



BCNET  
**CONNECT**  
HIGHER ED & RESEARCH TECH SUMMIT

# Digital Research Infrastructure in BC

# Outline

---

Jeff Albert, U Victoria

Roman Baranowski, UBC

Ryan Thomson, UBC

James Peltier, SFU

Megan Meredith-Lobay, SFU

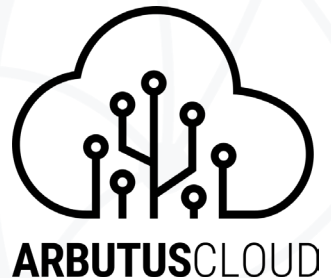
- Introduction to the topic
- What are the current DRI initiatives and systems in BC?
- What's coming next in DRI for BC?
- Questions



## Current DRI Initiatives and Services in BC

# Community Clouds

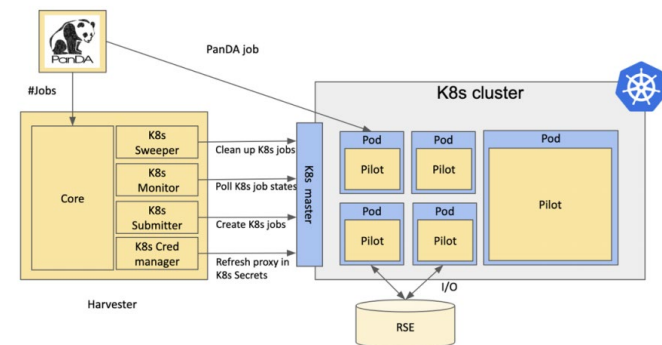
- BC hosts two national Community Cloud sites, Cedar Cloud at SFU and Arbutus Cloud at UVic
- Arbutus is the largest research cloud in Canada, and one of the largest in the world
- Built on open source software and commodity hardware to maximize research value for infrastructure investment
- Community Clouds provide a stable, centrally-funded foundation for cloud deployments that doesn't directly cost researchers, and that isn't subject to the volatile usage-based costs of Commercial Clouds
- Demand continues to soar upward for resources on these Community Clouds



# Kubernetes for Diverse Workloads

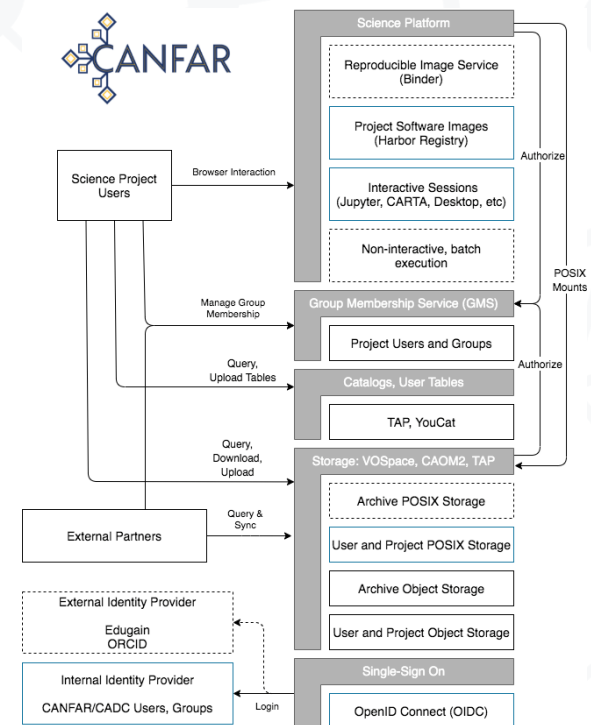
## ATLAS Tier-2 on Kubernetes at UVic Arbutus

- First ATLAS Tier-2 deployment on Kubernetes in the world
- Running for almost 2 years in production at scale
- Top-5 reliability out of over 150 Tier-2 sites worldwide



## CANFAR Science Platform on Kubernetes at UVic Arbutus

- Extends Kubernetes job/pod scheduling with a full user experience
- Interactive and Batch jobs with curated workload containers
- Identity and Access management, quotas, permissions
- Tenancy above the K8s layer



A single common interface layer to all clouds, community and commercial

Sets up low-friction workload mobility for true lock-in-resistant hybrid cloud

# Workload Orchestration Platform

Provide a unified view of research applications

- Support legacy workloads as well as containers
- Kubernetes isn't the only game in town
- Access to the infrastructure should be as standardized as possible allowing for leveraging economies of scale

Logging, Monitoring, Reporting, Alerting

- How do we provide a single pane of glass view into the infrastructure
- The view for researchers is often different administrators of the system
- How can we provide that

# Digital Research Infrastructure @ UBC

15,872 CPU cores (skylake + cascadelake Intel)

200 general purpose GPUs (V100 32gb)

EDR (100Gbps) Infiniband

1.3 PB Dell EMC Isilon w/ 192TB all-flash burst  
buffer

1.5 PB DDN Lustre

**HPC cluster:**  
[socketeye.arc.ubc.ca](https://socketeye.arc.ubc.ca)

# Digital Research Infrastructure @ UBC

**10 PB storage capacity**

Access through [Globus](#)

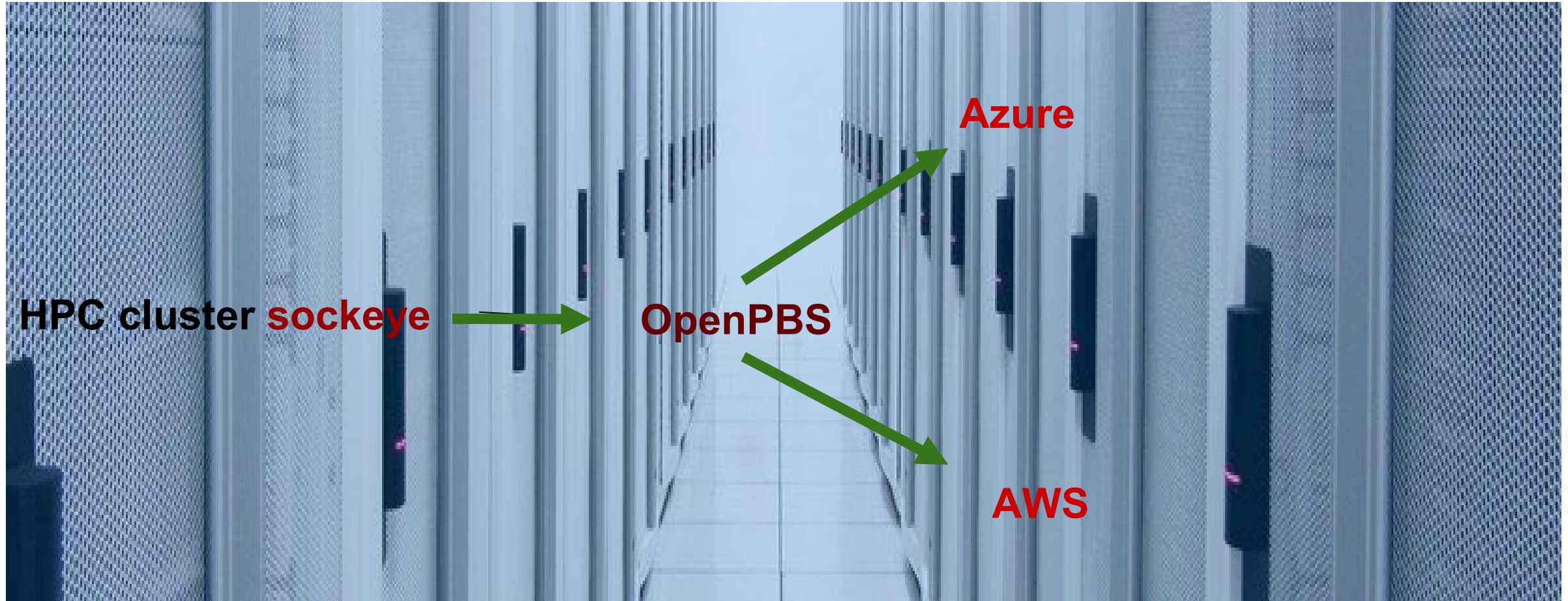
**S3 connectivity in progress**

**Object Storage: chinook**



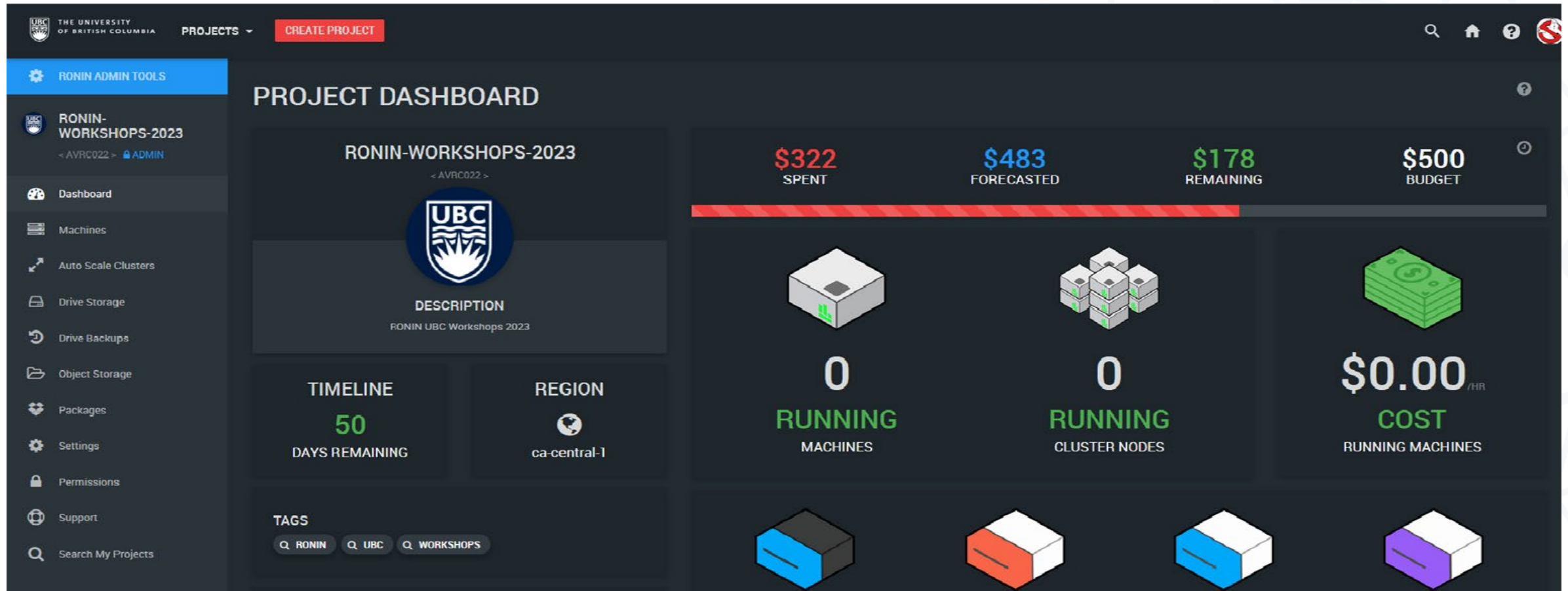
# Digital Research Infrastructure @ UBC

## Cloud Bursting



# Digital Research Infrastructure @ UBC

## Ronin Platform



# Digital Research Infrastructure @ UBC

## Ronin Platform

THE UNIVERSITY OF BRITISH COLUMBIA PROJECTS **CREATE PROJECT**

RONIN-ADMIN TOOLS

RONIN-WORKSHOPS-2023

RONIN-WORKSHOPS-2023 < AVRC022 > ADMIN

Dashboard

Machines

New Machine

Machine Summary

Auto Scale Clusters

Drive Storage

Drive Backups

Object Storage

Packages

Settings

Permissions

Support

UBC RONIN-WORKSHOPS-2023

### CREATE A NEW MACHINE

MACHINE COMPOSER

1. SOFTWARE 2. ADDRESS 3. MACHINE TYPE 4. STORAGE

OPERATING SYSTEMS

<b>Ubuntu 18.04</b> Canonical, Ubuntu, 18.04 LTS, amd64 ... <input type="button" value="SELECT"/>	<b>SUSE 15</b> SUSE Linux Enterprise Server 15 SP4 ... <input type="button" value="SELECT"/>	<b>Windows Server 2019</b> Microsoft Windows Server 2019 Core ... <input type="button" value="SELECT"/>
<b>Red Hat 9.0</b> Provided by Red Hat, Inc. <input type="button" value="SELECT"/>	<b>CentOS 7</b> CentOS-7-2111-20220825_1_x86_64 ... <input type="button" value="SELECT"/>	<b>Ubuntu 22.04</b> Canonical, Ubuntu, 22.04 LTS, amd64 ... <input type="button" value="SELECT"/>
<b>Ubuntu 20.04</b> Canonical, Ubuntu, 20.04 LTS, amd64 ... <input type="button" value="SELECT"/>	<b>Amazon Linux 2023</b> Amazon Linux 2023 AMI 2023.0.2023 ... <input type="button" value="SELECT"/>	

LAUNCH SUMMARY

**YOUR NEW MACHINE**

**\$0.00 / HOUR**

**STEP 1 - SOFTWARE**  
CHOOSE YOUR OPERATING SYSTEM OR SOFTWARE.

STEP 2 - ADDRESS  
ADDRESS YOUR MACHINE.

STEP 3 - MACHINE TYPE  
CHOOSE THE TYPE OF MACHINE YOU NEED.

# Digital Research Infrastructure @ UBC

## Ronin Platform

The screenshot shows the 'AUTO SCALE CLUSTER' creation wizard in the Ronin Platform. The interface is dark-themed and includes a sidebar with navigation options like 'RONIN ADMIN TOOLS', 'RONIN-WORKSHOPS-2023', 'Dashboard', 'Machines', 'Auto Scale Clusters', 'Drive Storage', 'Drive Backups', 'Object Storage', 'Packages', 'Settings', 'Permissions', 'Support', and 'Search My Projects'. The main content area is titled 'AUTO SCALE CLUSTER' and contains the following steps:

- 1. SELECT AN OPERATING SYSTEM**: Four options are shown: UBC (purple), ubuntu (red), CentOS (blue), and aws (orange). Each option has a 'SELECT' button.
- 2. SELECT A HPC SCHEDULER**: Three options are shown: SLURM Scheduler (blue, labeled 'PRODUCTION'), SGE Scheduler (red, labeled 'BETA'), and Torque Scheduler (blue, labeled 'BETA'). Each option has a 'SELECT' button.
- 3. ADDRESS YOUR CLUSTER**: A search box for a host name is shown with the example 'e.g. my-cluster'. A dropdown menu is open showing '.ubc-hpc.cloud' with the message 'Address is not available'. A 'SELECT' button is next to the dropdown.
- 4. SELECT YOUR HEAD NODE**: A 't3.small' instance is selected, showing '2 vCPUs' and '2 GiB RAM'. A 'CHANGE' button is next to it.
- 5. SELECT YOUR COMPUTE NODES**: A 't3.small' instance is selected, showing '2 vCPUs' and '2 GiB RAM'. A 'CHANGE' button is next to it.

At the bottom of the wizard, there are two buttons: 'CLOSE' and 'LAUNCH MY CLUSTER'.

# Research Software Development

- SFU currently has a research software development team as part of the Research Computing Group.
- Others have these internal groups that are often not well known
- Supporting research costs money.
  - How can we leverage grants in support of long term staffing rather than students
  - This is potentially a contentious issue
- Generates opportunities to talk about collaboration across institutions



BCNET  
CONNECT

What's Coming Next in DRI for BC?

# Infrastructure Above Infrastructure

- Researchers are increasingly looking for infrastructure-grade services (highly available, transparently scalable, globally accessible) at levels above the classic “IaaS” layer:
- Platform Services like Kubernetes, DASK, Kubeflow, Magic Castle, and more
- Turn-key Software Services like REDCap, Overleaf, JupyterLab, NextCloud, R-Studio, and others
- Providing access to raw compute, even with a managed operating system, doesn't meet the whole need
- Researchers don't all have the ability to build their own
- BC has emerged as a cluster of cloud technologists and community cloud compute and storage resources - we need to extend those strengths into the upper levels of the stack

# Access to DRI Training and Support as a Regional Differentiator

- National Service Delivery is continuing and growing under the Digital Research Alliance of Canada, but support and training remain largely an institutional and regional undertaking
- WestGrid, which once provided an umbrella offering of those services across the Western provinces, is no more
- BC's substantial concentration of DRI sites, technologists, and resources provides BC researchers with a great advantage - but access to support and training will determine how much of that advantage can be capitalized upon
- Now, at the beginning of this new chapter in Canadian DRI, is the moment for BC institutions to come together and work together to offer DRI training and support to empower all BC researchers



# Composable Infrastructure

- Dynamically (re)configure hardware resources on-demand via software
- Attach/detach storage, accelerators, networking and memory to/from servers
- Dedicated external PCIe switches and “directors”, managed via software
- PCIe enclosures are populated with devices (GPU, SCM, SSD, NIC, etc.)
- Servers populated with PCIe fabric HBA(s)
- Physical distance limitations (ie. “rack-scale”)
- Compute Express Link (CXL)

# Compute Express Link (CXL)

- CPU-to-device and CPU-to-memory interconnect over serial PCIe phys/electrical
- Config, link init and management, device discovery and enumeration, interrupts, DMA, and register I/O access using non-coherent loads/stores (CXL.io)
- Cache coherent system (CXL.cache) and device memory (CXL.mem) access
- CXL 1.1 is one host to  $\geq 1$  memory devices in a flat topology
- CXL 1.1 supported today by Intel Sapphire Rapids and AMD Zen 4 EPYC
- CXL  $\geq 2.0$  provide memory pooling, additional hierarchies/topologies, peer-to-peer device memory access and (some?) device sharing between hosts
- CXL 2.0 and 3.0 specifications published but no products yet

# Composable & CXL for BC DRI

We are evaluating how composable disaggregated infrastructure can help researchers accelerate time to discovery:

- Contributed (“condo”) systems
- Bespoke/non-standard configurations that may not have been supported/accepted previously
- Add resources/features to existing servers, potentially extending their useful lifetime
- Replace dedicated bigmem nodes and reduce overall memory footprint with memory pooling
- Replace dedicated local scratch devices in every node with storage pooling
- Reduce interconnect requirements (does every node need IB/OPA in a GP system?)
- Apply savings to purchase additional high demand resources (GPUs, etc.)
- HPC scheduling is already complex and composability may complicate that further

Questions?  
Comments?  
Observations?

Let's Talk About It!

