KubeCon | CloudNativeCon
North America 2021

RESILIENCE REALIZED

Kubernetes VMware User Group

Using GPUs with Kubernetes on vSphere

Myles Gray
Senior Technical Architect
VMware

Steven Wong
Open Source Software Engineer
VMware

Thursday October 14, 2021 - 11am Pacific
https://sched.co/lV8U

Agenda

Cloud workloads and GPUs

Benefits of GPUs as a pooled resource

GPU support for Kubernetes workloads on vSphere

Demo

GPUs + Kubernetes + vSphere: resources

How to participate in the Kubernetes VMware User Group

GPUs

What and Why?

Graphics Process Units, or GPUs, were originally designed as specialized compute devices for computer graphics and image processing. The are designed to rapidly manipulate data in a fast memory buffer. The highly parallel processing capability has proven to be more efficient than general purpose CPUs for algorithms that benefit from **processing large amounts of data in parallel**.

Common applications that utilize parallel processing include machine learning, modeling, image and video processing and recognition, modeling, graphics editing and more.

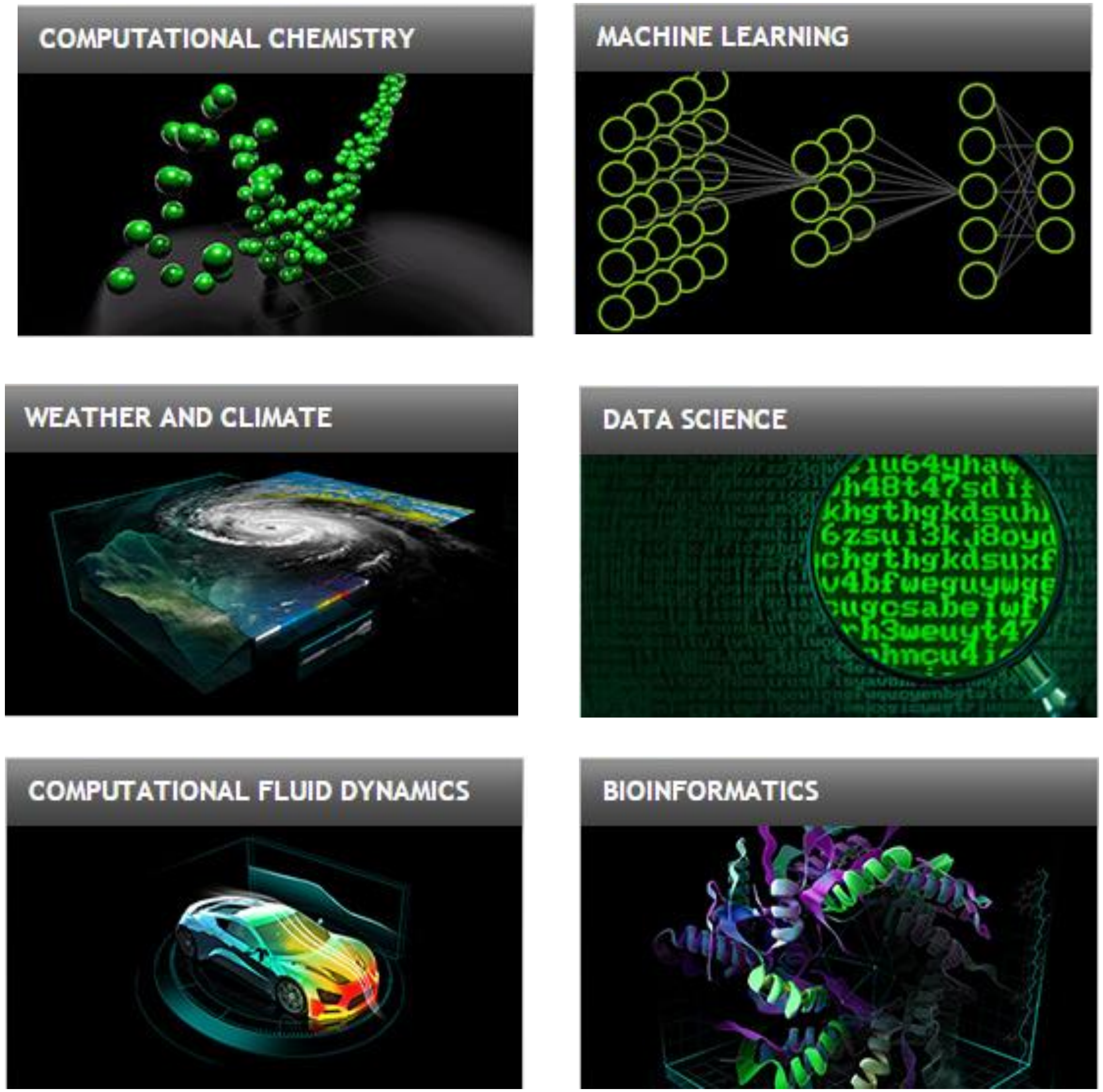


Image Source: Nviidia

Seymour Cray

“Which would you rather use: two strong oxen or 1024 chickens?”

This quip was viewed as obvious into the 90s – but today you want the chickens

Photo Bibliophile [CC BY-SA 3.0 (https://creativecommons.org/licenses/by-sa/3.0)]

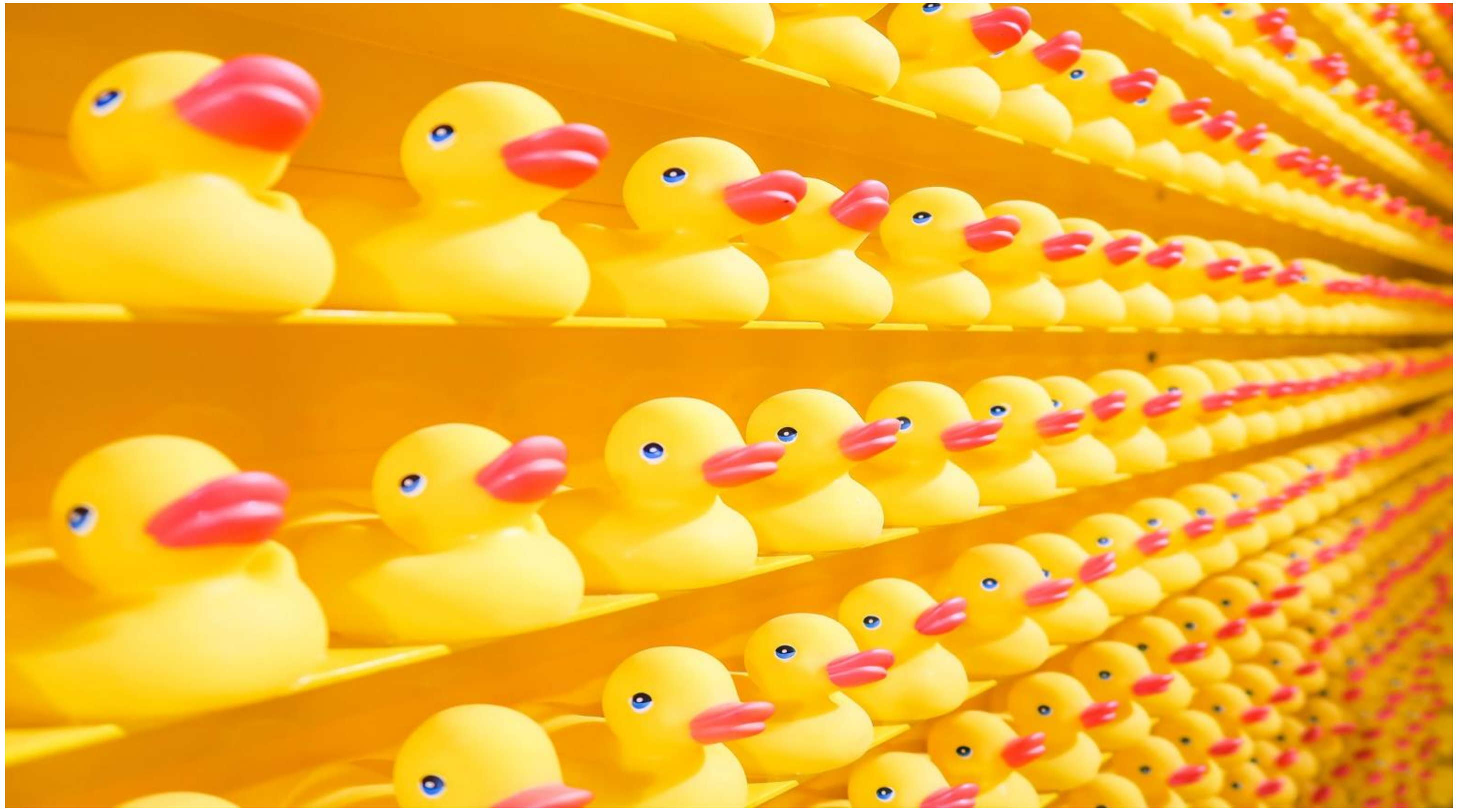
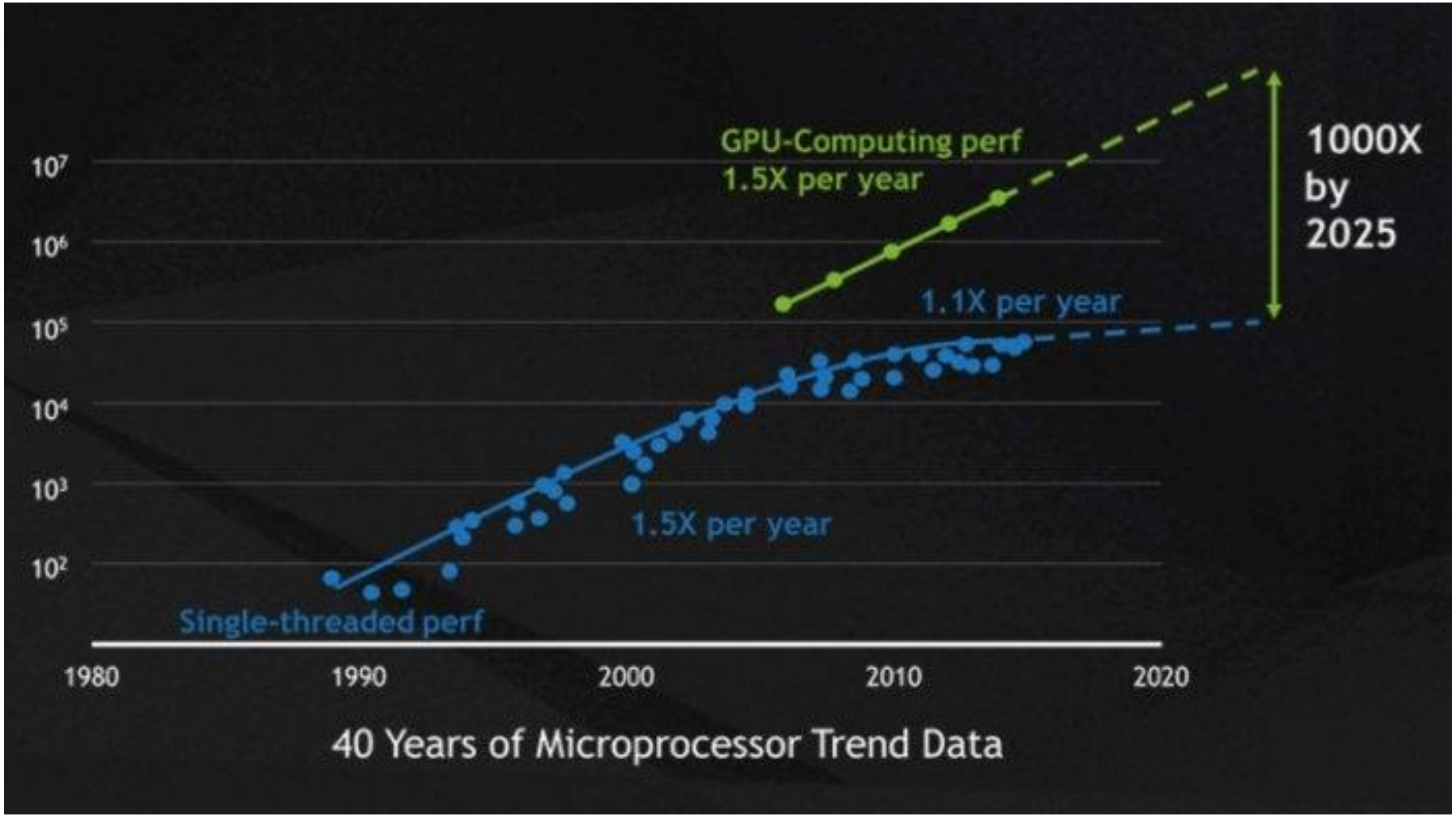


Photo Joshua Coleman on Unsplash

Why GPUs

Performance trajectory

GPU overall performance has been on a steeper growth curve that conventional CPUs.

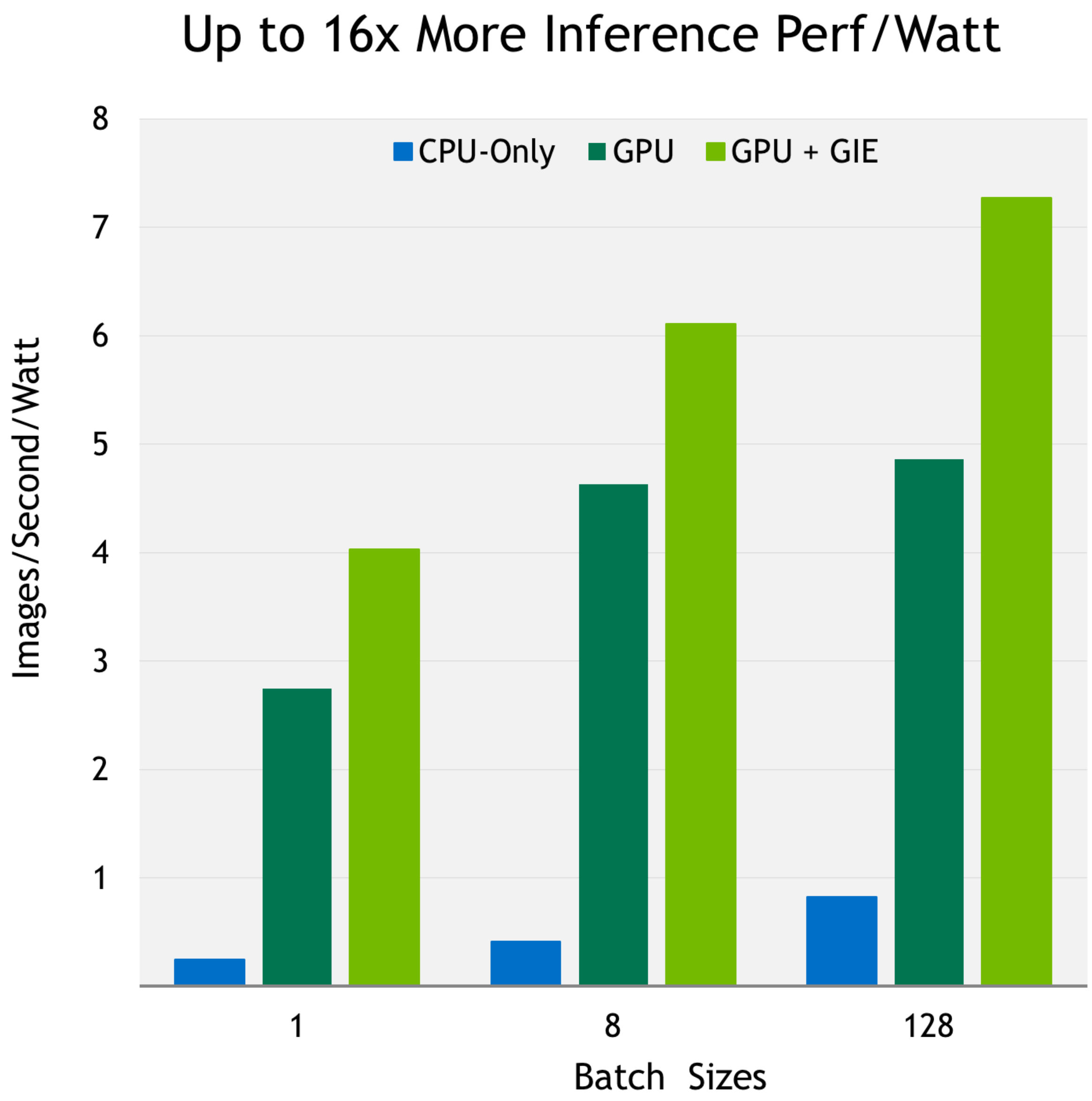


graph source: Nviidia

Why GPUs

Energy saving

Accomplish the same work using less power



graph source: Nviidia

Enable a GPU on Kubernetes

Ncidia

Graphics Process Units, or GPUs, were originally designed as specialize compute devices for computer graphics and image processing. The are designed to rapidly manipulate data in a fast memory buffer. The highly parallel processing capability has proven tomore efficient for algorithms that benefit from processing large amounts of data in parallel.

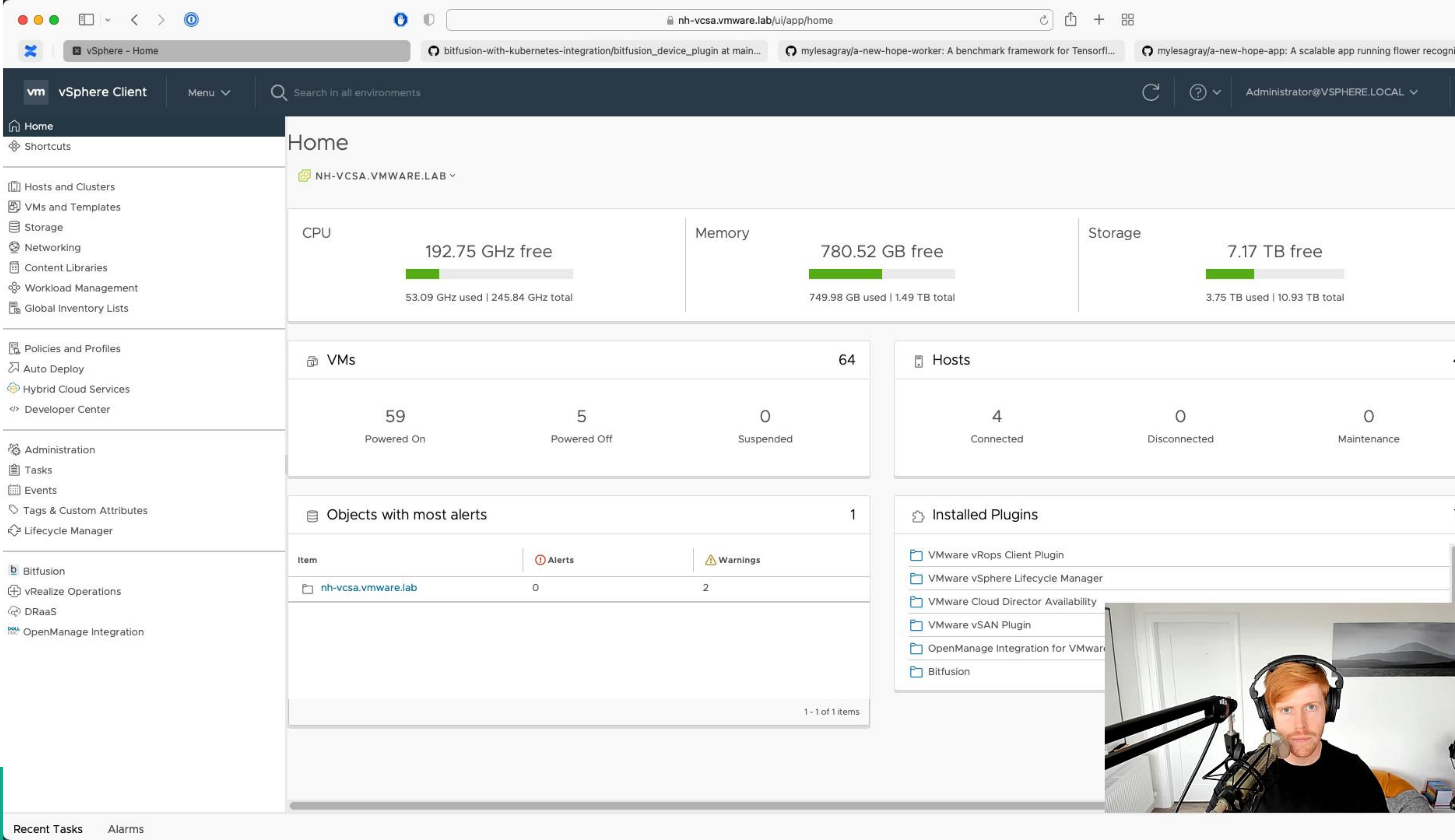
Common applications include machine learning, modeling, image and video processing and recognition, modeling, graphics editing and more.

CUDA is a platform model and API to allow software to use certain types of GPUs. It was deve

NVIDIA publishes CUDA enabled base containers

Demo

TensorFlow workload in Kubernetes
using a GPU resource pool



nh-vcsa.vmware.lab/ui/app/home

vSphere - Home | bitfusion-with-kubernetes-integration/bitfusion_device_plugin at main... | mylesagray/a-new-hope-worker: A benchmark framework for Tensorfl... | mylesagray/a-new-hope-app: A scalable app running flower recogni

Home

NH-VCSA.VMWARE.LAB

CPU
192.75 GHz free
53.09 GHz used | 245.84 GHz total

Memory
780.52 GB free
749.98 GB used | 1.49 TB total

Storage
7.17 TB free
3.75 TB used | 10.93 TB total

VMs 64

59	5	0
Powered On	Powered Off	Suspended

Hosts

4	0	0
Connected	Disconnected	Maintenance

Objects with most alerts 1

Item	Alerts	Warnings
nh-vcsa.vmware.lab	0	2

1 - 1 of 1 items

Installed Plugins

- VMware vRops Client Plugin
- VMware vSphere Lifecycle Manager
- VMware Cloud Director Availability
- VMware vSAN Plugin
- OpenManage Integration for VMwar
- Bitfusion

GPUs + Kubernetes + vSphere

resources

NVIDIA Container toolkit: https://docs.nvidia.com/datacenter/cloud-native/container-toolkit/install-guide.html

CUDA enables base containers, Deep learning: https://docs.nvidia.com/deeplearning/frameworks/user-guide/index.html#installdocker

Multi-user Jupyter on Kubernetes using Helm: https://z2jh.jupyter.org/en/latest/

Bitfusion: https://github.com/vmware/bitfusion-with-kubernetes-integration/tree/main/bitfusion_device_plugin

Where to experience more material like this and interact with other users

The Kubernetes VMware User Group

Kubernetes VMware User Group

What is it?

Similar to SIGs and Working Groups - intended to serve the needs of users running Kubernetes on particular platforms.

The VMware User group is the first (and currently only) K8s UG for a platform - covers running K8s on all VMware hypervisors.

Why is this important?

Create community culture among our users

- Users can help each other
- Users can help us make Kubernetes better – and strengthen user experience on our platforms:
  - Feature requests
  - Feedback + issue resolution

Who is involved?

Co-chairs

- Steven Wong, MAPBU CET
- Myles Gray, VMware Storage Tech Marketing, UK

User Co-leads

- Bryson Shepherd, Walmart
- Joe Searcy, T-Mobile

Kubernetes VMware User Group

User Group Meeting:

First Thursday each month 11am PT
calendar link

Link to join the group

- groups.google.com/forum/#!forum/kubernetes-ug-vmware

Link to join Slack channel (280+ Slack channel participants as c 020)

- https://kubernetes.slack.com/messages/ug-vmware

Speaker contact info

Deck link: https://sched.co/lV8U

Myles Gray
VMware

@mylesagray

Steve Wong
VMware
@cantbewong

Q&A:

https://kubernetes.slack.com/messages/ug-vmware

Get deck here:

Thank You

https://via.vmw.com/EZkQ