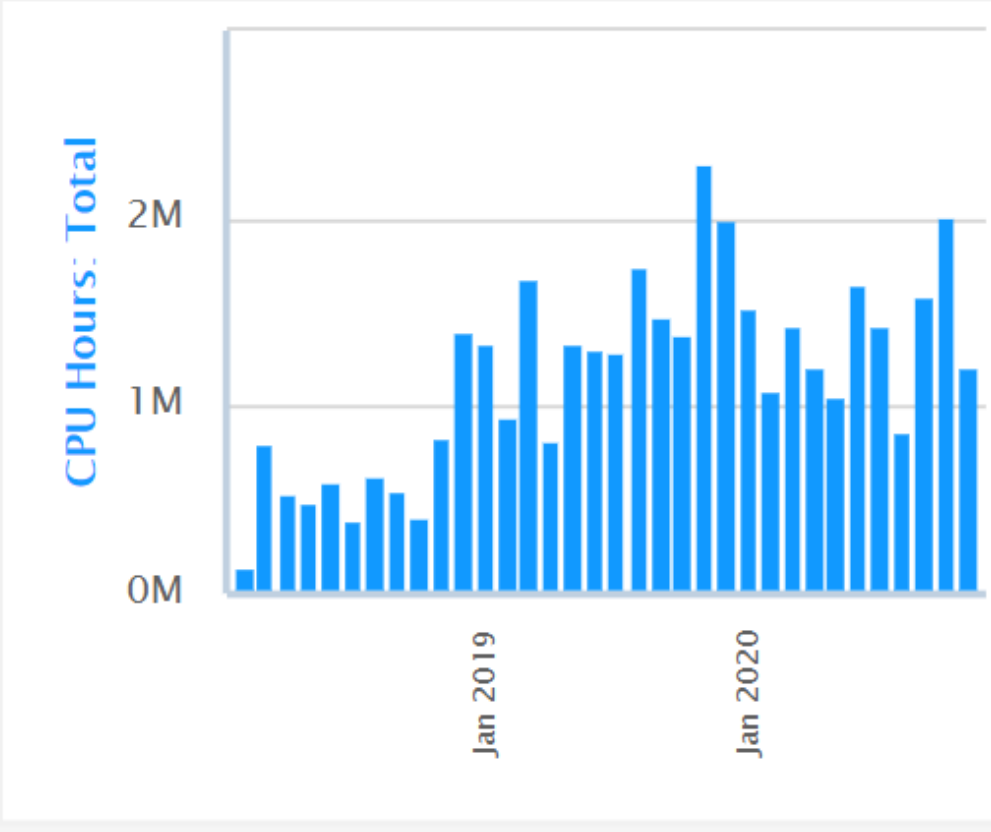
- Use of allocation started July 1, 2018
- Quick allocation request -- 1 quarter only, 20,000 wallclock core hours
- Normal Allocation request -- submitted at most quarterly for up to 4 quarters at a time
- Normal allocation requests are accepted 4 times per year, according to the following schedule
  - 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
- Projects
  - 105 projects
- VMs
  - 85 user requested VM

Activity		Jobs	CPU Time (h)	
Users:	PIs:	Total:	Total:	Avg (Per Job):
100	36	3,079,815	39,346,952.9	12.78

Total CPU Hours By Resource (Top 10)

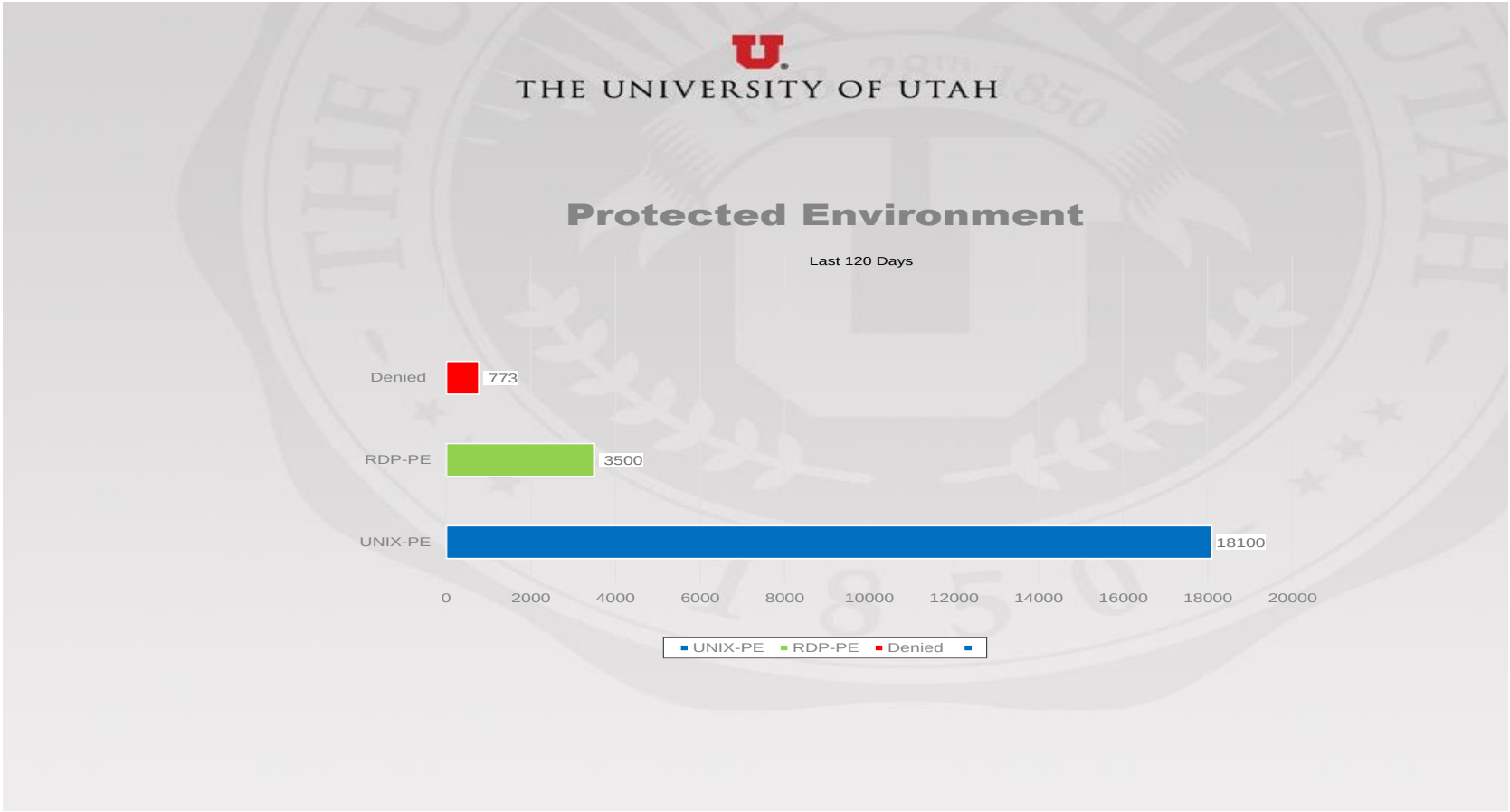


10 year hpc resource growth

Date	Cores	Nodes	RAM	Disk (TB)
02/2009	72	9 (8core, 16G)	144	6
10/2010	128	16 (8core, 16G)	256	28
04/2011	152	19 (8core, 16G)	368	34 home/group 5T scratch
02/2012	160	20 (8core, 16G)	384	34 home/group 5T scratch
04/2014	160	20 (8core, 16G)	384	34 home/group 5T scratch
10/2015	160	20 (8core, 16G)	34	34 home/group 5T scratch
04/2016	160	20 (8core, 16G)	384	34 home/group 5T scratch
03/2017	280	10 (28core,128G)	1280	34 home/group 5T scratch
12/2019	4292	157 (28 and 32 core, 128G to 384 G of ram), 2 nodes with 4 gtx 1080ti	26594	320T home/group, 160T nfs scratch

Unix, windows, denied DUO logins last 120 days

(does not include service accounts)



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

- 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)
  - 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
  - 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?
    - 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
- Do CHPC's HIPAAtraining (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 informaiton

Resources

These slides can be found via:
https://www.chpc.utah.edu/presentations/

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