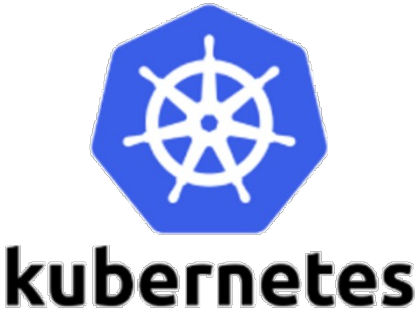


AI and Scientific Research Computing with Kubernetes

Monitoring your Work

A tutorial at PEARC24

July 22, Providence, Rhode Island

Presented by Mahidhar Tatineni and Dmitry Mishin
University of California San Diego – San Diego Supercomputer Center

Ref: Tutorials at PEARC, SC, 5NRP by Igor Sfiligoi, Dmitry Mishin, and Mahidhar Tatineni

Understand what you are requesting

List your requests

- Listing the pods the most useful
  `kubectl get pods`
  - Using the `-o wide`
    option provides good balance between detail and readability
- But remember that jobs are separate objects
  `kubectl get jobs`
- And let's not forget about storage
  `kubectl get pvc`

When pods don’t start

Monitor K8S internals

- Events provide good overview of what is happening
  `kubectl get events`
  - Usually better ordered `--sort-by=.metadata.creationTimestamp`

Make sure your requests are reasonable

- Check what’s available
  `kubectl get nodes`
- Use `-o wide` or `-o yaml` for detailed view

Description of pending/running pods

Get full information on pod

- `kubectl describe pod <pod name>`

Describe option can also be used to get node info

- `kubectl describe node <node name>`

Understanding running pods

Stdout and stderr can be accessed at any time

- `kubectl logs <pod name>`

Log into the pod for troubleshooting

- `kubectl exec`
- By default just runs a command, but can be made interactive with `-it -- /bin/bash`

Understanding the performance

Long-running pods report load to central collector

- `kubectl top pods`

You can also check what the node-level load is

- `kubectl top nodes <nodename>`

External tools

Most K8S clusters will have GUI tools

May differ between installations

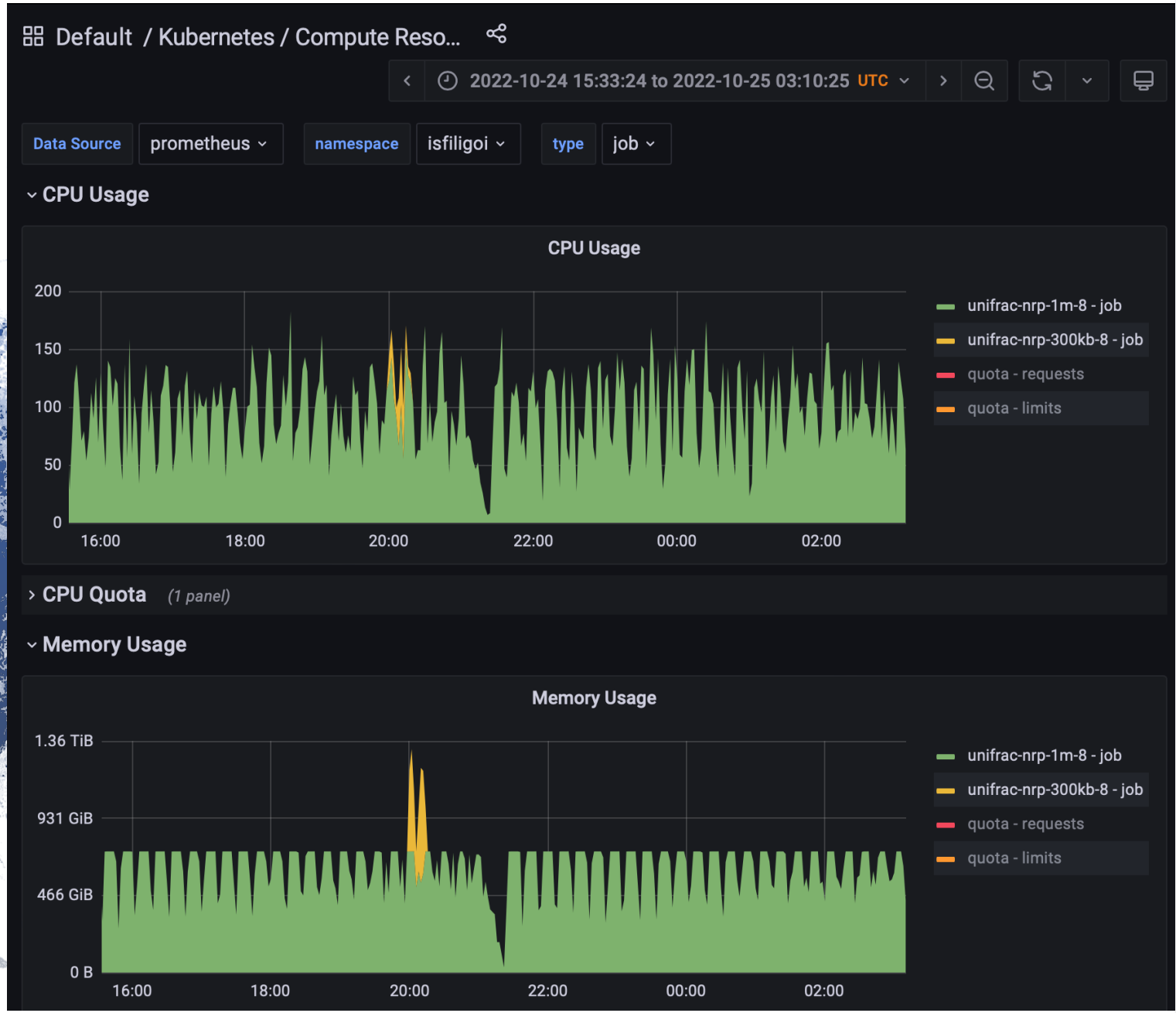
- There are many competing solutions available

Grafana often used for Web-based visualization

- Backend data often managed by Prometheus

https://prometheus.io/docs/introduction/overview/

Screenshot from Nautilus



Acknowledgements

This work was partially funded by
US National Science Foundation (NSF) awards
OAC-2112167, OAC-1826967, OAC-1541349,
OAC-2030508 and CNS-1730158.