

Omnibond

Research Computing &
Data Science
Solutions



Omnibond Background

The background image depicts a vast, futuristic alien landscape. In the foreground, a large, multi-tiered circular structure with a metallic, layered appearance dominates the left side. The ground is a mix of sandy terrain and dark, rocky outcrops. In the distance, several tall, thin, spire-like structures rise against a dark sky. A large, prominent planet or moon is visible in the upper center of the frame, surrounded by several smaller celestial bodies. The overall color palette is dark and moody, with shades of blue, grey, and brown.



Leadership Team

Omnibond

a customer-focused
software engineering and
support company

- Over 40 years combined experience in facilitating and supporting academic and corporate research in using the tools and technologies of advanced computing
- Experience at the working and technical, project and executive management levels at Clemson, Purdue and Miami Universities and the National Center for Supercomputing Applications located at UIUC
- Expertise in software development, systems integration, operations, applications support, data transmission, identity and access management, customer relations and research facilitation and engagement
- Founded ACI-REF <http://www.aciref.org> and CaRCC <http://carcc.org>
- Over three decades of funded projects from NSF, DoD, DoE, NSA, NIST and DARPA
-  **Internet.** Presidential Fellow & CSTAAC Committee Member

- **Identity & Security Management**
 - Passwordless MFA with OmniPasskey
 - NetIQ Identity Manager Connectors
 - Thousands of customers, sold through Novell/Micro Focus/OpenText, since early 2000's
- **Computer Vision & AI**
 - TrafficVision - AI based Automated Incident Detection (AID) & Data from existing cameras on roadways
 - BayTracker - Retail Vehicle Tracking and Timing
 - Port Observer - Drayage Queuing, AIS, Dashboard for Ports
- **Cloud HPC and Storage Orchestration**
 - CloudyCluster
 - OrangeFS
 - Eureka Project
 - Custom Cloud <-> On-Prem Integration

The background of the slide is a dark, semi-transparent image of a laptop screen. The screen shows a data dashboard with a line graph at the top and a pie chart below it. The text 'Software Products' is overlaid in large white font on the right side of the screen.

Software Products

Omnibond
a customer-focused
software engineering and
support company



Going the Extra Mile

Omnibond

a customer-focused software product engineering and support company

I have worked with Omnibond development and support for almost 15 years. They always deliver the highest level of support possible. I have been in Technical Support for 23 years and have worked with different companies across the world. Omnibond is at the top in response time and for going the extra mile to help our customers. I have no reservation in recommending them as a great company.

Reed Harrison

Provo, UT

Support Engineer

Novell/NetIQ/MicroFocus/Opentext

OmniPasskey - MFA Passwordless Shib Plugin



Enables Shibboleth to support the built in passkey standard on Apple, Google, and Microsoft Devices



We are looking for early access participants



CloudyCluster



Turn-Key Elastic HPC, HTC, & Storage in AWS

Create a familiar, secure & fully operational computational cluster in minutes, complete with:

Encrypted Storage: EBS, OrangeFS on PD

Compute: Job Driven Elastic Compute through CCQ (Spot, On-Demand, GPU)

Schedulers: Torque & SLURM with the CCQ Meta-Scheduler

Supports Billing Tags per job/user

Placement Policies, Advanced Networking

Supports Multiple Custom Images

End User Friendly UI: Open OnDemand

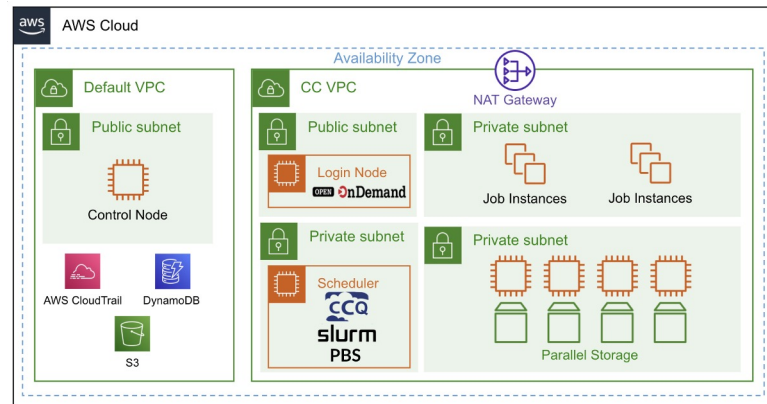
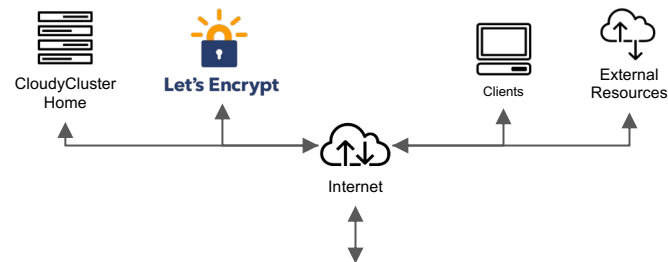
Includes Familiar Software

HPC Libraries: Boost, Cuda Toolkit, Docker, FFTW, FLTK, GCC, Gengetopt, GRIB2, GSL, HDF5, Intel MPI, Intel Runtimes, ImageMagick, JasPer, mpich, NetCDF, NumPy, Octave, OpenCV, OpenMPI, PROJ, R, Rmpi, SciPy, SWIG, WGRIB, UDUNITS, .NET Core, Singularity, Queue, Picard, xrootd, etc...

HPC Software: Amertools, ANN, ATLAS, BLAS, Blast, Blender, Burrows-Wheeler Aligner, CESM, GROMACS, JupyterLab, LAMMPS, NCAR, NCL, NCO, nwchem, OpenFoam, papi, paraview, Quantum Espresso, SAMtools, WRF, Galaxy, Vtk, Su2, Dakota, Gatk, etc...

ML Software: Mlpack, NuPIC, Octave, OpenCV, PICARD, Queue, Scikit-learn, Tensorflow

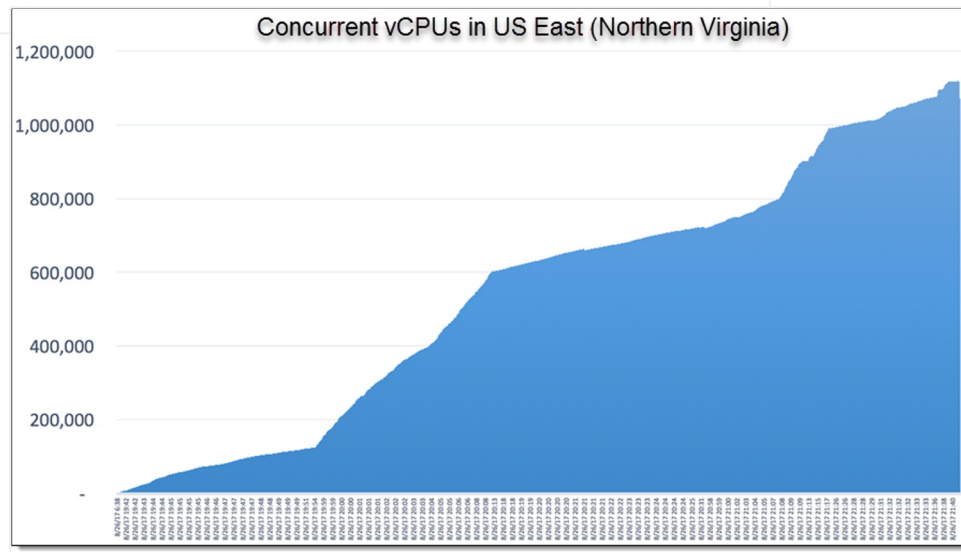
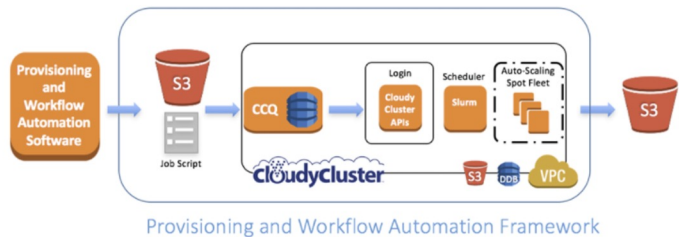
Sample Jobs: Intel Cluster Checker, Mpi_prime, wrf, etc...



Scaling on AWS

AWS News Blog

Natural Language Processing at Clemson University – 1.1 Million vCPUs & EC2 Spot Instances



Data Intensive Computing Ecosystems (DICE)
School of Computing, Clemson University

<https://aws.amazon.com/blogs/aws/natural-language-processing-at-clemson-university-1-1-million-vcpus-ec2-spot-instances/>

The Meta-Scheduler Approach

Scheduler Independent

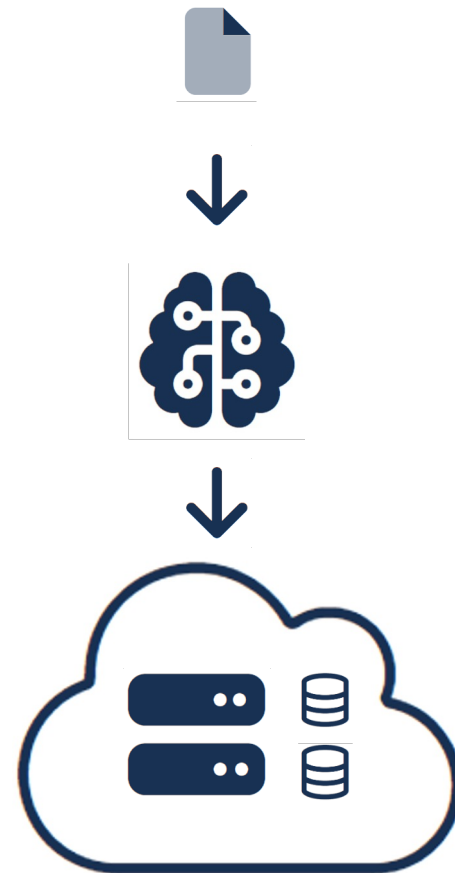
- Torque
- Slurm

Allow for Meta-Scheduler Directives

- Instance Type, each job can have a different instance type and billing model
 - Spot, Preemptible
- GPU
- Billing Tags/Labels
- Volume Type
- Custom Images for Different Jobs

Turns “Scheduler” into Dispatcher Handling:

- Instance and appropriate subnet provisioning
- Instance deletion



Turn-Key Elastic HPC, HTC, & Storage in GCP

Create a familiar, secure & fully operational computational cluster in minutes, complete with:

Encrypted Storage: GCS, OrangeFS on PD

Compute: Job Driven Elastic Compute through CCQ (Preemptable, On-Demand, GPU)

Schedulers: Torque & SLURM with the CCQ Meta-Scheduler

Supports Billing Labels per job/user

Placement Groups, BulkAPI, Tier1 Networking

Supports Multiple Custom Images

End User Friendly UI: Open OnDemand

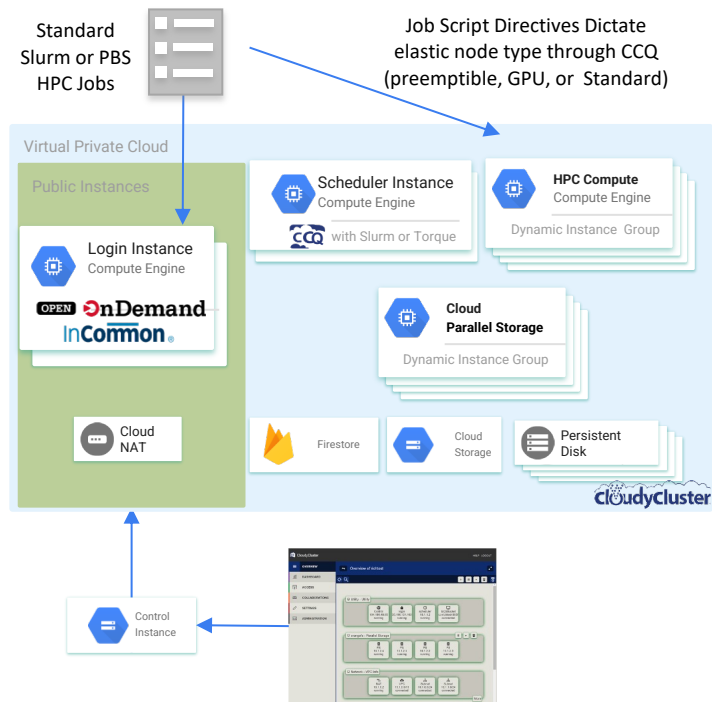
Includes Familiar Software

HPC Libraries: Boost, Cuda Toolkit, Docker, FFTW, FLTK, GCC, Gengetopt, GRIB2, GSL, HDF5, Intel MPI, Intel Runtimes, ImageMagick, JasPer, mpich, NetCDF, NumPy, Octave, OpenCV, OpenMPI, PROJ, R, Rmpi, SciPy, SWIG, WGRIB, UDUNITS, .NET Core, Singularity, Queue, Picard, xrootd, etc...

HPC Software: Amertools, ANN, ATLAS, BLAS, Blast, Blender, Burrows-Wheeler Aligner, CESM, GROMACS, JupyterLab, LAMMPS, NCAR, NCL, NCO, nwchem, OpenFoam, papi, paraview, Quantum Espresso, SAMtools, WRF, Galaxy, Vtk, Su2, Dakota, Gatk, etc...

ML Software: Mlpack, NuPIC, Octave, OpenCV, PICARD, Queue, Scikit-learn, Tensorflow

Sample Jobs: Intel Cluster Checker, Mpi_prime, wrf, etc...



Scaling on Google Cloud Platform

Google HPC Blog Post

Cloud against the storm: Clemson's 2.1 million VCPU experiment

<https://cloud.google.com/blog/topics/hpc/clemson-experiment-uses-2-1-million-vcpus-on-google-cloud>

Kevin Kissell, Technical Director,
Office of the CTO



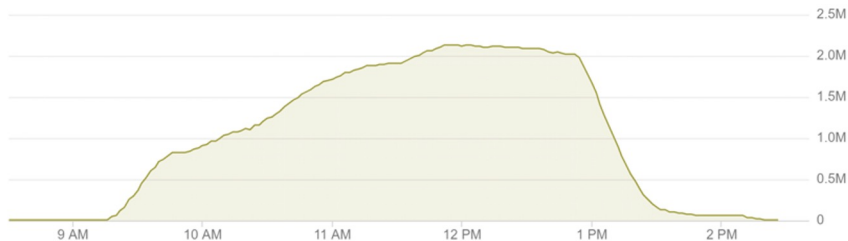
THE NEXT PLATFORM

Urgent HPC can Burst Affordably to the Cloud

<https://www.nextplatform.com/2020/01/08/urgent-hpc-can-burst-affordably-to-the-cloud/>



GCP CPU Core Ramp and Count



- 133,573 GCP Instances at peak
- 2,138,000 vCPUs at peak
- 6,022,964 vCPU hours

Processed 2,479,396 hours (~256TB) of video data

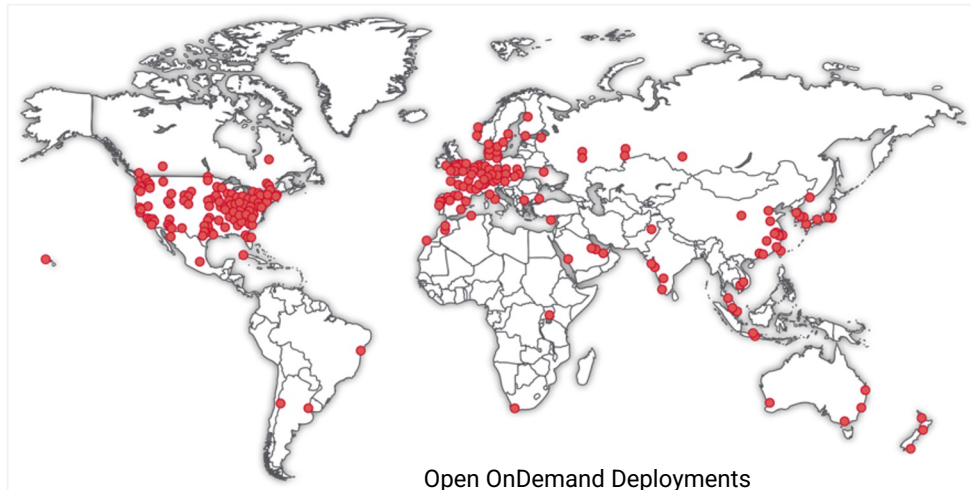
- ~4 hours of runtime
- ~1M vCPU within an hour
- ~1.5M vCPU within 1.5 hours
- 2.13M vCPU within 3 hours

Total Cost: \$52,598.64 USD

Average cost of \$0.008 USD per vCPU hour

Integration of OOD into CloudyCluster

- As part of the Cloudify Gateways program (Omnibond, OSC, U Buffalo, Va Tech)
- Full deployment automated as part of the CloudyCluster launch
- Working on project based UI initially to support data scientist project teams.
- The Collaboration Continues



CloudyCluster Files Jobs Clusters Interactive Apps

OPEN
OnDemand & CloudyCluster

OnDemand provides an integrated, single access point for all of your HPC resources.

Message of the Day

With Open OnDemand you are able to leverage a graphical user interface, while accessing the power of High Performance Computing With CloudyCluster. Other benefits include:

- Startup Interactive Resources, like Jupyter Notebook and Virtual Desktops
- File access which resembles Windows, MacOS and Linux file managers
- Job Composer tool to help build your job script files

["Your browser is the supercomputer: On Demand is a no-tears shortcut to research-computing"](#) from Matt Windsor of University of Alabama at Birmingham. A key phrase: "No experience necessary"

This material is based upon work supported by the National Science Foundation under grant numbers 1534949 and 1835725, and under active development by a team from the Ohio Supercomputer Center, U. of Buffalo CCR, and Virginia Tech

Relion & CRYO-EM

Working with the Cianfrocco Lab at University of Michigan to test data transfer and scalability to the Cloud for CRYO-EM projects using RELION

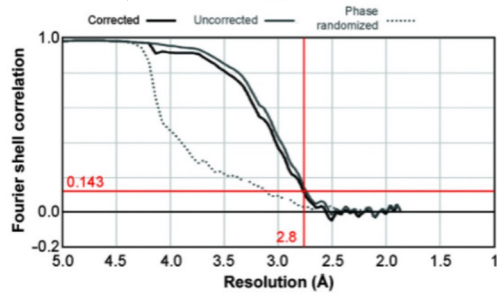
Also working with related Startups in the space that don't have access to large on prem clusters.

Initially Introduced as part of the Cloudify Program through the SGCI.

Cianfrocco is willing to consult with customers in conjunction with Omnibond and CloudyCluster

Another CRYO-EM project is in progress with the

ERN



High-resolution cryo-EM using beam-image shift at 200 keV

Jennifer N Cash¹, Sarah Kearns¹, Yilai Li¹, Michael A Cianfrocco¹

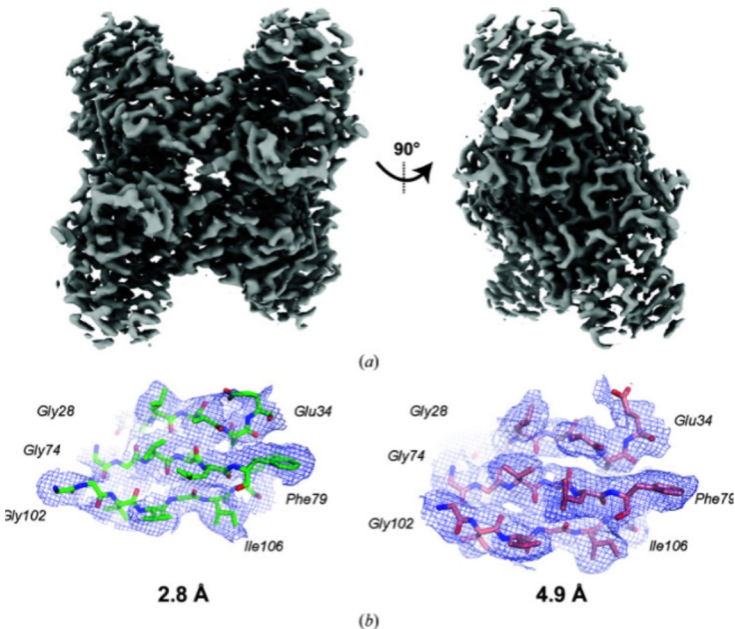
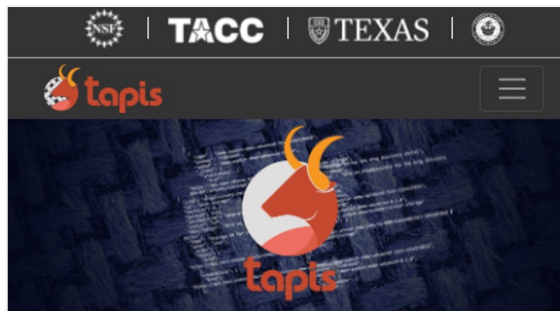


Figure 5 Final aldolase reconstruction at 2.8 Å resolution. (a) Sharpened aldolase reconstruction at 2.8 Å resolution. (b) Example densities and models for aldolase at 2.8 and 4.9 Å resolution. (c) FSC curve for the final reconstruction.

TAPIS - PEARC '22 Poster, Gateways '22 Paper



Tapis

Unify, simplify, automate and secure your interactions with advanced computing infrastructure.

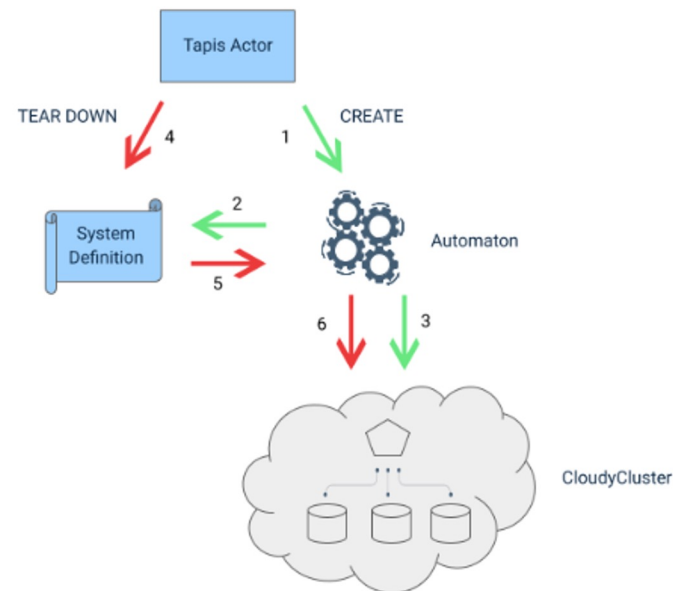
The Tapis Framework provides a hosted, unified web-based API for securely managing computational workloads across institutions so that experts can focus on their research instead of the technology needed to accomplish it.

Projects Using Tapis

- A2CPS
- Bridging Barriers
- CyVerse
- DesignSafe
- Drug Discovery Portal
- ECCO
- Ike 'Wai (at the University of Hawaii Manoa)
- iReceptor
- IR+
- NeuroNex 3DEM
- Planet Texas 2050
- Science Gateways Community Institute
- Synergistic Discovery and Design Environment (SD2E)
- VDJServer
- UTRC

Extending Tapis Workflow Management Framework with Elastic Google Cloud Distributed System using CloudyCluster by Omnibond

ERIC LAM, University of Hawaii, Manoa, USA
SEAN CLEVELAND, University of Hawaii - Information Technology Services - Cyberinfrastructure, USA
COLE MCKNIGHT, Omnibond Systems, USA
BOYD WILSON, Omnibond Systems, USA
RICHARD CARDONE, Texas Advanced Computing Center, USA
JOE STUBBS, Texas Advanced Computing Center, USA



<https://github.com/omnibond/automaton>

Rutgers & Cloud-Enabled Research

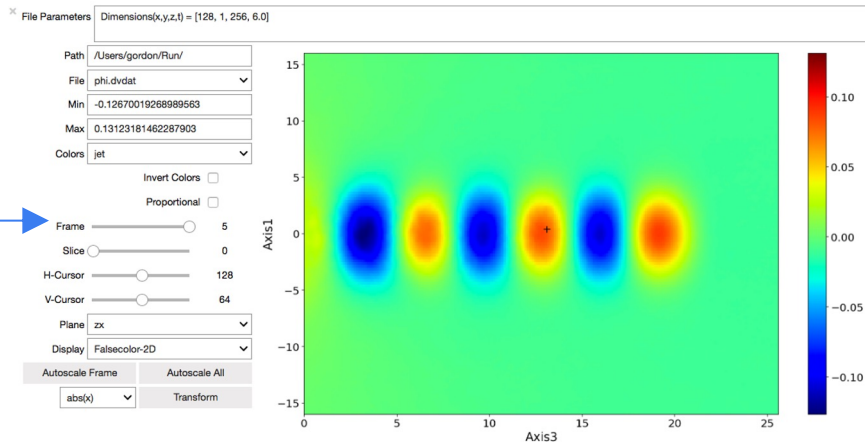
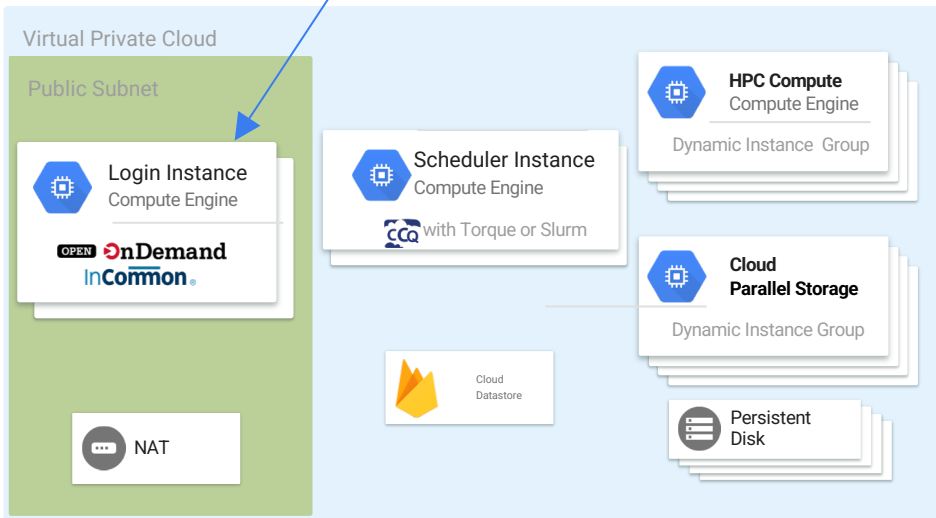
- 8,000+ faculty, 300 research centers
- Part of team's mission is pairing researchers with the **right tools for their projects**
- All faculty, staff, students have free access to our central HPC resources, but there are **constraints**
- Central IT manages accounting for granted and discounted cloud credits, including STRIDES-provided resources
- Cloud marketplaces are a **key differentiator** when considering cloud vs. on-prem capabilities
- **cloudycluster** is the primary recommendation for compute elasticity and scale



Wave Modeling in the Cloud (NRL)

Testing turboWave Modeling Scaling in CloudyCluster for NRL Briefing on Cloud Capabilities.

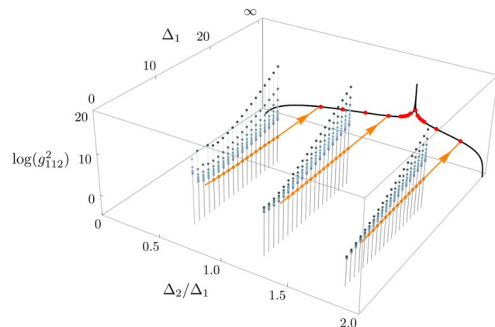
TurboWave
Workflow
HPC Job



SDPB Solver MPI Tuning

The Problem of Strong Coupling

Quantum field theory (QFT) is a universal language for theoretical physics, describing the Standard Model of particle physics, early universe inflation, and condensed matter phenomena such as phase transitions, superconductors, and quantum Hall fluids. A triumph of 20th century physics was to understand weakly coupled QFTs. However, weakly interacting systems represent a tiny island in theory space and cannot capture many of the most interesting physical phenomena.



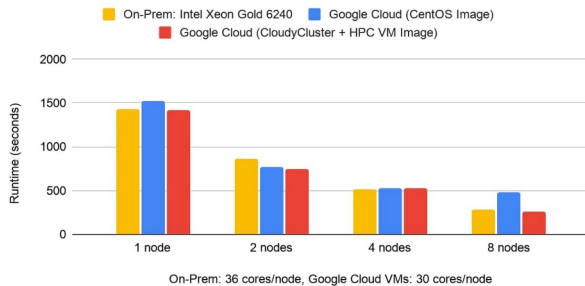
<https://bootstrapcollaboration.com/>

Our Collaboration

Recently, members of our collaboration discovered new bootstrap techniques that apply in general dimensions. In the past few years we have applied these techniques to a wide variety of seemingly unrelated problems: to perform the world's most precise analysis of the 3d Ising model, to constrain strongly coupled theories of physics beyond the Standard Model, to aid in classifying superconformal field theories, to derive locality and black hole thermality in models of quantum gravity, and to prove irreversibility of renormalization group flows. We believe this is the beginning of a much larger enterprise, crossing traditional boundaries between string theory, condensed matter physics, and phenomenology, and making strong connections to modern mathematics and computer science.

Customer Story: Scaling SDPB solver on Google Cloud

Runtime comparison: On-Prem vs. Google Cloud (Lower is better)



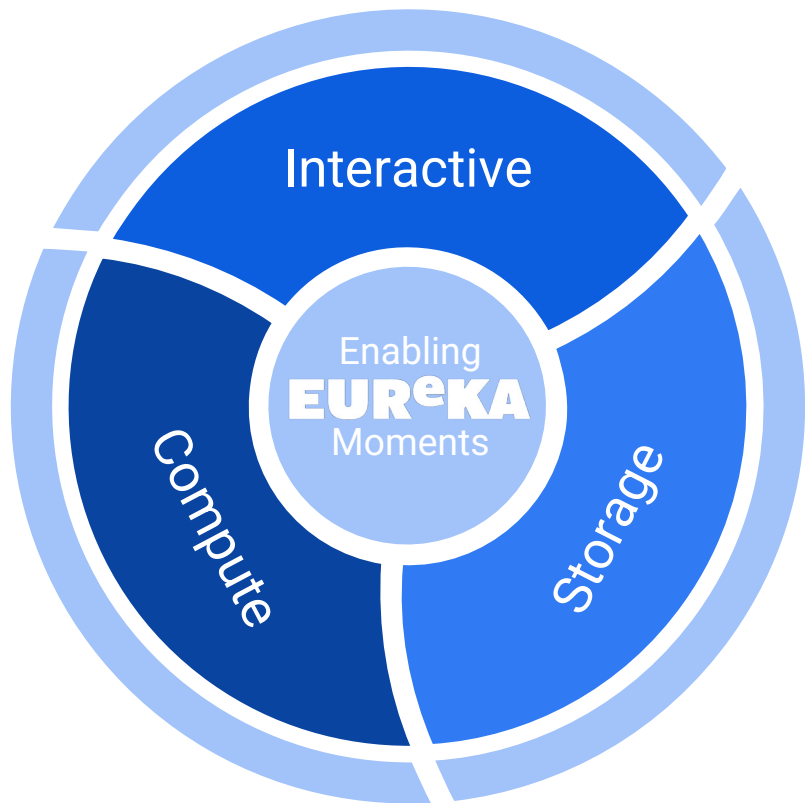
<https://cloud.google.com/blog/topics/hpc/introducing-hpc-vm-images>

To expand the collaboration's computation capabilities, Walter Landry wanted to see how SDPB would scale on Google Cloud. Working with Omnibond's **CloudyCluster** and leveraging the HPC VM image, Landry achieved comparable performance and scaling to an on-premises cluster at Yale, based on Intel Xeon Gold 6240 processors and Infiniband FDR.

Project Eureka Overview



Project Eureka Vision

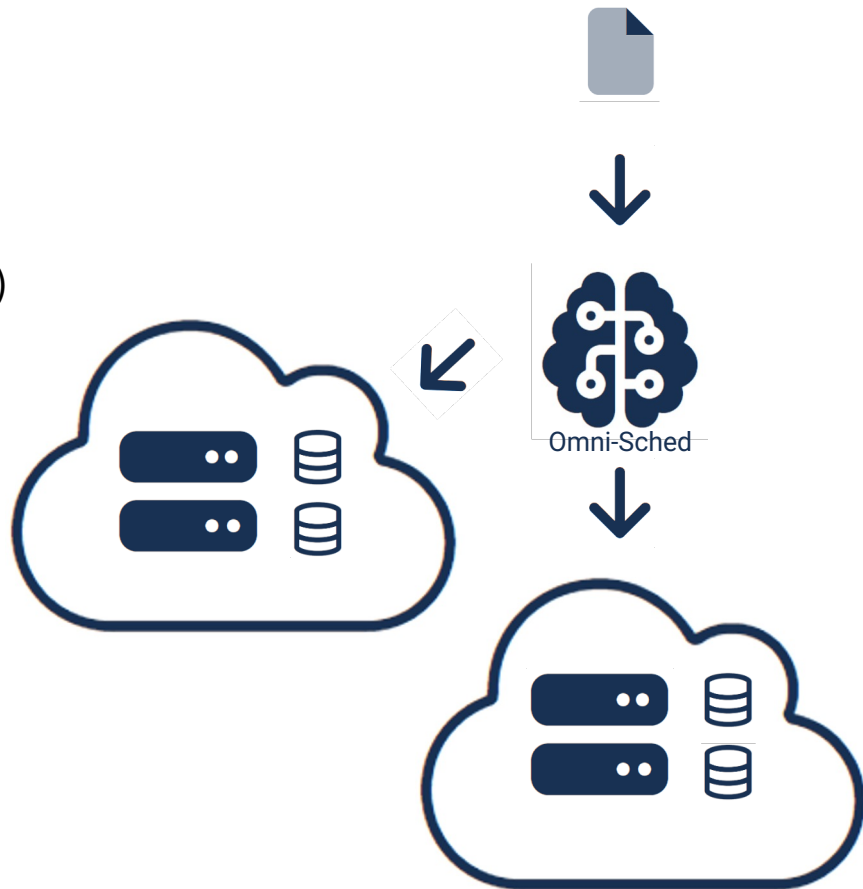


- **Interactive Apps**
 - Applications & Launchers
 - API Applets & SaaS Apps
 - Project Focused
- **Computational Apps**
 - Compute Anywhere (HPC, AI, & Beyond)
 - Enable Cloud Specialties
 - Simplify Compute and Storage Interactions
- **Storage Integration**
 - Integrate Diverse Storage Resources
 - Collaborate First
 - Project Level Data Lifecycle

The Omni-Scheduler Approach

Expand meta-scheduler concept (OmniSched) to handle:

- Dynamic Scratch
 - Driven by job Directives
- Data Staging
 - Driven by Job Directives (Input Data & Results)
- Beyond HPC
 - Data Science & Engineering Apps
 - HPC adjacent (publishing, websites, etc..)
 - Generative AI
 - Enable Apache Spark
 - K8s
- Scheduling Data Movements
 - iRODS integration
- Cross Cloud - leveraging Constellation / TerraForm
 - Enable On-Prem (TF -> kvm)
- Job routing between locations



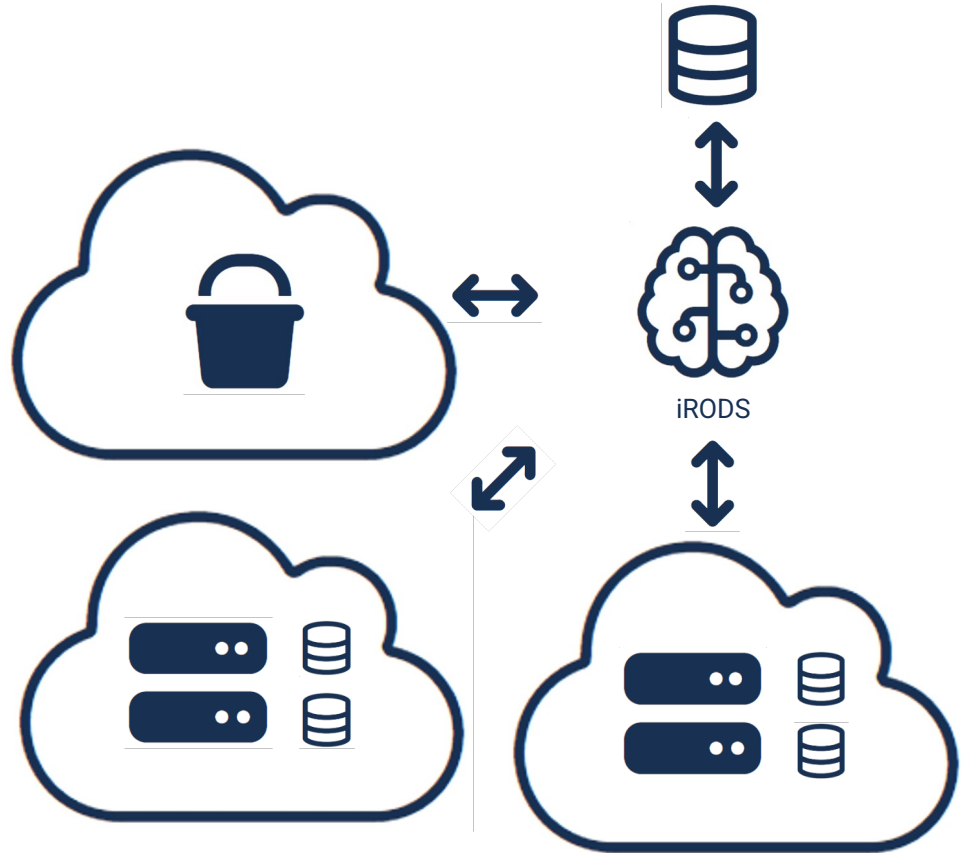
The Integrated Data Management

In addition to job based directives

Data Management Capabilities

- Replicate
- Archive
- Ingest
- Tier
- Publish

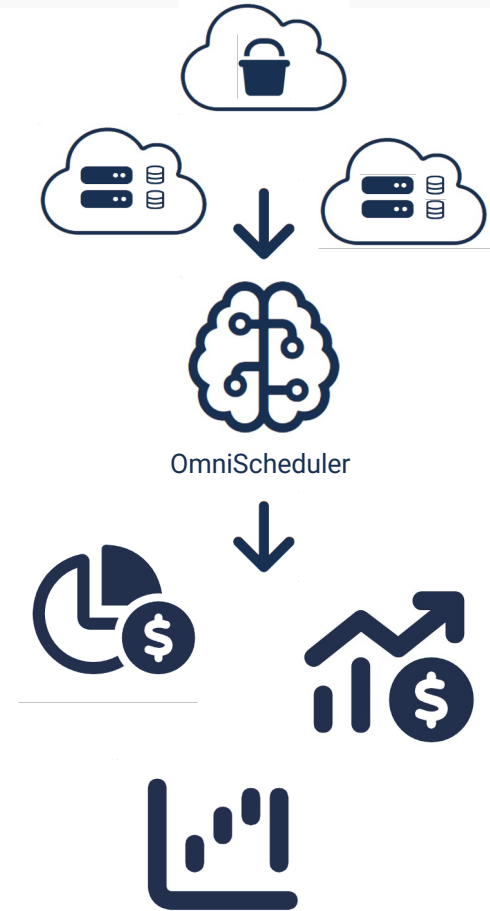
All configured in Eureka/OOD
with iRODS



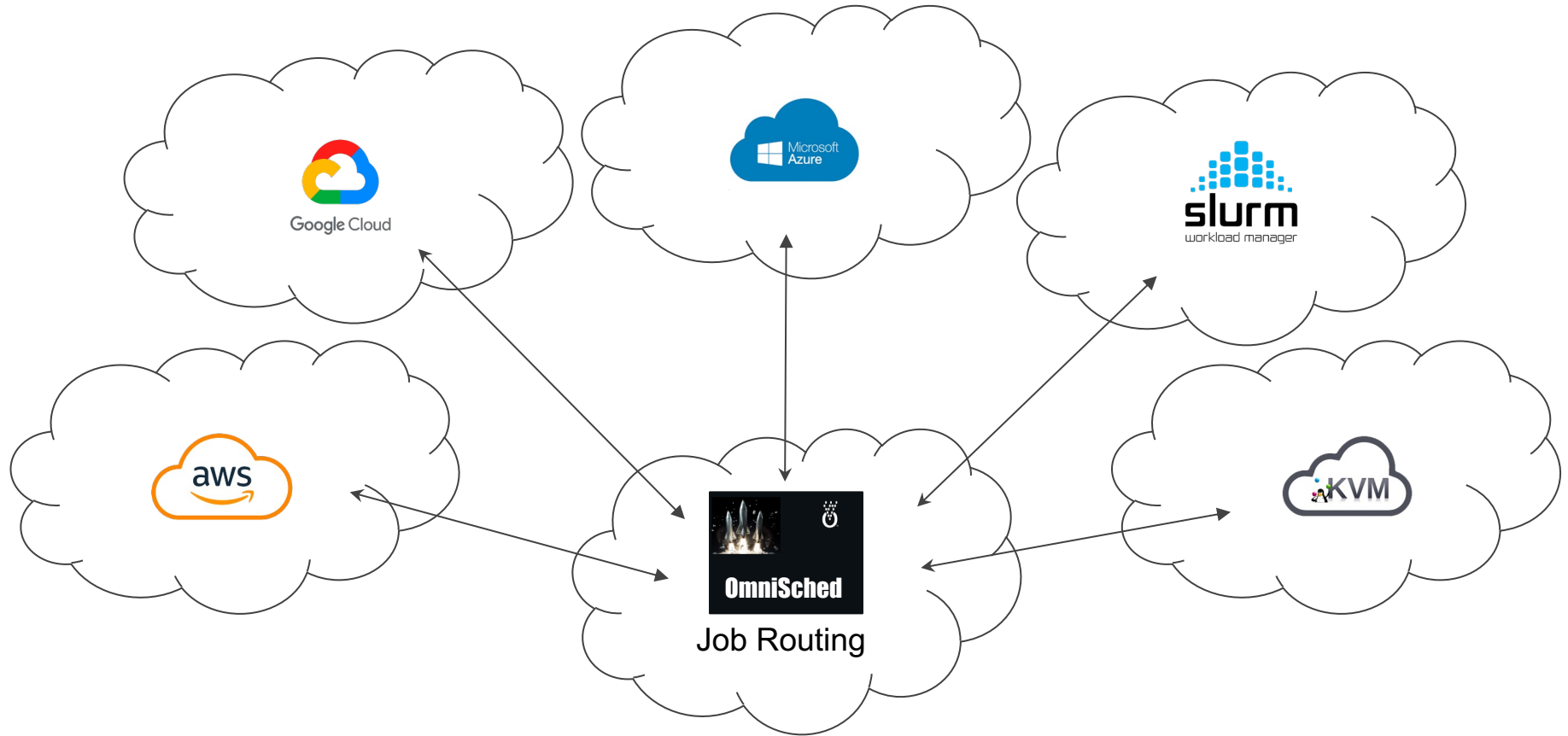
The Integrated Budget Management

Budget Management

- Person
- Project
- Department
- Limits with Actions
- Tag Integration with Cloud Providers Billing

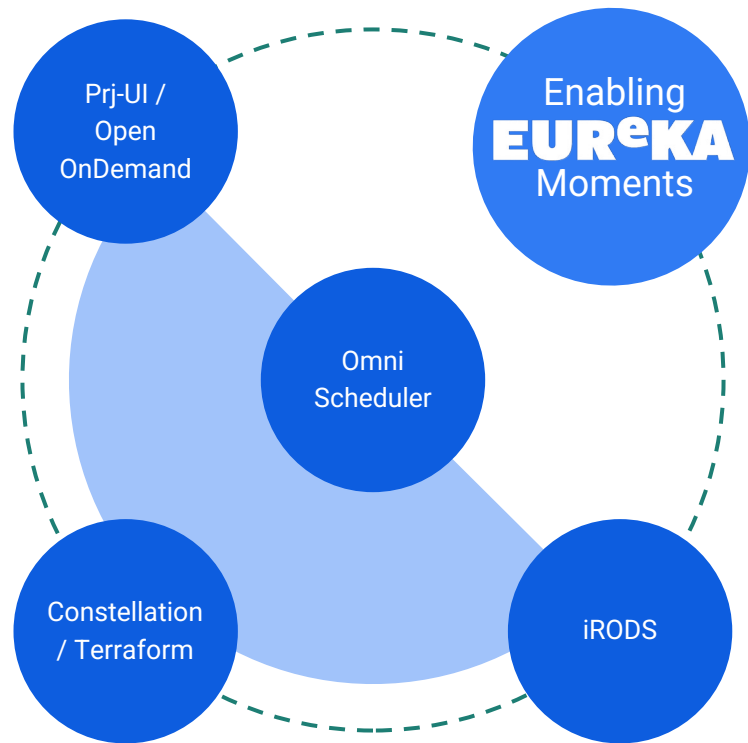


Multi-Cloud Architecture



Technology

- **Project UI with Open OnDemand**
 - HPC Industry Standard
 - Project-UI - Simplify & Empower
- **iRODS**
 - Metadata Driven Data Management & Movement
 - Designed for Diverse Storage
- **Constellation Driven Terraform**
 - Vendor Supported Interfaces
 - Multi-Cloud Enabler
- **Omni-Scheduler**
 - Coordinates UI, Storage, & Compute

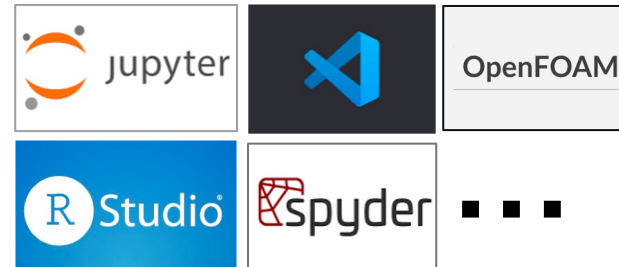


Overarching Architecture

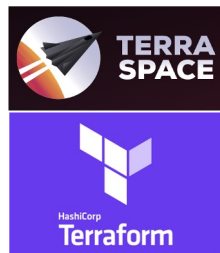
Project Based UI



Interactive and HPC Apps



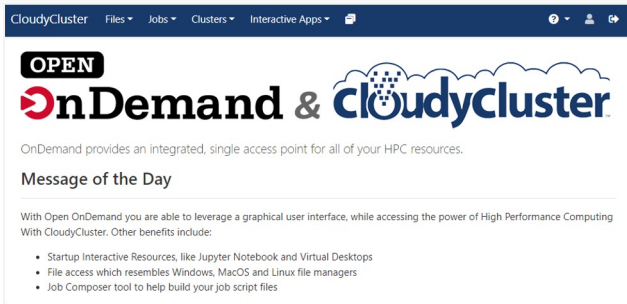
Cloud Integration



Integrated Storage



Eureka User Experience



OPEN
OnDemand & CloudyCluster

OnDemand provides an integrated, single access point for all of your HPC resources.

Message of the Day

With Open OnDemand you are able to leverage a graphical user interface, while accessing the power of High Performance Computing With CloudyCluster. Other benefits include:

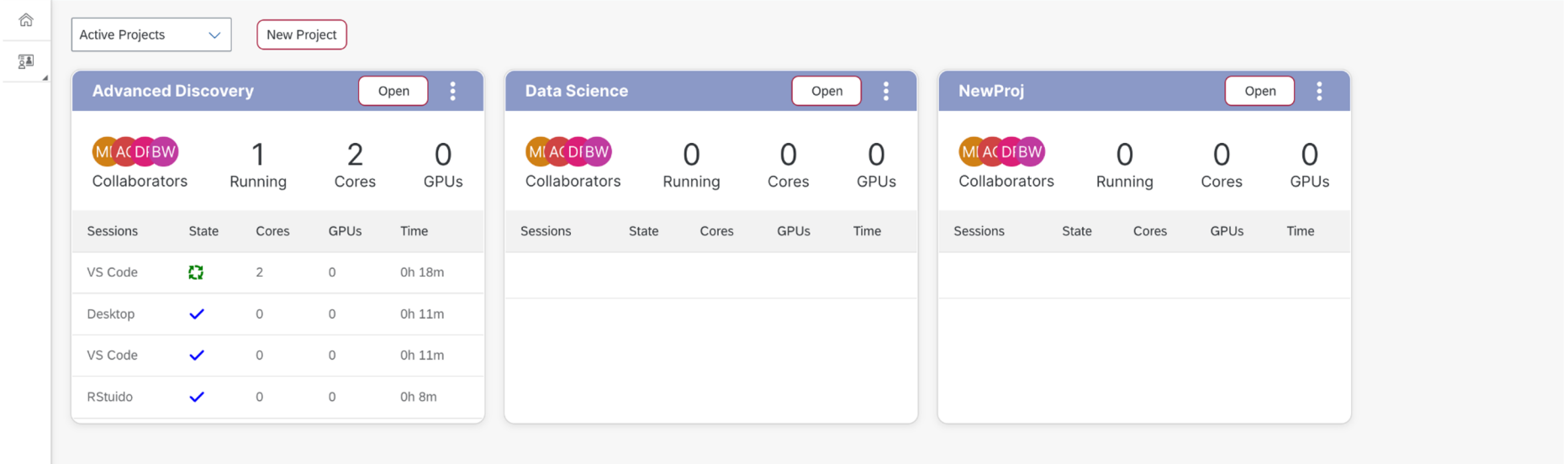
- Startup Interactive Resources, like Jupyter Notebook and Virtual Desktops
- File access which resembles Windows, MacOS and Linux file managers
- Job Composer tool to help build your job script files

This material is based upon work supported by the [National Science Foundation](#) under grant numbers 1534949 and 1835725, and under active development by a team from the [Ohio Supercomputer Center](#), [U. of Buffalo CCR](#), and [Virginia Tech](#)



Open OnDemand Deployments

Data Science Platform



Active Projects

Advanced Discovery

Open

MIACDFBW

Collaborators: 1 Running, 2 Cores, 0 GPUs

Sessions	State	Cores	GPUs	Time
VS Code		2	0	0h 18m
Desktop		0	0	0h 11m
VS Code		0	0	0h 11m
RStudio		0	0	0h 8m

Data Science

Open

MIACDFBW

Collaborators: 0 Running, 0 Cores, 0 GPUs

Sessions	State	Cores	GPUs	Time
----------	-------	-------	------	------

NewProj

Open

MIACDFBW

Collaborators: 0 Running, 0 Cores, 0 GPUs

Sessions	State	Cores	GPUs	Time
----------	-------	-------	------	------

Open OnDemand Deployments



Don't see an organization?
The more the merrier!
Let us know any that belong on the list



openondemand.org/orgs

iRODS Consortium Members



National Institute of
Environmental
Health Sciences



Universiteit Utrecht



TEXAS ADVANCED COMPUTING CENTER



Bibliothèque
et Archives
nationales



Maastricht University



university of
 groningen



SOFTIRON



INES
Centre Informatique National
de l'Enseignement Supérieur



Research Computing
UNIVERSITY OF COLORADO BOULDER



wellcome
sanger
institute



Touching lives through innovation



Value from Innovation



emnbond™



3 Anonymous

Omnibond Services

for Research Organizations

1 **Cloud Image Customization**

for seamless operation
- For a majority of workloads the cloud image can look like on premise image

3 **Job Cloud Billing Integration**

- From an auditing or bill back perspective

5 **Data & Results Integration**

2

Integrating Cloud <-> On-Prem

- Providing the ability to make the cloud simplified to the end researcher.

4

IDM Integration

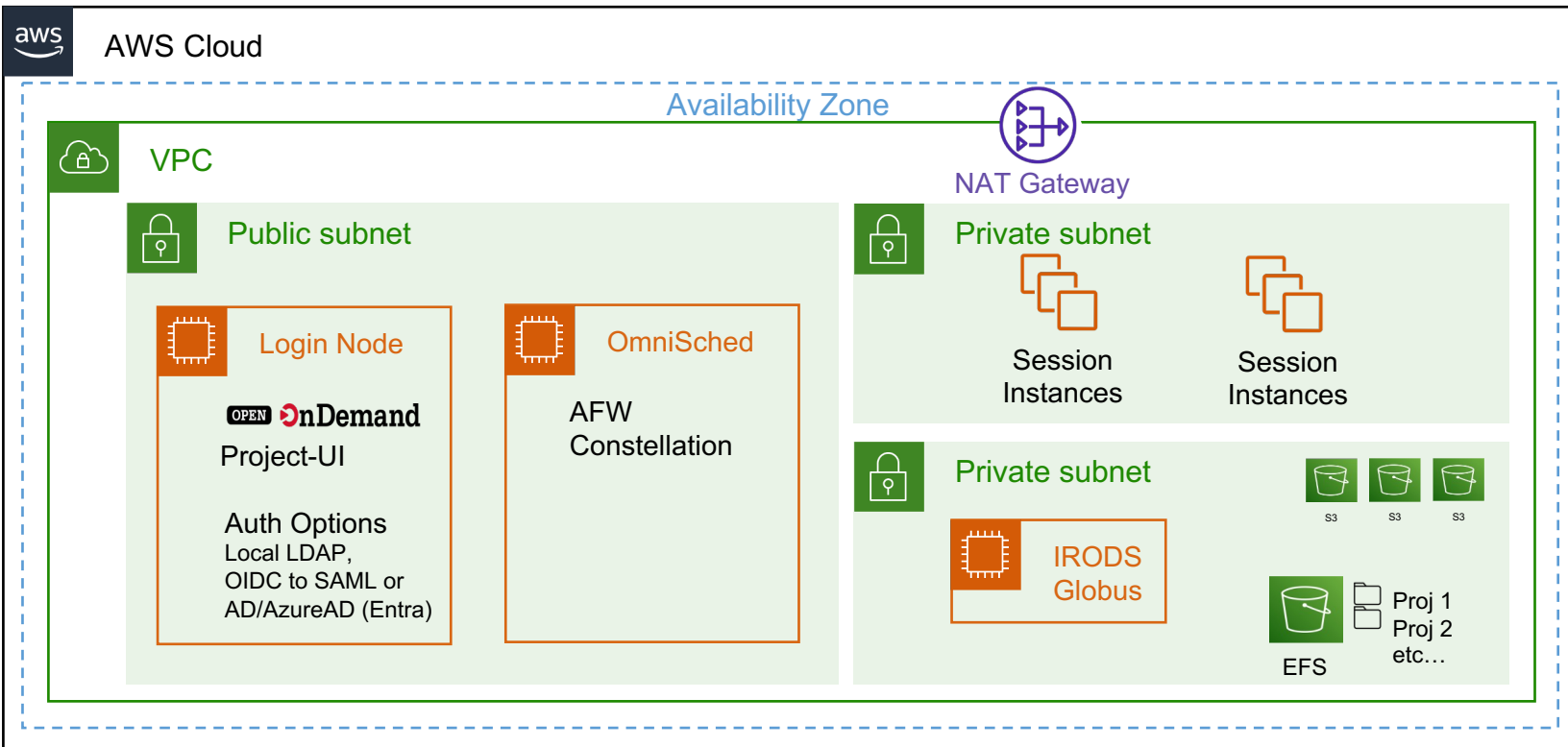
on Premise and Cloud HPC including HPC System Consolidation

6

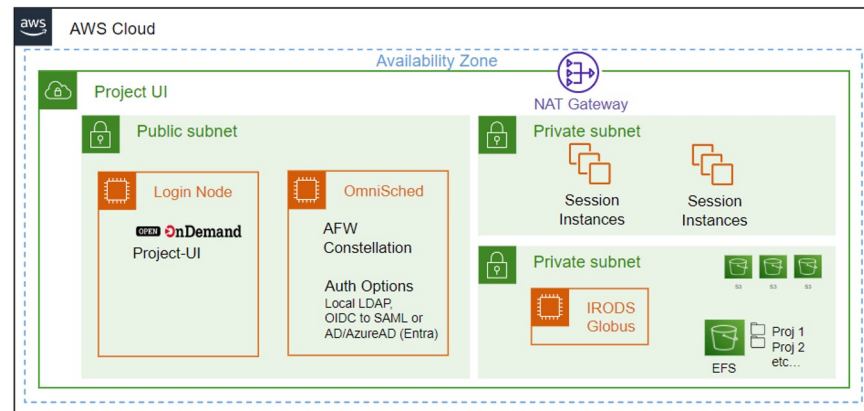
Training & Tools

to train the Research Computing Support Staff

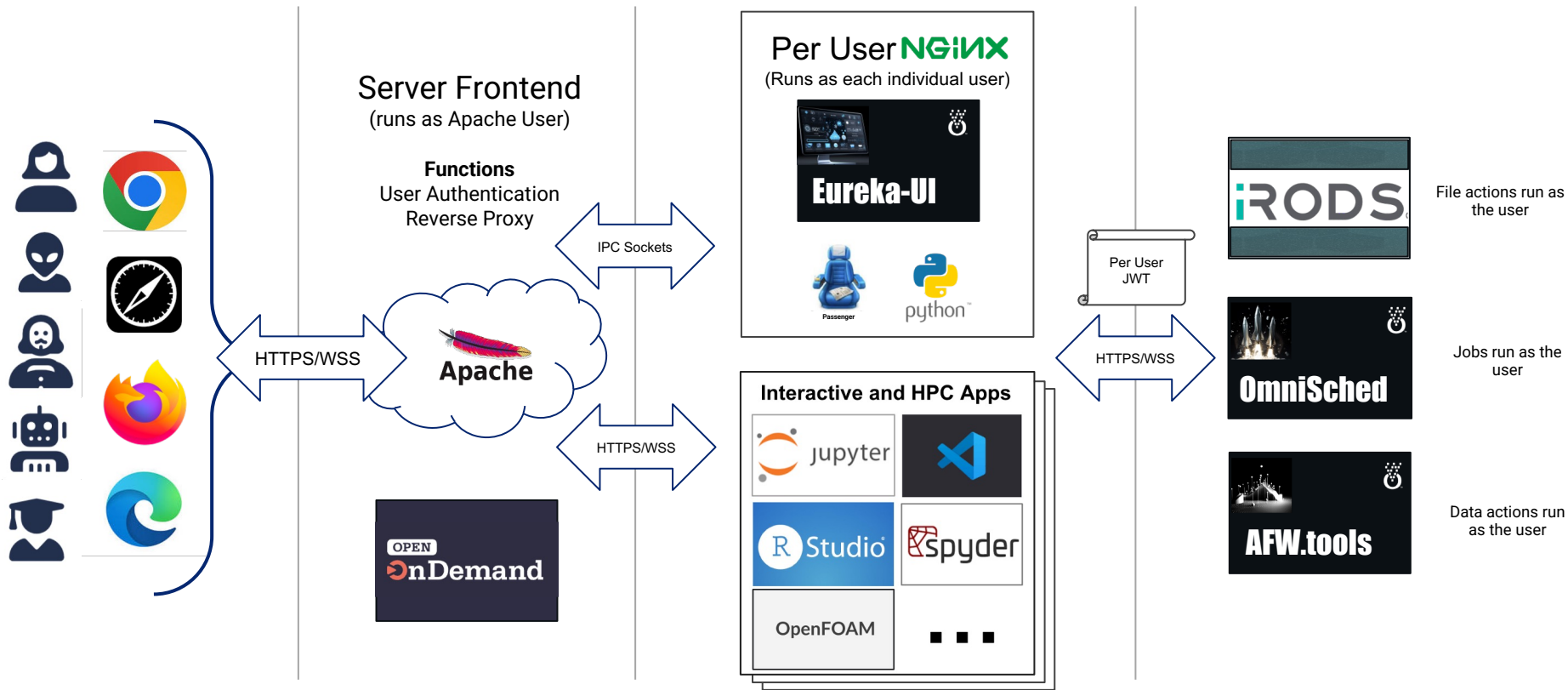
Deployment Architecture (AWS Example)




Identity Architecture (Using AWS as an Example)




User-Level Security Architecture






Project Eureka Demo/Screenshots





Projects Overview UI

☰  Data Science Platform BW


Active Projects

Advanced Discovery

 Collaborators: 1 Running: 2 Cores: 2 GPUs: 0


Sessions	State	Cores	GPUs	Time
VS Code		2	0	0h 18m
Desktop		0	0	0h 11m
VS Code		0	0	0h 11m
RStudio		0	0	0h 8m

Data Science

 Collaborators: 0 Running: 0 Cores: 0 GPUs: 0

Sessions	State	Cores	GPUs	Time
----------	-------	-------	------	------

NewProj

 Collaborators: 0 Running: 0 Cores: 0 GPUs: 0

Sessions	State	Cores	GPUs	Time
----------	-------	-------	------	------

Project UI

☰ **Data Science Platform** BW

🏠
📁
📄
⚙️

Advanced Discovery

Launchers ▾ Add Launcher Show Fewer

D Desktop ★ ⋮

2 Cores 4 RAM 0 GPU Launch

V VS Code ★ ⋮

2 Cores 4 RAM 0 GPU Launch

R RStudio ★ ⋮

2 Cores 4 RAM 0 GPU Launch

J Jupyter ★ ⋮

2 Cores 4 RAM 0 GPU Launch

S Spyder ★ ⋮

2 Cores 4 RAM 0 GPU Launch

S Stata ★ ⋮

2 Cores 4 RAM 0 GPU Launch

S Spyder2 ★ ⋮

2 Cores 4 RAM 0 GPU Launch

D OpenFOAM Small ★ ⋮

2 Cores 4 RAM 0 GPU Launch

D OpenFOAM Large ★ ⋮

32 Cores 64 RAM 0 GPU Launch

Active Sessions ▾

Runtime: 19h 57m
RStudio 2023.03.0 Running ⋮
Job Id: 6

2 Cores 4 RAM 0 GPU End Connect

Runtime: 19h 57m
Jupyter 4.0.0 Running ⋮
Job Id: 7




2 Cores 4 RAM 0 GPU End Connect

Runtime: 19h 24m
OpenFOAM Small Running ⋮
Job Id: 8










2 Cores 4 RAM 0 GPU End Connect

⌵

Future Storage Collections

  **Data Science Platform** 

Discover Innovation

	Name	Target	Resource Path	External Path	Storage Type	Automation	
	ProjectDirName		projects/projectDirName/	pvfs2://ofs-1.3334:/shared/projects/ProjectDirName	Posix	None	
	Bucketname		projects/Bucketname/	s3://Bucketname.s3.amazonaws.com/		Object	
	Replicated						
	KickBucket		projects/KickBucket/	s3://storage.googleapis.com/bucket/KickBucket.	Object	Tiered	

Add Storage Resource

>

Manager

The screenshot displays the 'Data Science Platform' interface. At the top, there is a navigation bar with a hamburger menu, the platform name, and a user profile icon labeled 'BW'. Below the navigation bar, the main content area is titled 'Discover Innovation'. On the left side, there is a sidebar with navigation icons and a list of storage collections: 'prj-storage', 'Folder1', 'Folder2', 'S3 Bucket-1', and 'GCS Bkt'. The 'prj-storage' collection is expanded, showing a table of files. The table has columns for 'name', 'last_write_time', 'size', 'type', and 'logical_path'. Two files are listed: 'OmniPasskey.png' and 'Screenshot 2023-05-14 at 7.31.42 PM.png'. A callout box with a pointer highlights the 'GCS Bkt' collection, with the text 'Additional Storage Collections' inside.

Data Science Platform

Discover Innovation

Upload

prj-storage > Folder1 >

name	last_write_time	size	type	logical_path
OmniPasskey.png	1684165227	3821822	data_object	prj-storage/Folder1/OmniPasskey.png
Screenshot 2023-05-14 at 7.31.42 PM.png	1684165200	205944	data_object	prj-storage/Folder1/Screenshot 2023-05-14 at 7.31.42 PM.png

S3 Bucket-1

GCS Bkt

Additional Storage Collections