Kubernetes Implementation into HPC

Livermore Computing

Jessica Gomes, Jonathan Victorino, Harry Yee
HPC CEA Interns

August 10, 2017

LLNL-PRES-XXXXXX

Lawrence Livermore National Laboratory

Meet Team Kubes

Harry Yee
UCLA

Jessica Gomes
University of Illinois at Urbana-Champaign

Jonathan Victorino
Stanford University

Objectives

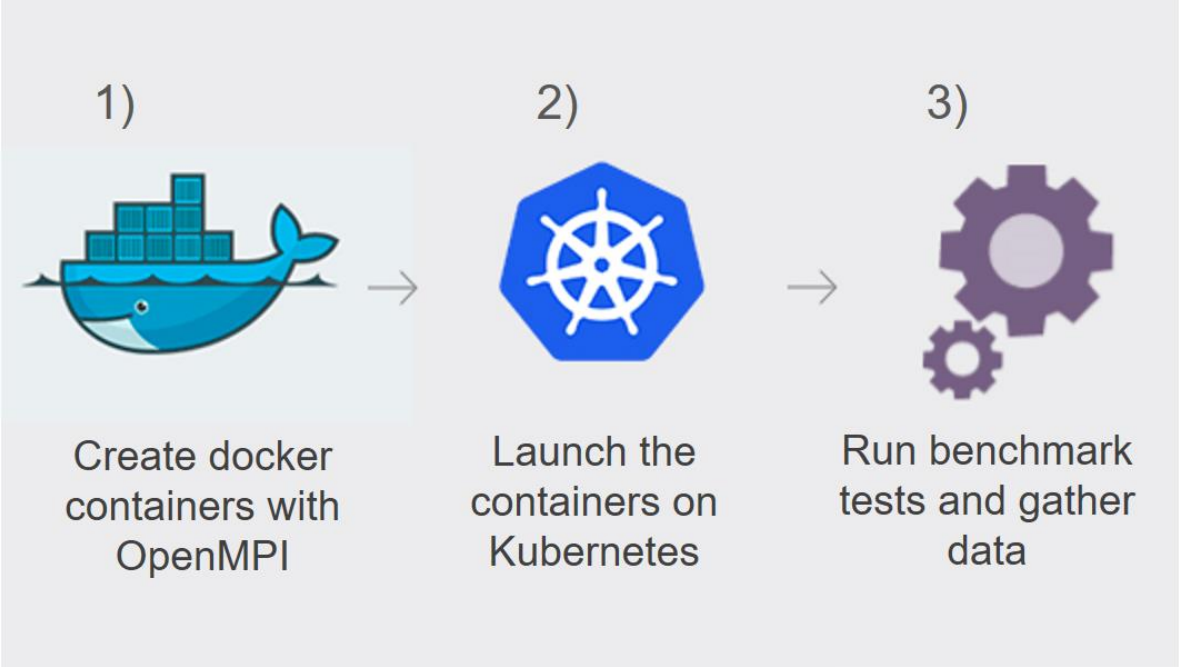
I. Setup a Kubernetes cluster using Docker containers

II. Run LINPACK benchmark tests comparing Kubernetes and bare metal cluster

III. Determine performance overhead of running the cluster in Kubernetes

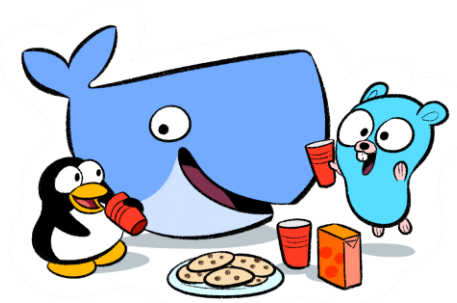
IV. Automate Kubernetes builds and deployments using Puppet

Our Project


1)
2)
3)
Create docker containers with OpenMPI
Launch the containers on Kubernetes
Run benchmark tests and gather data

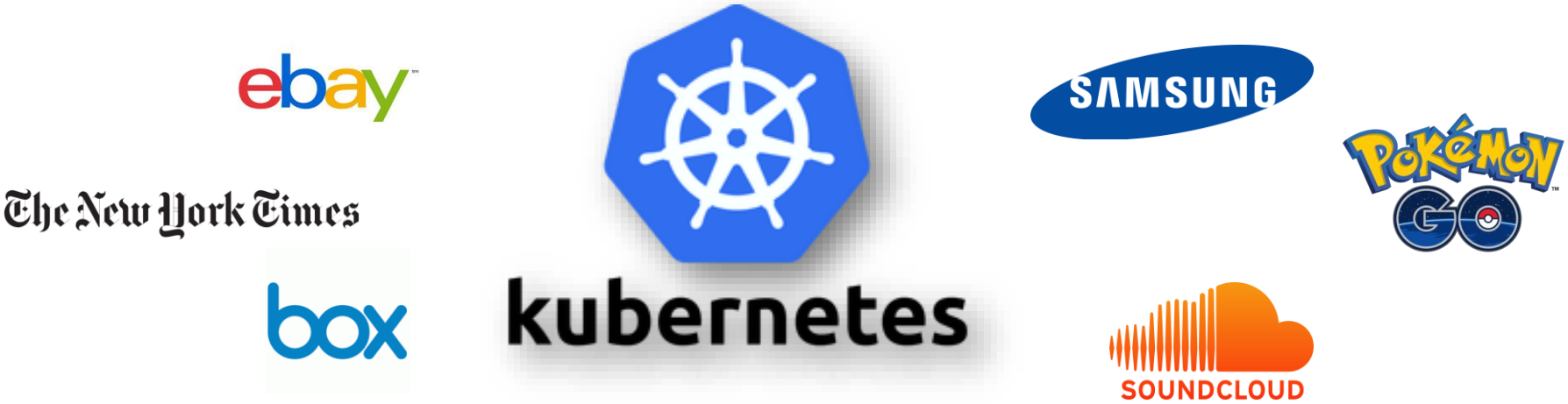
Docker

- World's leading software container platform
- **Container**: stand-alone package that includes everything needed to run it
- Easily create images for containers using Dockerfiles
- Use Docker Hub to automatically update the containers



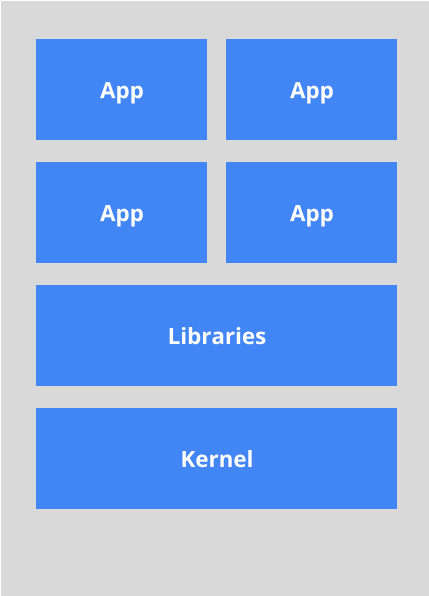
What is Kubernetes?

- Open-source system for automating deployment, scaling, and management of containerized applications
- Schedule and deploy any number of Docker container replicas onto a node cluster



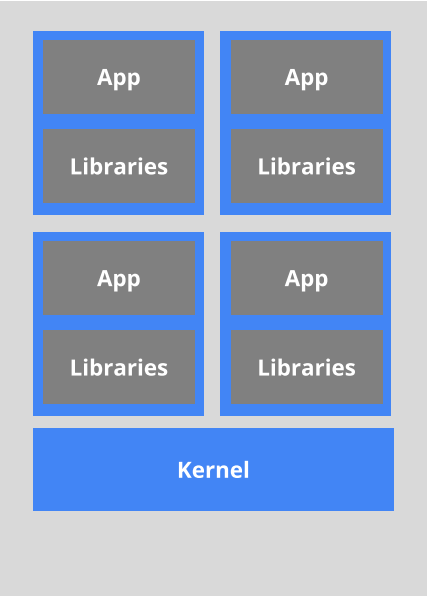
Kubernetes Container Visualization

The old way: Applications on host



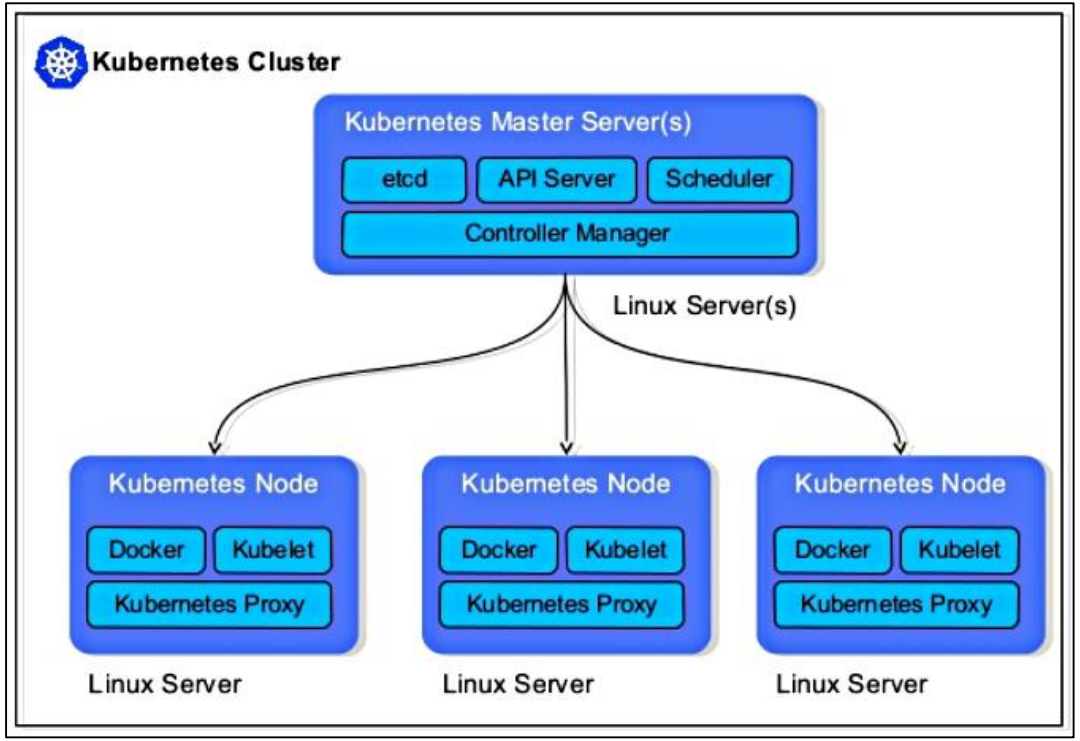
Heavyweight, non-portable
Relies on OS package manager

The new way: Deploy containers



Small and fast, portable
Uses OS-level virtualization

Kubernetes Architectural Overview



Installation Process

puppet

root@boroni:~/boron_puppet/manifests

```
    ensure => "file",
    owner => "root",
    group => "root",
    mode => "644",
    source => "/root/boron_puppet/tftpboot/pxelinux.cfg/default",
}


file { '/etc/dhcp/dhcpd.conf':
    ensure => "file",
    owner => "root",
    group => "root",
    mode => "644",
    source => "/root/boron_puppet/etc/dhcp/dhcpd.conf",
}

$packages = ['ntp', 'openmpi', 'openmpi-devel', 'slurm', 'munge',
 'nfs-utils', 'sudo', 'yum-utils', 'device-mapper-persistent-data', 'lvm2']
package { $packages:
    ensure => installed,
}

exec {'dockerInstall':
    command => "/root/boron_puppet/scripts/dockerInstallScript.sh",
    #unless => 'test -e /usr/bin/docker',
}

package {'docker-ce':
    ensure => installed,
}

exec {'kubeInstall':
    command => "/root/boron_puppet/scripts/kubeInstallScript.sh",
    #unless => 'test -e /usr/bin/kubelet && test -e /usr/bin/kubeadm',
}

$kubepackages = ['kubelet', 'kubeadm']
package { $kubepackages:
    ensure => installed,
}
```

Puppet Manifest File

root@boroni:~/boron_puppet/scripts

```
#!/bin/bash

# Docker installation
sudo yum-config-manager --add-repo
https://download.docker.com/linux/
centos/docker-ce.repo

sudo yum makecache fast
sudo yum -y install docker-ce
sudo systemctl start docker
sudo systemctl enable docker
```

Docker Installation Script

root@boroni:~/boron_puppet/scripts

```
#!/bin/bash

# Kubectl Installation
curl -LO https://storage.googleapis.com/kubernetes-release/release/$(curl -s
https://storage.googleapis.com/kubernetes-release/release/stable.txt)/bin/linux/amd64/kubectl

# Change Permissions
chmod +x ./kubectl
sudo mv ./kubectl /usr/local/bin/kubectl

# Install Packages Kubelet and Kudeadm
touch /etc/yum.repos.d/kubernetes.repo
sudo echo "[kubernetes]
name=Kubernetes
baseurl=https://packages.cloud.google.com/yum/repos/kubernetes-el7-x86_64
enabled=1
gpgcheck=1
repo_gpgcheck=1
gpgkey=https://packages.cloud.google.com/yum/doc/yum-key.gpg
       https://packages.cloud.google.com/yum/doc/rpm-package-key.gpg" >
       /etc/yum.repos.d/kubernetes.repo

# Disable SELinux Enforcement
setenforce 0
~
```

Kubernetes Packages Installation Script

.yaml Files

Branch: master ▾ MPI-Kubernetes-Cluster / master.yaml — Find file | Copy path

U-THE-LAB\dixon30 Reduced CPU amount to 15 vs 16 — 9603fcf 23 hours ago

2 contributors @h @j

Executable File | 65 lines (64 sloc) | 1.59 KB — Raw | Blame | History

```yaml
apiVersion: v1
kind: Service
metadata:
  name: mpi-master
  labels:
    service: mpi-master
spec:
  type: NodePort
  ports:
  - port: 22
    nodePort: 30001
    protocol: TCP
  selector:
    app: mpi-master-deployment
---
apiVersion: apps/v1beta1
kind: Deployment
metadata:
  # Unique key of the Deployment instance
  name: mpi-master-deployment
spec:
  # 3 Pods should exist at all times.
  replicas: 1
  template:
    metadata:
      labels:
        # Apply this label to pods and default
        # the Deployment label selector to this value
        app: mpi-master-deployment
```

Branch: master ▾ MPI-Kubernetes-Cluster / worker.yaml — Find file | Copy path

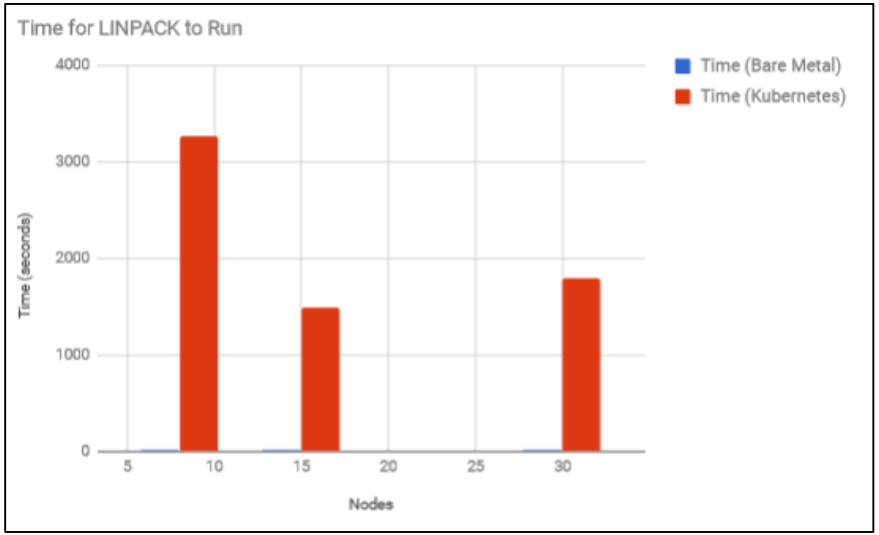
U-THE-LAB\dixon30 Reduced CPU amount to 15 vs 16 — 9603fcf 23 hours ago

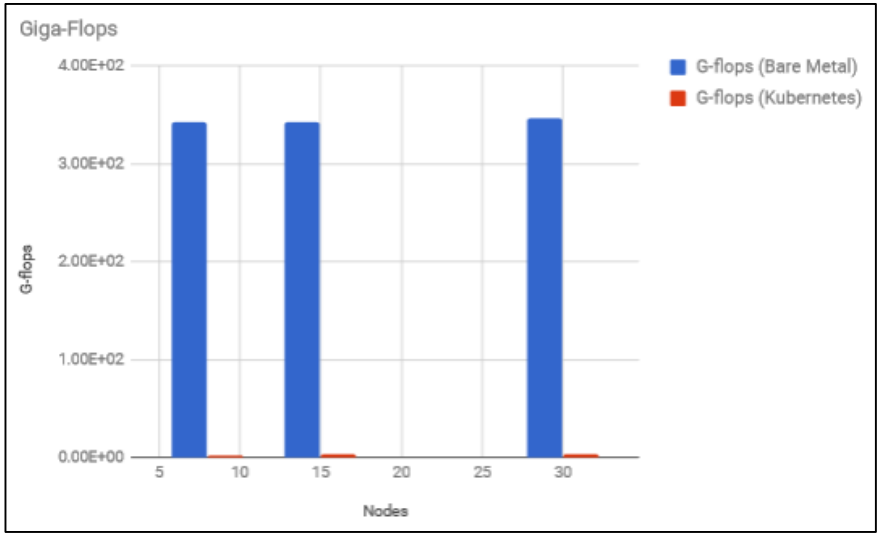
2 contributors @h @j

Executable File | 39 lines (38 sloc) | 988 Bytes — Raw | Blame | History

```yaml
apiVersion: apps/v1beta1
kind: Deployment
metadata:
  # Unique key of the Deployment instance
  name: mpi-worker
spec:
  # 3 Pods should exist at all times.
  replicas: 5
  template:
    metadata:
      labels:
        # Apply this label to pods and default
        # the Deployment label selector to this value
        app: mpi-worker
    spec:
      volumes:
        - name: ssh-keys
          secret:
            secretName: ssh-key-secret
        # - name: test-volume
        #   persistentVolumeClaim:
        #     claimName: test-pvc
      containers:
      - name: kube-mpi
        # Run this image
        image: javawolfpack/base-centos-mpi
        imagePullPolicy: Always
        ports:
          - containerPort: 22
```

LINPACK Benchmarks





	Time (seconds)	G-flops
Bare Metal	15.39	3.47E+02
Kubernetes	1796.654	2.99E+00

Conclusion

- Kubernetes
  - Large performance overhead on clusters
  - Extra setup steps
  - Not ideal for HPC
- Use bare metal cluster
- After further testing, the bottleneck was found to be due to a networking issue
- Time and G-flops were comparable for containers running on the same node to the bare metal configuration

Future Plans

- Use Puppet to automate builds and deployments
- Automatically ssh from worker to master
- Run more Benchmark tests on Kubernetes cluster
- Find out how to fix the bottleneck and improve communication between containers


Lawrence Livermore
National Laboratory