THE GEORGE WASHINGTON UNIVERSITY
WASHINGTON, DC

Research Computing and Cloud Resources at GWU: Use Cases and Solutions

CLARK GAYLORD
DIRECTOR, RESEARCH TECHNOLOGY SERVICES
CGAYLORD@GWU.EDU

THE GEORGE WASHINGTON UNIVERSITY
WASHINGTON, DC

Cloud consultancy and deployments

- Numerous engineered solutions using cloud resources.
- Retain each environment's configuration for reusability, reconfiguration, redeployment
- Engage researchers in consultancy and research solutions
- Examples:
  - Controlled Unclassified Information environment
  - Departmental research computing and data processing environment
  - Dynamic provisioning system for more common use cases
  - Project-specific data management, database hosting

THE GEORGE WASHINGTON UNIVERSITY
WASHINGTON, DC

Protected data environment

- Controlled unclassified information (similar for HIPAA, etc.)
- Biomedical research for defense application
  - "Secure workstations" in cloud
  - Connect with laboratory instruments
- Project requirement for data sharing with other institutions
  - Federated identity: Globus
  - Separate AWS VPC from the workstation environment

THE GEORGE WASHINGTON UNIVERSITY
WASHINGTON, DC

Department/lab research computing

- Several applications, servers, workstations
- Supporting various databases, teaching
- Applications and workflows: R Studio Server, Shiny, Galaxy
- Instead of several physical servers or virtual machines in data center, hosted in AWS cloud, some on-demand

THE GEORGE WASHINGTON UNIVERSITY
WASHINGTON, DC

"Dynamic cloud platform"

- Based on Amazon's Service Workbench
  - Useful for specified templates to deploy
  - Some data sharing between deployments can be supported
- Stand-alone VMs
  - Long running instances, perhaps needing desktop experience (not batch friendly)
  - Student web sites
- Templated solutions are the key
  - Could include HPC or Galaxy

THE GEORGE WASHINGTON UNIVERSITY
WASHINGTON, DC

Project specific database hosting

- Many projects for a single research program
- Project-specific requirements in terms of performance, cost, retention
- Often security controls are required
  - Configurable network access
- Leverage AWS RDS with ancillary compute and storage resources as needed

THE GEORGE WASHINGTON UNIVERSITY
WASHINGTON, DC

Various other example research platforms

Note: all these are "on premise" support efforts

- Multi-TB MySQL server for data mining of large dataset (Visual Cognition)
- Specialized storage and virtual desktop environment (School of Public Health)
- Database and compute environment for orchestrated pipelines (BioCompute)
- Lab specific engineering support (Nanofabrication and Imaging Center)
- Numerous “boutique” clusters dedicated to research groups or labs (Engineering, Arts & Sciences, Medicine)

SURA WORKSHOP
TULANE UNIVERSITY

Tulane

SAAS BROKER

Researcher Need

- NIH recognizes the value of conducting studies across ***multiple centers and*** mine research data holistically
- Tulane Primate Center was awarded to act as a ***data coordination center for a multi-site/center study*** working on data collection and analysis platform for COVID non-human subject research
- Task Order needed a ***highly secured platform*** stood up within ***2-3 months (FEDRAMP/FISMA)***

Approach

- Tulane IT needs a ***highly secured full stack Electronic Data Capture platform*** with little lead time.
- ***SaaS vendor*** is the only solution with the ***rich requirements*** in ***features and cybersecurity and short timeline***.
- IT will provide resources to administer the platform going forward after the 1st study as a ***full managed service***.

Technical Delivery

- Tulane had already completed an in-depth review of different EDC vendor choices.
- After selecting a vendor, it brokered a relationship with the SaaS provider to host a unified data collection platform with the required security to allow sharing research data across the institutions.
- Professional services collaborated with Tulane Primate Center PI & IT and delivered the solution in less than 2 months

Business Value

- This managed clinical research SaaS service allows a faster start up for future clinical research opportunities with predictable cost structure.
- Tulane can provision new studies in weeks instead of months because of the existing approved master services agreement & skillset.
- Becomes a solid stepping-stone to create an RFP to extend this solution to become a Tulane wide Life Science Clinical Full-Service Platform

Tulane

CMMC In The Microsoft Cloud

Secure Enclave in Microsoft Azure

THE UNIVERSITY OF SOUTHERN MISSISSIPPI

USM Research Requirements

The University of Southern Mississippi in preparation for Cybersecurity Maturity Model Certification (CMMC) requirements including those that will be required per the DoD DFARS 252.204-7012 Interim Rule has initiated this project to enable researchers to continue supporting DoD and other initiatives that will require this compliance.

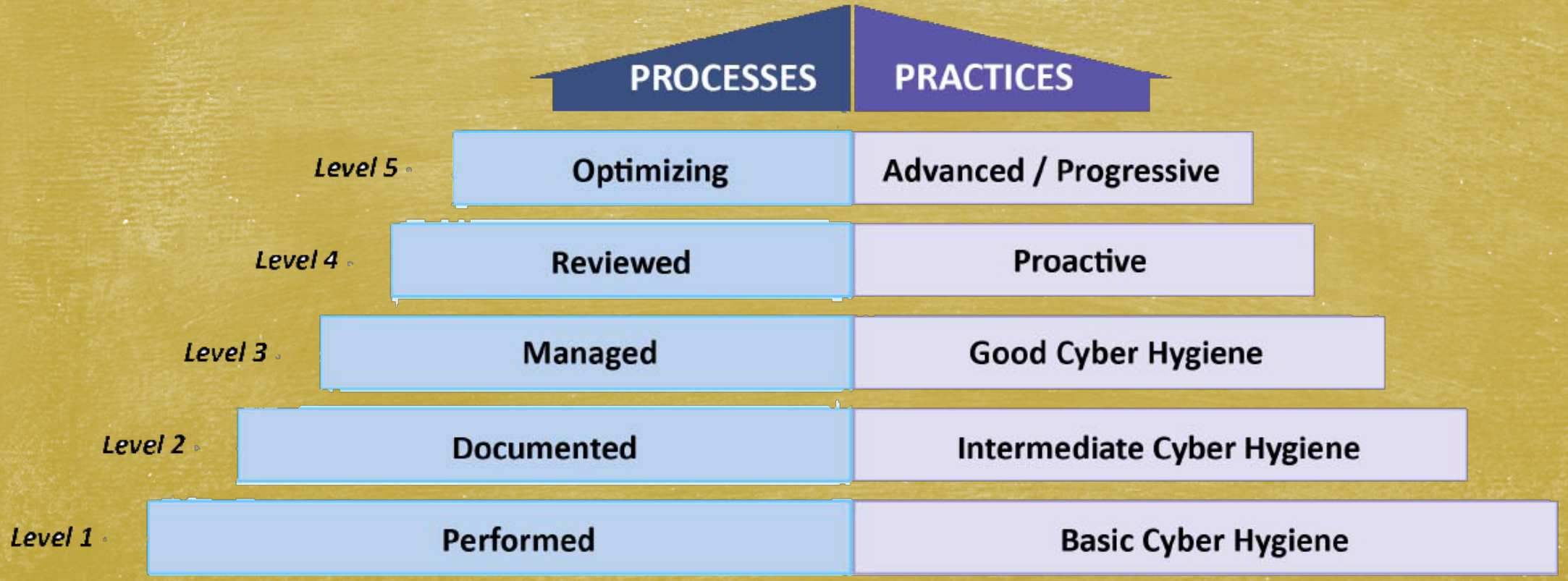
THE UNIVERSITY OF SOUTHERN MISSISSIPPI

USM Research Requirements

The University of Southern Mississippi is seeking a scalable Microsoft Office 365 GCC High messaging and collaboration platform for researchers requiring CMMC and NIST 800-171 compliance, enabling them to accept federally funded grant awards subject to these requirements. The platform MUST be DFARS -7012 and CMMC Level 3 compliant.

THE UNIVERSITY OF SOUTHERN MISSISSIPPI

CMMC Levels



https://www.thecoresolution.com

THE UNIVERSITY OF SOUTHERN MISSISSIPPI

CMMC Level 3 Chosen

Processes: Managed

Level 3 requires that an organization establish, maintain and resource a plan demonstrating the management of activities for practice implementation.
The plan may include information on missions, goals, project plans, resourcing, required training, and involvement of relevant stakeholders.

https://www.preveil.com/blog/what-are-the-5-levels-of-cmmc/

THE UNIVERSITY OF SOUTHERN MISSISSIPPI

CMMC Level 3 Chosen

Practices: Good Cyber Hygiene

Level 3 focuses on the protection of CUI and encompasses all of the security requirements specified in NIST SP 800-171 as well as 20 additional practices to mitigate threats. Any contractor with a DFARS clause n their contract will need to at least meet Level 3 requirements. Note that DFARS clause 252.204-7012 applies and specifies additional requirements beyond NIST SP 800-171 security requirements such as incident reporting.

https://www.preveil.com/blog/what-are-the-5-levels-of-cmmc/

THE UNIVERSITY OF SOUTHERN MISSISSIPPI

USM Research Requirements

DoD contracts currently require that projects implement controls to safeguard government owned data. USM is not easily able to comply with the government requirements in our current network environment.
Two possible approaches for addressing the information security requirements around DoD contracts are as follows:

- USM can physically deploy a separate network on campus
- USM can setup a secure enclave on the internet

THE UNIVERSITY OF SOUTHERN MISSISSIPPI

USM Bid Process

The University of Colorado, Boulder is seeking qualified quotes from Bidders for the CMMC MS O365 GCC-High Tenant Project

Microsoft O365 GCC High Platform-Office 365 Govt G5 License Procurement & Fast Track Implementation

University of Colorado
Boulder | Colorado Springs | Denver | Anschutz Medical Campus

THE UNIVERSITY OF SOUTHERN MISSISSIPPI

USM Bid Process

https://www.coloradobids.us/colorado-bids/bids-NBD15675706752464107.htm

Search for "University of Colorado" "PSC-U-1411"

https://oit.colorado.edu/oit-projects

University of Colorado
Boulder | Colorado Springs | Denver | Anschutz Medical Campus

THE UNIVERSITY OF SOUTHERN MISSISSIPPI

USM Vendor Selection

Microsoft
US Partner Award 2020
SUMMIT7
Security and Compliance

2 Parade St NW
Huntsville, AL 35806
256.585.6868
info@summit7.us
cmmc@summit7.us

https://www.summit7.us

THE UNIVERSITY OF SOUTHERN MISSISSIPPI

What Is Included In The Solution

- Office 365 GCC High Licensing – recurring annually
- Azure Government Subscription – recurring monthly
- Office 365 GCC High Backup – recurring annually
- CMMC Level 3 Windows 10 MFA – setup cost, then recurring annually
- CMMC Level 3 Office 365 GCC High – one time cost to configure tenet
- CMMC Level 3 Secure Enclave – one time cost to architect enclave

THE UNIVERSITY OF SOUTHERN MISSISSIPPI

First Things First

- Apply for Office 365 Government GCC High
- This process can take 15 business days to receive a Category 3 eligibility from Microsoft, once received it is valid for 3 years
- There are 3 types of documentation you can submit for eligibility:
  - Submit a contract that calls out ITAR, Export Controlled, CUI or DFARS 7012 as requirements
  - Submit a Department of State DS-2032 form
  - Sponsorship letter signed by a government official

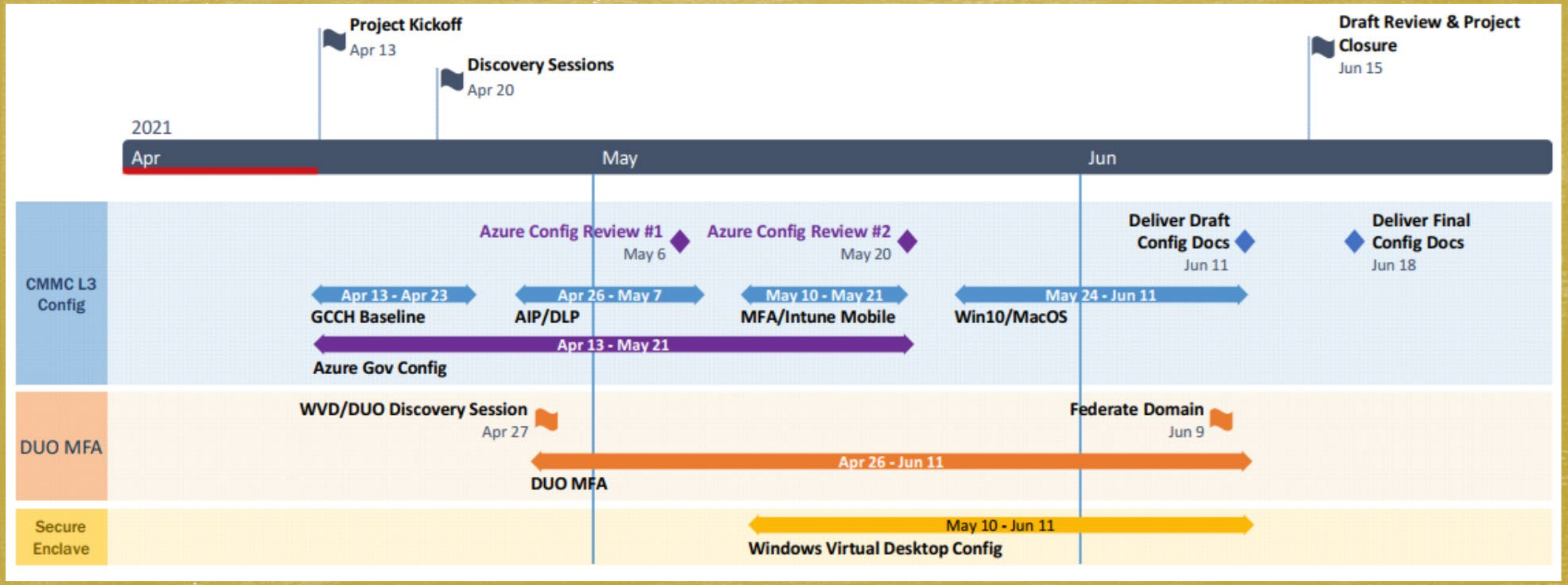
THE UNIVERSITY OF SOUTHERN MISSISSIPPI

Project Kickoff

- Select and create a name for your GCC High tenant
- Multi-Factor Authentication Setup
  - DUO Security Federal Configuration
  - DUO ADFS & Federation
- Cloud Backup for Office 365
  - AvePoint Federal Cloud Service

THE UNIVERSITY OF SOUTHERN MISSISSIPPI

Project Timeline



Office 365

- Office 365 Tenant CMMC Level 3 Configuration
  - Office 365 Tenant
  - Azure Identity Management
  - SharePoint Online
  - Exchange Online
  - Advanced Threat Protection (ATP)
  - Intune Enterprise Mobility (Mobile Devices: Phones, Tablets)
  - Intune Windows 10 Devices (Computers: Laptops, Desktops)
  - Azure Information Protection
  - Label Creation (CUI, Company Proprietary, [External] Proprietary, Public)

THE UNIVERSITY OF SOUTHERN MISSISSIPPI

Azure

- Azure Tenant CMMC Level 3 Configuration
  - Deploy Azure CMMC L3 Baseline Configuration
  - Deploy and Configure Next Generation Cloud Firewall - Azure Firewall
  - Deploy and Configure 3 Virtual Networks, 4 Network Security Groups
  - Build 2 Domain Controllers to CMMC L3 Standard
  - Build AD Connect Server and Sync with Azure Active Directory
  - Deploy and Configure Office 365 GCC High Backup (AvePoint)
  - Implement the Approved AD Design to CMMC Level 3

THE UNIVERSITY OF SOUTHERN MISSISSIPPI

Contact Information

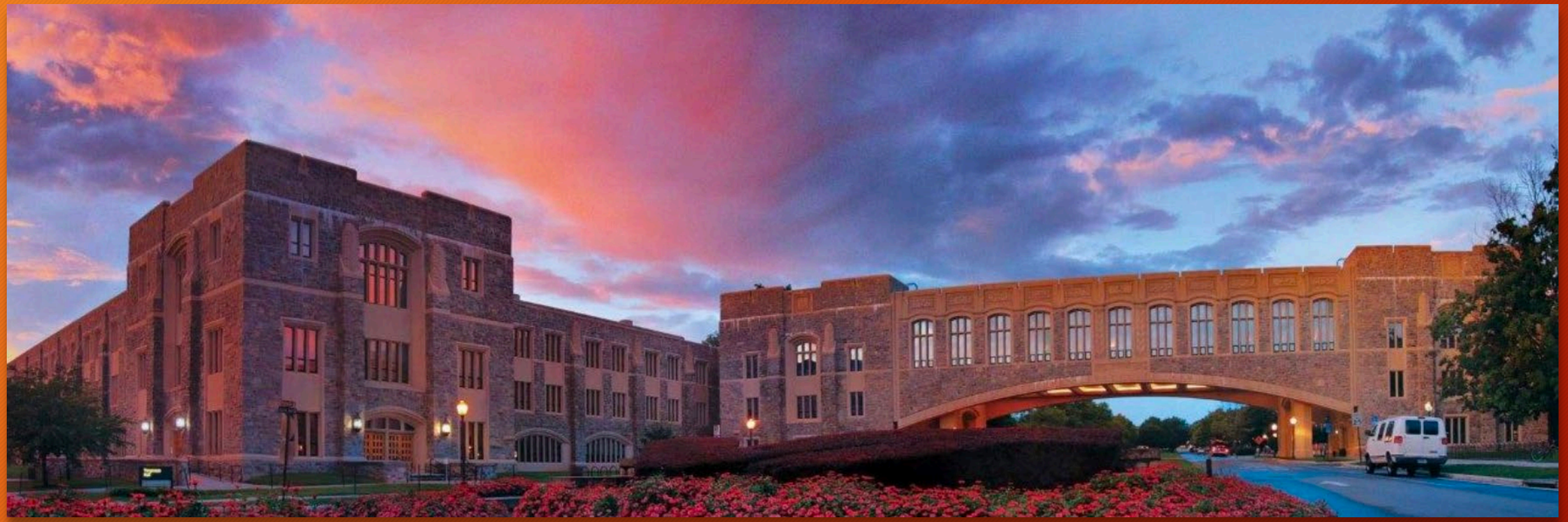
David J. Sliman

david.sliman@usm.edu

601-266-4190

https://www.usm.edu/itech/

THE UNIVERSITY OF SOUTHERN MISSISSIPPI



Open OnDemand as a single pane of glass

Robert Settlage
Advanced Research Computing @Virginia Tech
April 2021

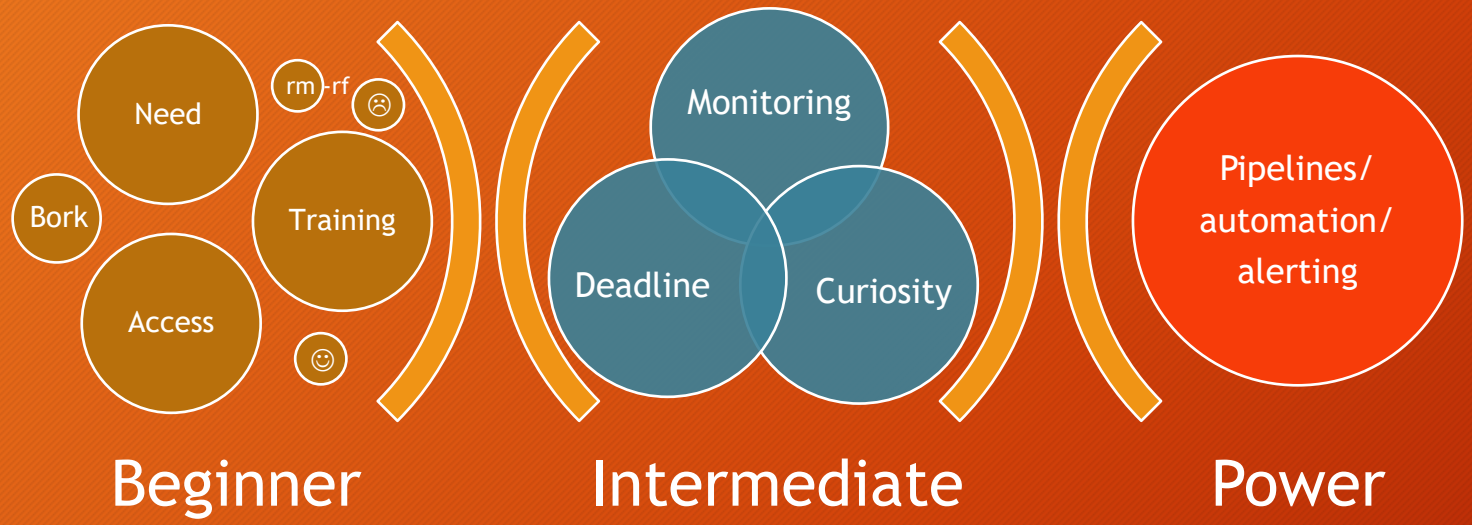
OPEN OnDemand

University at Buffalo Center for Computational Research

Ohio Supercomputer Center

VIRGINIA TECH

User spectrum and experience



GOAL: Facilitate and speed the progression

Account request

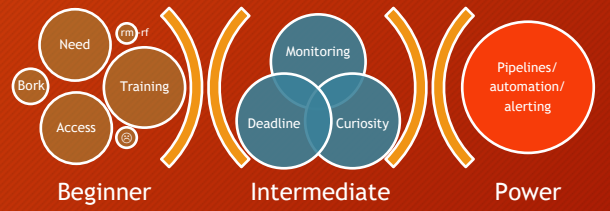
Resource request

Job request

Result viewing

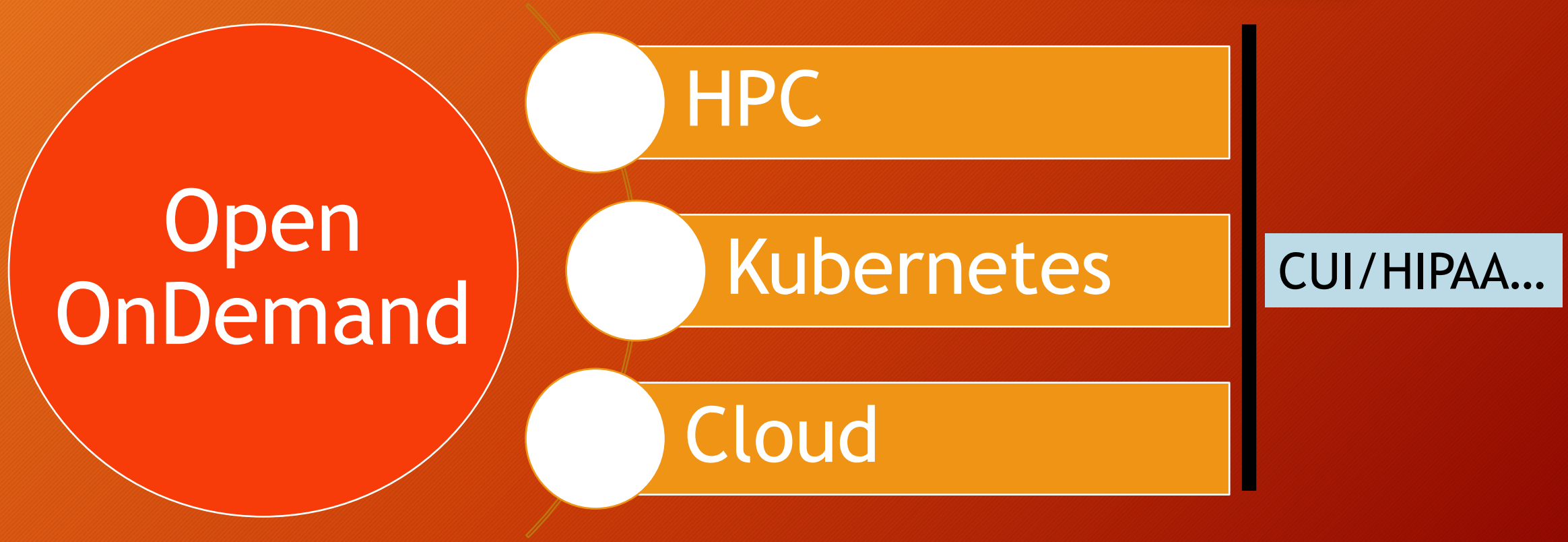
Resource utilization

Usage monitoring

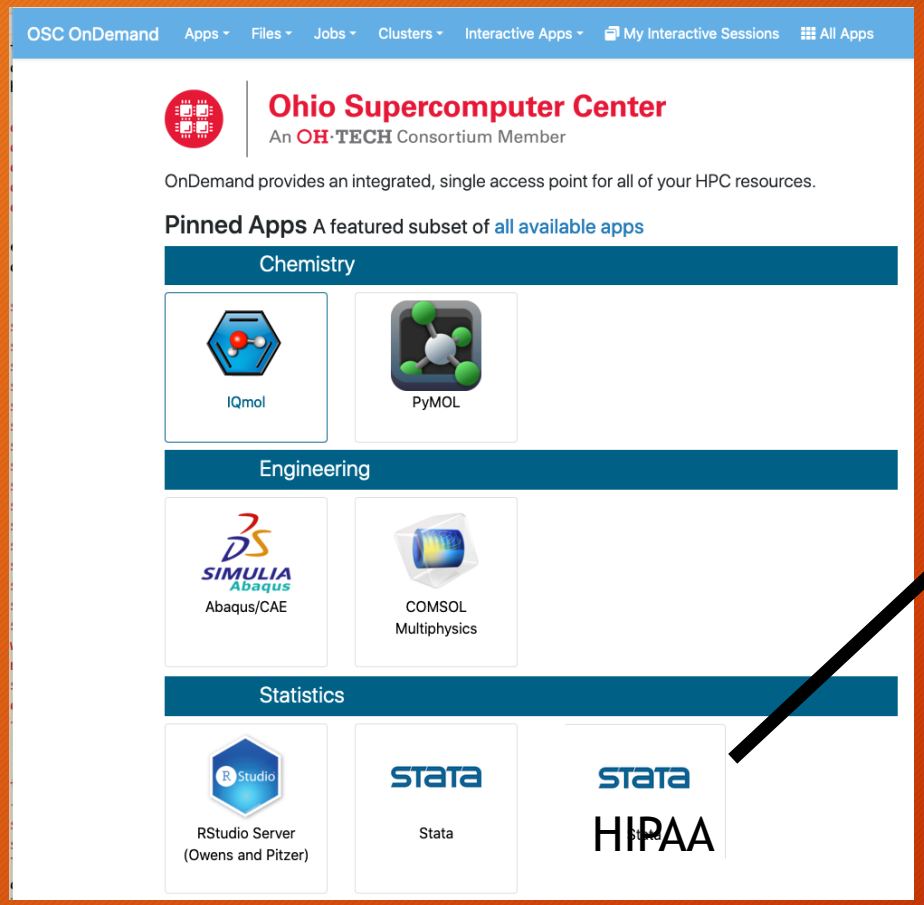


ColdFront | Open OnDemand | XDMoD | ColdFront

Unify the user experience



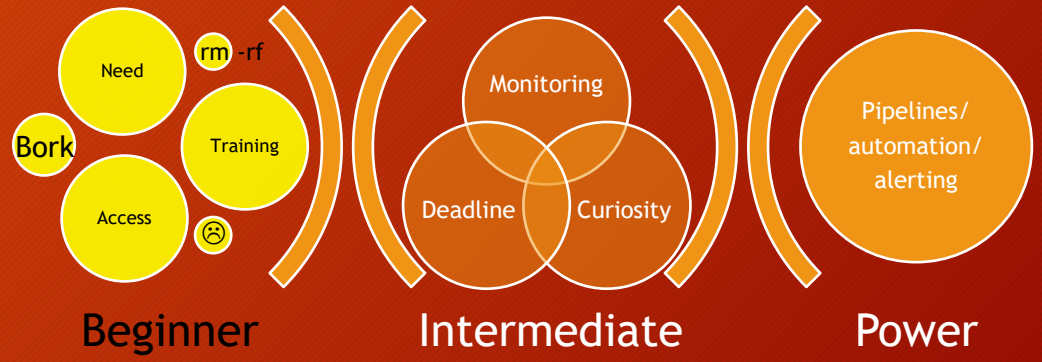
Open OnDemand 2.0 – user perspective



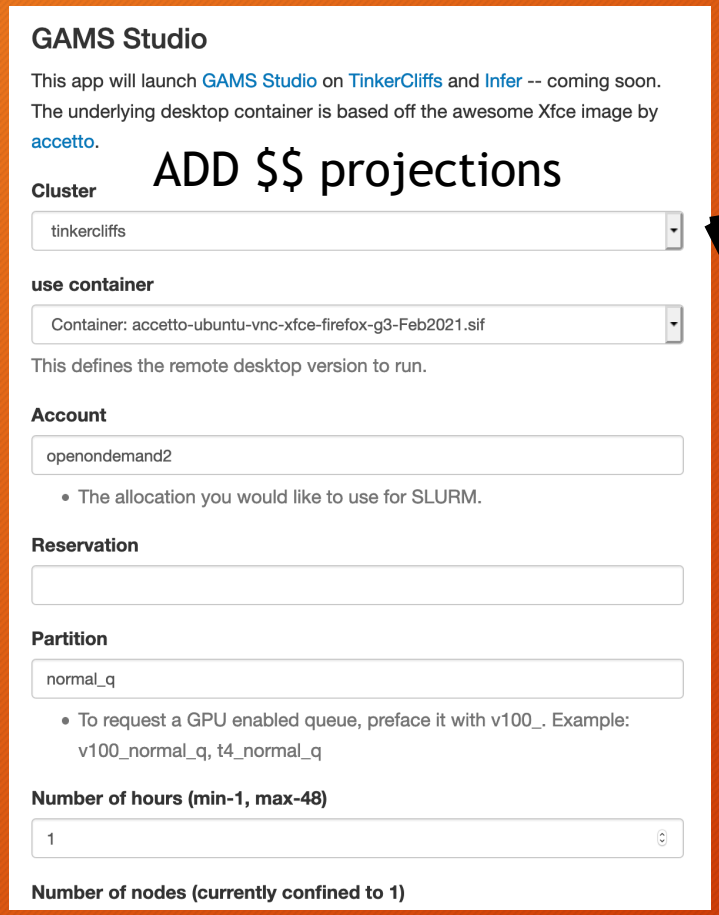
What and how matter. Where is a distraction.

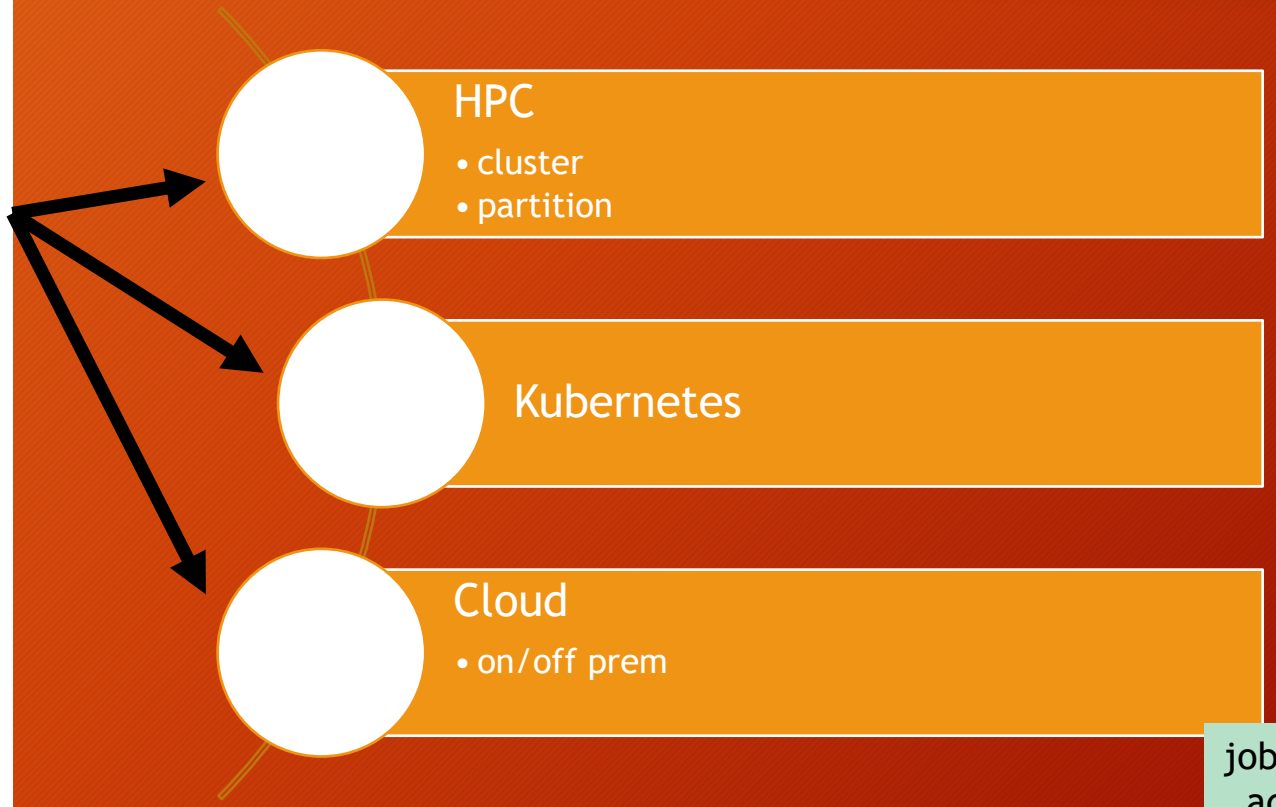
- Browser based HPC access
- Zero user install
- Site customizable

- ACL controls visibility/access to apps – CUI/HIPAA/Class/….



Where is a (sometimes) optional choice, with consequences

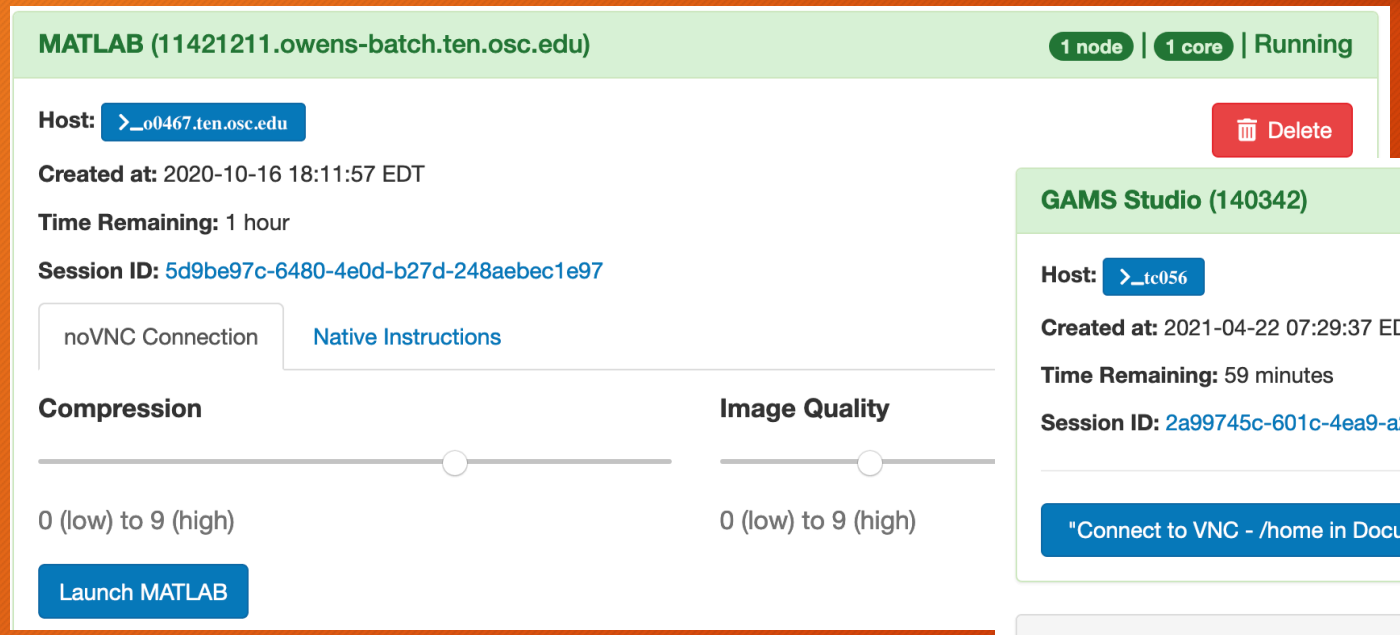


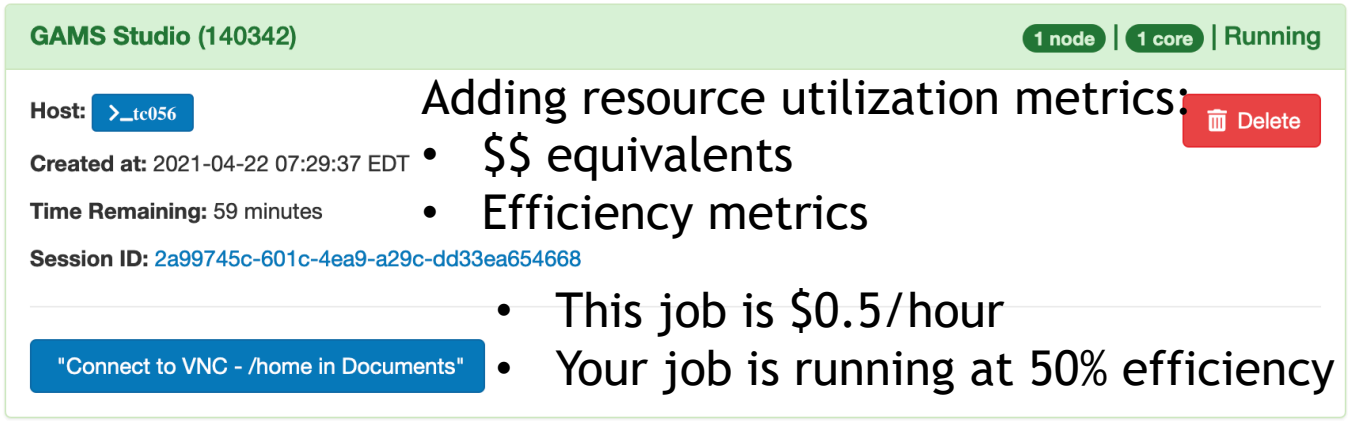


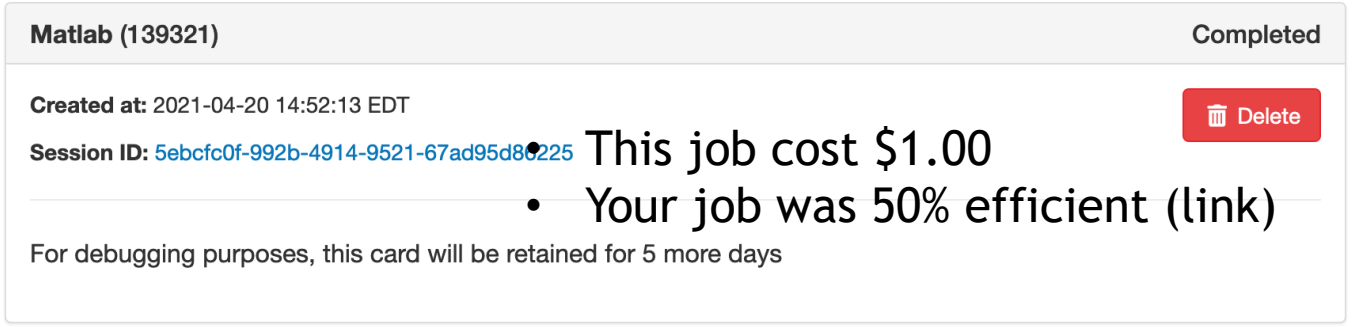
$

```
job:
  adapter: "kubernetes"
  cluster: "ood-dev"
  bin: "/usr/local/bin/kubectl"
```

Open OnDemand: running jobs







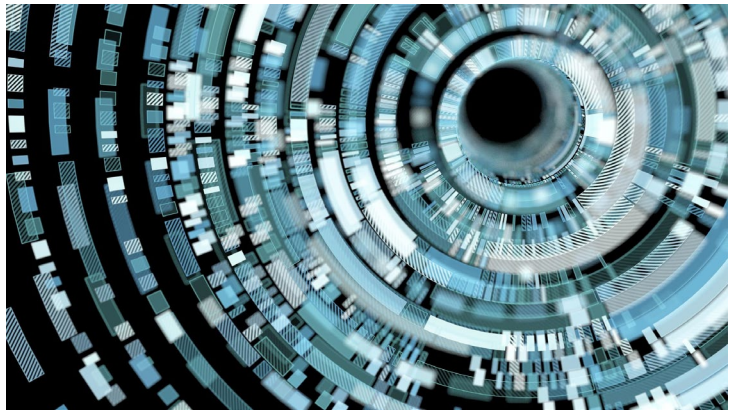
- Reverse proxy setup automated
- NoVNC
  - Image/connection settings
- Direct SSH to node
- View only sharable link

Adding resource utilization metrics:

- $$ equivalents
- Efficiency metrics
  - This job is $0.5/hour
  - Your job is running at 50% efficiency
  - This job cost $1.00
  - Your job was 50% efficient (link)

Questions?

Robert Settlage
Advanced Research Computing @Virginia Tech
Oct 2020



RESEARCH COMPUTING IN THE CLOUD

EXAMPLES OF CLOUD USAGE FOR RESEARCH AT USC

UNIVERSITY OF
South Carolina

Research Computing using Cloud Services

- Campus IT uses AWS and Azure
  - Mostly storage for AWS, Azure for Microsoft Services
- UofSC Research Computing usually Cloud agnostic (AWS, GCP, Azure, OCI)
- Challenges in billing end users
- As a state institution, required to use resellers (Onyx, Enquizit)
- Challenges in building cost models for charging grants

- Free resources for teaching AWS
  https://aws.amazon.com/education/awseducate/

Cloud Compute

SAS Viya in GCP for HIV ML Analysis

Need:
Researcher to explore and identify new patterns/combinations of predictors associated with HIV medical care utilization.

Approach:

Ingest and analyze data sets by combining both Statistical and Machine learning approaches.

Technical Delivery:
Delivered SAS Viya package SaaS on GCP by building core infrastructure utilizing automation and custom machine types to minimize cost.

Business Value:
HIV researcher can simply sign up, log in and get to work and tune the solution to their requirements and focus on solving analytic challenges and quickly realize value. Overall solution cuts infrastructure costs by providing on-demand allocations and services.

Delivering Secure Healthcare Insight in AWS

Need:

Provide analysis, core data, and visualizations to healthcare providers, insurance companies, and state government agencies to guide health care decisions, optimize cost. Includes massive healthcare provider data sets, like treatments and outcomes, to inform and optimize resourcing in specific hospital and provider locations. Must be highly secure and scalable.

Approach:

Design and build PaaS to support dynamic and scalable data volume, access, and usage requirements while building framework for modular application structure.

Technical Delivery:

Delivered fully automated PaaS on AWS by building automation, secured core framework and infrastructure, automated data and development pipelines, and provided training and knowledge transfer to use and manage platform.

Business Value:

Enabled research organization efficiently collect data and use provided pipeline to provide insight through analysis and data visualizations, support dynamic scaling to as many users as needed, and trained developers in AWS framework. The flexible pipeline allows for the import of any dataset. The researchers Used pipeline infrastructure to provide COVID data to the state of South Carolina.

Univ of South Carolina

Metagenomics on GCP

Need:

Researcher needed to perform bioinformatics analysis on enormous metagenomics environmental datasets to investigate areas like climate change on ecosystems over time.

Approach:

Build pipeline and software tools to parallelize and distribute work on massive scale.

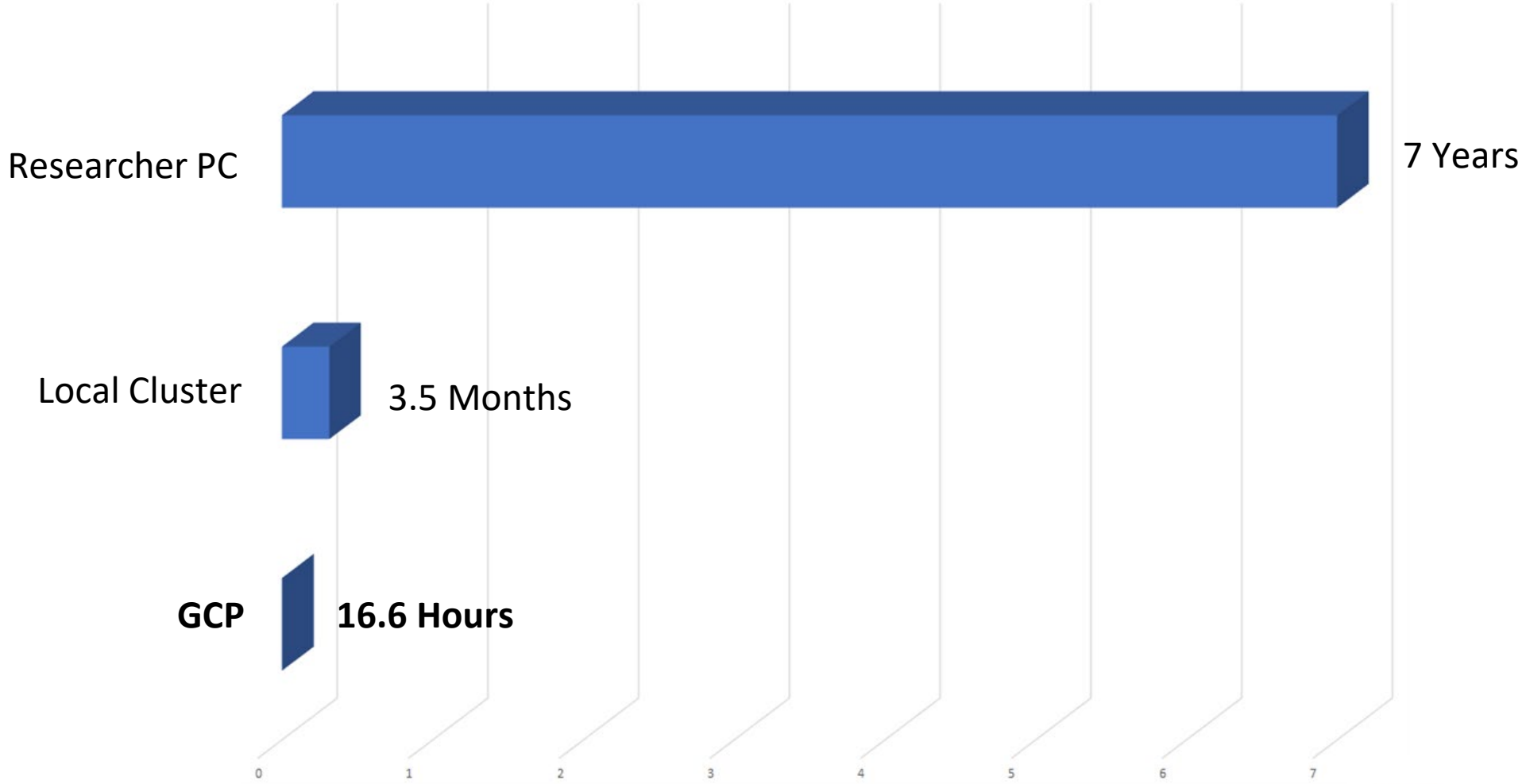
Technical Delivery:

Developed automated HPC tools within GCP to automatically scale to size of data set and efficiently leverage persistent storage features. Minimize spin-up and spin-down to cut costs.

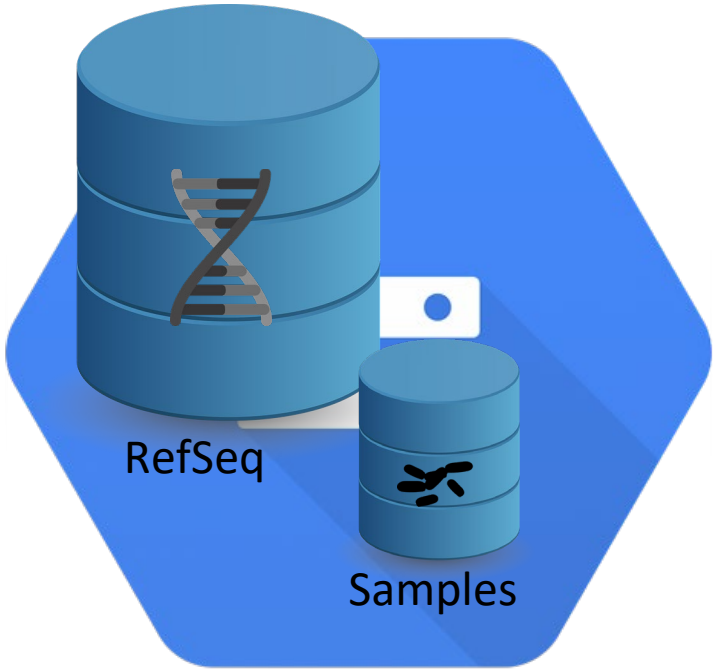
Business Value:

Scaled HPC style-job to massive scale, not available on any local resources. Cut computation time down from months to hours.

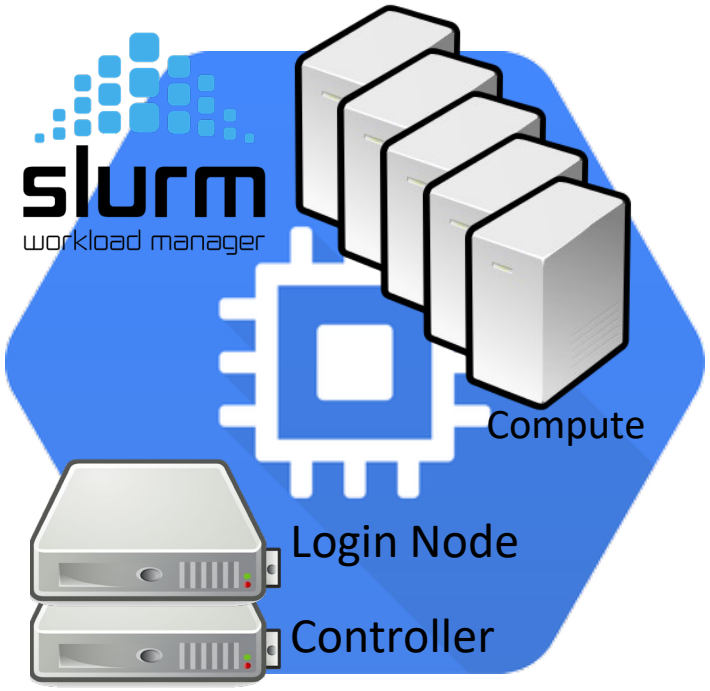
Time to Compute



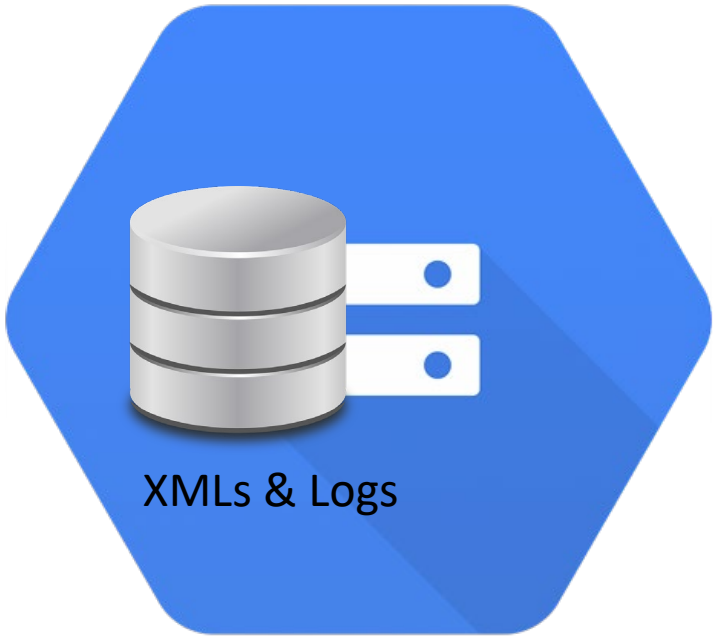
Google Cloud Platform



Attached Persistent SSD
(Read Only)



Dynamic HPC Cluster



Google Cloud Bucket

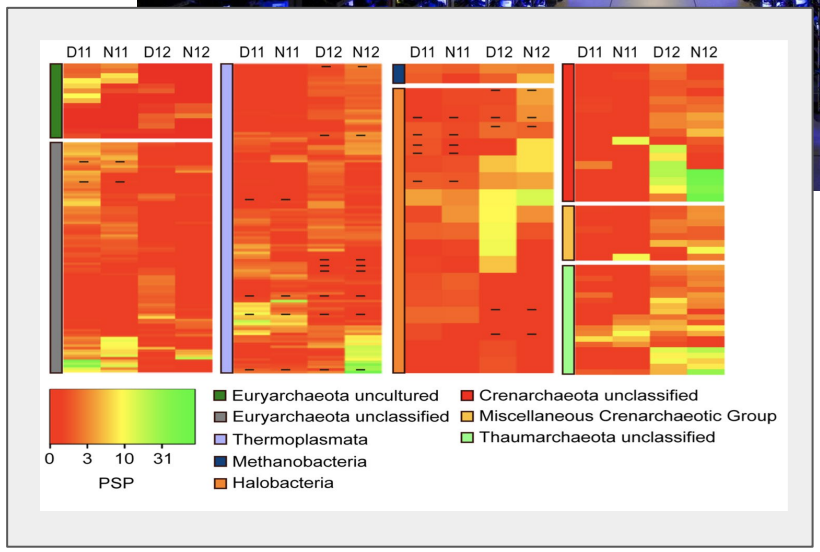
124,352 CORES

3886 NODES

16.6 HOURS

Challenges and Findings

- Enormously successful
- Encountered I/O bottlenecks during testing
  - Tried cluster NFS storage
  - Tried Elastifile storage
  - Tried copying to local disk
- Slurm FUN!
- Hyperthreading added run time



Cloud Storage

Research Computing Cloud Storage

Qumulo Shift for AWS S3

- Replication of the qumulo based filesystem, holding our user home and department/lab shares to buckets on S3
- Allows RC to offer fast, seamless replication of user data to AWS for cloud processing or archive

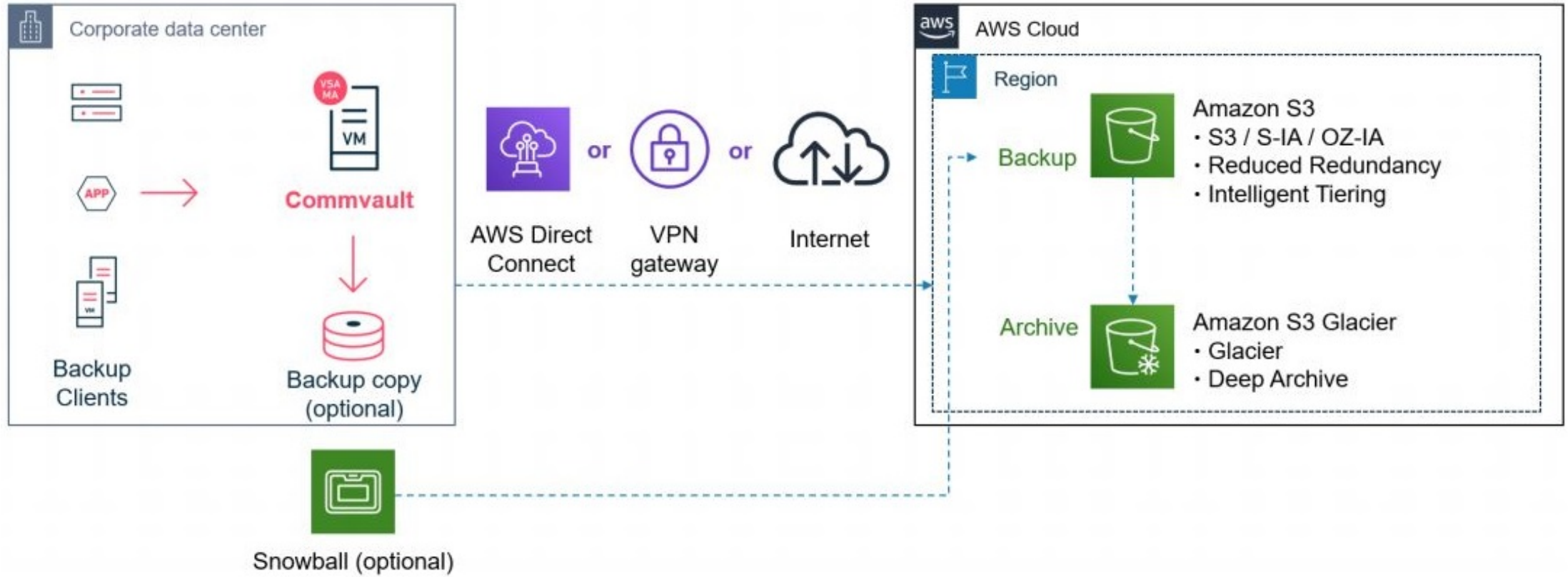
Globus Data Transfer Network

- Data transfer nodes connect UofSC research computing sites with research institutions globally.
- Allows high speed, secure data transfers between globus endpoints.
- Globus for Amazon S3 provides an instant link from S3 to Globus endpoint storage locations
- Log in and seamlessly move and share data stored in S3 across any storage resources using a unified, intuitive interface

UofSC Restricted Data File Share in AWS

- AWS FSx for Windows
- Encrypted in Transit
- Restrict ingress through on campus firewalls, AWS security groups, isolated subnets for each FSx file system
- FSx access limited through on-campus VPN or virtual desktops with no internet access
- DFS Namespaces to create user friendly alias and flexibility to change source in the future
- Enable VSS (Volume Shadow Service) on FSx to enable user to perform data recovery operations

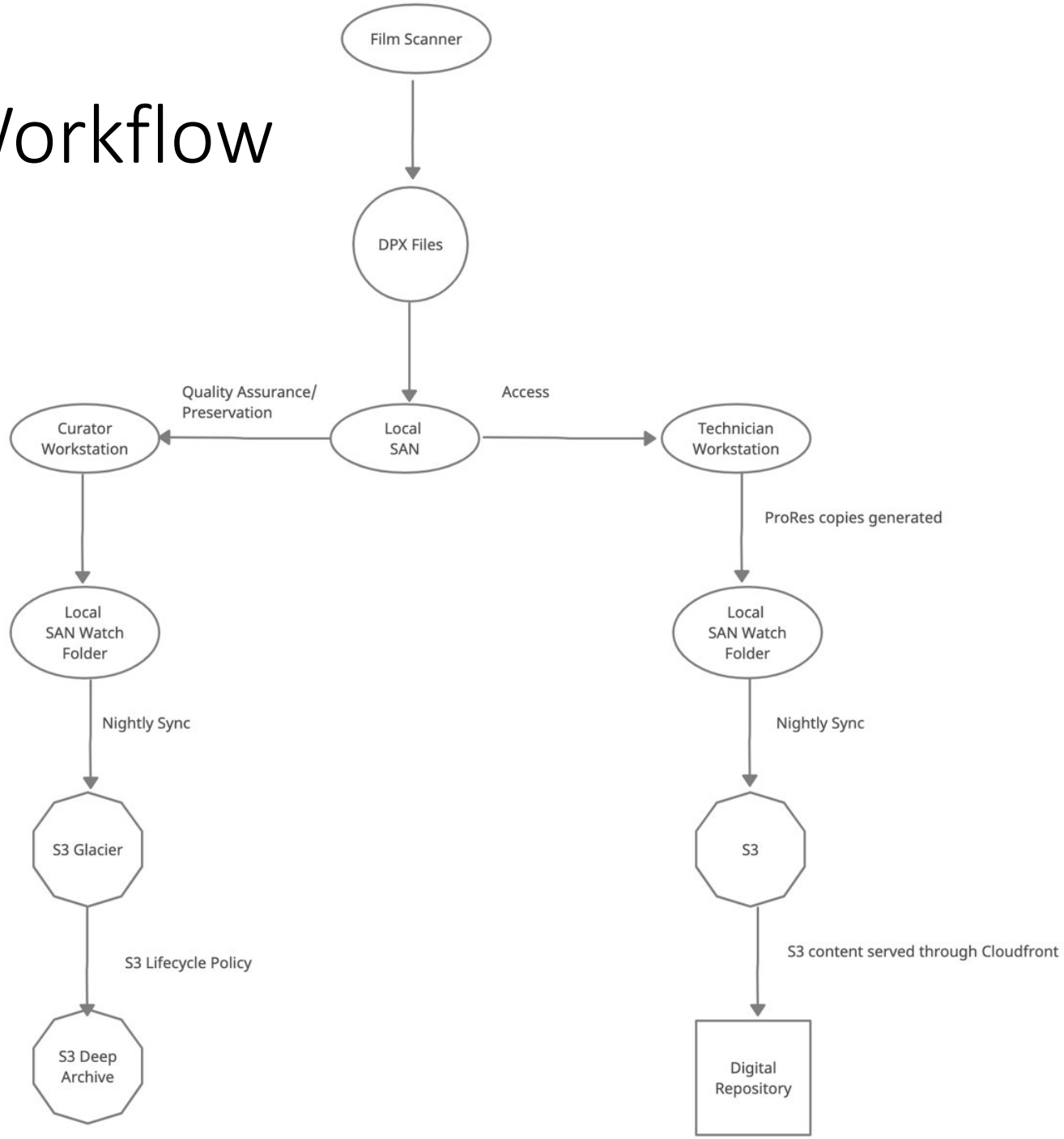
UofSC Commvault Off-Site Backups



- Commvault direct to S3 backups bypassing need for on-site storage
- Use S3 Lifecycle Policies to archive into Glacier
- Use Commvault retention policies to retain data through end of life

MIRC Film Digitization Workflow

- Physical media is digitized via on site film scanner
  - Preservation level DPX files
  - Stored on local SAN
- Preservation level files
  - File quality is verified and DPX files are bagged
  - Files are sent to sync watch folder in local SAN
  - Files are moved into S3 Glacier nightly
    - awscli in a cron job
  - Files are moved to Deep Archive after 14 days
    - Via a lifecycle policy
- Access level files
  - ProRes access copies are generated from DPX masters
  - Access copies are sent to sync watch folder in local SAN
  - moved into S3 nightly
    - awscli in a cron job
  - S3 bucket can be accessed by a CloudFront distribution
  - Streaming URL generated from CloudFront is integrated into digital repository



Our future in the cloud…

- Federated HPC cluster bursting through Slurm (and Bright)
- Expand Science Gateways
- Connection to on-prem secure enclave to AWS, GCC High in Azure for collaboration tools like Teams
- Kubernetes
- Use cloud for hosting our workshops