

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*.

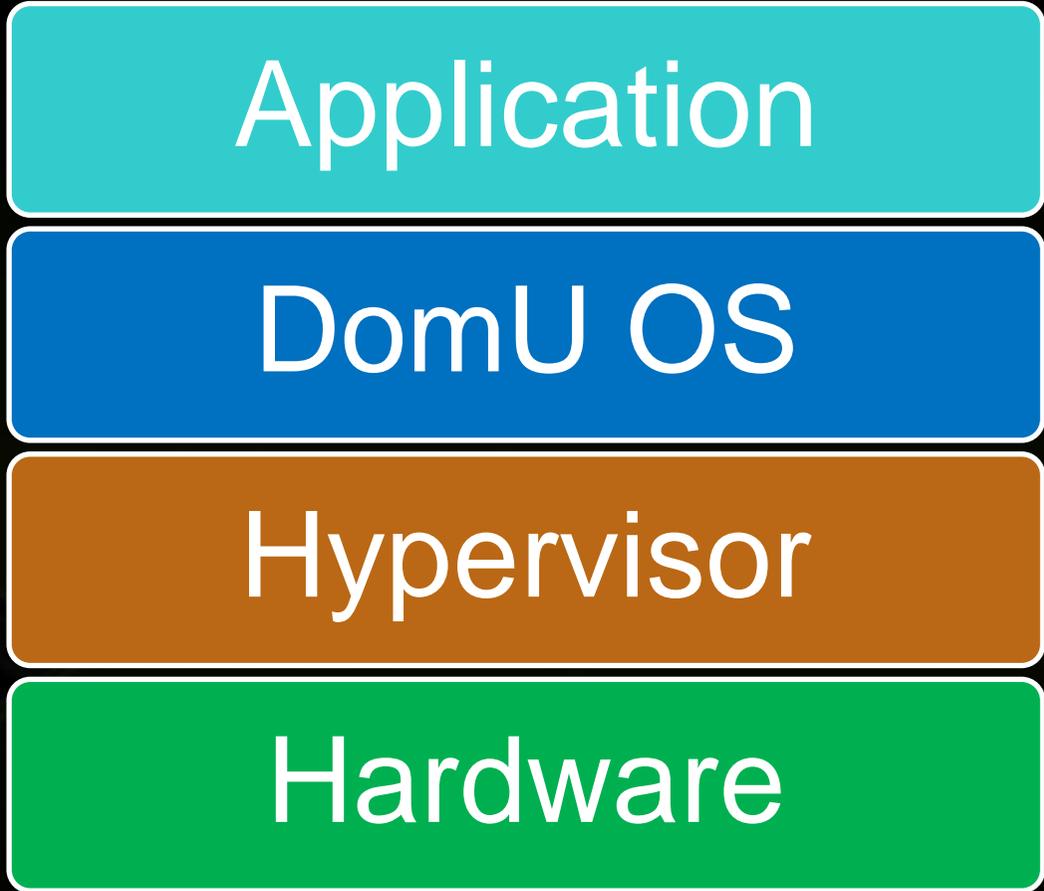


A Brief Introduction to Amazon EC2 with Tesla

Cloud Computing



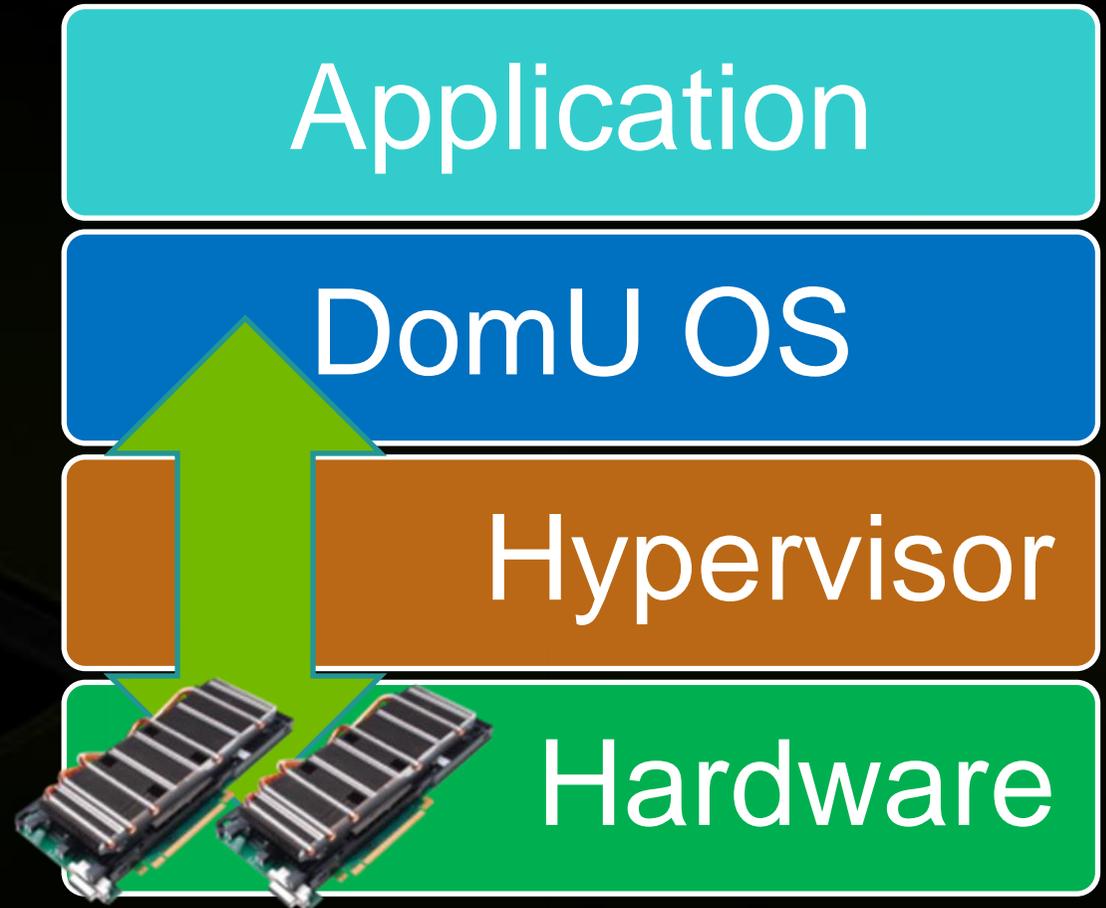
- Processing is delivered on-demand via Internet.
- Typically hosted on a virtualized substrate.
- Originally targeted at web solutions like LAMP.



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!**



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 Started 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

The screenshot shows the AWS Management Console interface for Key Pairs in the US East region. The 'Create Key Pair' button is highlighted with a red circle. The main content area displays a message: "You do not have any key pairs defined. Click the Create Key Pair button to download a new private key." Below this message is a "Create Key Pair" button. The navigation pane on the left includes sections for INSTANCES, IMAGES, ELASTIC BLOCK STORE, and NETWORKING & SECURITY, with "Key Pairs" selected under NETWORKING & SECURITY. The footer contains copyright information and links for Feedback, Support, Privacy Policy, and Terms of Use.

Step 4: Add SSH to a Security Group



Security Groups



The screenshot shows the AWS Management Console interface for Security Groups. The navigation pane on the left includes sections for INSTANCES, IMAGES, ELASTIC BLOCK STORE, and NETWORKING & SECURITY. Under NETWORKING & SECURITY, 'Security Groups' is highlighted. The main content area shows a table of Security Groups with one item, 'default'. Below this, the 'Allowed Connections' table is displayed, with the 'SSH' rule highlighted by a red oval. The 'SSH' rule is configured with Protocol: TCP, From Port: 22, To Port: 22, and Source: 0.0.0.0/0.

Connection Method	Protocol	From Port	To Port	Source (IP or group)	Actions
All	icmp	-1	-1	default group	Remove
All	tcp	0	65535	default group	Remove
All	udp	0	65535	default group	Remove
SSH	TCP	22	22	0.0.0.0/0	Save

Step 5: Launching Your First Instance



AMIs

AMI ID:
ami-aa30c7c3

The screenshot shows the AWS Management Console interface for Amazon Machine Images (AMIs). The 'Launch' button is circled in red. The table below shows the details of the selected AMI.

Name	AMI ID	Source	Owner	Visibility	Status
Cent OS GPU	ami-aa30c7c3	amazon/EC2 CentOS 5.5 GPU HVM AMI	amazon	Public	avail

1 EC2 Amazon Machine Image selected

EC2 Amazon Machine Image: ami-aa30c7c3

Description

AMI ID: ami-aa30c7c3

Name: EC2 CentOS 5.5 GPU HVM AMI

Description: EC2 CentOS 5.5 GPU HVM AMI

Source: amazon/EC2 CentOS 5.5 GPU HVM AMI

Owner: amazon (206029621532)

Visibility: Public

Product Code:

State: available

Kernel ID: -

RAM Disk ID: -

Image Type: machine

Architecture: x86_64

Platform: Cent OS

Step 5 (continued): Setting Instance Type



Instance Type
Cluster GPU
(cg1.4xlarge)

The screenshot shows the AWS Management Console interface in a Windows Internet Explorer browser. The main window is titled "Request Instances Wizard" and is currently on the "INSTANCE DETAILS" step. The wizard is a modal dialog box with a "Cancel" button in the top right corner. The "Instance Type" dropdown menu is open, showing "Cluster GPU (cg1.4xlarge, 22 GB)" selected. A red arrow points from the text on the left to this dropdown. Other visible options in the dropdown include "Request Spot Instances" and "Launch Instances Into Your Virtual Private Cloud". The "Number of Instances" is set to 1, and the "Availability Zone" is set to "No Preference". The "Continue" button is visible at the bottom right of the wizard. The background shows the AWS console navigation pane with "Instances" selected.

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

The screenshot shows the AWS Management Console interface. The 'My Instances' section is active, displaying a table with one instance. The 'Instance Actions' dropdown menu is highlighted with a red circle. Below the table, the details for the selected EC2 instance are shown.

Name	Instance	AMI ID	Root Device	Type	Status	Security Groups	Key Pair Name	Monitoring	Virtualizat
empty	i-cac7a5a7	ami-aa30c7c3	ebs	cg1.4xlarge	running	default	My_AWS_Key	disabled	hvm

1 EC2 Instance selected

EC2 Instance: i-cac7a5a7

Description	Monitoring	Tags
AMI ID: ami-aa30c7c3	Zone: us-east-1d	
Security Groups: default	Type: cg1.4xlarge	
Status: running	Owner: 210258467358	
VPC ID: -	Subnet ID: -	
Virtualization: hvm	Placement Group:	
Reservation: r-533ab939	RAM Disk ID: -	
Platform: -	Key Pair Name: My_AWS_Key	
Kernel ID: -	Monitoring: disabled	

Our new Instance

More Success!



```
root@ip-10-17-162-196:~  
Main Options  VT Options  VT Fonts  
bash-3.2$  
bash-3.2$ ssh -i My_AWS_Key.pem root@ec2-184-72-177-152.compute-1.amazonaws.com  
  
  __|  __|_  ) CentOS  
  _|  (  _| /   v5.5  
  __| \__|__| HVMx64 GPU  
  
Welcome to an EC2 Public Image  
Please view /root/README  
:-)  
  
[root@ip-10-17-162-196 ~]# grep Model /proc/driver/nvidia/cards/*  
/proc/driver/nvidia/cards/0:Model:      Tesla M2050  
/proc/driver/nvidia/cards/1:Model:      Tesla M2050  
[root@ip-10-17-162-196 ~]# cat /proc/driver/nvidia/version  
NVRM version: NVIDIA UNIX x86_64 Kernel Module  260.19.12  Fri Oct  8 11:17:08 P  
DT 2010  
GCC version:  gcc version 4.1.2 20080704 (Red Hat 4.1.2-48)  
[root@ip-10-17-162-196 ~]#
```

Next Steps

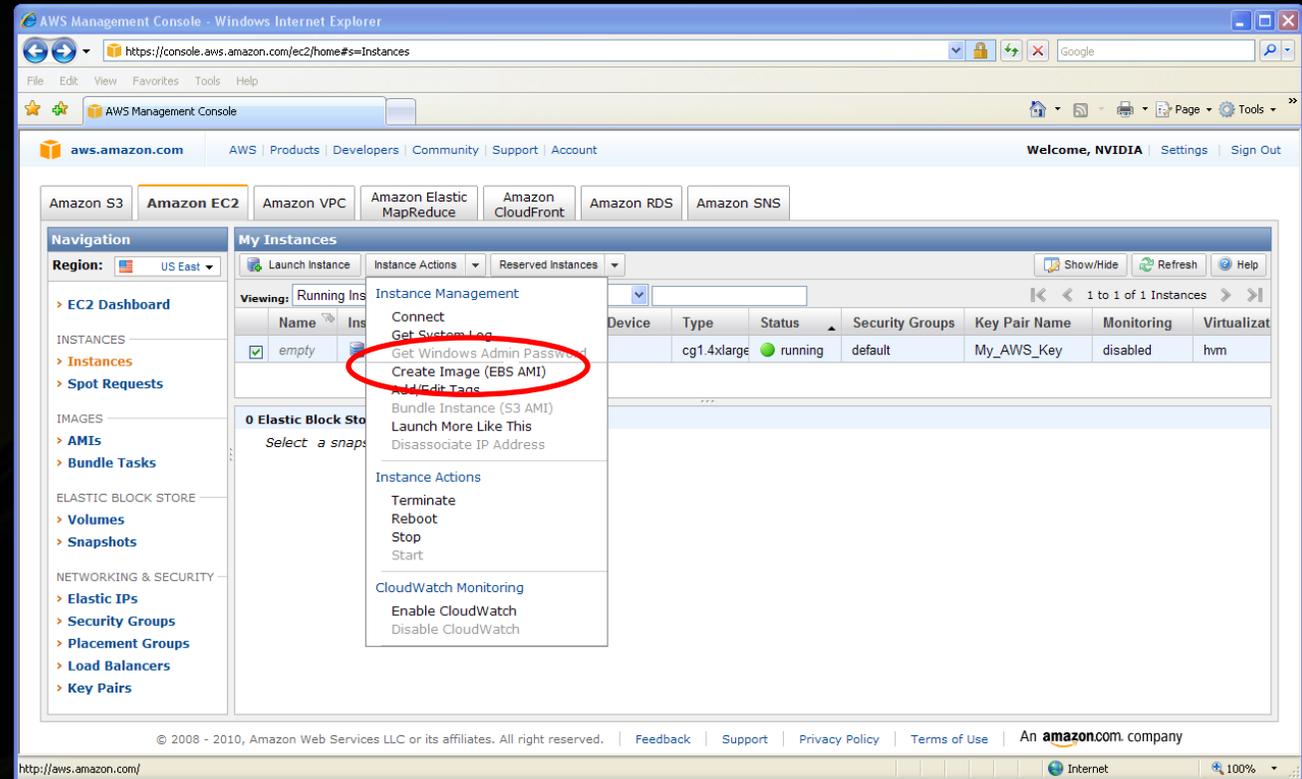
I Need a Customized OS



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.