EE 542 – Laboratory Assignment

Instructor: Young H. Cho

T.A.: Yue Shi

Due date: October 10 at 11:59pm

**GPU cluster lab.**

**Please note that in order to do this lab, in order to do this lab, you need request for a limit increase on the type of g2.2xlarge instance type.**

**In this lab:**

1. **You will learn how to build a GPU cluster with multiple machine instances, with each having multiple GPUs. You will run a simple CUDA-aware MPI program which is provided as a sample program by NVIDIA.**
2. **Read the advanced topic** **about GPU cluster for deep learning.**

**What to turn in:**

**Go through the entire tutorial. Show you have successfully run the sample program on the GPU cluster. Submit a slide and video demo.**

**Background:**

**AWS Multi-GPU instance: AWS GPU instance**s: <https://docs.aws.amazon.com/dlami/latest/devguide/gpu.html>

An AWS GPU instance having multiple GPUs is a Multi-GPU instance. Programming could be done with CUDA itself.

**GPU cluster:** Create a cluster of instances with multiple GPU instances. Need to use CUDA-aware MPI(Message Passing Interface).

AWS instance Images with CUDA installed.

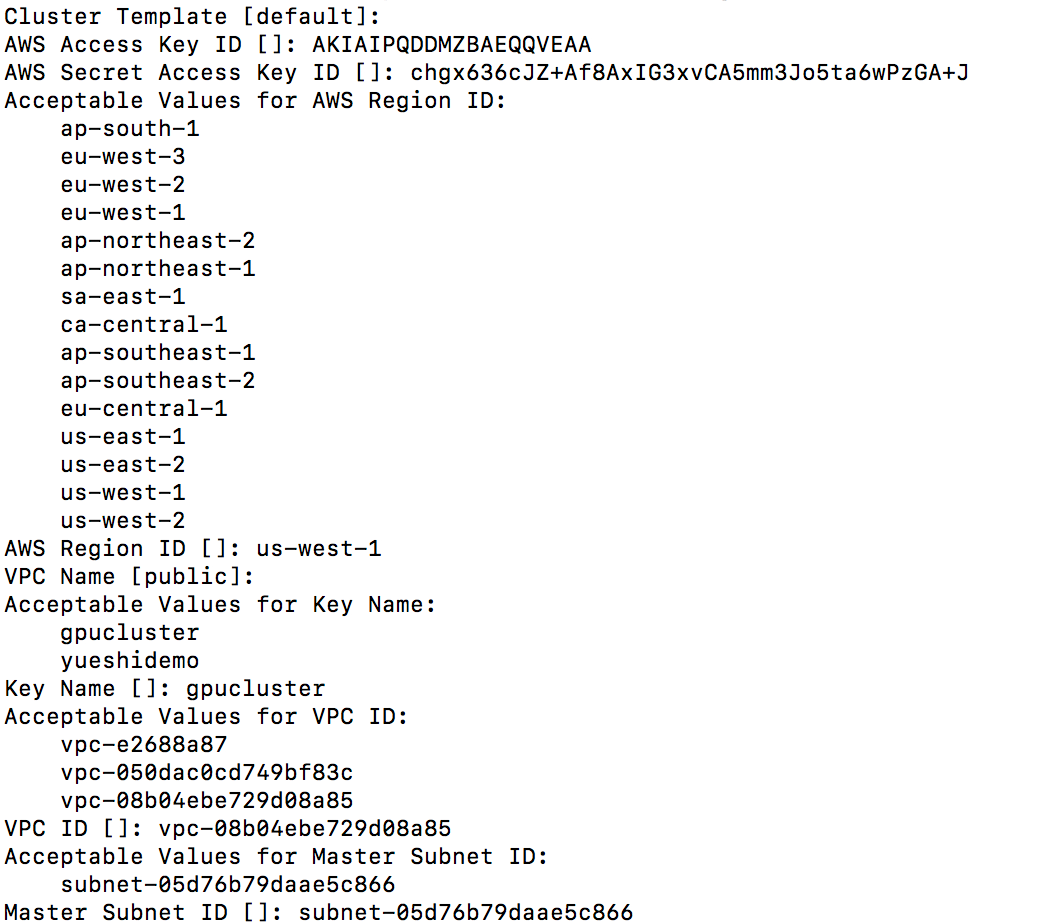
|  |
| --- |
| Deep Learning AMI (Ubuntu) Version 14.0 - ami-0d840505a2c54bf18 |
| Deep Learning AMI (Amazon Linux) Version 14.0 - ami-01a483b8eb6c73261 |
| Deep Learning Base AMI (Ubuntu) Version 10.0 - ami-06d4e2be03258ae38 |
| Deep Learning Base AMI (Amazon Linux) Version 11.0 - ami-067f9f9caa65e31f6 |

**Steps:**

1. Build a HPC with two GPU instances as the compute nodes.

Follow the guide here: <https://d1.awsstatic.com/Projects/P4114756/deploy-elastic-hpc-cluster_project.pdf>. However, you should customize your configure file to have the GPU instances.

1. Below is an example of running the command: ***cfncluster configure***



1. Modify the configuration file. The highlighted part should be added to your configuration file:

|  |
| --- |
| **[aws]**  **aws\_region\_name = us-west-1**  **aws\_access\_key\_id = AKIAIPQDDMZBAEQQVEAA**  **aws\_secret\_access\_key = chgx636cJZ+Af8AxIG3xvCA5mm3Jo5ta6wPzGA+J**  **[cluster default]**  **vpc\_settings = public**  **key\_name = gpucluster**  **custom\_ami = ami-0d840505a2c54bf18**  **initial\_queue\_size = 2**  **max\_queue\_size = 2**  **compute\_instance\_type = g2.2xlarge**  **master\_instance\_type = g2.2xlarge**  **[vpc public]**  **master\_subnet\_id = subnet-05d76b79daae5c866**  **vpc\_id = vpc-08b04ebe729d08a85**  **[global]**  **update\_check = true**  **sanity\_check = true**  **cluster\_template = default**  **[aliases]**  **ssh = ssh {CFN\_USER}@{MASTER\_IP} {ARGS}** |

1. Connect to the cluster master.
2. Run the sample CUDA-aware MPI with the GPU cluster created.

Reference:

[1] Cuda-aware MPI: <https://devblogs.nvidia.com/introduction-cuda-aware-mpi/>

[2]. GPU cluster for deep learning. Read the blog below if you want to apply deep learning on gpu clusters for your final project.

<https://aws.amazon.com/blogs/machine-learning/scalable-multi-node-deep-learning-training-using-gpus-in-the-aws-cloud/>