intel

cloudycluster by Omnibond

Google Cloud

SGCI

globus online

TACC
TEXAS ADVANCED COMPUTING CENTER

XSEDE

HPC in the City: St. Louis HACKATHON

SC21
St. Louis, MO | science & beyond.

Google and CloudyCluster Training

Join the HPC in the City Discord using this QR Code!

Google Cloud

globus online

TEXAS ADVANCED COMPUTING CENTER

HACKATHON

Google and CloudyCluster Training

Agenda

- Introductions
- Hackathon Objective
- Deliverables and Resources
- General Information
- Google Cloud Platform
- Cloudycluster

Presenter: Je'aime Powell

Organizers

Alex Nolte - *University of Tartu*
alexander.nolte@ut.ee

Boyd Wilson - *Omnibond*
boyd@omnibond.com

Amy Cannon - *Omnibond*
amycannon@omnibond.com

Je'aime Powell - *TACC*
jpowell@tacc.utexas.edu

Linda Hayden - *ECSU*
haydenl@mindspring.com

HPC in the City: St. Louis HACKATHON

http://hackhpc.org/hpcinthecity

SC21 St. Louis, MO science & beyond.

The Objective of HPC in the City: St. Louis

The hackathon aims to harness the resources, skills, and knowledge found in the HPC community in an effort to provide applied exposure towards students from 2-4 year post-secondary educational institutions. In short, the hackathon will provide HPC skills and training while targeting problems that directly affect the participants.

- Develop knowledge about solutions to identified issues affecting St. Louis through application of data analysis/presentation or management.

Student Outcomes

- Increased familiarity with data science in the cloud
- Experience collaborative software engineering
- Develop professional communication skills

Student Deliverables and Resources

Deliverables:

- **Source code Including Comments**
- **PDF of presentation**
  - Team members with pictures
  - Use of HPC technology in the project
  - Regional (St. Louis) implications of the project
- **Github Repository Link**
  - README.md with project description

Resources:

- **Google Cloud (Provided Credits)**
- **Cloudy Cluster**
- **Most Commonly Used**
  - Python
  - Jupyter Notebooks
  - Node.Js (JavaScript)
  - Repl.it (Collaborative Environment)
  - HTML
- **Discord**
  https://discord.com/invite/rSXasYKDwE

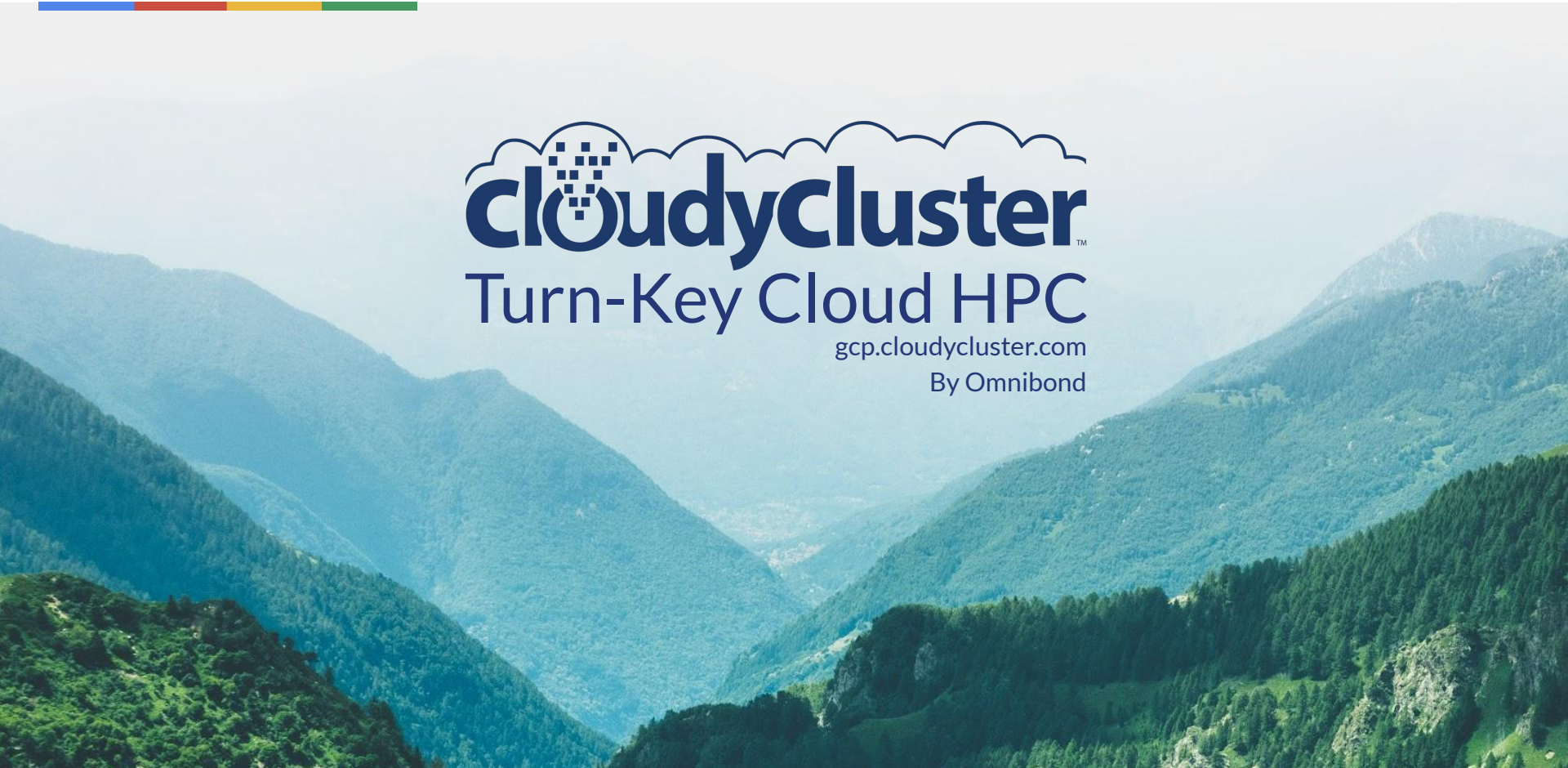
Join the HPCHack Discord using this QR Code!

General Information (the 3 T's)

- **Teams**
  - 4-5 Students
  - 1 Primary Mentor
  - 1 Specialist/Staff
- **Time**
  - November 4th - 8th
    - 11/4@~6pm ET Event Start
      - Team formation
    - 11/[5-8] @ 11 ET & 6pm ET- Checkins
    - 11/8@6pm ET-Final Presentations
- **Topic Examples**
  - Data Analysis of COVID 19
  - Economic disparities and their effects on college participation
  - Genomics, Molecular Dynamics, or Weather Modeling in the Cloud.
  - Social Justice
  - AI-based Crowd Status
  - Public Data Management
  - Graduation Rates
  - Broadband Access
  - Insurance vs. Public Health Resilience

Presenter: Boyd Wilson

HPC Resource Platforms


CloudyCluster™
Turn-Key Cloud HPC
gcp.cloudycluster.com
By Omnibond

Omnibond - Introduction

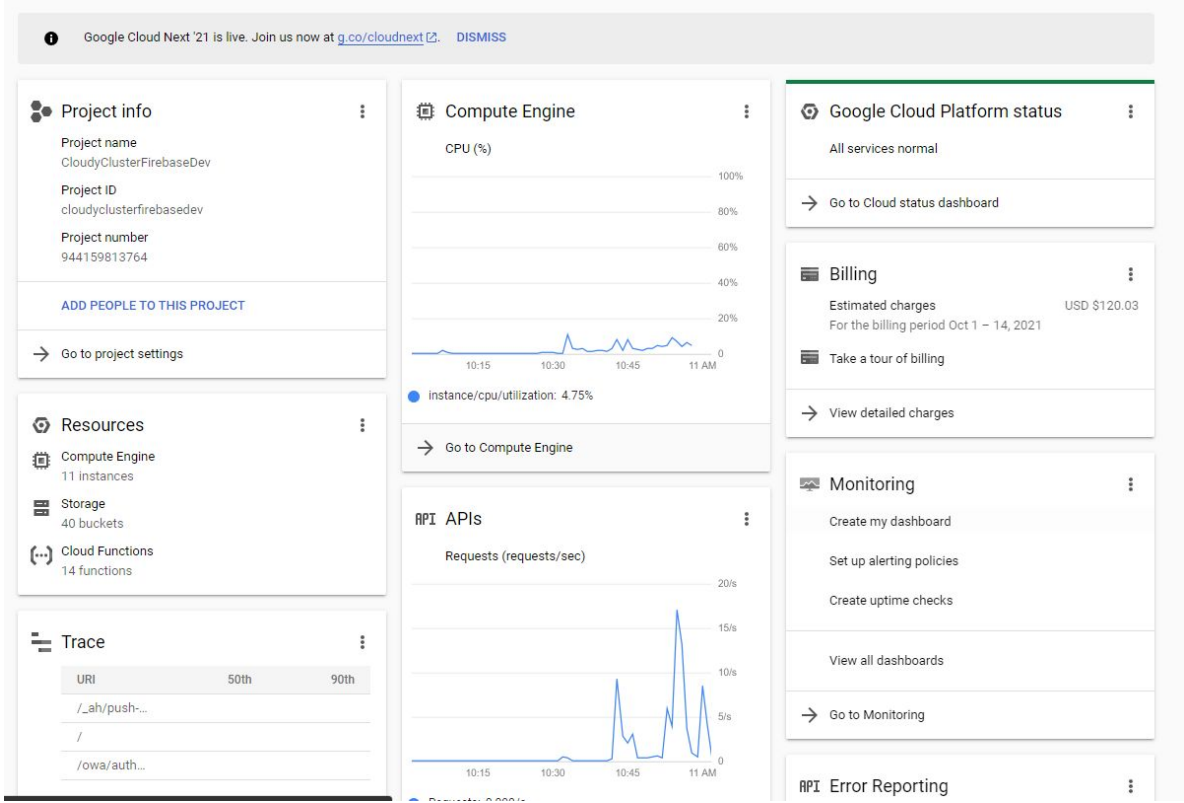
Leadership Team

- Former COO at NCSA
- CIO at Purdue and Clemson Universities
- Sw CTO - Clemson University
- Director Computing Engineering School Miami Ohio
- Consultant to Tribal Colleges and Minority Serving Institutions on cyberinfrastructure
- Founded ACI-REF http://www.aciref.org and CaRCC http://carcc.org
  (Emphasize the human element)

Software Products

- Cloud Orchestration - **CloudyCluster**
- Computer Vision & AI based **TrafficVision** analytics system.
- Parallel File System - **OrangeFS** included upstream in the Linux Kernel
- Identity Management - **Identity Manager Connectors**
  - products installed in thousands of customers worldwide

Google Cloud Console



Turn-Key HPC with CloudyCluster

CloudyCluster = Automated Cloud CyberInfrastructure

- Automated Deployment of fully functioning HPC/HTC environment
  (Complete with VPC, firewall config, scheduler, login, parallel & object storage)
- Provides users a familiar HPC & HTC experience to on-prem
- Over 300 HPC, HTC, and AI packages and libraries pre-configured
- Job scripts determine instances used (Standard, Preemptible, GPU)
- Slurm or Torque fronted by the powerful CCQ meta-scheduler
- Parallel and Object storage Options
- Operates in the customer's own GCP account (eliminating 3rd party BAAs)
- HPC / HTC job integration with GCP Billing Labels
- Automatically leverages GCP Placement Policies for jobs using Supported (C2) Instances

cloudycluster

Self-Service Elastic HPC & HTC

You Create a fully operational & secure computation cluster in minutes, complete with:

Encrypted Storage: GCS, OrangeFS on PD

Compute: Job Driven Elastic Compute through CCQ

Scheduler: Torque & SLURM with the CCQ Meta-Scheduler

Includes over 300 packages and libraries used in HPC including:

HPC Libraries:

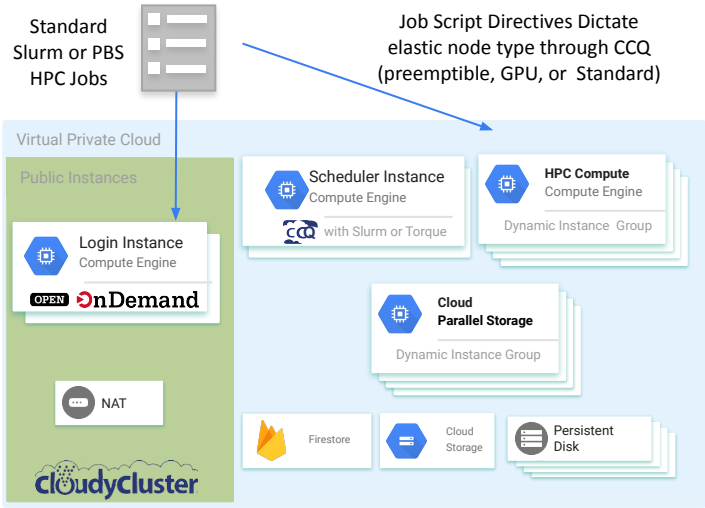
Boost, Cuda Toolkit, Docker, FFTW, FLTK, GCC, Gengetopt, GRIB2, GSL, Hadoop, HDF5, ImageMagick, JasPer, mpich, NetCDF, NumPy, Octave, OpenCV, OpenMPI, PROJ, R, Rmpi, SciPy, SWIG, WGRIB, UDUNITS, .NET Core, Singularity, Queue, Picard, and xrootd

HPC Software:

Ambertools, 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, and Gatk

ML Software:

Mlpack, NuPIC, Octave, OpenCV, PICARD, Queue, Scikit-learn, Tensorflow and Theano



Customization: *You can easily add your own software to a custom Image.*

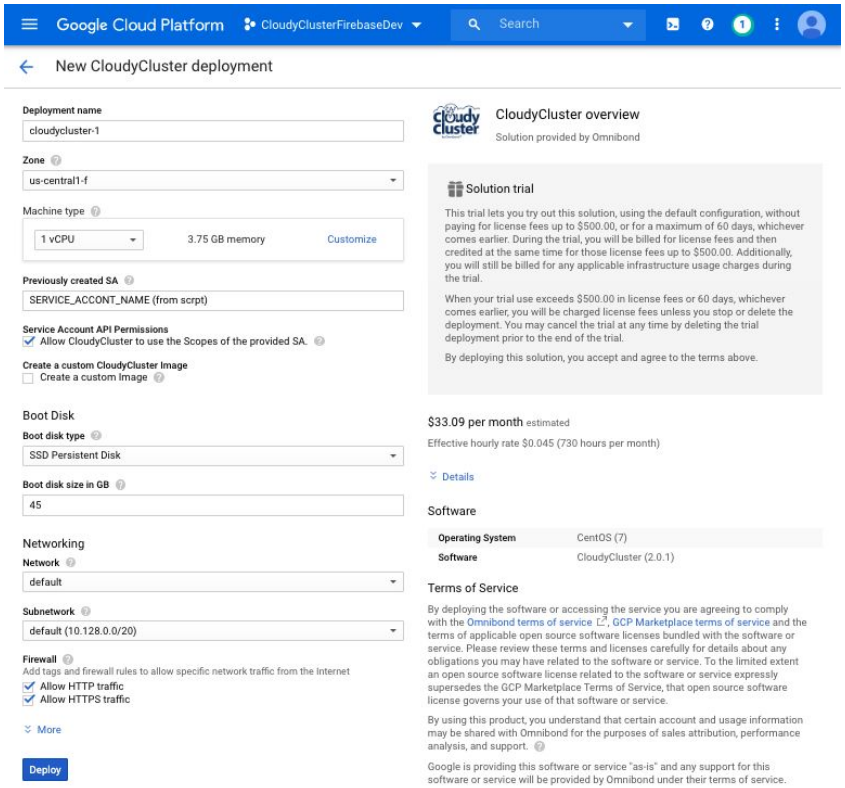
Consumption, Deployment & Engagement

CloudyCluster can be launched easily from the GCP Marketplace:

- Launch easily from the GCP Console
- Pricing is about 5% of on-demand costs
- Can be sold as a subscription for select customers

The CloudyCluster Team Can Help with:

- HPC & HTC Discussions (Consulting and Pre-Sales)
  - Workloads that work well in the Cloud
  - Migrating HPC & HTC Jobs that can run in the cloud and on premises
  - Cloud HPC Cost Estimates
  - Proof of Concept (POC) workflow
- Additional Services available for tighter campus integration



Google Cloud

CloudyCluster

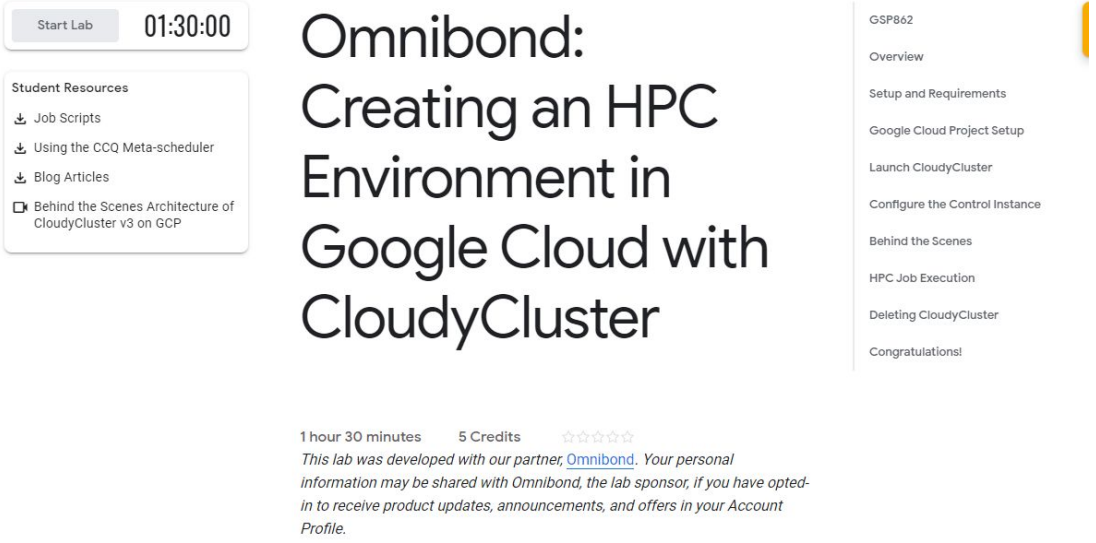
Online Training - Qwiklabs

Qwiklabs

Here is the link to apply for credits with qwiklabs:

https://www.cloudskillsboost.google/focuses/21221?parent=catalog

Here is the link to the CloudyCluster qwiklab:
https://edu.google.com/programs/credits/training/?modal_active=none



Google Cloud

cloudycluster

HPC Workloads for Google Cloud

If you look at the Campus resources a majority (>90%) of the jobs will run on <9 instances (given 30 cores per instance).

The MPI latency between <9 instances is only impactful on a very small number of job types.

A majority of workloads run on a campus cluster will be supported in GCP

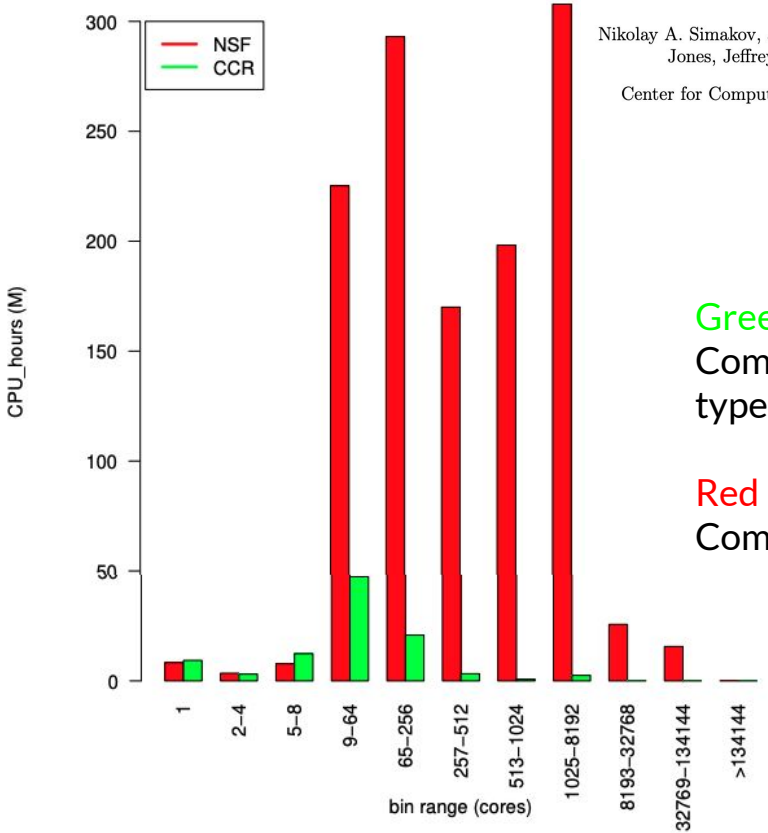
The larger NSF resources require more investigation

A Workload Analysis of NSF's Innovative HPC Resources Using XDMoD

Nikolay A. Simakov, Joseph P. White, Robert L. DeLeon, Steven M. Gallo, Matthew D. Jones, Jeffrey T. Palmer, Benjamin Plessinger, and Thomas R. Furlani

Center for Computational Research, University at Buffalo, Buffalo NY USA 14203

January 16, 2018



Green represents Campus Computing Resources and the job types (# cores)

Red represents National NSF Computing Resources

Google Cost Comparison

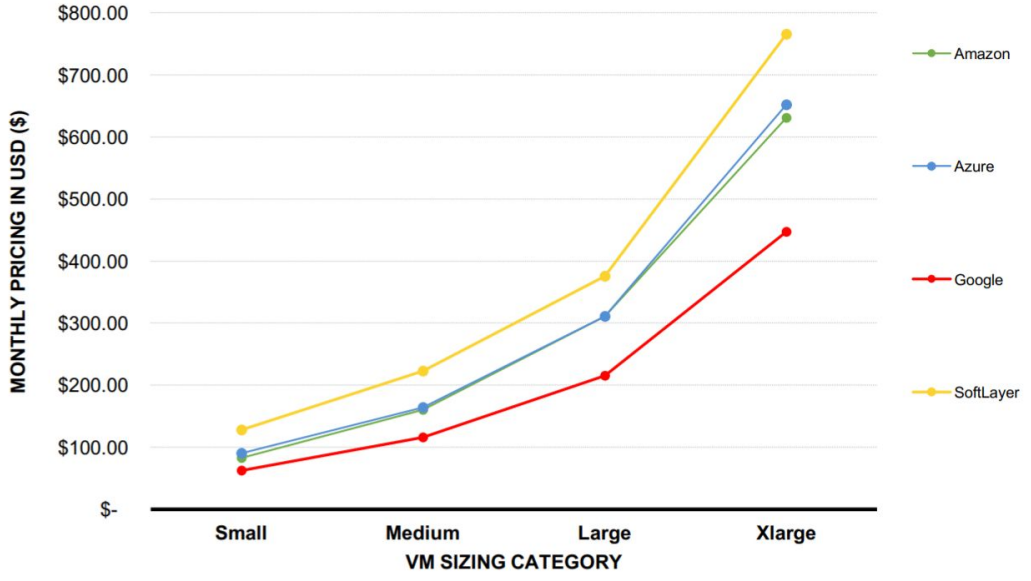


Table 7A: Monthly Cost of VMs Across CSPs

	Small	Medium	Large	Extra Large
Amazon	$83.00	$160.27	$310.54	$631.08
Azure	$90.19	$163.92	$310.65	$652.02
Google	$62.24	$115.98	$214.96	$446.92
SoftLayer	$128.00	$223.00	$376.00	$766.00

Source: Kinsta

Next Evolution in Cloud Pricing

Google Cloud Subscription Agreement for Public Sector

Provides the “right to use” any GCP product for a period of time for a defined, specified use case for a fixed price

Provides you with predictability of costs - no overage fees

Reduce your implementation risk - use whatever service you need, as much as you need, for your project

We’re in this with you! **Google Cloud has “skin in the game”** with you - we want your project to be successful

Google Cloud

Google Working Toward Optimal HPC Performance

Google Cloud

Blog Menu

HPC

Introducing HPC VM images—pre-tuned for optimal performance

Pavan Kumar
Product Manager

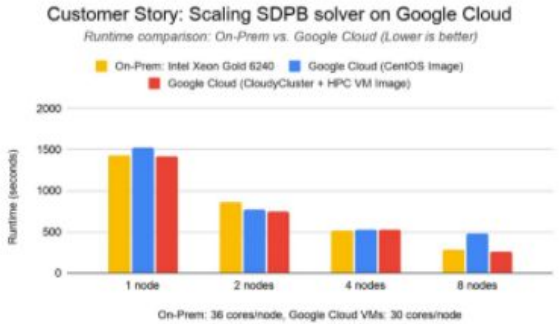
Jason Zhang
Software Engineering Manager

February 2, 2021

Customer story: Scaling SDPB solver using CloudyCluster and HPC VM image

Walter Landry is a research software engineer in the Caltech Particle Theory Group working with the international Bootstrap Collaboration. The collaboration uses SDPB, a semidefinite program solver, to study Quantum Field Theories, with application to a wide variety of problems in theoretical physics, such as early universe inflation, superconductors, quantum Hall fluids, and phase transitions.

To expand the collaboration's computation capabilities, Landry wanted to see how SDPB would scale on Google Cloud. Working with Omnibond 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.



https://www.googlenewsapp.com/introducing-hpc-vm-images-pre-tuned-for-optimal-performance/

HTC Scaling

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

THENEXTPLATFORM

Urgent HPC can Burst Affordably to the Cloud

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

- 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

CLEMSON UNIVERSITY

Google Cloud

cloudy cluster by Omnibond

intel

TrafficVision

burwood group, inc.

Architecture - Video

DEMO - Video

1. Control Instance Creation
2. Environment Creation

DEMO - Video

3. Job submission through the CCQ Meta-Scheduler to SLURM
4. CCQ instance clean up once the Jobs have completed

Questions and Concerns

Next Training Sessions:

- Data to Dashboard - [*10/21/21*]

- Beginning to End Project Example - [*10/28/21*]

Schedule:
https://jeaimehp.github.io/HackHPC-HPCintheCity21/

Presenter Contact Information:

Boyd Wilson (Omnibond) - boyd@omnibond.com