Tetris: Predictive Pod Placement Strategy for Kubernetes

CSC 724 Spring 2019

Project Mid-Review

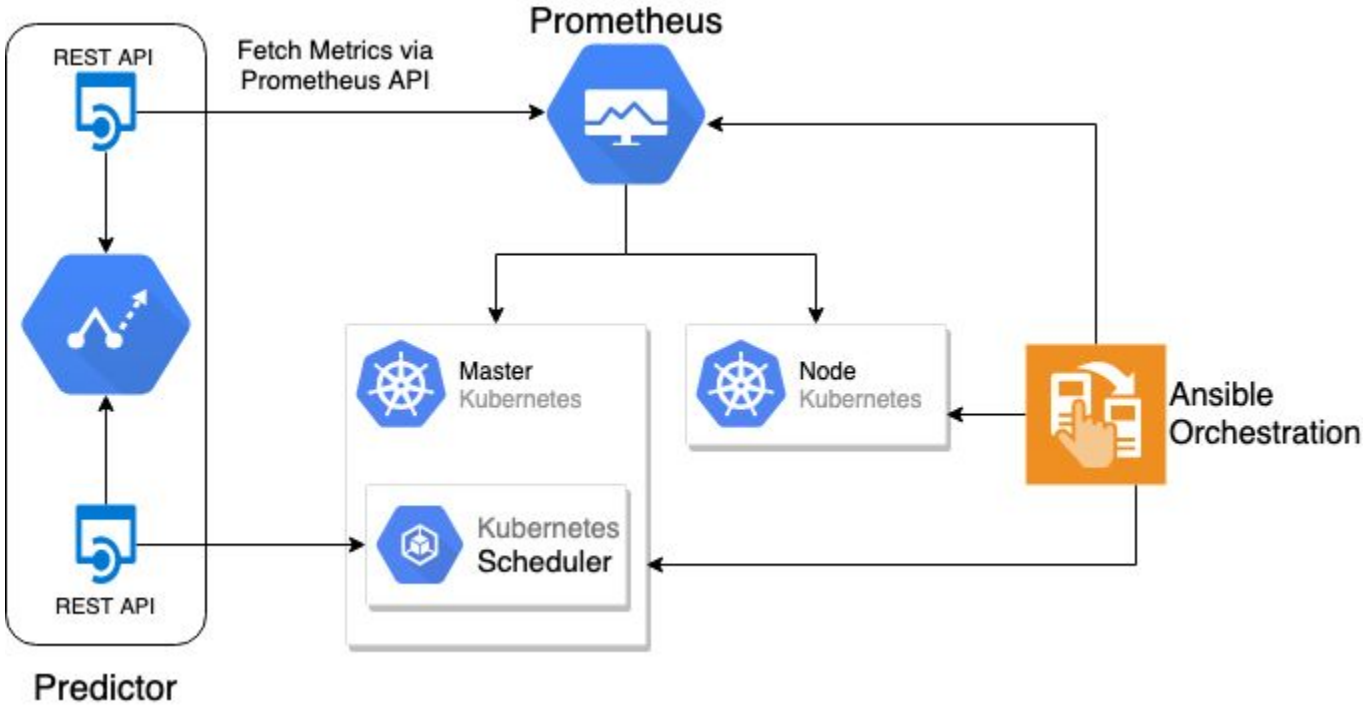
Kyle Martin (kdmarti2)

Vasudev Bongale (vabongal)

Sohail Shaikh (sashaikh)

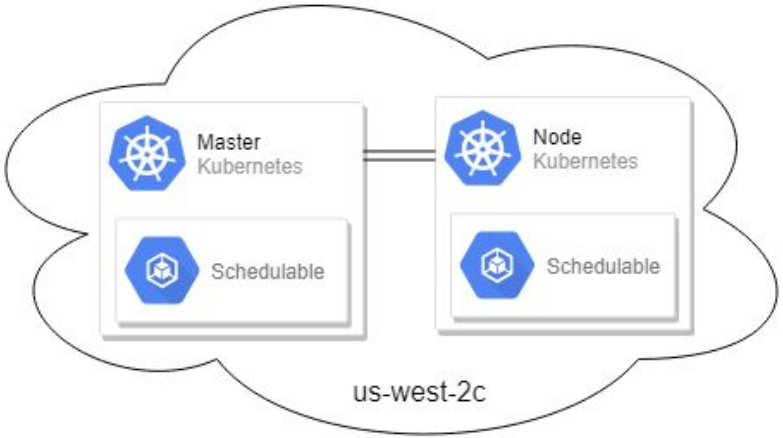
28-March-2019

Architecture

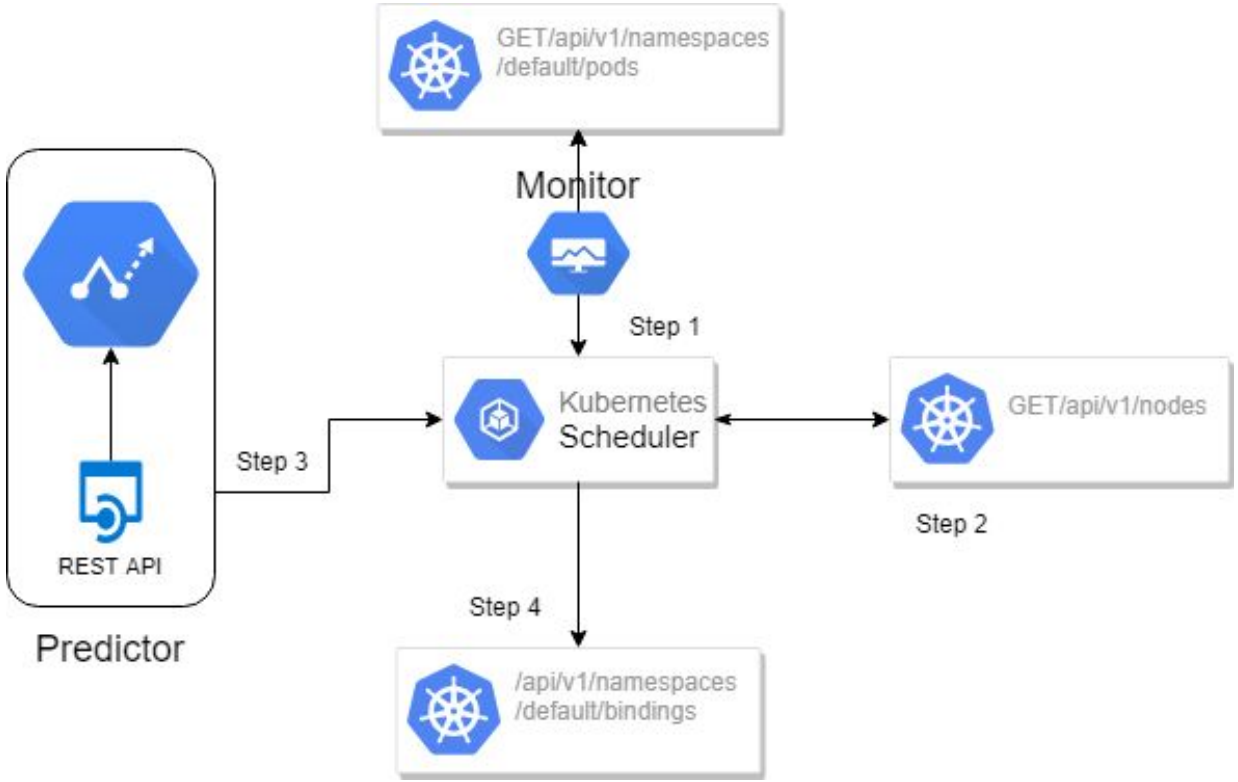


Kubernetes Setup

- AWS EC2 Cloud Computing
  - 2 t2.xlarge compute nodes
- Custom Ubuntu 18.04 image
- Ansible for rapid redeployment



Tetris Scheduler

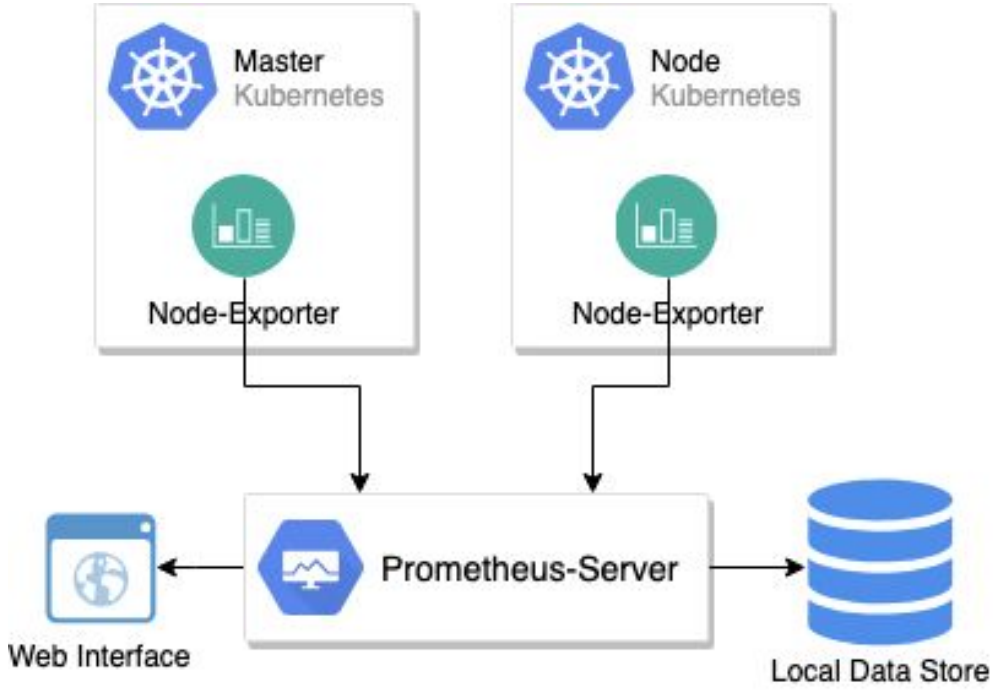


Prometheus Setup

- Prometheus Server
- Node-exporter
- Scrape metrics
  - Interval: 5 seconds
- Setup using Ansible

Before this:

kube-prometheus

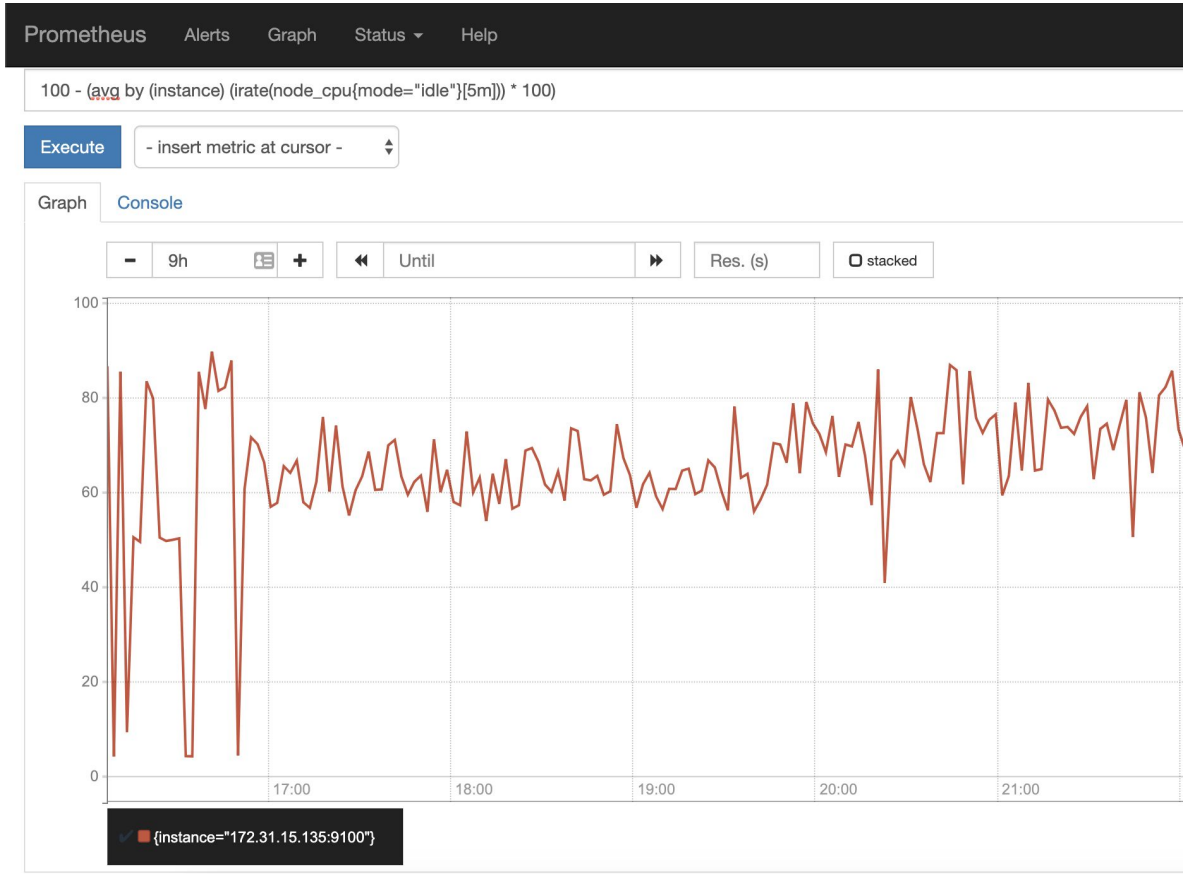


Prometheus Queries

- CPU Usage: "100 - (avg by (instance) (irate(node_cpu{mode='idle'}[5m])) * 100)"
- Free Memory: "100 * (1 - ((node_memory_MemFree + node_memory_Cached + node_memory_Buffers) / node_memory_MemTotal))"
- Current Number of IO operations*: "node_disk_io_now{device='xvda'}"

* Xiao Qin, Hong Jiang, Yifeng Zhu, and David R Swanson. 2003. Dynamic load balancing for I/O-intensive tasks on heterogeneous clusters. In International Conference on High-Performance Computing Springer, 300–309.

CPU Usage %


Prometheus
Alerts
Graph
Status
Help
100 - (avg by (instance) (irate(node_cpu{mode="idle"}[5m])) * 100)
Execute
- insert metric at cursor -
Graph
Console
9h
Until
Res. (s)
stacked
100
80
60
40
20
0
17:00
18:00
19:00
20:00
21:00
{instance="172.31.15.135:9100"}

Collected Metrics

```
In [12]: prometheusUtils.run_query_range("node_load15{instance!='localhost:9100'}",
 1553555405, 1553561712, 800)
Out[12]:
[('172.31.15.135:9100', '1553555405', '0.57'),
 ('172.31.15.135:9100', '1553556205', '0.27'),
 ('172.31.15.135:9100', '1553557005', '0.16'),
 ('172.31.15.135:9100', '1553557805', '0.09'),
 ('172.31.15.135:9100', '1553558605', '0.21'),
 ('172.31.15.135:9100', '1553559405', '0.16'),
 ('172.31.15.135:9100', '1553560205', '0.41'),
 ('172.31.15.135:9100', '1553561005', '0.51'),
 ('172.31.15.164:9100', '1553559405', '0.05'),
 ('172.31.15.164:9100', '1553560205', '0.08'),
 ('172.31.15.164:9100', '1553561005', '0.07')]
```

APIs

- Components communicate using REST APIs.
- Prometheus Module
  - Get metrics from prometheus-server
  - Prometheus API
    - GET /api/v1/query
    - GET /api/v1/query_range
- Scheduler Queries a Flask API to interact with the module.

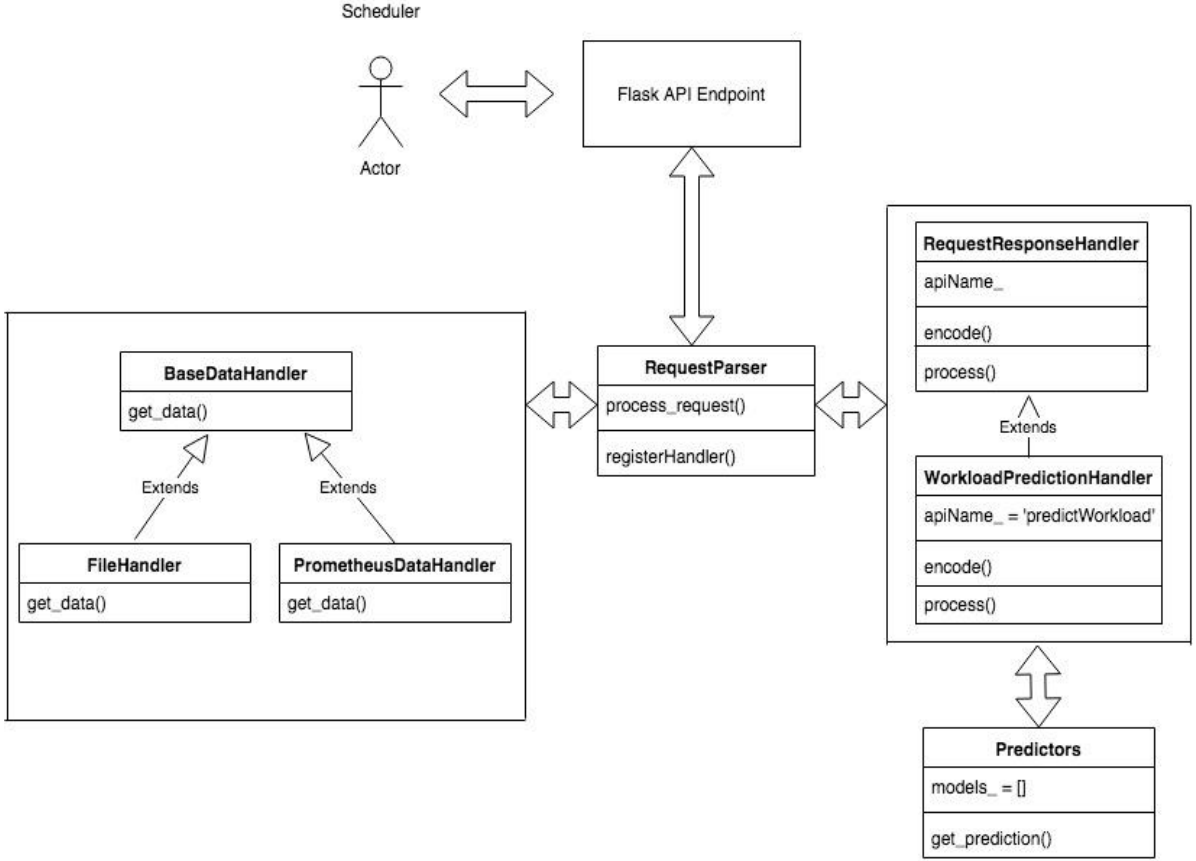
Training Data Collection

1. Three different K8s stress-ng deployments
   a. Three different classes of stress-ng were used:
   b. Randomized number of workers, cpu load, load slice, time out values

cpu	stress-ng --cpu $workers --cpu-load $cpu_load --cpu-load-slice $load_slice -t $time
mem	stress-ng --sequential 1 --class vm --timeout $time_out
io	stress-ng --hdd $hdd_workers --utime $utime_workers --utime-fsync -t $timeout

2. Each container was deployed for 6 hrs.
3. Stats exported by Prometheus Node Exporter

Prediction Module



Request/Response

Request format:

```
{
   "apiName" : "predictWorkload",
   "model" : "cpu",
}
```

Response format:

```
[
    [
        {"workload": {"172.31.15.135:9100": 64.09697330707029}},
        {"workloadType": "cpu"},
        {"unit": "percent"}
    ]
]
```

Result: Tetris Scheduler

```
INFO:root:Starting to Schedular pods for tetis in namespace: default
INFO:root:Event Triggered!!! Phase: Pending scheduler_name: tetris-s
cheduler-cpu
INFO:root:Getting Nodes!!!
WARNING:root:Node k8master is not available
WARNING:root:Node k8master is not available
WARNING:root:Node k8master is not available
INFO:root:Found Node k8master:172.31.15.135 is Ready!
WARNING:root:Node k8node1 is not available
WARNING:root:Node k8node1 is not available
WARNING:root:Node k8node1 is not available
INFO:root:Found Node k8node1:172.31.15.164 is Ready!
{'Data': [('172.31.15.135:9100', 73.87563408636008)], 'Choice': 'k8m
aster'}
INFO:root:Putting cpu-stressor-5ffcf97c8-bqwlp on k8master in namesp
ace: default
WARNING:root:Recieved ValueError for Null target, Ignoring because h
ttps://github.com/kubernetes-client/python/issues/547#issuecomment-4
55362558
```

```
ubuntu@K8master:~$ kubectl get deployments
NAME           READY   UP-TO-DATE   AVAILABLE   AGE
cpu-stressor   1/1     1            1           4m49s
```

Progress

February 26, 2019 – March 26, 2019

Period: 1 month

Overview

13 Active Pull Requests

47 Active Issues

12	1	35	12
Merged Pull Requests	Proposed Pull Request	Closed Issues	New Issues

6 Open ✓ 0 Closed

Presentations and Final Paper
Updated 2 days ago

Experiments and Evaluations
Updated 26 days ago

Testing
Updated 26 days ago

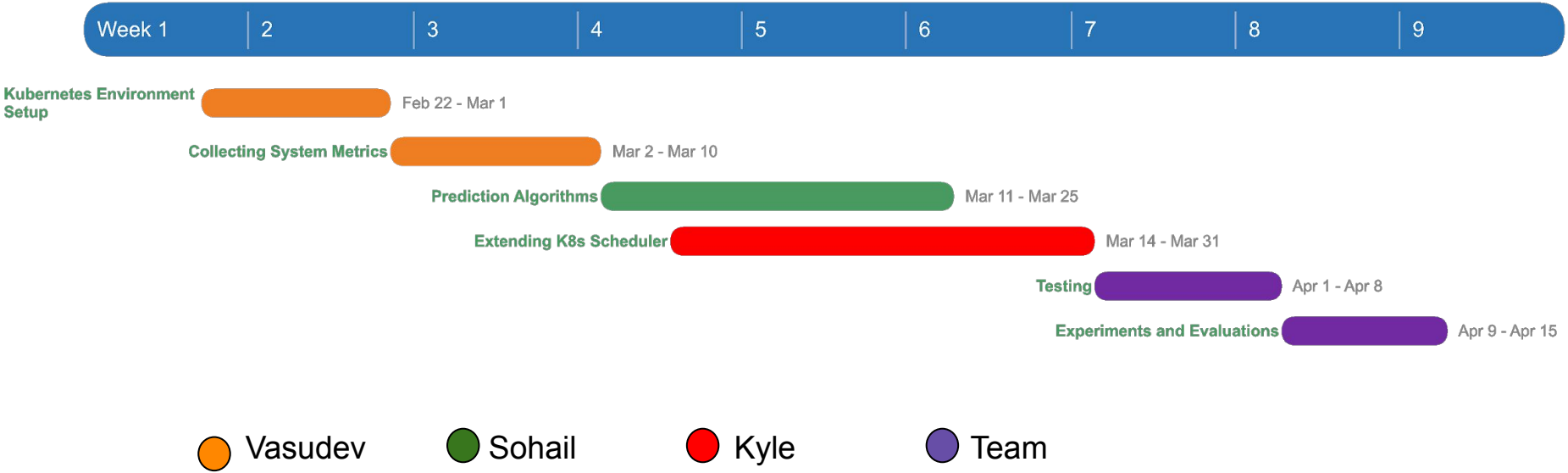
Extending K8s Scheduler
Updated 17 minutes ago

Prediction Algorithms
Updated 17 hours ago

Setup and Metrics Collection
Updated just now

From: https://github.ncsu.edu/vabongal/csc724-k8s/projects

Project Roadmap - On Track



Questions?