GPU Cloud Computing 101: Getting Started

Dale Southard
dsouthard@nvidia.com

About the Speaker and You

[Dale] is a senior solution architect with NVIDIA (I fix things). I primarily cover HPC in Gov/Edu/Research and cloud computing. In the past I was a HW architect in the LLNL systems group designing the vis/post-processing solutions and on-call for capability systems.

[You] are here because you are interested in what Amazon's EC2 GPU announcement means for *High Performance Computing*, and how you can get started *today*.
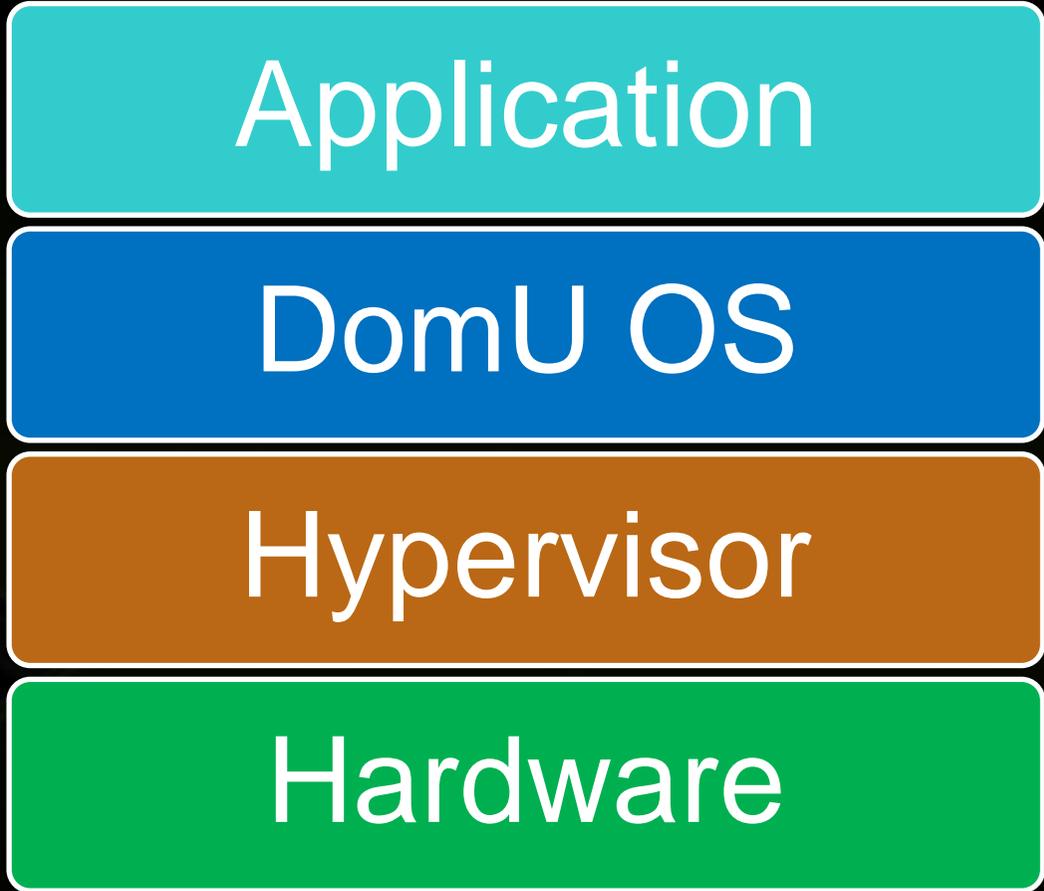
A Brief Introduction to Amazon EC2 with Tesla

Cloud Computing

- **Processing is delivered on-demand via Internet.**

- **Typically hosted on a virtualized substrate.**

- **Originally targeted at web solutions like LAMP.**

| Application |
| DomU OS |
| Hypervisor |
| Hardware |

Amazon EC2 Cluster GPU Instances w/Tesla

- **Utilizes HVM with IOMMU pass-through.**

- **DomU OS and applications do direct IO with the GPU.**

- **100% of Tesla goodness!**

Application

DomU OS

Hypervisor

Hardware
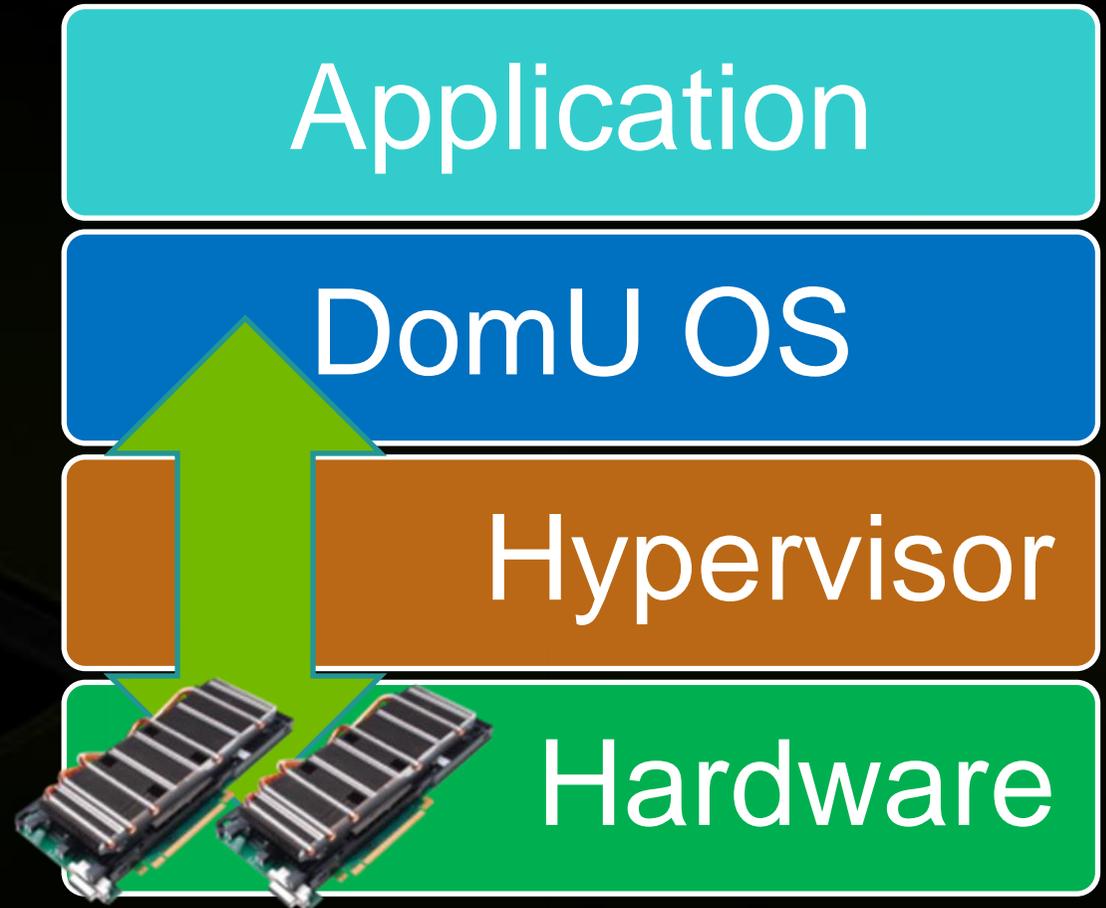
The Cluster GPU Quadruple XL Instance

- **HVM hosted 64-bit platform**
- **Dual quad-core processors**
- **22GB of memory**
- **10GbE networking**
- **Two NVIDIA Tesla m2050 GPUs**

$2.10 per hour

Getting Stared on Amazon EC2 with Tesla

Some Nomenclature

EC2 – Amazon Elastic Compute Cloud

AMI – Amazon Machine Image, your virtual OS disk in the cloud

EBS – Elastic Block Storage, presents to VM as a block device

Instance – a running virtual machine (think of a node)

Key Pair – a public/private key pair used to login to instances

Security Group – manages the firewall settings on instances

Step 1: Create an AWS account

http://aws.amazon.com

Step 2: Sign in to the EC2 Console

http://aws.amazon.com/console

Step 3: Create a Key Pair

Key Pairs

Step 4: Add SSH to a Security Group

Security Groups

Step 5: Launching Your First Instance

AMIs

AMI ID:
ami-aa30c7c3

Step 5 (continued): Setting Instance Type

Instance Type
Cluster GPU
(cg1.4xlarge)

Step 5 (continued): Configure and Launch

- The next two screens can be left default

- For Create Key Pair, you can use the existing Key Pair

- For Configure Firewall, use the Security Group with SSH added

- Launch!

Success!

Instances

Our new Instance

More Success!

Next Steps

I Need a Customized OS

Modify the running AMI, then select **Create Image.**

The customized AMI will be Available for instancing from the AMI's pane.

Amazon also provides tools and documentation for creating AMIs from scratch using standard Linux packaging techniques.
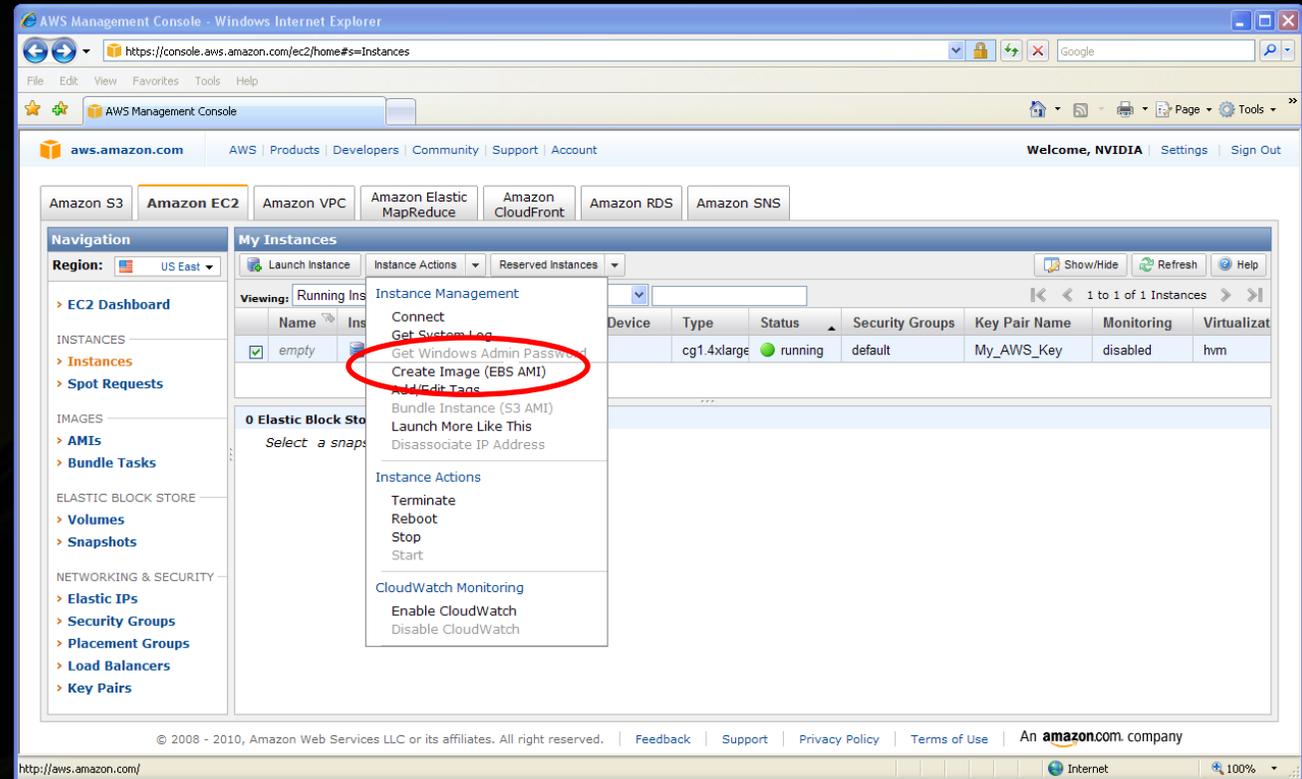
I Need Access to my Data

Amazon EC2 has multiple options for cloud-hosted storage

- **Amazon EBS volumes**

- **Amazon S3 (Simple Storage System)**

- **Amazon AWS Import/Export**

- **Public datasets http://aws.amazon.com/publicdatasets**

I Prefer Command Line Tools

- **Amazon provides EC2 API and AMI tools at:**

  http://aws.amazon.com/developertools

- **API tools are for controlling instances**

- **AMI tools are for bundling machine images**

- **Require a generating a private key and an X.509 certificate**

- **There is also an API available for several languages**

I Need Multiple Nodes

- **EC2 Cluster Instances have 10GbE interconnect**

- **Cluster Placement Groups are used for grouping instances**

- **Placement Groups can be created in the Web or shell interfaces**

- **Instances are launched into existing Placement Groups**

I Can't Setup EC2 Because I'm Already on Bourbon Street and Can't Hear This Talk

Cycle Computing

booth #4638

I Have Other Questions about EC2

http://aws.amazon.com/documentation

Summary and Questions

Go Instance!

1. Create an AWS Account at http://aws.amazon.com/

2. Sign into the EC2 console at http://aws.amazon.com/console

3. Create a Key Pair for ssh.

4. Add SSH to a Security Group to open the firewall.

5. Launch ami-aa30c7c3 using instance cg1.4xlarge.

6. Enjoy 1TF of Tesla GPU goodness at $2.10/hr.