

Hands-on Tutorial Deploying Kubernetes and JupyterHub on Jetstream

1st Andrea Zonca
San Diego Supercomputer Center
University of California, San Diego
San Diego, USA
zonca@sdsc.edu

2nd Kenneth Yoshimoto
San Diego Supercomputer Center
University of California, San Diego
San Diego, CA, USA

3rd Aaron Culich
Berkeley Institute for Data Science
University of California, Berkeley
Berkeley, CA, USA

4th Richard Signell
Coastal and Marine Geology Program
USGS
Woods Hole, MA, USA

Abstract—Jupyter Notebooks have become a mainstream tool for interactive computing in every field of science. Being a web-based platform, they are naturally suited to serve as companion application to Science Gateways: a scientist can use a Jupyter Notebook in their browser to pre-process inputs, launch a job on the Science Gateway via web API and then access, analyze, plot and postprocess the job outputs, without ever worrying about setting up and keeping updated their software environment. The JupyterHub project provides a multiple-user platform for Jupyter Notebooks and it is very easy to install and configure on a single server. However, when we need to provide computational resources to a large pool of users, we need to distribute the users on a cluster of machines, the best way to achieve scalability is thanks to the container orchestration platform Kubernetes. In this tutorial we will work through the installation of Kubernetes on a set of Jetstream Virtual Machines, setup persistent storage and install a bare-bone JupyterHub deployment using the zero-to-jupyterhub recipe provided by the Jupyter team. Then we will customize the setup configuring authentication (XSEDE, Globus or Github), choosing our preferred software environment for the users via Docker. Finally we will show how to execute computational jobs, either interfacing with the web APIs of a test gateway to submit jobs or launching a pool of workers on Kubernetes and execute a distributed computation (using dask).

has pointers to Github repositories containing all the necessary software and configuration files.

Index Terms—kubernetes, jupyterhub

I. MATERIAL

All the steps of the tutorial are available in 3 web pages:

- Deploy Kubernetes on Jetstream with Kubespray¹
- Explore a Kubernetes deployment on Jetstream with Kubespray²
- Deploy JupyterHub on Kubernetes deployment on Jetstream created with Kubespray³

The tutorial is designed to be executed on other user accounts than the ones available during the conference and

Presented at Gateways 2018, University of Texas, Austin, TX, September 25-27, 2018. <https://gateways2018.figshare.com>

¹<https://zonca.github.io/2018/09/kubernetes-jetstream-kubespray.html>

²<https://zonca.github.io/2018/09/kubernetes-jetstream-kubespray-explore.html>

³<https://zonca.github.io/2018/09/kubernetes-jetstream-kubespray-jupyterhub.html>