The State of
Research Computing and Data (RCD)
at UK and Beyond

James Griffioen

Director
UK Center for Computational Sciences

CCS RCD Seminar Series
January 30, 2024

Center for
Computational Sciences

Some History and Terminology

(The transition from HPC to RCD)

- For decades universities have been creating Centers dedicated to supporting research that involves massive computations
- These Centers have historically been called Supercomputing Centers or High Performance Computing (HPC) Centers that focus on running a very large computation (a.k.a. “job”) as fast as possible.
- Over time researchers began to (also) need High Throughput Computing (HTC) resources that could complete as many (not necessarily large) jobs as fast as possible (in parallel).
- More recently researchers from all academic disciplines are beginning to use Research Computing and Data (RCD) referring to the need to run computations and store big data on “something bigger than a desktop or laptop computer”.

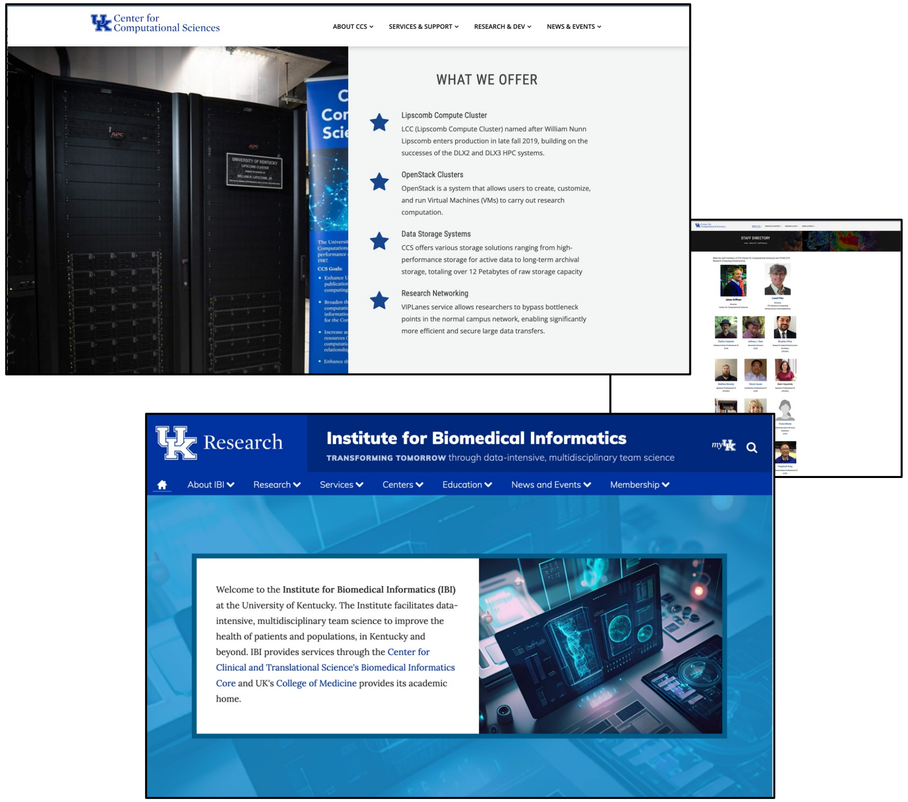
CCS

(The Center For Computational Sciences)

CCS provides RCD support to all faculty, staff, and (associated) students at UK

and the Commonwealth (as part of the KY Cyberteam)

- CCS is tightly integrated with the ITSRCI (ITS Research Computing Infrastructure) group within ITS which is led by Lowell Pike.
- CCS also works closely with IBI-EDC (Institute for Biomedical Informatics Enterprise Data Center Services) led by Jeff Talbert



The Sea-change Sweeping RCD

- Fields that historically have not needed computational capabilities, now do.
- Specialized hardware (and devices/instruments) are advancing very quickly
- Complex software systems and platforms have become readily available and relatively easy to use
- Big data (and AI) is changing everything
  - Big data sets are readily available and new data is being produced at a staggering rate by things like IoT devices (including cell phone apps), research instruments, networked measurement equipment (e.g., sensors), logging systems (e.g., web transactions, OS/network events, business transactions), public and commercial data compilations/repositories, etc.
  - Big data provides massive amounts of information useful to research and is driving the need to perform advanced search/filtering, data analytics, data-driven computations, simulation/emulation, AI (training/inferencing), data streaming, data sharing, etc.
  - Big data is driving rapid advances in the design and scale of compute, storage, and network systems

Implications of The RCD Sea-change

- Major increase in the number of Research Domains, which implies:
  - Major increase in the total number of Researchers
  - Huge percentage of First-time RCD Users
  - Significant increase in the software needed to support research across a wide range of disciplines
- Major increase in Demand for Resources per User, which implies:
  - Major increase in the Total Number of Resources needed
- Widespread use of AI, which implies:
  - Increased demand for certain types of resources: GPUs, Storage, AI software and pipelines, Compliance and Regulatory Guidelines and Procedures, etc.
- Significant increase in the use of Data Transfer Services, which implies:
  - Increase pressures on network bandwidth
- Significant increases in …
- In short, there are many new needs and requirements for hardware, software, documentation, support, etc. (and we haven't even mentioned the challenges of supporting protected data)

CCS By The Numbers

- Over 1200 CCS Users
  - From 65 Departments
  - Across 12 Colleges
  - Over 250 PIs/Research Groups
- Support 7 Computational Clusters with more than
  - 450 Nodes
  - 34 Thousand Cores
  - 300 TB Total Memory
  - 2.2 PetaFLOPs of Total Compute Capacity
- Handle over 12 Million Compute Jobs/Year
- Providing more than 285 million Compute Hours/Year
- More than 28 PB Total Storage Space consisting of multiple storage systems with more than
  - 300 TB of high-speed SSD Storage
  - 6.5 PB of Parallel File System Storage
  - 5 PB of high-speed Object Storage
  - 16.5 PB of long-term (NAS) Storage
  - And high-speed access to Remote Tape Storage
- Supporting Multiple DTN Nodes with
  - 10 and 40 Gbps links to the Internet
  - 100 and 200 Gbps links to clusters and storage systems
- Over 800 Software Packages and Libraries

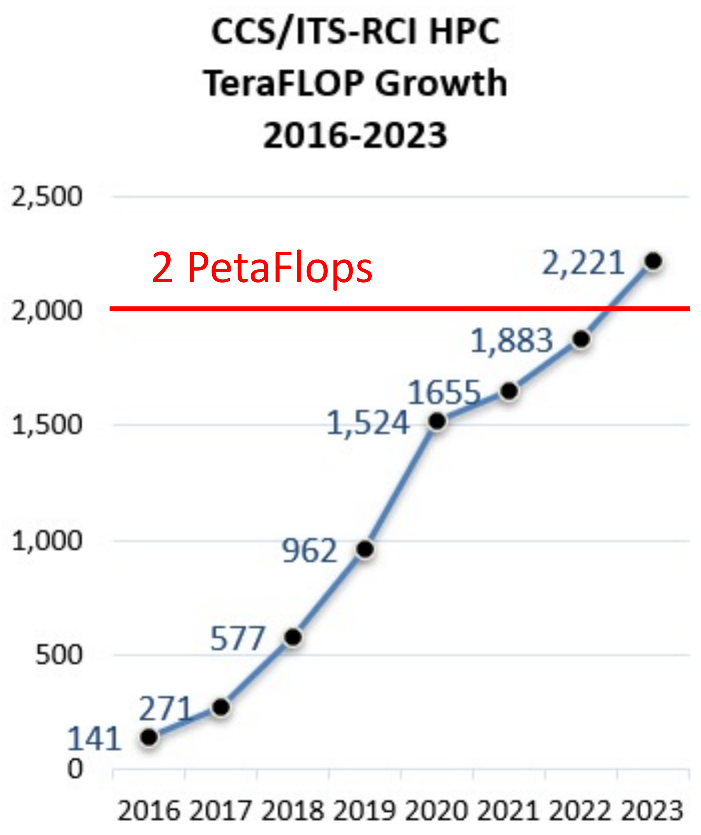
Talk Outline

- The RCD Sea-change
- CCS Resources
  - CCS Compute Resources
  - CCS Storage Resources
  - CCS Network Resources
  - CCS Condo Resources
- CCS Support
  - Documentation and Getting Help
  - Account Setup and Onboarding
  - User Portals
  - Software
  - Usage Information
- Other Resources
  - ACCESS
  - FABRIC
  - NAIRR

Talk Outline

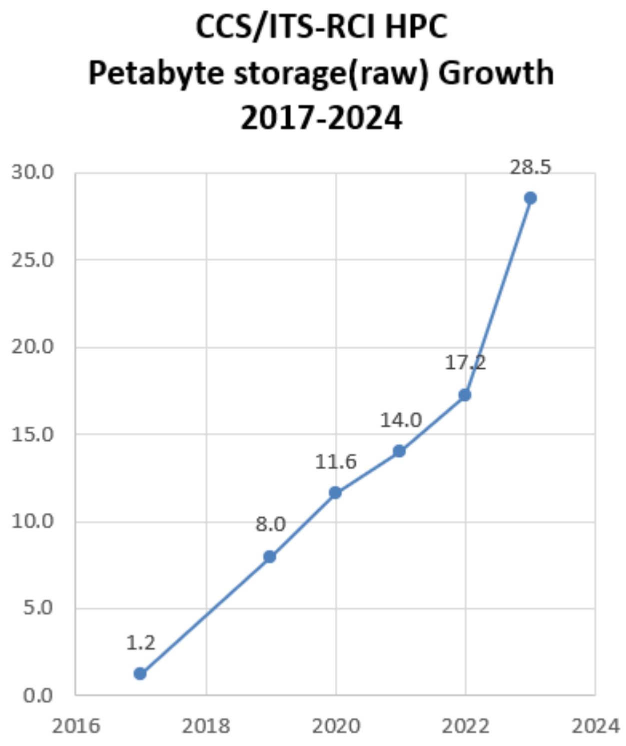
- The RCD Sea-change
- CCS Resources
  - CCS Compute Resources
  - CCS Storage Resources
  - CCS Network Resources
  - CCS Condo Resources
- CCS Support
  - Documentation and Getting Help
  - Account Setup and Onboarding
  - User Portals
  - Software
  - Usage Information
- Other Resources
  - ACCESS
  - FABRIC
  - NAIRR

Computational Growth



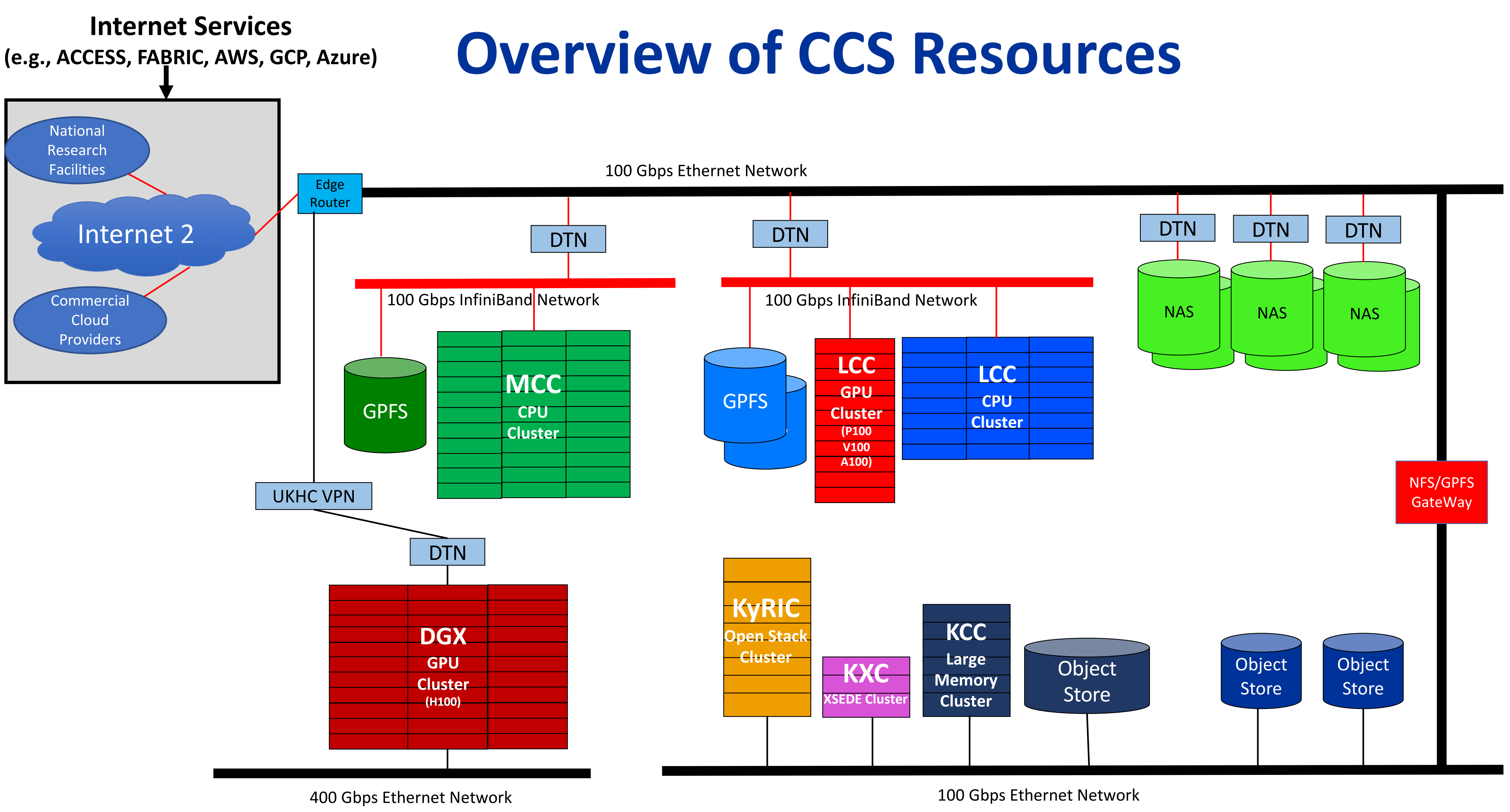
- Surpassed 2 PetaFLOPs with MCC addition
- Growth by expansion (not Moore's Law)
  - 33,700 Cores (4 Intel and 1 AMD Generations)
- Most growth in GPUs (for ML/DL and growing number of HPC codes)
  - 34 nodes, 128 P100/V100/A100s
- Much of the growth is funded by the condo model
- And more "GPUs" to come!

Storage Growth



- Growth by expansion of scratch and long-term storage systems
  - Total raw storage now exceeds 28 PB
- Most growth in NAS storage and growing
  - 16.5PB NAS, 6.5PB GPFS, 5.3PB Object Store
- Much of the NAS storage growth is funded by the condo model
- With more storage to come!

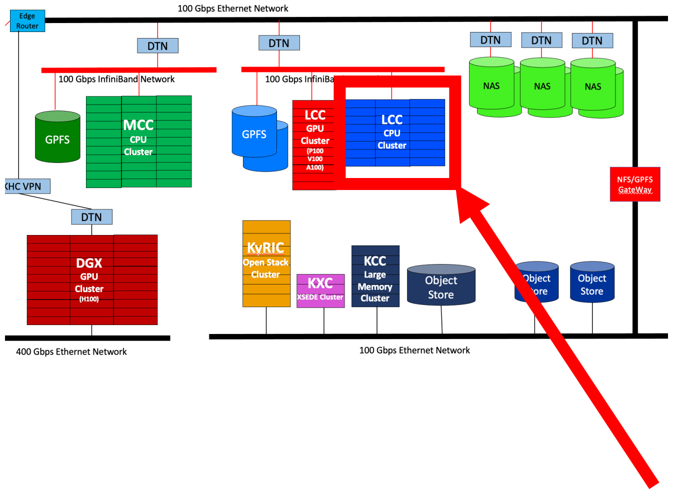
Overview of CCS Resources



CCS Compute Resources

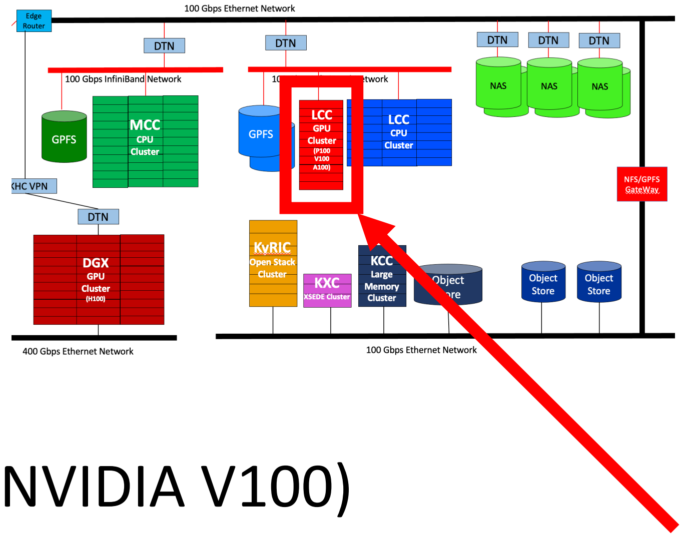
LCC-CPU (HPC cluster)

• Cluster Name:	Lipscomb Compute Cluster (LCC)
• Purpose/Use:	Ideal for Parallel HPC applications requiring large numbers of CPUs/Cores and/or the need for low-latency communication between processes (e.g., MPI applications)
• CPU Architecture:	Intel Skylake and Intel Cascade
• Attached Storage:	GPFS
• Total TFLOPs:	481 TFLOPs
• Total Nodes:	118 nodes
• Total Cores:	7,168 cores
• Cores per Node:	32 (Skylake) 48 (Cascade)
• Memory per Node:	192 GB
• Total Disk:	4.1 PB (Raw) 2.6 PB (Usable)



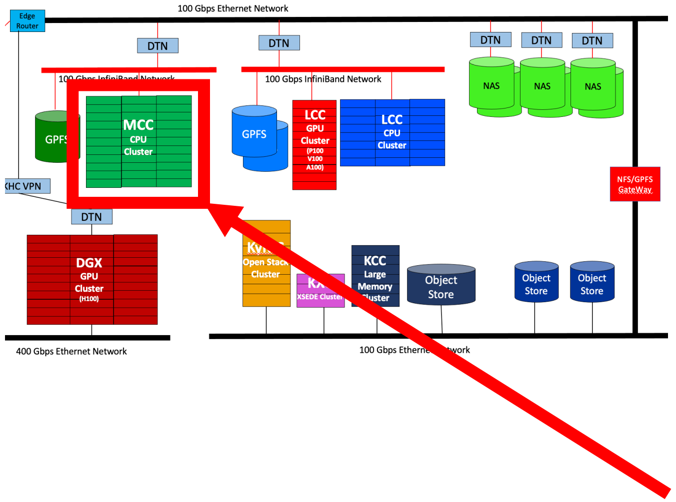
LCC-GPU (HCP Cluster)

- Cluster Name: Lipscomb Compute Cluster (LCC)
- Purpose/Use: Ideal for AI applications or Scientific or Image Processing applications optimized to run on GPUs
- CPU Architecture: Intel Skylake and Intel Cascade
- Attached Storage: GPFS
- Total TFLOPs: 818 TFLOPs
- Total Nodes: 30 nodes
- Total Cores: 816 cores
- Cores per Node: 32 (Skylake) 48 (Cascade)
- Memory per Node: 192 GB
- Total GPUs: 120 (48 NVIDIA P100 and 72 NVIDIA V100)



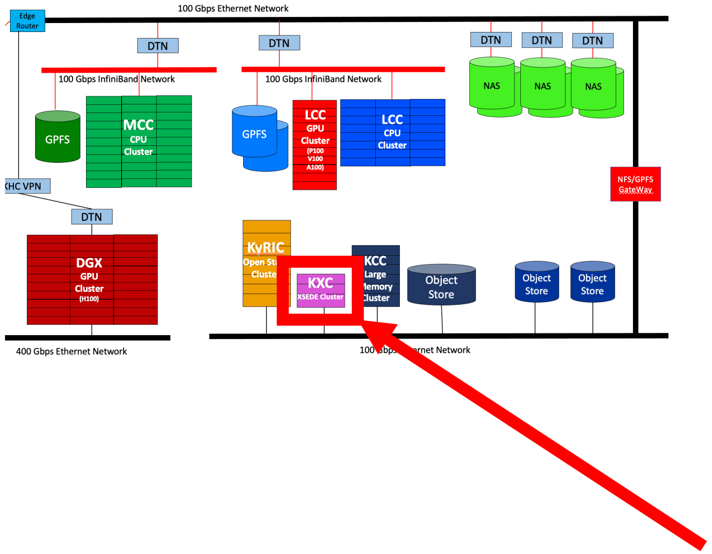
MCC-CPU (HPC Cluster)

- Cluster Name: Morgan Compute Cluster (MCC)
- Purpose/Use: Ideal for Parallel HPC applications requiring large numbers of CPUs/Cores and/or the need for low-latency communication between processes (e.g., MPI applications)
- CPU Architecture: AMD Rome EPYC 7702
- Attached Storage: GPFS
- Total TFLOPs: 737 TFLOPs
- Total Nodes: 180 nodes
- Total Cores: 23,500 cores
- Cores per Node: 128
- Memory per Node: 512 GB per node
- Total Disk: 2.4 PB (Raw) 1.8 PB (Usable)



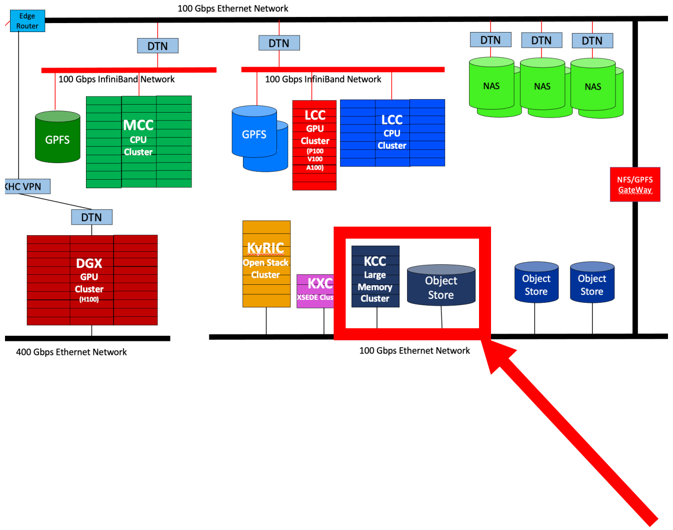
KXC (HPC cluster)

• Cluster Name:	KyRIC XSEDE Cluster (KXC)
• Purpose/Use:	Reserved for ACCESS (XSEDE) user applications requiring large amounts of memory
• CPU Architecture:	Intel Broadwell
• Attached Storage:	NAS Storage
• Total GFLOPs:	6,720 GFLOPs
• Total Nodes:	5 nodes
• Total Cores:	200 cores
• Cores per Node:	40
• Memory per Node:	3 TB per node
• Total Disk:	300 TB



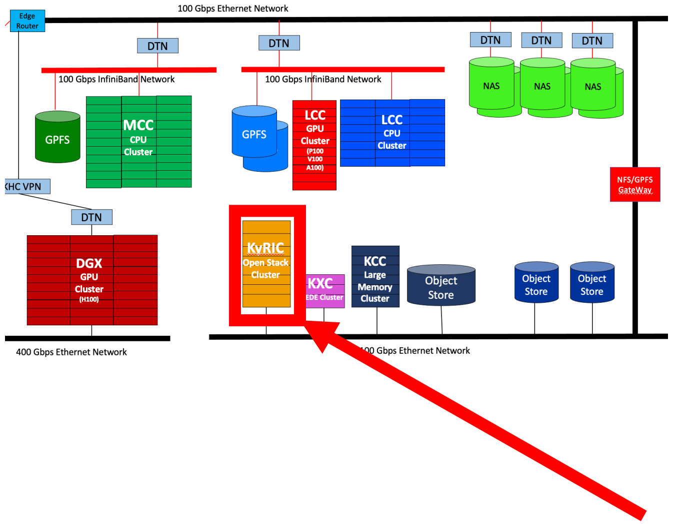
KCC (HPC cluster)

- Cluster Name: KyRIC Compute Cluster (KCC) – being incorporated into LCC
- Purpose/Use: Ideal for applications requiring very large amounts of memory (e.g., bioinformatics and other big data applications)
- CPU Architecture: Intel Broadwell
- Attached Storage: CephFS
- Total GFLOPs: 26,880 GFLOPs
- Total Nodes: 20 nodes
- Total Cores: 800 cores
- Cores per Node: 40
- Memory per Node: 3 TB per node
- Total Disk: 1.3PB (usable)



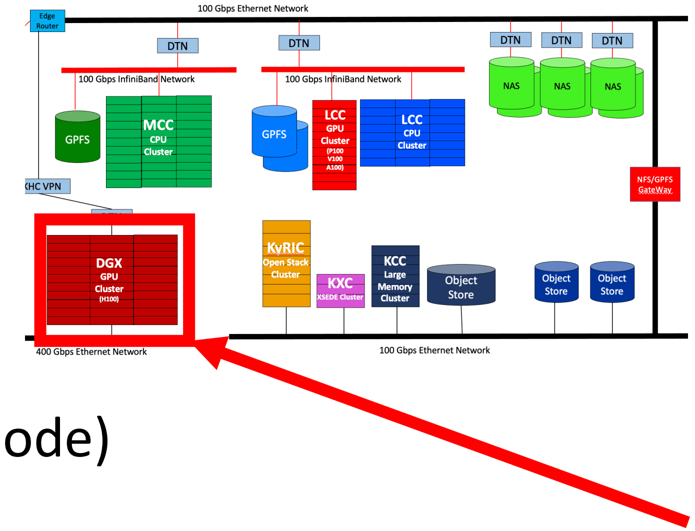
KyRIC OpenStack Cluster

- Cluster Name: KyRIC OpenStack Cluster
- Purpose/Use: Ideal for **VMs** that require very large amounts of memory (e.g., big data applications or servers managing big data sets)
- CPU Architecture: Intel Broadwell
- Disk Storage: Ceph Block Storage
- Total GFLOPs: 33,600 GFLOPs
- Total Nodes: 25 nodes
- Total Cores: 1,000 cores
- Cores per Node: 40
- Memory per Node: 3TB per node
- Total Disk: 100 TB



DGX-GPU (HCP Cluster)*

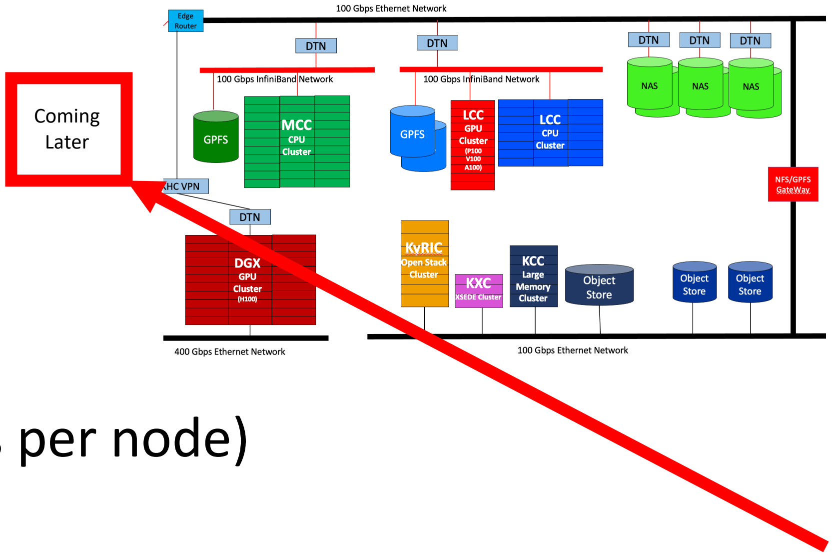
- Cluster Name: DGX MRI Compute Cluster (hosted by IBI/EDC)
- Purpose/Use: Ideal for AI applications or Image Processing applications optimized to run on GPUs. Can be used for HIPAA applications
- CPU Architecture: Intel Xeon
- Disk Storage: NAS storage via NFS
- Total TFLOPs: 1,360 TFLOPs
- Total Nodes: 5 nodes
- Total Cores: 1,120 cores
- Cores per Node: 224
- Memory per Node: 2TB per node
- Total GPUs: 40 NVIDIA H100 (8 GPUs per node)
- Total Disk: 1+ PB storage



* Supported by NSF MRI Grant #2216140 (PI – Talbert, Co-PIs – Risko, Moseley, Bumgardner, Griffioen)

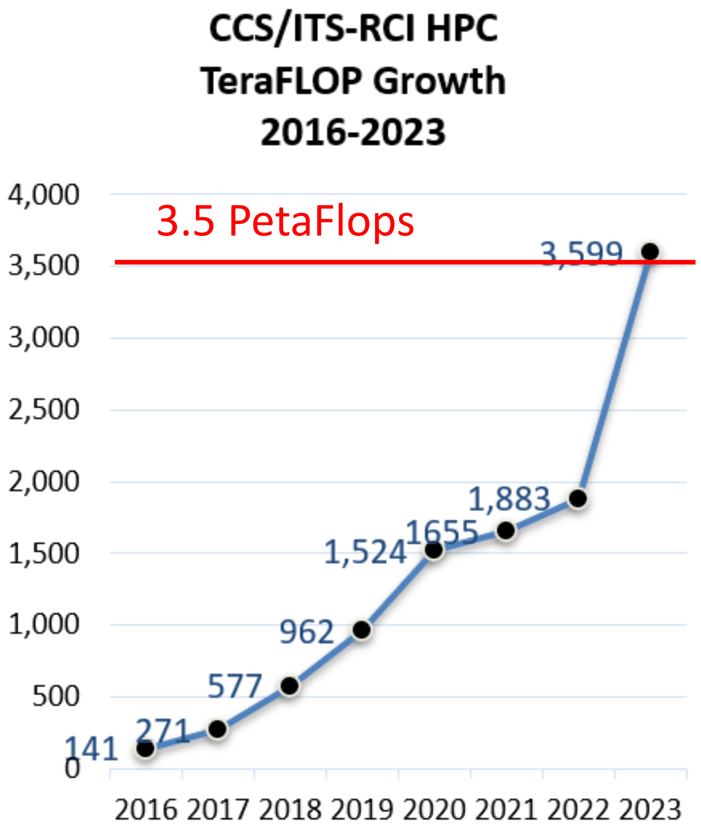
ECC-GPU (ECC)* – coming later this year

- Cluster Name: EduceLab CYBER Cluster (ECC)
- Purpose/Use: Ideal for Heritage Science Applications (e.g. AI applications on big data collected from specialized instruments/devices)
- CPU Architecture: Intel Xeon
- Disk Storage: GPFS
- Total TFLOPs: 680 TFLOPs
- Total Nodes: 5 nodes
- Total Cores: 320 cores
- Cores per Node: 64
- Memory per Node: 512 GB per node
- Total GPUs: 20 NVIDIA H100 (4 GPUs per node)
- Total Disk: 2+ PB storage



* Supported by NSF Midscale Grant #2131940 (PI – Seales, Co-PIs – Baker, Balk, Reyes-Centeno, Smith)

Computational Growth with DGX and ECC



- UK's total computational capabilities increased to 3.6 PFLOPS with DGX and will reach 4.2 PFLOPs with ECC later this year
- Massive increase in GPUs available for AI applications
- Each H100 has 80 GB of RAM enabling processing of much larger AI data sets

CCS Storage Resources

GPFS Storage

- GPFS provides large-scale high-speed parallel file system storage
- Purpose/Use: Temporary "Scratch" Storage (90 days)
- HPC cluster nodes access GPFS files at high speed and low latency over an (EDR/HDR) InfiniBand Network.
- Total GPFS File Storage is 4.7 PB usable (6.4 PB raw)
- LCC GPFS file systems (on 100 Gbps EDR):
    1. Lenovo GPFS (DSS-G) parallel file system 1: 1.3PB usable (1.9PB raw)
    2. Lenovo GPFS (DSS-G) parallel file system 2: 1.6PB usable (2.2PB raw)
- MCC GPFS files systems (on 200 Gbps HDR):
    1. Lenovo GPFS (DSS-G) parallel file system -1: 1.8PB usable (2.3PB raw)

Ceph Storage

- Ceph provides large-scale high-speed parallel object storage
- Purpose/Use: High-speed access to (long-lived) large data sets
- Unlike files, objects are intended to be streamed (read/written) in their entirety
- There is no hierarchy/directory structure and instead supports “buckets”
- File systems can be built on top of Ceph – e.g., CephFS
- Clients stream data to/from Ceph via the popular S3 API
- High speed access to objects requires a high-bandwidth network path from the client to the server and a parallel file transfer service (e.g., rclone).
- Total Ceph Object Storage is 3.5 PB usable (5.3 PB raw)
- CCS has three Ceph Object Store systems:
  1. KyRIC Ceph System (1.8 PB usable capacity) – supports Ceph Block Storage for OpenStack
  2. PKS2 Ceph System (600 TB usable capacity, 1.3 PB raw)
  3. CoT Ceph System (600 TB usable capacity, 1.3 PB raw)

Network Attached Storage (NAS)

- NAS provides large-scale network accessible file storage
- Purpose/Use: Long-term storage of very large data sets
- Data is not intended to be accessed directly by CCS computational clusters (or any other machine)
- Instead, users typically copy data from the NAS to GPFS, use it, and then write it back to NAS
- Some users have used sshfs to ”mount” their data on CCS clusters
- Any computer anywhere can access NAS data using these methods
- Transfer speeds depend on network path bandwidth and protocols, but can be as high as 12 Gbps
- Total NAS Storage is 13.2 PB usable (16.5 PB raw)
- CCS has several NAS systems. Many are condo systems.

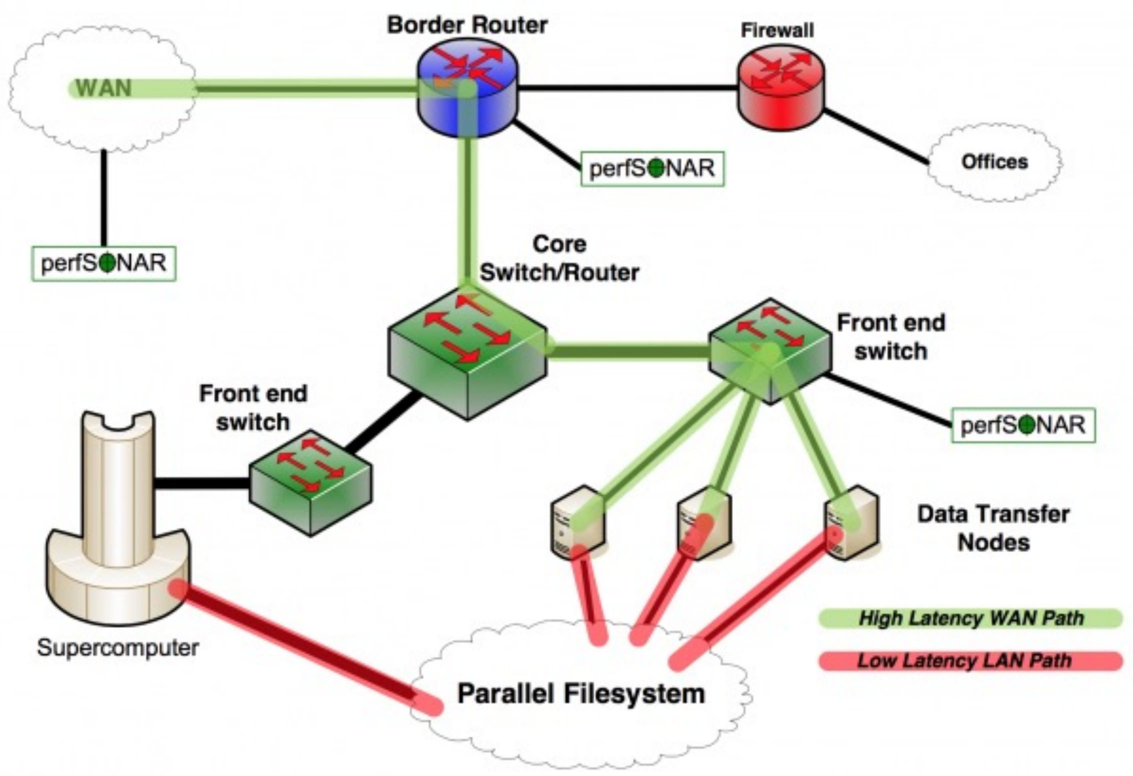
Tape Storage

- CCS is a member of OURRstore – a (sort of) free backup service hosted by the University of Oklahoma
- Purpose/Use: Archival or Backup storage
- Users purchase tapes and ship them to the University of Oklahoma
- User can then save data to their tapes
- One copy of the user's data is written to a tape that remains at the University of Oklahoma
- A second copy of the user's data is written to a tape that is mailed (postal service) to the user's institution
- Saved files must meet certain requirements
- High-speed transmission of data to OURRstore is possible via Globus or rclone
- Total OURRstore Tape Storage – depends on the # tapes purchased by the user
- Also note that UK ITS provides a tape backup service for a fee

CCS Network Resources

DTN Nodes and Cluster Interconnects

- Clusters are mostly located in Science DMZs
- DTN nodes connect to the Internet via 10 and 40 Gbps connections
- DTNs support Globus and other parallel transfer services (e.g., rclone).
- Most internal cluster networks are 100 Gbps Infiniband. A few are 100 Gbps Ethernet.



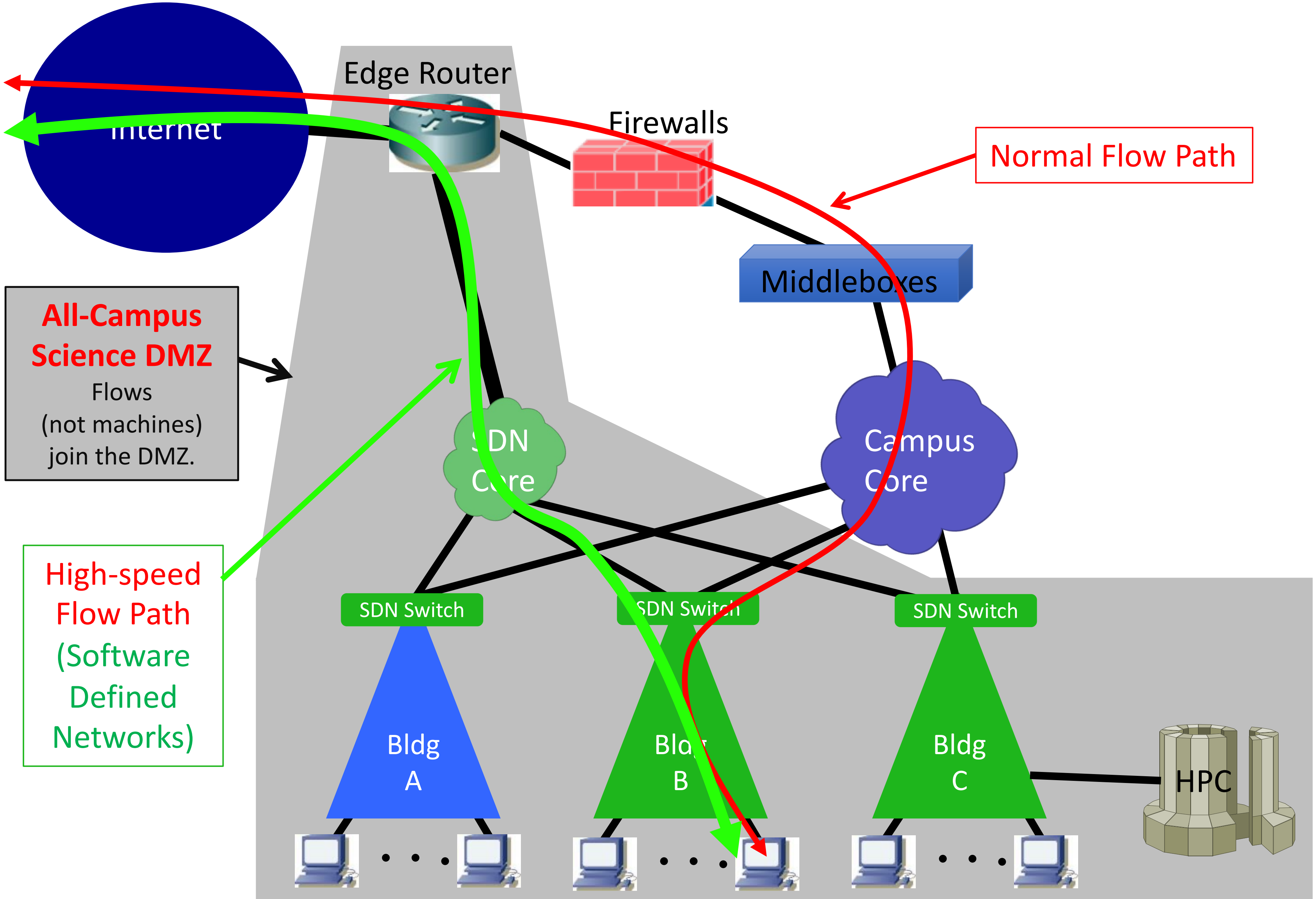
ESnet Science DMZ Architecture

(https://fasterdata.es.net/science-dmz/science-dmz-architecture/)

Globus

- Globus is a high-speed parallel file transfer service
- It offers an easy-to-use GUI to transfer data between Globus Endpoints
- It can be used to transfer data between CCS DTNs connected to CCS clusters and data storage systems, as well as user desktops/laptops
- It can also be used to transfer data between CCS DTN nodes and other sites on the Internet (e.g., OURRstore, ACCESS supercomputing centers, cloud providers, etc.)
- Globus also supports shared files and guest collections which are useful for collaboration

UKY All-Campus High-Speed Research Network

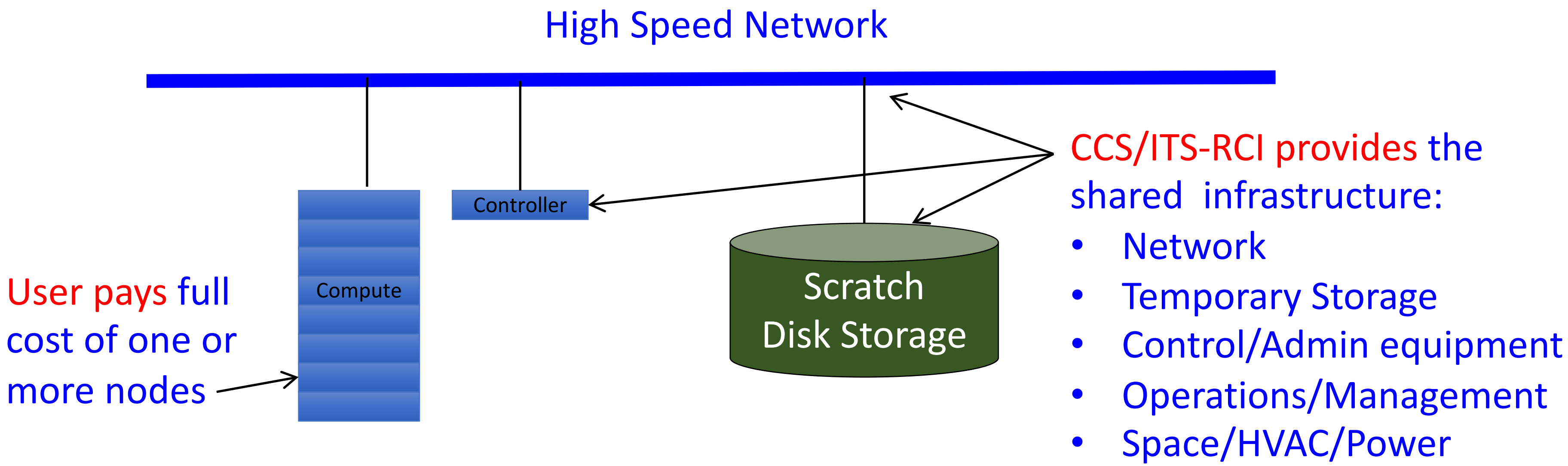


CCS Condo Resources

Competition for Resources

- Problem:
  - Expanding User Community
  - Increasing competition for limited resources
  - The need for new types of CI resources (not just RCD compute clusters)
  - Growth of (costly/inefficient/insecure) independent research infrastructures
- Idea: Condo Model
  - Researchers pay for equipment to be operated by RCD center and shared with other researchers when idle.
  - RCD center as part of its normal activity pays for staff to purchase, install, operate, administer, maintain, upgrade, secure, etc. the equipment on behalf of the researcher. RCD center pays for space, power, network connectivity, heating/cooling, fault-tolerant features, key infrastructure services, etc.

UKy Condo Model



Designed to encourage participation

- Significant CCS/ITS-RC investment in the shared parts
- Researchers purchase equipment – at cost
- Guaranteed priority access to purchased equipment (or equivalent)
- Unused Condo resources available for open access
- Now also offering Condo Storage – 100’s TBs storage, supported for 5 years

Allocations

- Allocation refers to a portion of a resource designated for use by a particular user or group of users.
  - Compute capabilities for some amount of time
  - Storage space for some amount of time
  - VM usage for some amount of time
- Reason/Purpose for an Allocation:
  - Condo: allocations based on the condo purchase
  - Educational: a small allocation used for students in an approved class
  - CCS Discretionary: awarded by CCS for critical/urgent unmet needs
  - Open Access: unlimited allocations to unused resources given to everyone

Allocation Priority Levels

Allocation Type	Priority
Condo User Allocation	1
Discretionary Allocation (CCS)	1
Education Allocation	2
Open Access (unlimited) Allocation	3

(Priority 1 is the highest priority)

- Allocations at Levels 1-2 are guaranteed (modulo scheduling constraints and allocation’s expiration date)
- Allocations at Level 3 are not guaranteed

Talk Outline

- The RCD Sea-change
- CCS Resources
    - CCS Compute Resources
    - CCS Storage Resources
    - CCS Network Resources
    - CCS Condo Resources
- $\boxed{\text{CCS Support}}$
    - Documentation and Getting Help
    - Account Setup and Onboarding
    - User Portals
    - Software
    - Usage Information
- Other Resources
    - ACCESS
    - FABRIC
    - NAIRR

Documentation
and
Getting Help

CCS Web Pages + Helpful Documentation

CCS Webpages:
- https://www.ccs.uky.edu/

CCS General Documentation and Support:
- https://ukyrcd.atlassian.net/wiki

CCS Cyberteam:
- https://ukyrcd.atlassian.net/wiki/spaces/KyCyberteam/overview

Cyberinfrastructure Partnership:
- https://kycyberteam.cyberinfrastructure.org/

More HELP:
- https://ask.cyberinfrastructure.org/

Find Help with:
- HPC Systems
- Storage
- Software Lists
- Policies
- Events
- Project Wikis
- Help Desk
- OpenStack Guide
- Getting Started
- File Transfers
- SLURM
- Condo Model
- Google Drive
- Educational Usage
- Containers

Account Setup

Requesting an Account

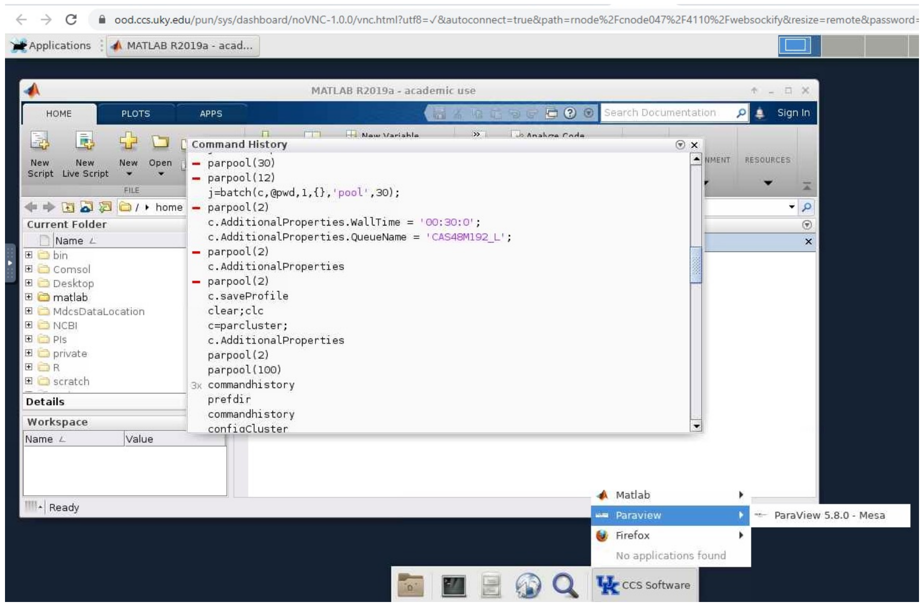
- Any faculty, staff, or student may request a User Account via a support ticket: https://ukyrcd.atlassian.net/servicedesk/
- We also collect details about the PI (faculty/staff) such as project detail, group research activities on the cluster, computational methodologies used, software needs, initial resource request if known(core hrs/storage, GPU, CPU), type of resources needed (HPC/Cloud/VMs), etc.
- We use this information to help find the most appropriate CCS resource (cluster) for the user.

User Portals

User Portals

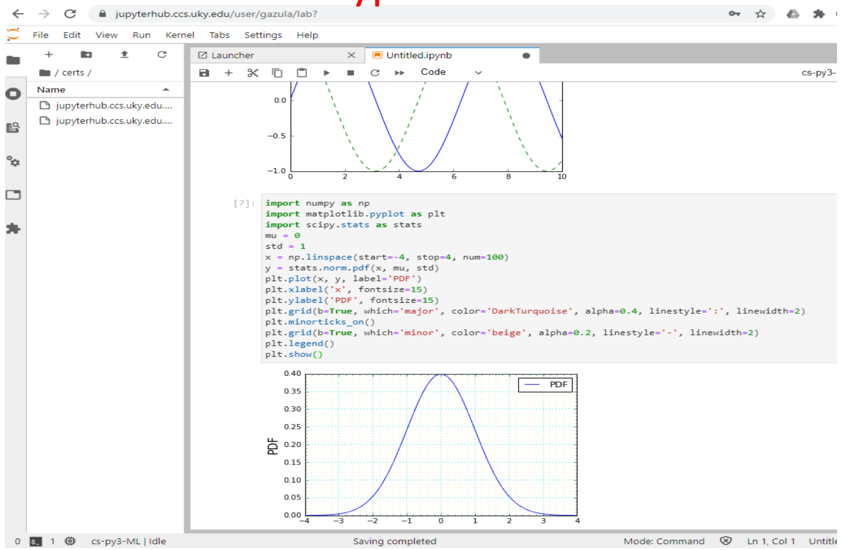
- **Open On-Demand**
  - *Open OnDemand* provides advanced web and GUI interfaces to make HPC easier to use and avoid traditional (terminal) command line interfaces.
  - MCC/LCC's *Open On-Demand* supports popular GUI based applications tools such as MATLAB, R Studio, Paraview, and others through web browsers.
  - https://ood.ccs.uky.edu/pun/sys/dashboard

Open On-Demand



- **JupyterHub**
  - *JupyterHub* is a multi-user server for *Jupyter Notebooks*.
  - Jupyter Notebooks are interactive computing environments used for data science
  - Notebooks contain *narrative text, live code, visualizations* in a single document that can be shared with other users for reproducibility and collaboration.
  - https://jupyterhub.ccs.uky.edu/hub/login

JuypterHub



Software

Example Cluster Software

- CCS supports more than 800 software packages and scientific libraries
- We support GUI-based applications using Open OnDemand
  - MATLAB, R Studio, Jupyter notebooks, ANSYS, Paraview, SAS, etc.
- Globus for web-based data transfers including sharing of data.
- Popular debugging tools
  - Totalview, gdb, hpctoolkit

Partial list of supported software:

- http://ukyrcd.ccs.uky.edu/confluence/CCS/HPC+Software+List

For containers:

- http://ukyrcd.ccs.uky.edu/confluence/CCS/Conda+and+Containers+on+LCC
- http://ukyrcd.ccs.uky.edu/confluence/CCS/Software+list+for+singularity+containers+for+conda+packages+in+the+MCC+cluster

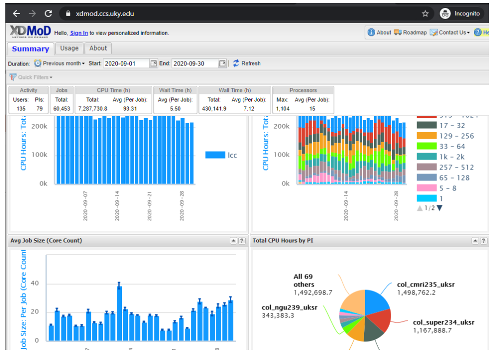
Virtual Machines (VMs) vs Containers

- On-prem Cloud Support
  - CCS/ITSRCI supports on-prem cloud services via OpenStack
  - Dynamically instantiate right-sized Virtual Machines (VMs)
  - User can control and customize the VM (any OS or applications)
  - Can leverage pre-defined VM images
- LCC/MCC support for Containers
  - Full support for Singularity Containers (lightweight virtualization)
  - Many advantages, but primarily useful for packaging software and reproducibility
    - Can easily pull containerized software from shared repositories and run in many environments (e.g., LCC/MCC nodes, desktop/laptop, cloud VM, etc)
    - Can scale research problem, starting small and moving to larger problems
    - Much of the LCC/MCC software has been containerized

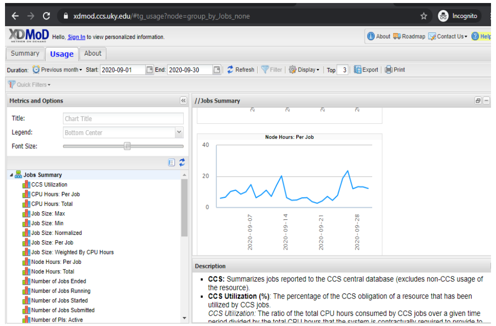
Usage Information

Usage/Performance Monitoring

- Useful to both users and CCS/ITSRCI staff for operations
- Open XDMoD
  - HPC monitoring tools to track users/groups, allocations, jobs, resources, etc. to understand how the supercomputer is being used.
  - Provides comprehensive metrics for CPU hrs. consumed, jobs executed, load on queues, individual user and group usage stats, etc.
  - https://xdmod.ccs.uky.edu/



- Prometheus/ThanOS/Grafana
  - Primarily useful to CCS/ITSRCI staff for operations
  - Provides detailed performance information about every node in the cluster (e.g., CPU load, memory usage, disk usage, power consumption, etc.)
  - Helps to quickly identify and resolve performance issues
  - https://monitor.ccs.uky.edu

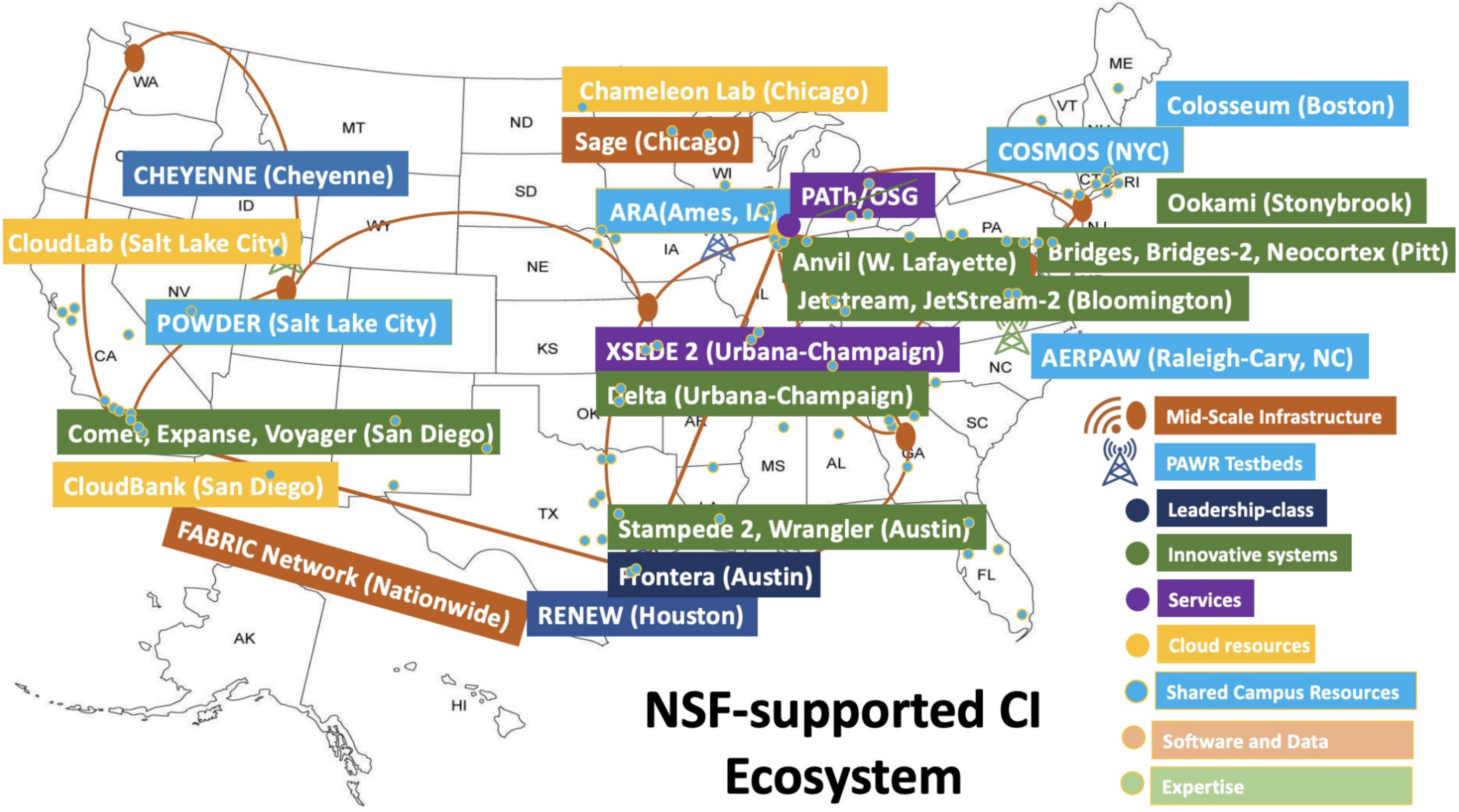


- LOKI: Promtail/Grafana/RSYSLOG
  - Primarily useful to CCS/ITSRCI staff for operations
  - Provides detailed information about system events (e.g., log messages)
  - Helps quickly identify errors, outages, inappropriate usage, attacks, etc.

Talk Outline

- The RCD Sea-change
- CCS Resources
  - CCS Compute Resources
  - CCS Storage Resources
  - CCS Network Resources
  - CCS Condo Resources
- CCS Support
  - Documentation and Getting Help
  - Account Setup and Onboarding
  - User Portals
  - Software
  - Usage Information
- Other Resources
  - ACCESS
  - FABRIC
  - NAIRR

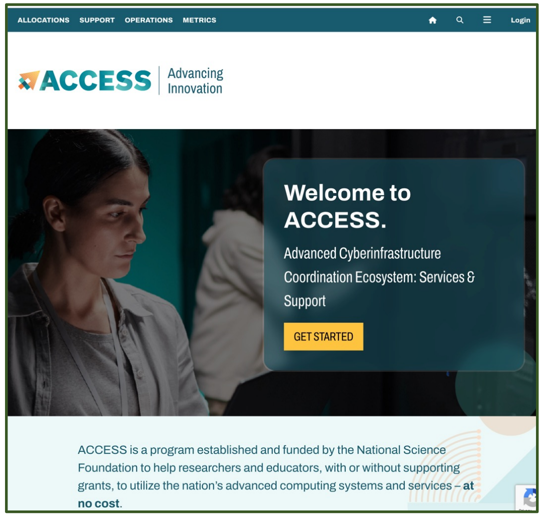
ACCESS



NSF-supported CI Ecosystem

Source https://ci-compass.org/assets/464282/2022_03_01_parashar_mfci_workshop.pdf

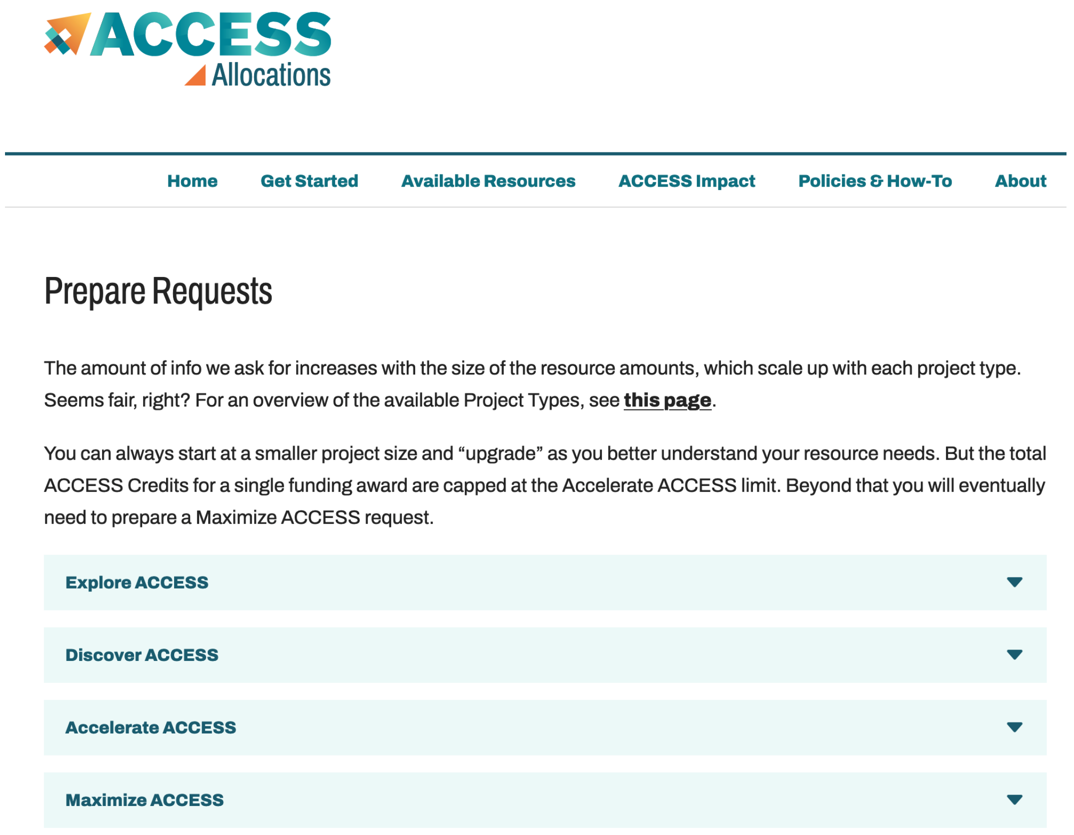
NSF ACCESS



- ACCESS is the follow-on to the XSEDE Program
- ACCESS provides an Interface to NSF-funded Cyber-Infrastructure including national supercomputers, large-scale storage systems, cloud computing resources, and leading-edge networks.
- ACCESS provides help with:
  - Operations – User (account) management for Resource Providers (RPs)
  - Allocations – Awarding users time on resources
  - Support – User support and help services
  - Metrics – Monitoring and measurement services
- Information about ACCESS can be found at: https://access-ci.org/

ACCESS Allocations

- There are multiple ways to request an ACCESS Allocation (ranging from very easy to more involved)
- Even the smallest Allocations provide significant compute time
- ACCESS supports easy-to-use User Interfaces including Open On Demand
- Allocations are transferrable
- There is a wide range of help and support available, including a Recommender System for new users.

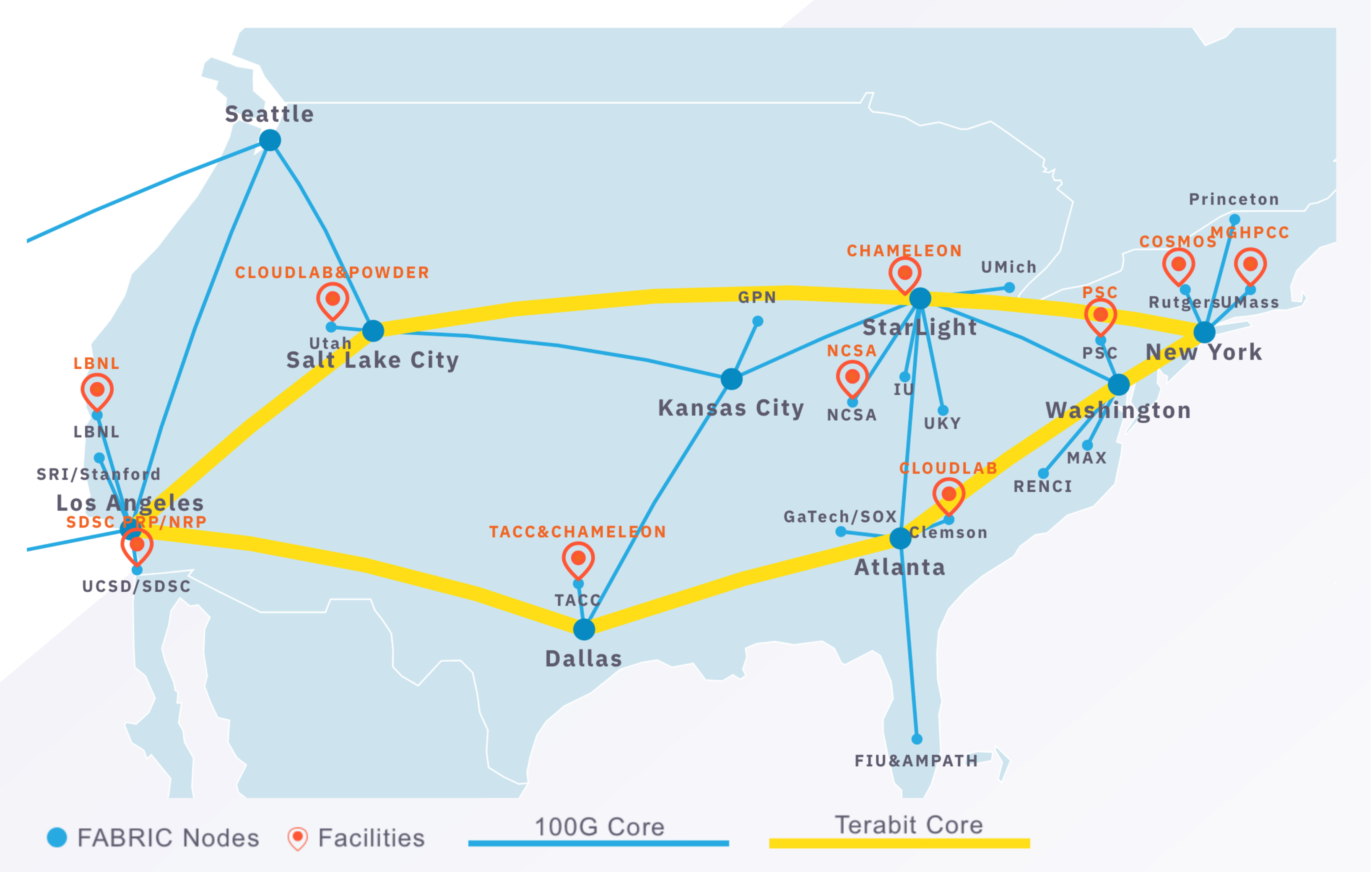


FABRIC

NSF FABRIC Network

- NSF FABRIC Network (https://whatisfabric.net) is an international network testbed with advanced data processing capabilities that can handle, transmit, and share massive amounts of data in completely new ways
- The FABRIC network itself is programmable (not just the end systems)
- Each FABRIC “router” has characteristics of a small supercomputer
- FABRIC has a TeraCore ring that spans the continental U.S. offering transmission speeds of 1.2 Tbps
- FABRIC interconnects many NSF CyberInfrastructure facilities (including ACCESS resources)
- FABRIC recently transitioned from “construction” to “operations” and is now taking on users.

FABRIC Network



NAIRR

National AI Research Resource (NAIRR)

- A NAIRR Pilot was recently launched that makes AI resources available to researchers (https://nairrpilot.org/)
- NAIRR is a multi-agency initiative that involved “connecting U.S. researchers to responsible and trustworthy Artificial Intelligence (AI) resources, as well as the needed computational, data, software, training, and educational resources to advance research, discovery, and innovation.” (https://nairrpilot.org/about)
- Pilot resources include computational allocations for research and education at facilities such as (Summit, Delta, Frontera, and Neocortex) as well as data sets, trained models, etc.

Thank you

Questions?

https://www.ccs.uky.edu