

**UTSouthwestern**  
Medical Center

BioHPC

---

# BioHPC OnDemand: Extending an HPC cluster to cloud infrastructures

Paniz Karbasi

BioHPC – UTSW

IBM Spectrum Scale User Group Meeting

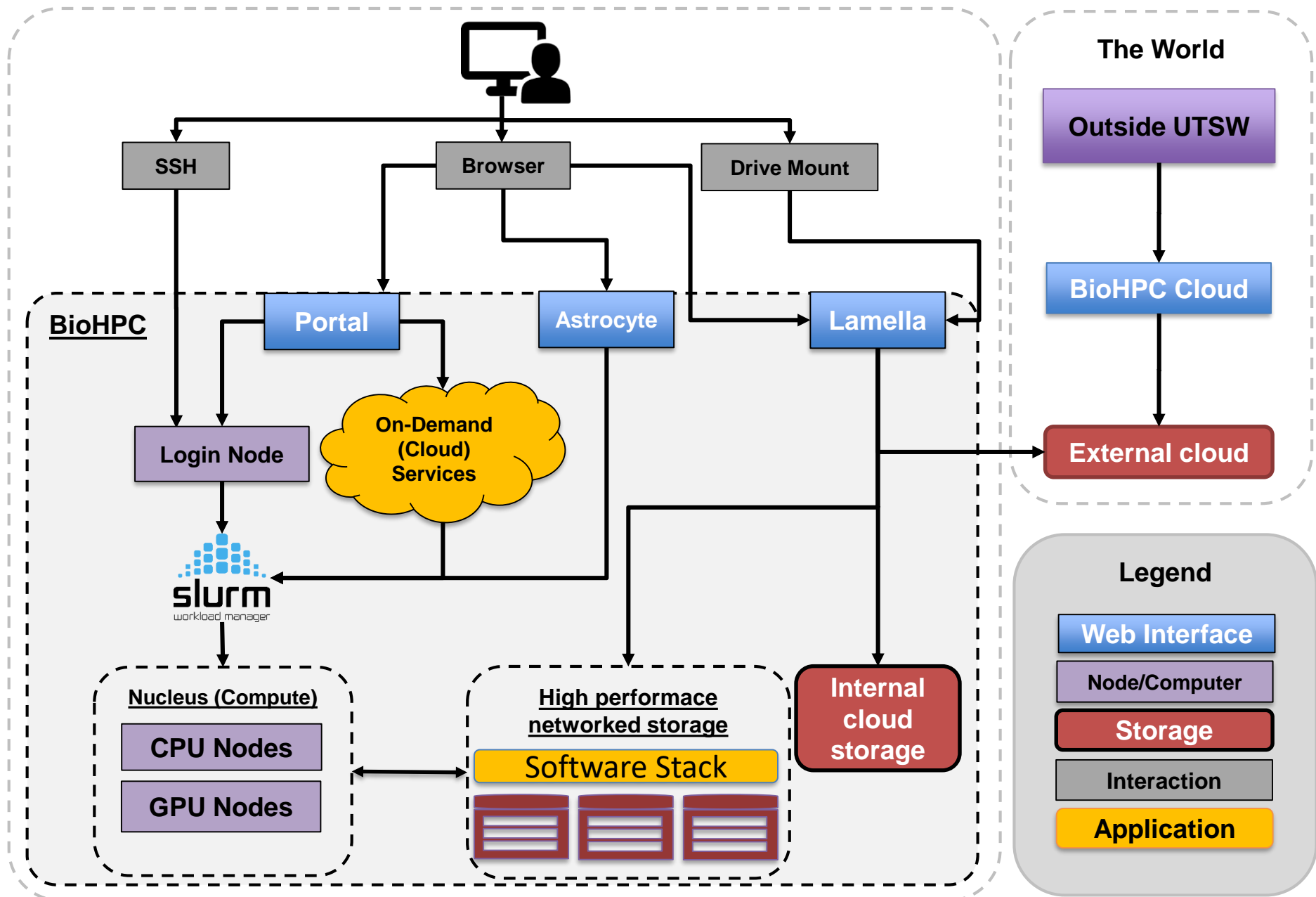
SC 2022



- Cluster name: Nucleus
- Cluster size: 800 nodes of cutting-edge GPU and CPU nodes
- 50 PB of high-throughput storage (360-400 GB/s)
- Two side HPC: 1.6 Tb/s RDMA
- 2000 users
  - Hot users 250 per months

### Our vision:

- Make science to be possible at UTSW
- Lower the barrier for scientists to get access to HPC resources



# Terminal OnDemand

## Nucleus Login Shell

This terminal provides a login session to the BioHPC Nucleus cluster. Please login using your BioHPC password, your username is automatically detected.

For safety, the terminal is only active when you mouse is over it. Remember to log-out of your session (CTRL-D or exit) when you are finished!

```
Last login: Thu Sep  1 13:23:43 2022

Welcome to BioHPC-Nucleus Supercomputer
UT Southwestern Medical Center

Based on RedHat Enterprise Linux Server 7
Cluster ID: Nucleus

-----

** Unauthorized use/access is prohibited. **

If you log on to this computer system, you acknowledge your awareness
of and concurrence with the UT Southwestern Medical Center Use Policy. The
University will prosecute violators to the full extent of the law.

-----

Use the following commands to adjust your environment:

'module avail'           - show available modules
'module add <module>'    - adds a module to your environment for this session
'module initadd <module>' - configure module to be loaded at every login

-----

Current BioHPC Storage Quotas:

FILE      |      SPACE USAGE      |      NUMBER OF FILES      |
SYSTEM   |  USED  SOFT  HARD  |  USED  SOFT  HARD  |
```

# Web application OnDemand

## BioHPC OnDemand

**BioHPC OnDemand** is a new service from BioHPC, providing easy one-click access to web-based scientific applications that can run on the compute or GPU nodes of our Nucleus cluster. There's no need to install software on your laptop or desktop and you can analyze your data faster with BioHPC.



### NVIDIA DIGITS

DIGITS from NVIDIA is a web-based platform allowing BioHPC users to easily harness the power of modern deep learning toolkits, including Caffe, Torch, and Tensorflow. DIGITS provides a web interface to upload and explore datasets; define, train, and test models; interactively examine predictions. DIGITS can use multi-GPUs and allows easy setup of jobs with different parameters, minimizing coding so that you can concentrate on your applying deep learning to your data.



### Jupyter - Interactive Notebooks

The Jupyter project provides web-based notebooks for interactive coding in many languages, including Python, R, and Julia. Create beautiful notebooks incorporating code, plots, tables, and documentation.

- OnDemand Jupyter
- OnDemand DIGITS
- OnDemand RStudio
- OnDemand JupyterLab
- OnDemand BisQue
- OnDemand Applications
- OnDemand CryoSPARC
- OnDemand Clara

portal, which will run on the powerful  
Launch a session from the portal and

## Applications OnDemand

This service allows you to launch a graphical environment on the Nucleus cluster running your selected module.

You can access these sessions via your browser. However, smoother access is available by connecting with a VNC client (WebGUI/WebGPU) or the DCV client (Windows DCV sessions only). If you prefer to connect using a VNC client you may, using the VNC address and password that are displayed for each session. On Windows the VNC link will open with RealVNC / TightVNC or other VNC clients you may have installed.

*Jobs shown 'PENDING' are waiting for a free node on the cluster. You can connect to jobs once they are 'RUNNING'. Reload the page to refresh the status of your webGUI jobs.*

Jobs run for a maximum of 20 hours. To end a job earlier than this please click the cancel link.

### Current OnDemand Applications



Please select your module if it exists in the above list or visit the [module browser page](#) and select a module to run.

You have no current webGUI sessions

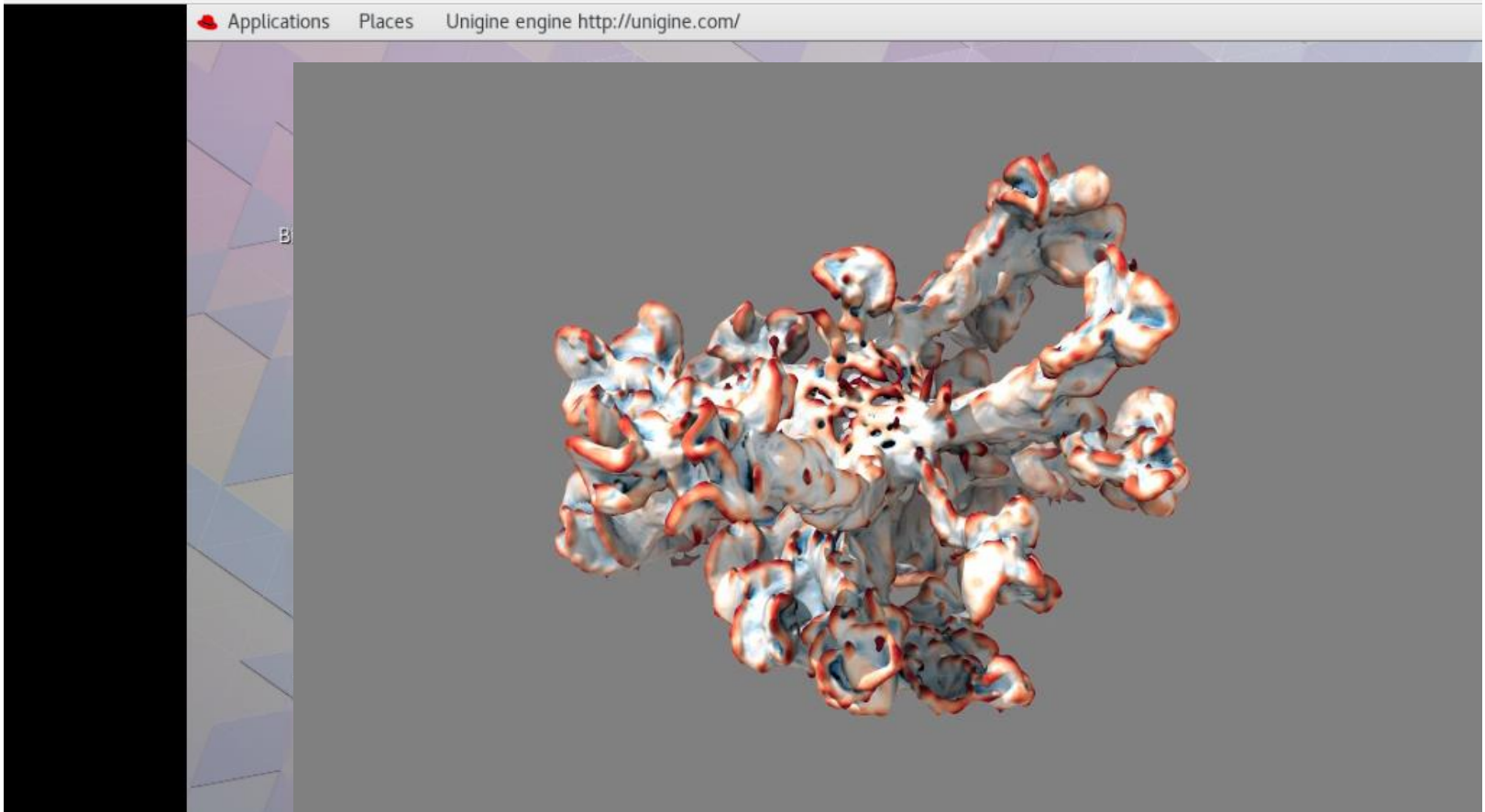
For best performance we recommend TurboVNC for webGUI and webGPU sessions, and the NICE DCV client for webWinDCV sessions.

TurboVNC Client Download: [\[Windows\]](#) [\[Mac OSX\]](#) [\[Linux 64-bit .deb\]](#) [\[Linux 64-bit .rpm\]](#) (Version 2.0.91)

NICE DCV Client Download: [\[Windows\]](#) [\[Mac OSX\]](#) [\[Linux .tar.gz\]](#)

## Remote webGPU OnDemand

<https://portal.biohpc.swmed.edu:4443/guacamole/#/client/Y29ubmVjdGlvbl8xAGMAanNvbG?data=AfUXfJ0W5k8QpyUtQwPjdZkVGsw1TT1HoyY59xKSLHUJ6fb2%2FM8%2BuMT%2FWMTZagcr9>



Reference: Drs. Gaudenz Danuser and Meghan Driscoll @ UTSW

## Storage OnDemand: Lamella

The screenshot displays the Lamella Storage OnDemand web interface. The browser address bar shows the URL: <https://lamella.biohpc.swmed.edu/index.php/apps/files/?dir=/&fileid=1060850393>. The interface includes a left sidebar with navigation options: All files, Recent, Favorites, Shares, Tags, External storages, and Deleted files. The main content area shows a file list with columns for checkboxes, file names, share icons, ellipsis menus, sizes, and dates. The files listed are:

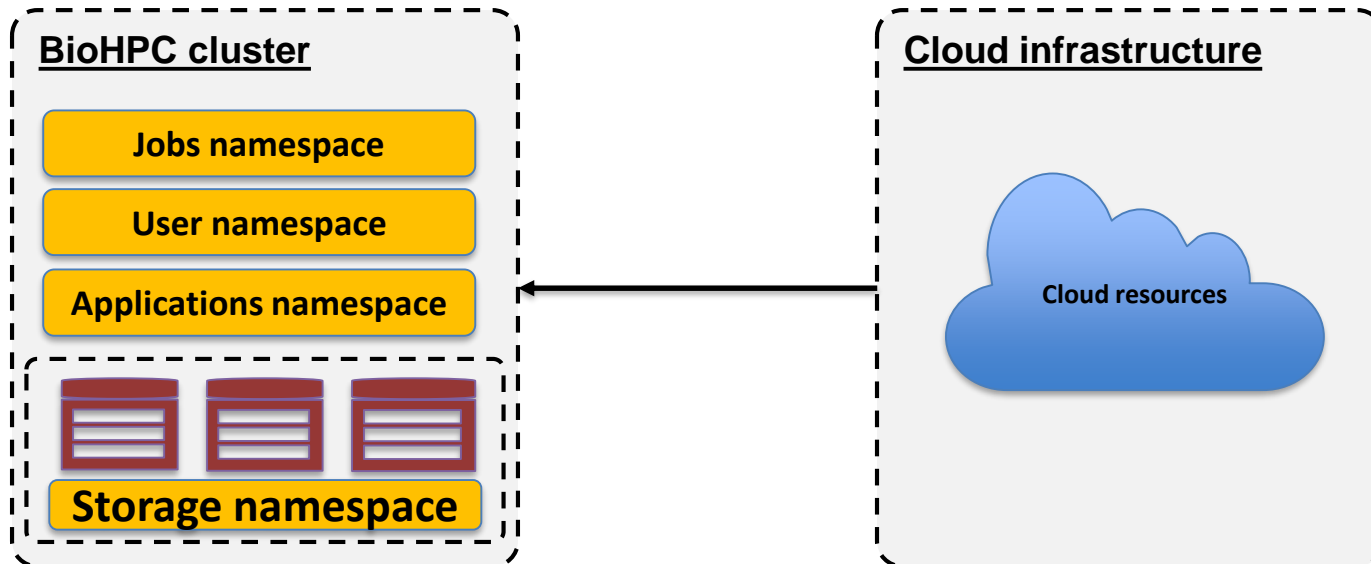
File Name	Size	Last Modified
Microsoft.txt	< 1 KB	8 years ago
dsf.txt	< 1 KB	3 years ago
project	Pending	7 days ago
Cloud	Pending	a year ago
home	Pending	a month ago
work	Pending	4 months ago

Below the file list, it indicates "15 folders and 20 files" with a "Pending" status. On the right, a sidebar for the selected "project" folder shows details: "project", "★ < 1 KB, 7 days ago", and "Tags". It also features tabs for "Activity", "Comments", and "Sharing". A search input field labeled "Name..." is present, along with options to "Share link" and "Add to a project".

PM: Xiaochu Lou

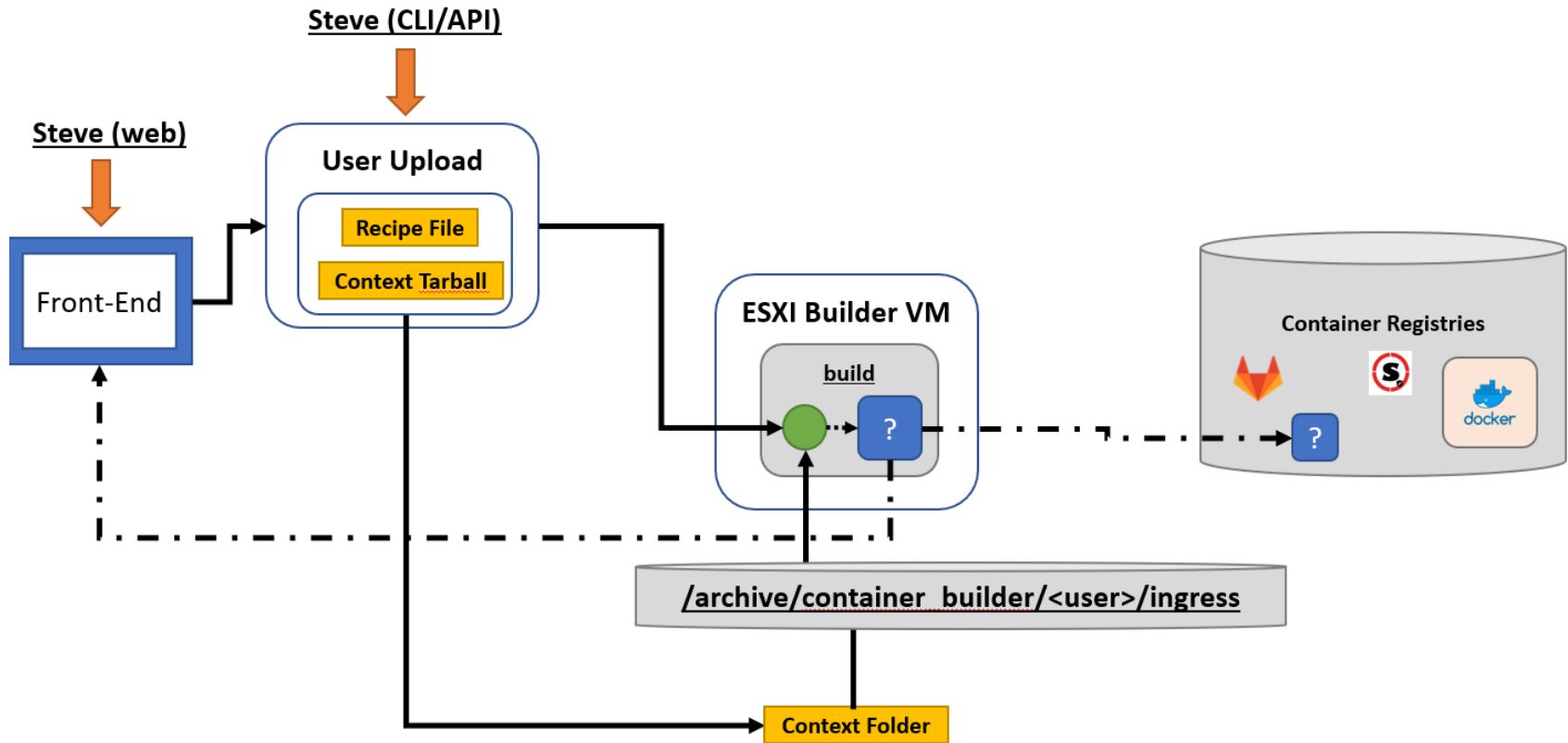


## Extend BioHPC OnDemand to cloud infrastructures



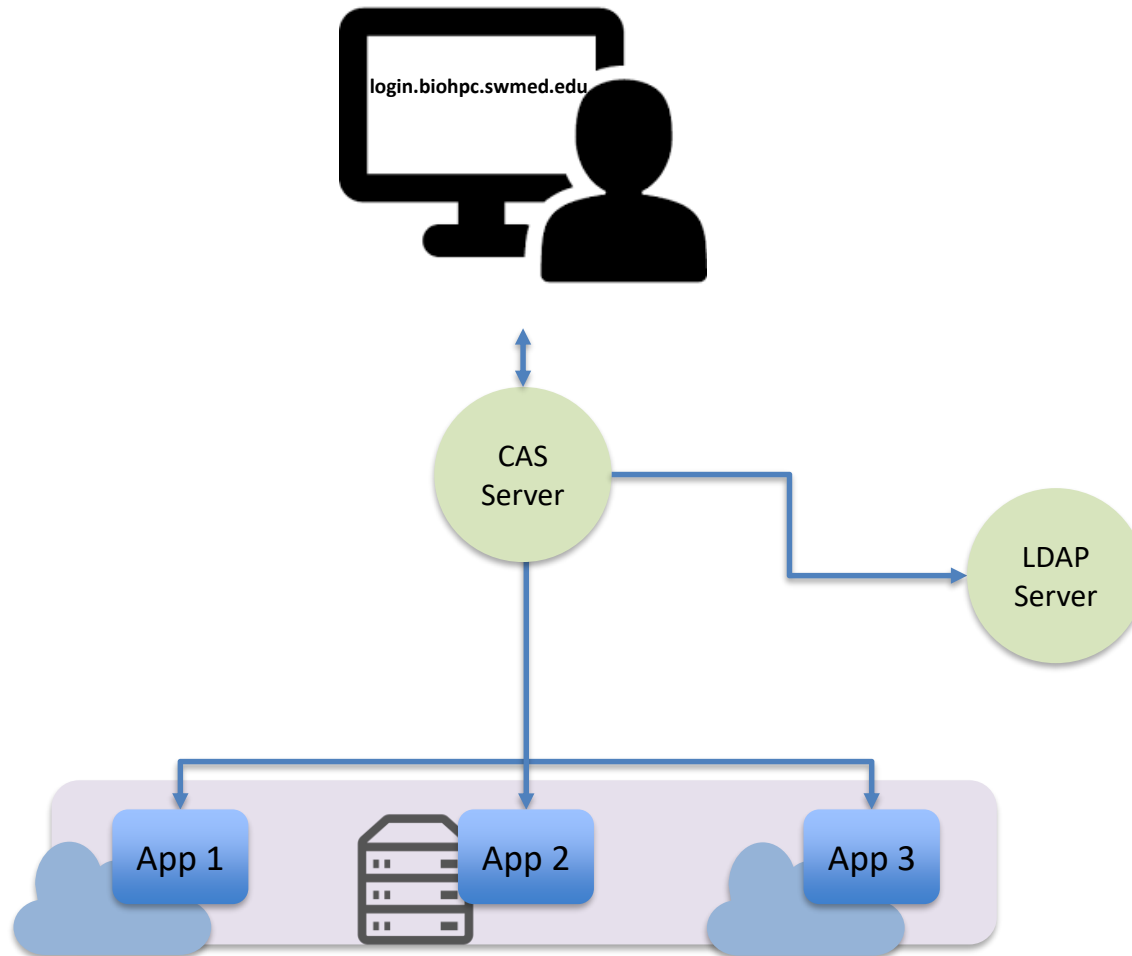
Extend on-premise cluster to outside infrastructure with minimizing refactoring the existing workflow.

### POC – Container Constructor Flow

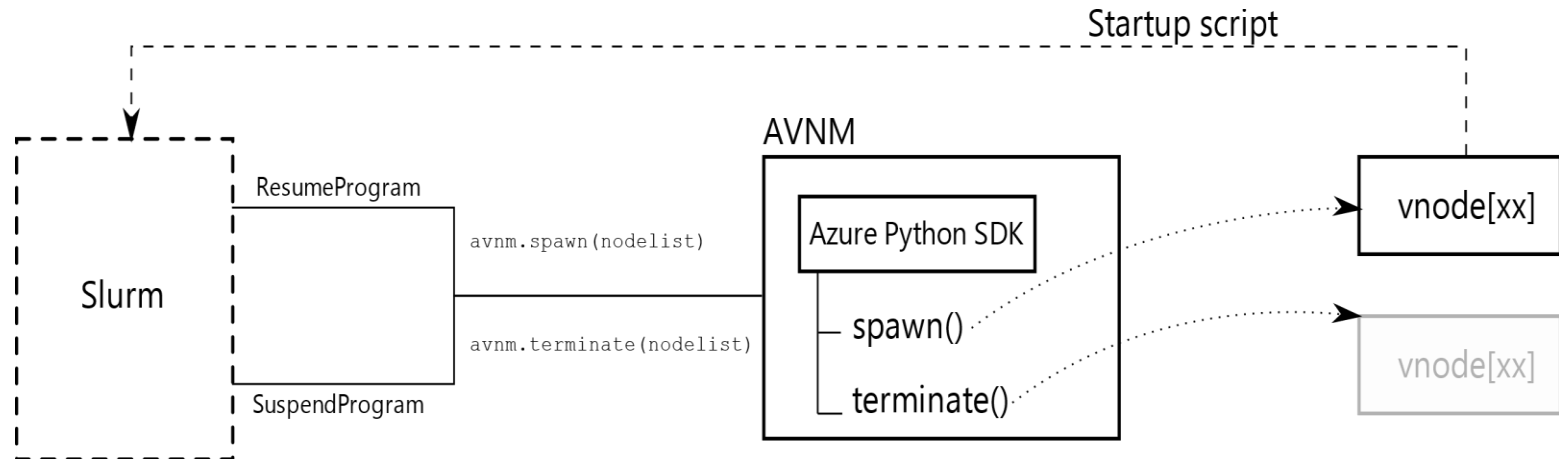


PM: Devin O’Kelly

## POC – CAS Integration with LDAP

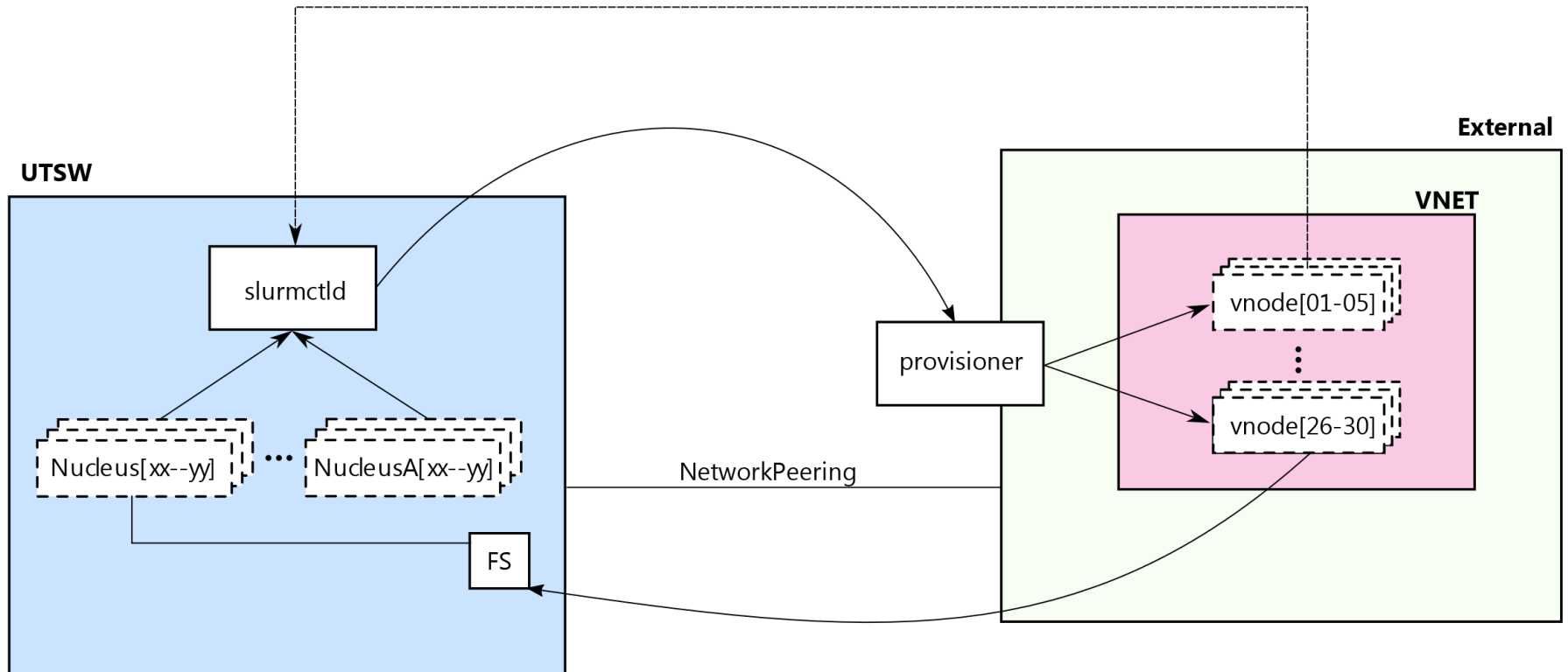


## POC – Cluster Extension to Azure



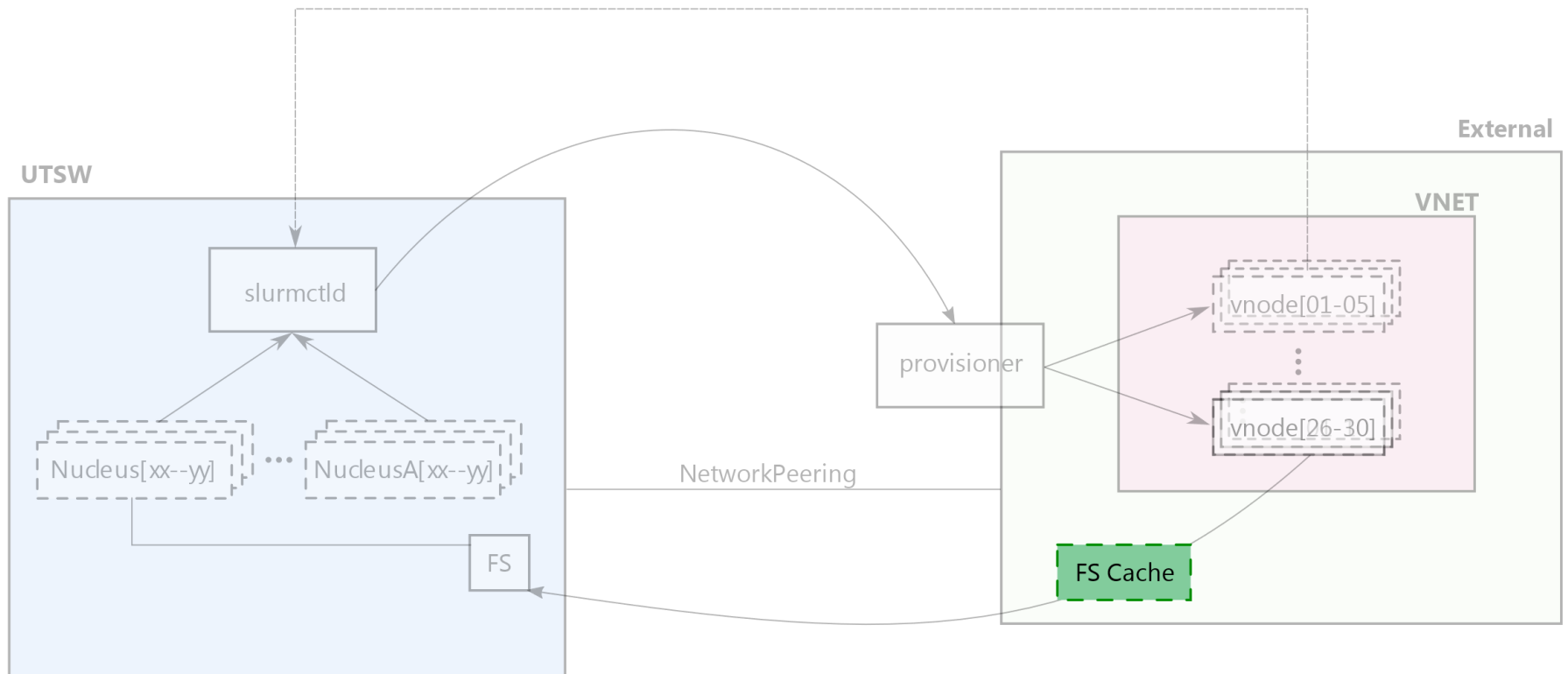
## POC - Cluster Extension to any external location

Extend on-premise cluster to outside infrastructure with minimizing refactoring the existing workflow.

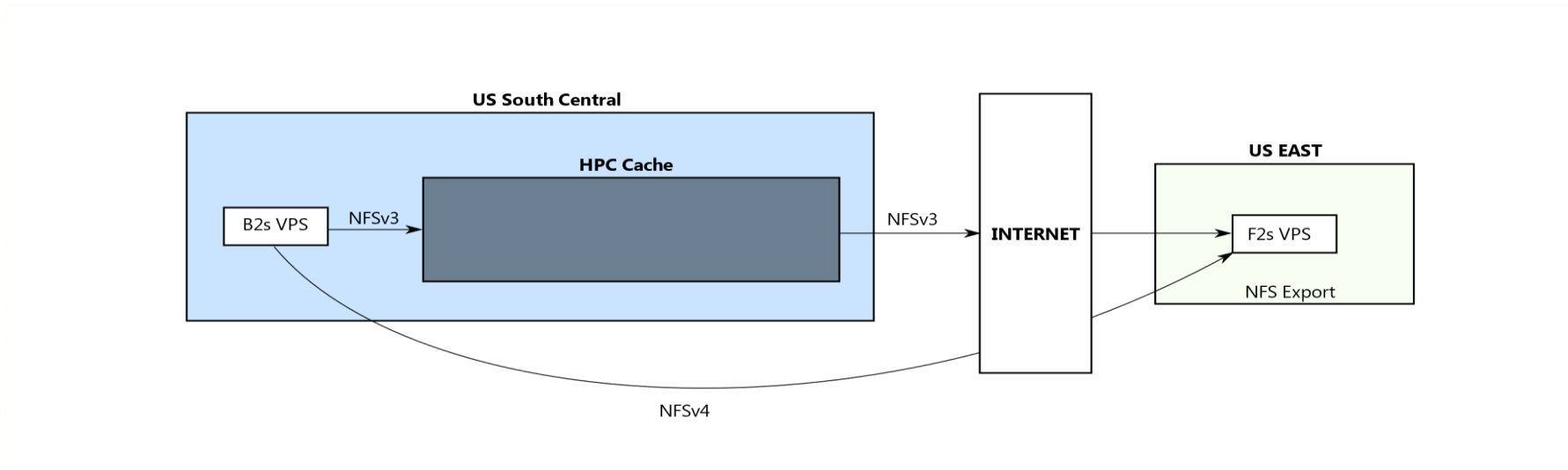


## Extend BioHPC OnDemand to cloud infrastructures: Storage namespace mapping

Extend on-premise cluster to outside infrastructure with minimizing refactoring the existing workflow.

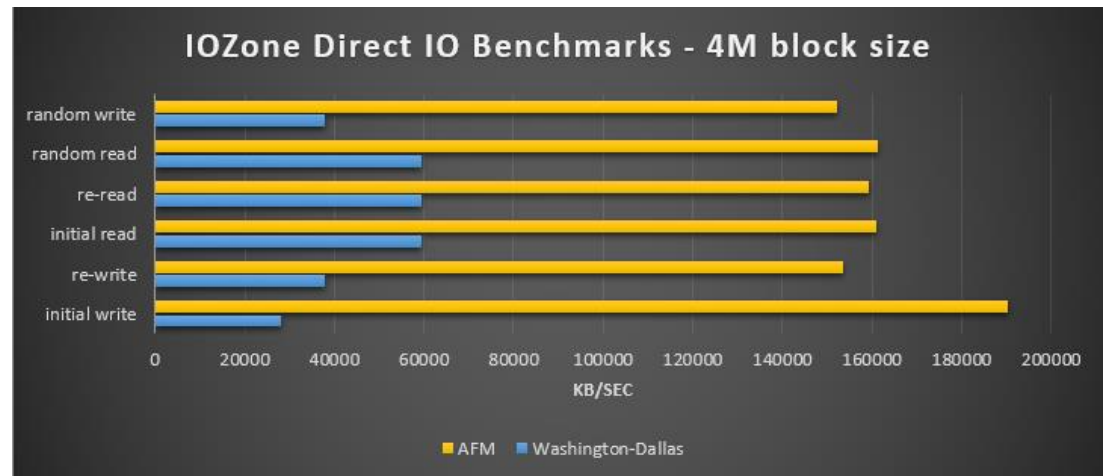
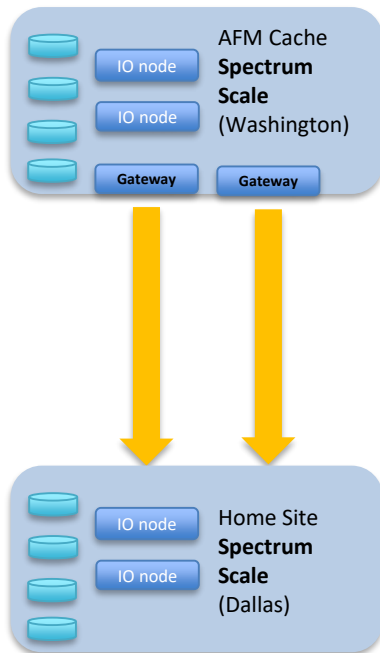


## POC – Microsoft HPC Cache



Direct Mount	MS HPC Cache
16.95 MB/s	4.25 MB/s

## POC – IBM Spectrum Scale AFM as HPC/FS Cache



IBM Cloud sites: Dallas (primary) and Washington



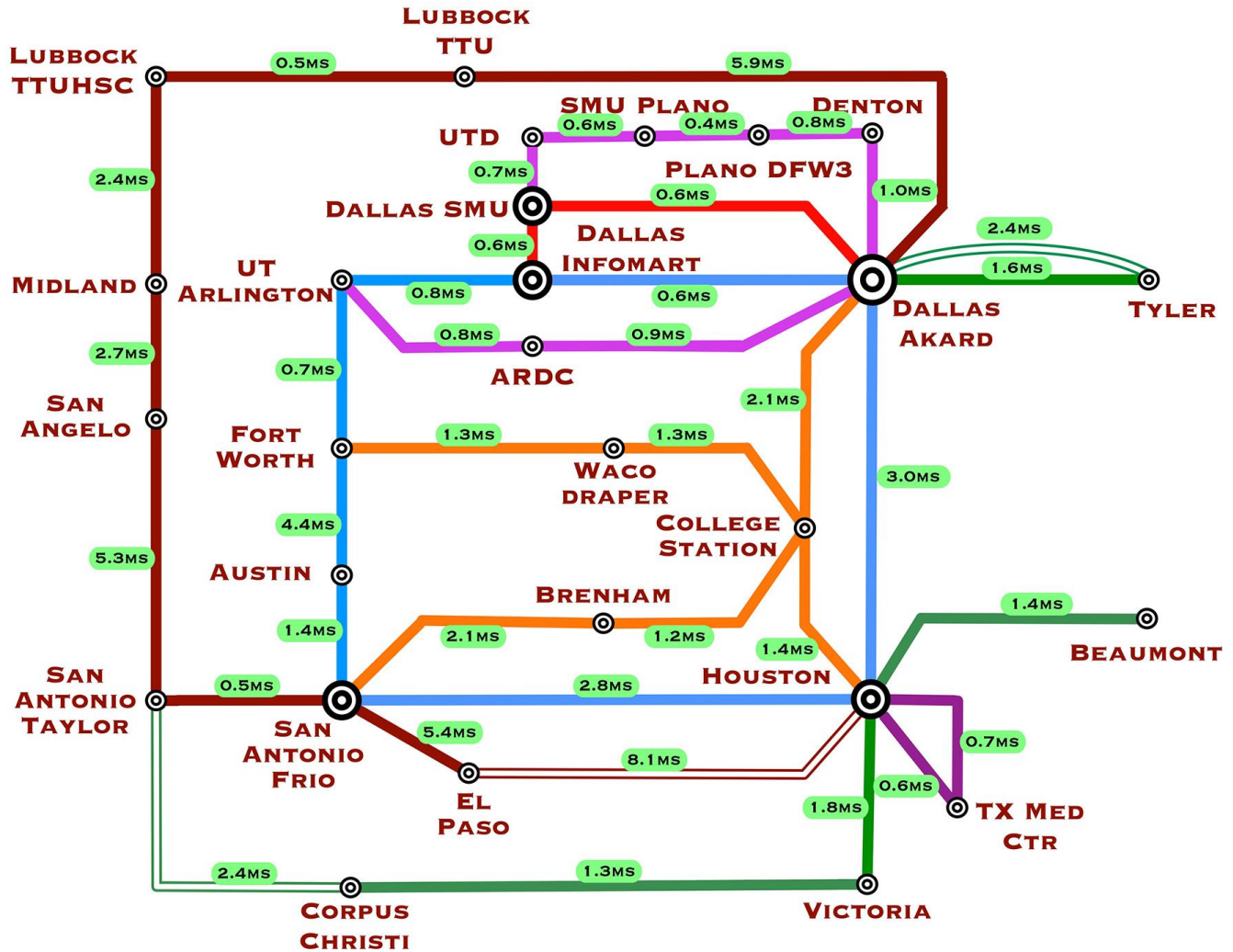
## How about the network extension to cloud

### Network latency

Unit (s)	Multiple	Symbol
$10^{-12}$	1 picosecond	ps
$10^{-9}$	1 nanosecond	<b>ns</b>
$10^{-6}$	1 microsecond	$\mu$ s
$10^{-3}$	1 millisecond	<b>ms</b>

1.3us + 5us/km over fiber  
(every kilometer of fiber )

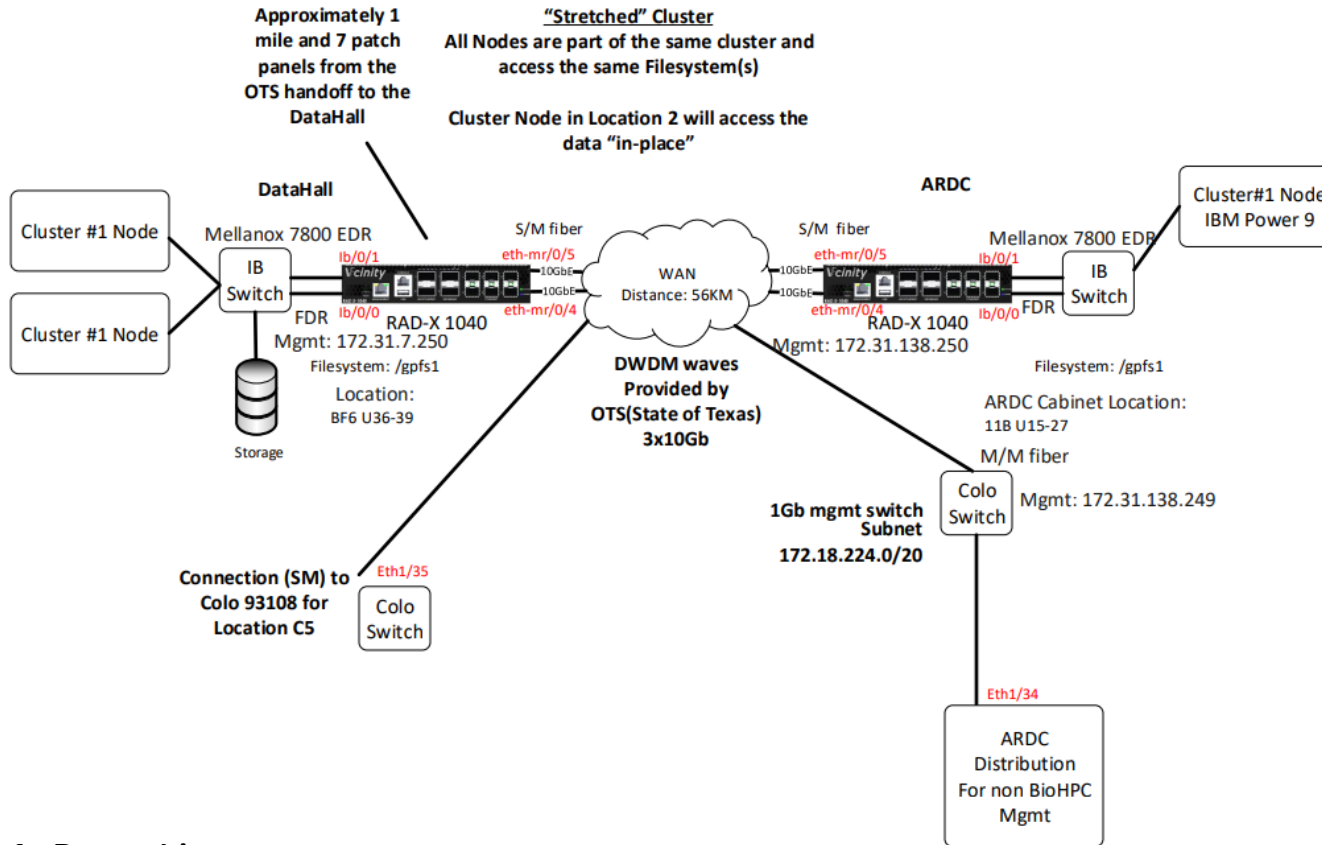
# LEARN - Texas State DWDM network



# Latency testing with Nvidia – RDMA via DWDM

## UTSW BioHPC ARDC Expansion Feasibility Test

May 11<sup>th</sup>, 2020

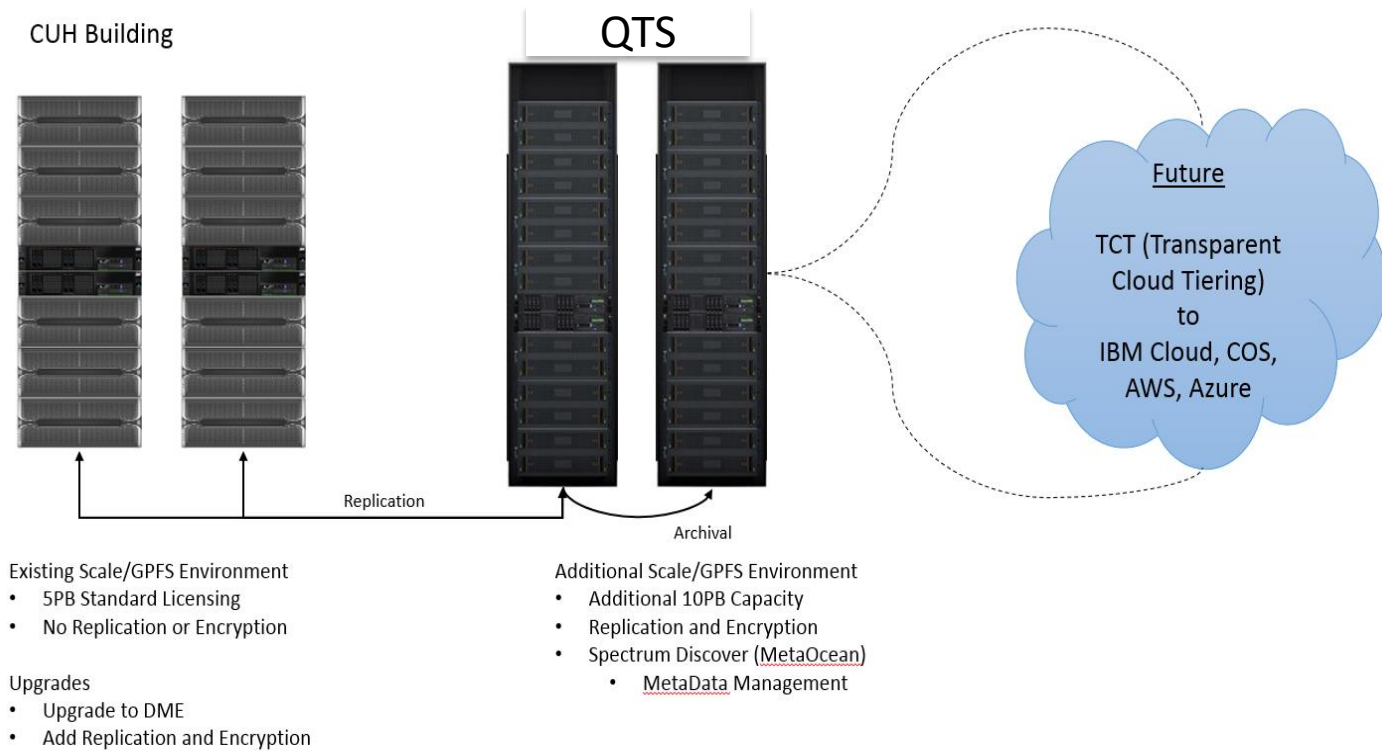


PM: Peng Lian

## ARDC Benchmark

Networks	bandwidth tests			latency tests		
	iperf	ib_read_bw	ib_write_bw	ping (64B-64kB)	ib_read_lat (2KB - 1MB)	ib_write_lat (2KB - 1MB)
ARDC						
<b>ib/0/0.0</b>	9.21 Gbits/sec	9.03 Gbits/sec	N/A	N/A	N/A	N/A
<b>ib/0/1.0</b>	9.24 Gbits/Sec	9.06 Gbits/sec	9.06 Gbit/sec	1.42-1.8 ms	1.2-1.7ms	0.6 - 1.6 ms
UTSW						
<b>10GbE</b>	9.86 Gbits/sec	N/A	N/A	0.21 - 0.37 ms	N/A	N/A
<b>Bass IB</b>	43.4 Gbit/sec	48.2 Gbit/sec	46.3 Gbit/sec	0.12 - 0.47 ms	0.0027 - 0.19 ms	0.0014 - 0.18 ms
<b>Bass-CUH IB</b>	21.2 Gbit/sec	36.6 Gbit/sec	36.6 Gbit/sec	0.25 - 0.51 ms	0.033 - 0.26 ms	0.013 - 0.23 ms

## POC – Spectrum Scale Transparent Cloud Tiering



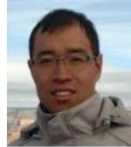
## Meet the BioHPC team



**Liqiang Wang**  
Director



**Paniz Karbasi**  
*Ph.D. Electrical  
and computer  
engineering*



**Peng Lian** *Ph.D.*  
*Biomedical  
Engineering,  
Bioinformatics*



**Xiaochu Lou** -  
*Ph.D.*  
*Biochemistry*



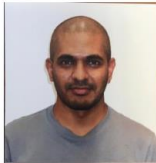
**Felix Perez**  
M.S, Computer  
Science



**Devin O'Kelly**  
*Ph.D., Biomedical  
and Molecular  
Imaging*



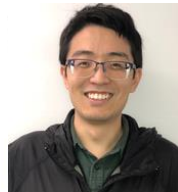
**Hung Le**  
*Ph.D.,  
Computational  
Chemistry*



**Suresh  
Pannerselvam**  
*Ph.D., Medical  
Sciences*

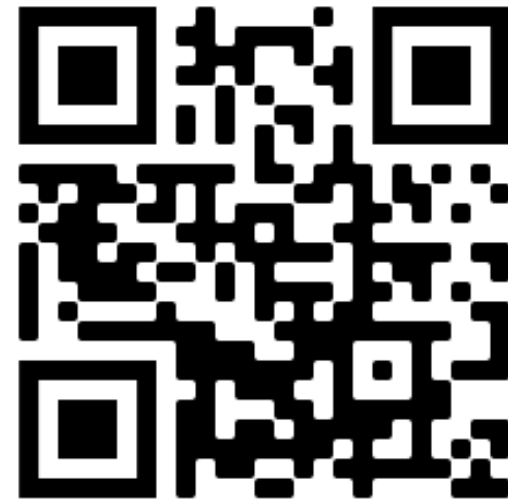


**Kenny Huynh**  
*B.S., Computer  
Science*



**Xueyan Li**  
*M.S.,  
Computational  
Geophysics*

Find out more  
about us!



Thank you.