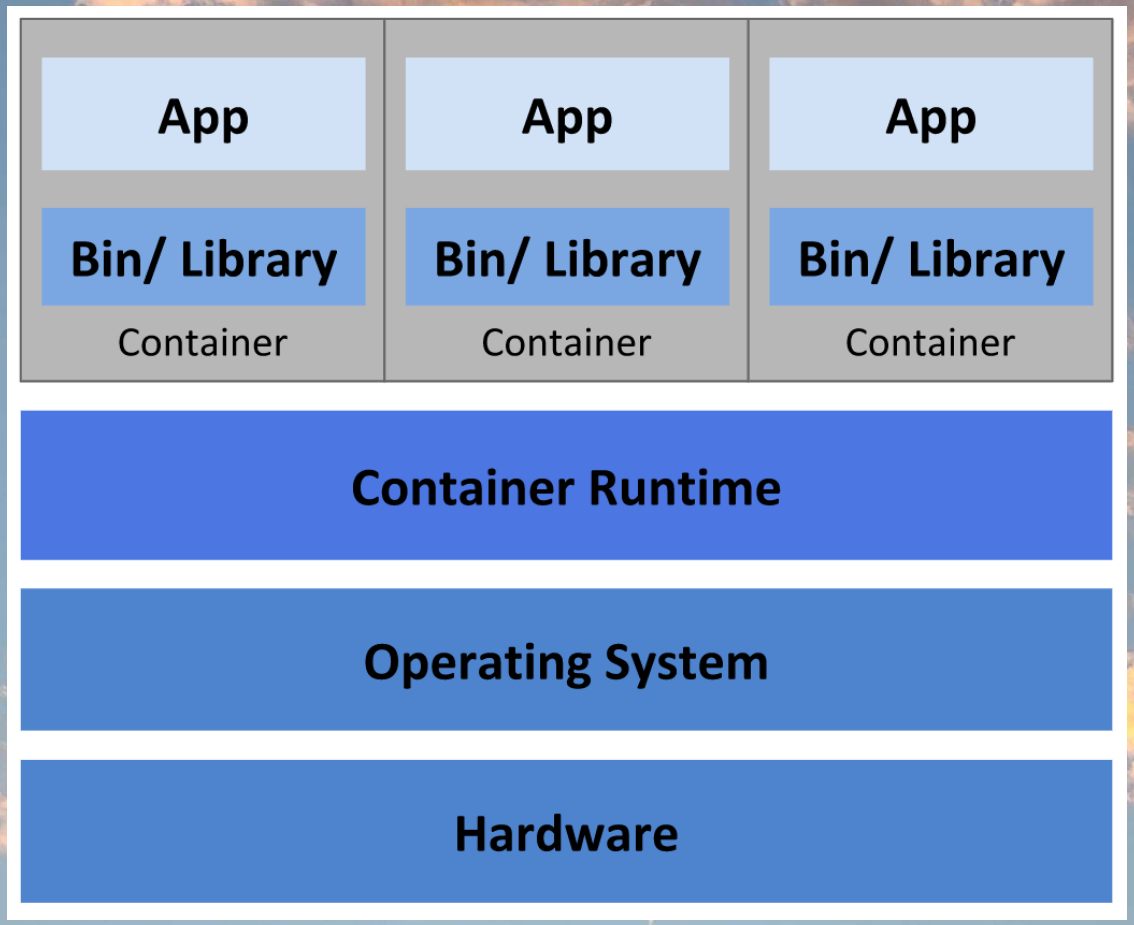
DSA 8430
Parallel Computing for Data Analytics

Kubernetes

Module Topics

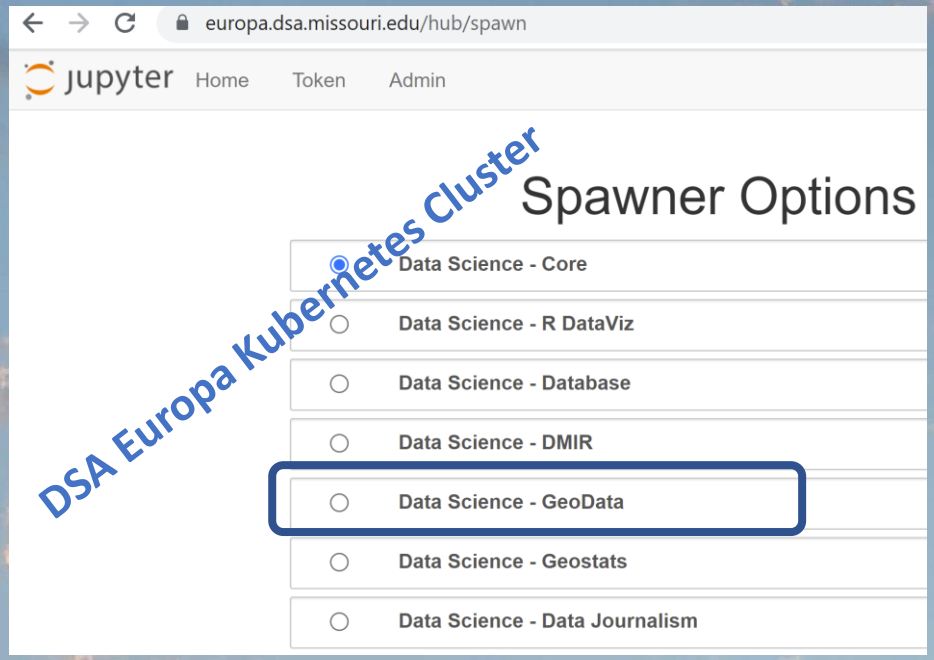
- Kubernetes
- Jobs on the Nautilus system

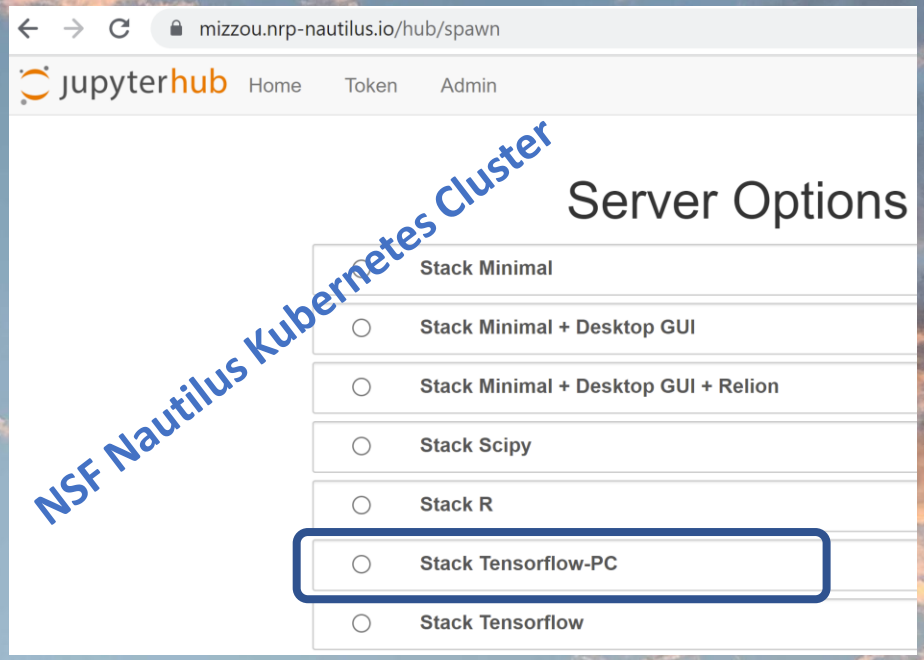
Kubernetes – A foundation of Containers



Kubernetes Cluster - Containers

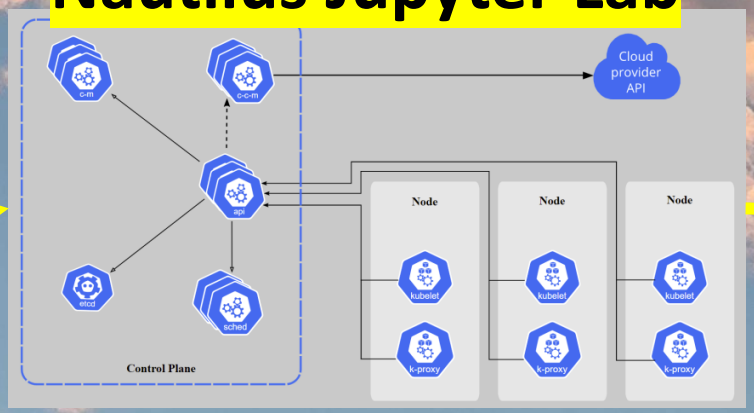
- You have been using two Kubernetes Clusters and the associated containers in this class!





Kubernetes – Inception Style

Nautilus Jupyter Lab



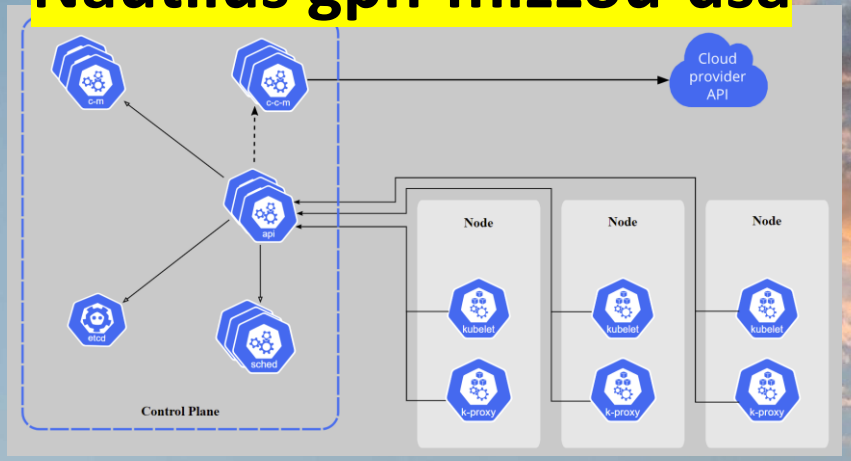
Use Stack Datascience

Browser UI to Jupyter Lab

kubectl CLI

Your local computer

Nautilus gpn-mizzou-dsa



Kubernetes Concepts

- **Pods** - smallest deployable units of computing that you can create and manage in Kubernetes
- **Jobs** - creates one or more **Pods** and will continue to retry execution of the Pods until a specified number of them successfully terminate
- **Persistent Storage** - piece of storage in the cluster that has been provisioned by an administrator or dynamically provisioned using Storage Classes

`kubectl` – CLI for Kubernetes

- Allows you to run commands against Kubernetes clusters. You can use kubectl to deploy applications, inspect and manage cluster resources, and view logs
- Can control remote Kubernetes clusters from any number of environments (Linux, Windows, or Mac)

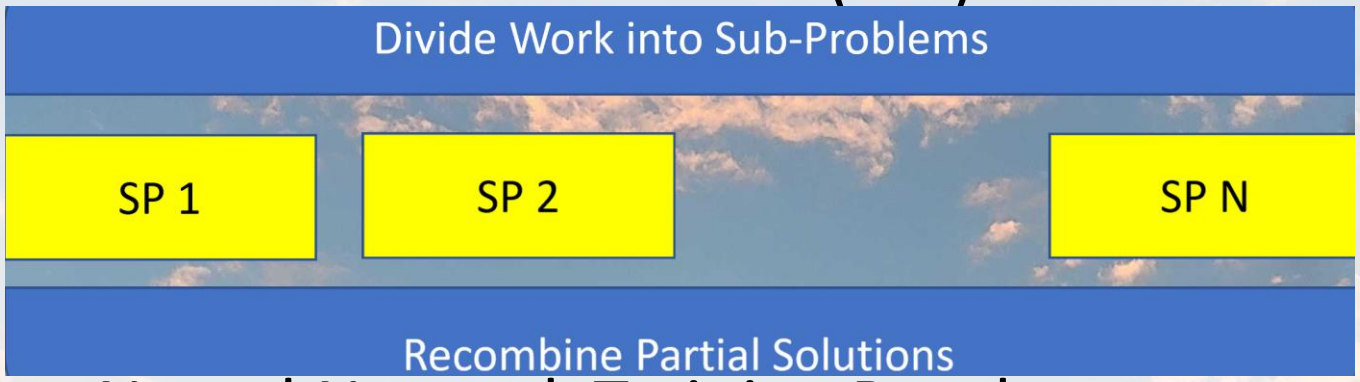
Learning Activities

- Lab – Setup Kubernetes tools on Nautilus Jupyter Lab
- Practice – Attaching persistent storage to Pods
- Exercise – Parallelizing Deep Learning Training and Validation on Kubernetes

M5 Exercise – Divide & Conquer

- Using Template to mass produce experiment YAML files.
- Multi-job Submission

Remember me (M2)!



- Deep Neural Network Training Results – Aggregation

Pay Attention to Detail!

Start early, work methodically!

Take notes as you work through things!