




# Kubernetes and HPC


A horizontal row of five small, solid-colored dots in the order: blue, green, yellow, orange, and red.

Daniel Gruber  
The UberCloud

# Kubernetes Intro: Native Batch Job Submission

```
kubectl apply -f job.yaml
```

```
1  apiVersion: batch/v1
2  kind: Job
3  metadata:
4    name: sleep
5  spec:
6    template:
7      spec:
8        containers:
9          - name: sleep
10           image: busybox
11           command: ["sleep", "100"]
12          restartPolicy: Never
```

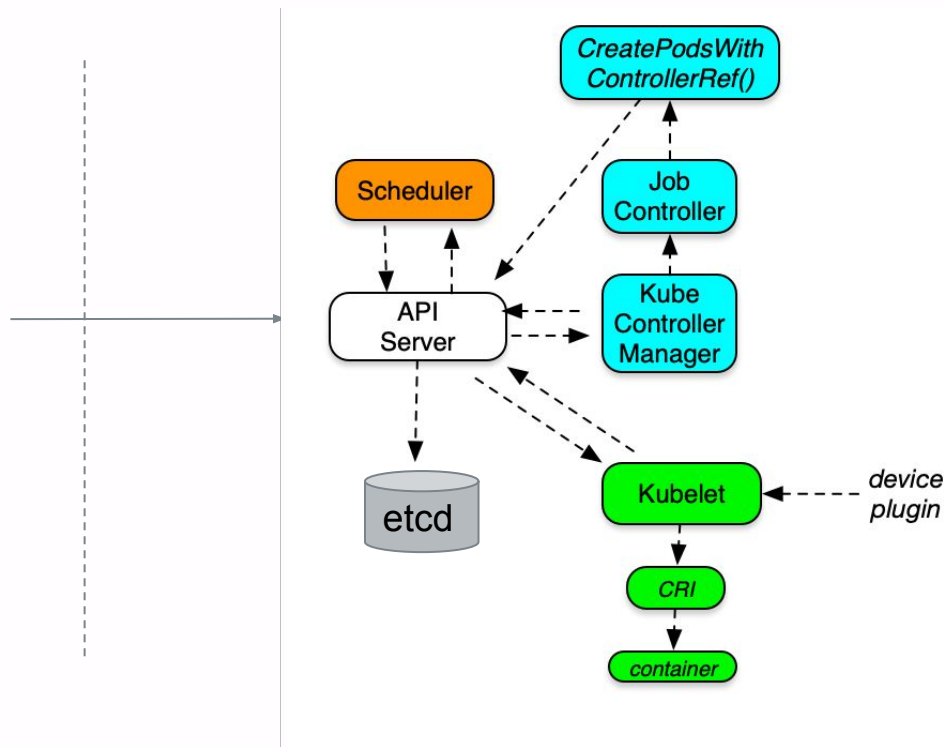


# Kubernetes Intro: Native Batch Job Submission

```
kubectl apply -f job.yaml
```

```
1  apiVersion: batch/v1
2  kind: Job
3  metadata:
4    name: sleep
5  spec:
6    template:
7      spec:
8        containers:
9          - name: sleep
10           image: busybox
11           command: ["sleep", "100"]
12           restartPolicy: Never
```

Deployment of more complex applications: *helm, Operators*



# Kubernetes and HPC

## HPC Requirements

- Fast job submission times
- Multi-node jobs (distributed memory)
- Support for GPUs, FPGAs
- Easily accessible by HPC applications
- Job queueing capabilities
- Job prioritization
- Job pre-emption

## Kubernetes features

- ❖ *kubectl*, API
- ❖ No direct MPI Support
- ❖ GPUs supported via *device plugin*
- ❖ AI
- ❖ Pods, since 1.9
- ❖ 1.14
- ❖ 1.14

# Kubernetes and HPC (continued)

## HPC Requirements

- Advance reservation
- License management capabilities
- Support for Infiniband, Omnipath
- Full hardware utilization
- Resource control
- Accounting and reporting
- Storage

## Kubernetes features

- ❖ No
- ❖ No, Scheduler Extenders
- ❖ device plugin
- ❖ More daemons, typically virtualized, CPU manager
- ❖ Memory and CPU resources, Resource quotas
- ❖ Yes
- ❖ PVs and PVCs, storageclass, hostpath

# References

- Kubeflow: <https://github.com/kubeflow/kubeflow>
- MPI Operator: <https://github.com/kubeflow/mpi-operator>
- Poseidon: <https://github.com/kubernetes-sigs/poseidon>
- CPU Manager: <https://kubernetes.io/docs/tasks/administer-cluster/cpu-management-policies/>
- Kube batch: <https://github.com/kubernetes-sigs/kube-batch>
- Volcano: <https://github.com/volcano-sh/volcano>
- Slurm Operator: <https://github.com/sylabs/slurm-operator>
- Lustre: <https://github.com/kvaps/kube-lustre>
- DRMAA2: <https://github.com/dgruber/drmaa2os>
- qsub for k8s: <https://github.com/dgruber/qsub>
- RDMA: <https://github.com/Mellanox/k8s-rdma-sriov-dev-plugin>
- Univa's Command: <http://www.univa.com/products/navops.php>
- UberCloud: <https://www.theubercloud.com/>



# Thank you

Daniel.Gruber@TheUberCloud.com

Some of my strange thoughts: [www.drmaa2.org](http://www.drmaa2.org) (--> <https://www.gridengine.eu>)



<https://github.com/dgruber>