Chapter 1. Kubernetes on NVIDIA GPUs: CSP Installation Guide

1.1. Using Kubernetes with Amazon Web Services

1.2. Using Kubernetes with Google Cloud Platform
Chapter 1.
KUBERNETES ON NVIDIA GPUS: CSP INSTALLATION GUIDE

This document serves as a step-by-step guide to getting started with using Kubernetes on NVIDIA GPUs available in cloud infrastructure provided by Amazon Web Services (AWS) and Google Cloud Platform (GCP).

Kubernetes is an open-source platform for automating deployment, scaling and managing containerized applications. Kubernetes on NVIDIA GPUs includes support for GPUs and enhancements to Kubernetes so users can easily configure and use GPU resources for accelerating workloads such as deep learning. This document serves as a step-by-step guide to getting started with using Kubernetes on NVIDIA GPUs available in cloud infrastructure provided by Amazon Web Services (AWS) and Google Cloud Platform (GCP).

1.1. Using Kubernetes with Amazon Web Services

On the Amazon Web Services (AWS) platform, NVIDIA makes available a customized Amazon Machine Image (AMI) with the software needed to use NVIDIA GPUs. For more information on the software included in the AMI, see the NVIDIA Volta Deep Learning AMI Release Notes.

1. Use the procedures in Using NGC with AWS Setup Guide to launch virtual machine (VM) instances on AWS using the NVIDIA AMI.
2. After connecting to the instance, follow the instructions provided in the Kubernetes on NVIDIA GPUs Installation Guide to get started with setting up a Kubernetes cluster on AWS.

1.2. Using Kubernetes with Google Cloud Platform

On Google Cloud Platform (GCP), NVIDIA provides a cloud image with the software needed to use NVIDIA GPUs. For more information on the software included in the AMI, see the NVIDIA GPU Cloud Image for Google Cloud Platform Release Notes.
1. Use the procedures in Using NGC with Google Cloud Platform Setup Guide to launch GPU instances on GCP.

2. After connecting to the instance, follow the instructions provided in the Kubernetes on NVIDIA GPUs Installation Guide to get started with setting up a Kubernetes cluster on GCP.
Notice

THE INFORMATION IN THIS GUIDE AND ALL OTHER INFORMATION CONTAINED IN NVIDIA DOCUMENTATION REFERENCED IN THIS GUIDE IS PROVIDED “AS IS.” NVIDIA MAKES NO WARRANTIES, EXPRESSED, IMPLIED, STATUTORY, OR OTHERWISE WITH RESPECT TO THE INFORMATION FOR THE PRODUCT, AND EXPRESSLY DISCLAIMS ALL IMPLIED WARRANTIES OF NONINFRINGEMENT, MERCHANTABILITY, AND FITNESS FOR A PARTICULAR PURPOSE. Notwithstanding any damages that customer might incur for any reason whatsoever, NVIDIA’s aggregate and cumulative liability towards customer for the product described in this guide shall be limited in accordance with the NVIDIA terms and conditions of sale for the product.

THE NVIDIA PRODUCT DESCRIBED IN THIS GUIDE IS NOT FAULT TOLERANT AND IS NOT DESIGNED, MANUFACTURED OR INTENDED FOR USE IN CONNECTION WITH THE DESIGN, CONSTRUCTION, MAINTENANCE, AND/OR OPERATION OF ANY SYSTEM WHERE THE USE OR A FAILURE OF SUCH SYSTEM COULD RESULT IN A SITUATION THAT THREATENS THE SAFETY OF HUMAN LIFE OR SEVERE PHYSICAL HARM OR PROPERTY DAMAGE (INCLUDING, FOR EXAMPLE, USE IN CONNECTION WITH ANY NUCLEAR, AVIONICS, LIFE SUPPORT OR OTHER LIFE CRITICAL APPLICATION). NVIDIA EXPRESSLY DISCLAIMS ANY EXPRESS OR IMPLIED WARRANTY OF FITNESS FOR SUCH HIGH RISK USES. NVIDIA SHALL NOT BE LIABLE TO CUSTOMER OR ANY THIRD PARTY, IN WHOLE OR IN PART, FOR ANY CLAIMS OR DAMAGES ARISING FROM SUCH HIGH RISK USES.

NVIDIA makes no representation or warranty that the product described in this guide will be suitable for any specified use without further testing or modification. Testing of all parameters of each product is not necessarily performed by NVIDIA. It is customer’s sole responsibility to ensure the product is suitable and fit for the application planned by customer and to do the necessary testing for the application in order to avoid a default of the application or the product. Weaknesses in customer’s product designs may affect the quality and reliability of the NVIDIA product and may result in additional or different conditions and/or requirements beyond those contained in this guide. NVIDIA does not accept any liability related to any default, damage, costs or problem which may be based on or attributable to: (i) the use of the NVIDIA product in any manner that is contrary to this guide, or (ii) customer product designs.

Other than the right for customer to use the information in this guide with the product, no other license, either expressed or implied, is hereby granted by NVIDIA under this guide. Reproduction of information in this guide is permissible only if reproduction is approved by NVIDIA in writing, is reproduced without alteration, and is accompanied by all associated conditions, limitations, and notices.

Trademarks

NVIDIA and the NVIDIA logo are trademarks and/or registered trademarks of NVIDIA Corporation in the Unites States and other countries. Other company and product names may be trademarks of the respective companies with which they are associated.

Copyright

© 2018 NVIDIA Corporation. All rights reserved.