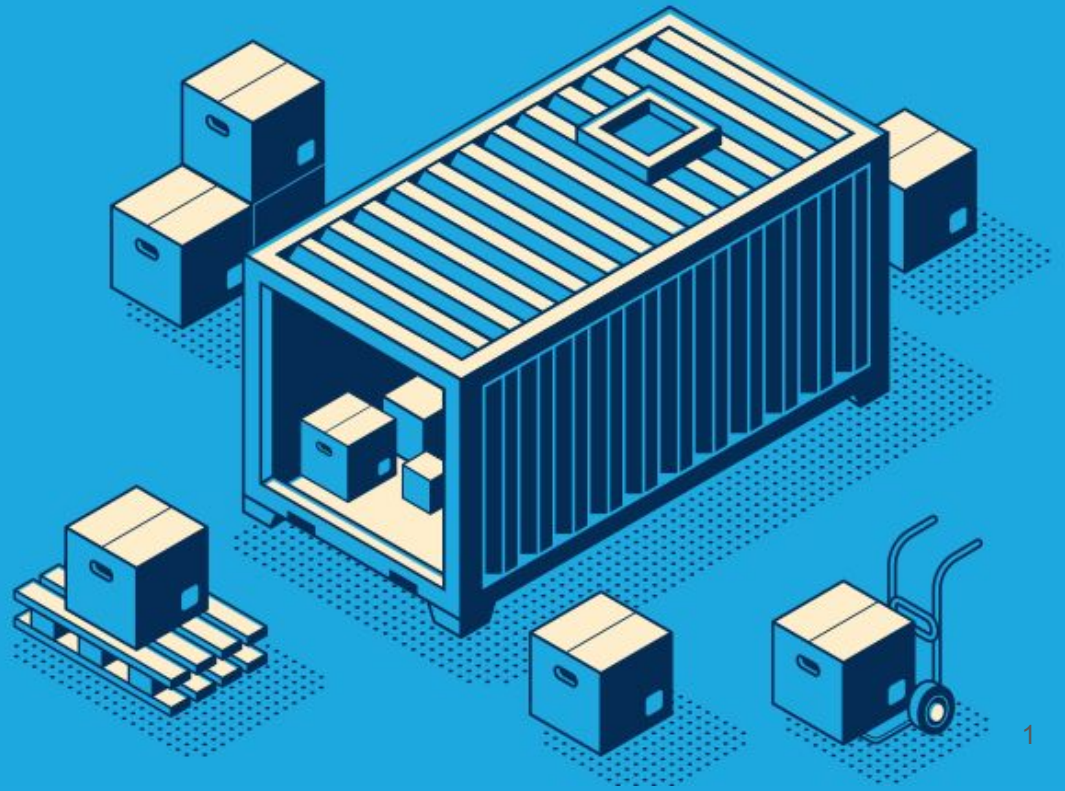
Kubernetes

Vertical Pod Autoscaler Operator (VPA)



The Team

Selen

Laxmi

Apoorva

Akshay

Sindhu

Mentors (Redhat Operate-First)

Humair Khan & Anand Sanmukhani

Recap of previous sprints

(Sprint 1-4)

Sprint 1

- Understanding problem, getting access from RedHat for setting up tools

Sprint 2

- Install VPA Operator, test using a dummy workload, setup Grafana

Sprint 3

- Setup VPA to act on RC app and visualized test results on Grafana

Sprint 4

- Tested VPA by sending workloads for different scenarios to study behavior

What we achieved in THIS sprint

(Sprint EDINBURGH)

Grafana

1. Added filters like namespace, container and pod
2. How we made it customizable for any namespace

What we achieved in THIS sprint

(Sprint EDINBURGH)

VPA - Load simulation

1. Memory test cases
    1. Day and night
    2. Over utilization
    3. under utilization
2. CPU test cases
    1. Day and night
    2. Gradual increasing

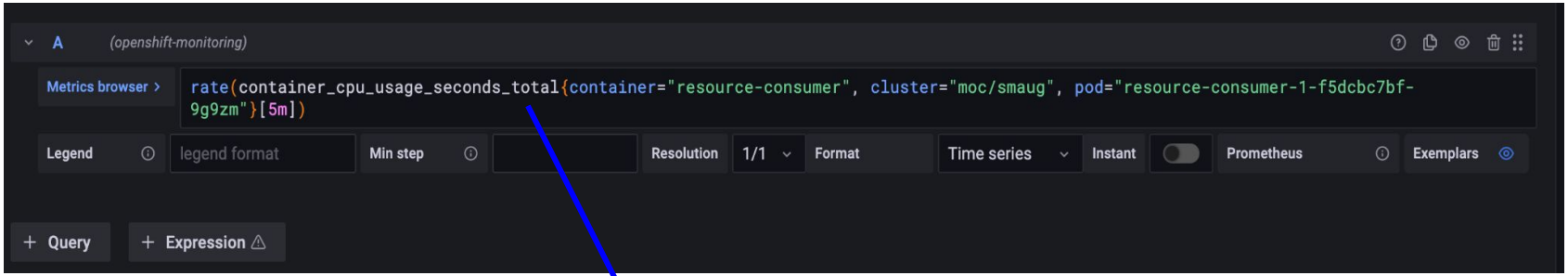
What we achieved in THIS sprint

(Sprint EDINBURGH)

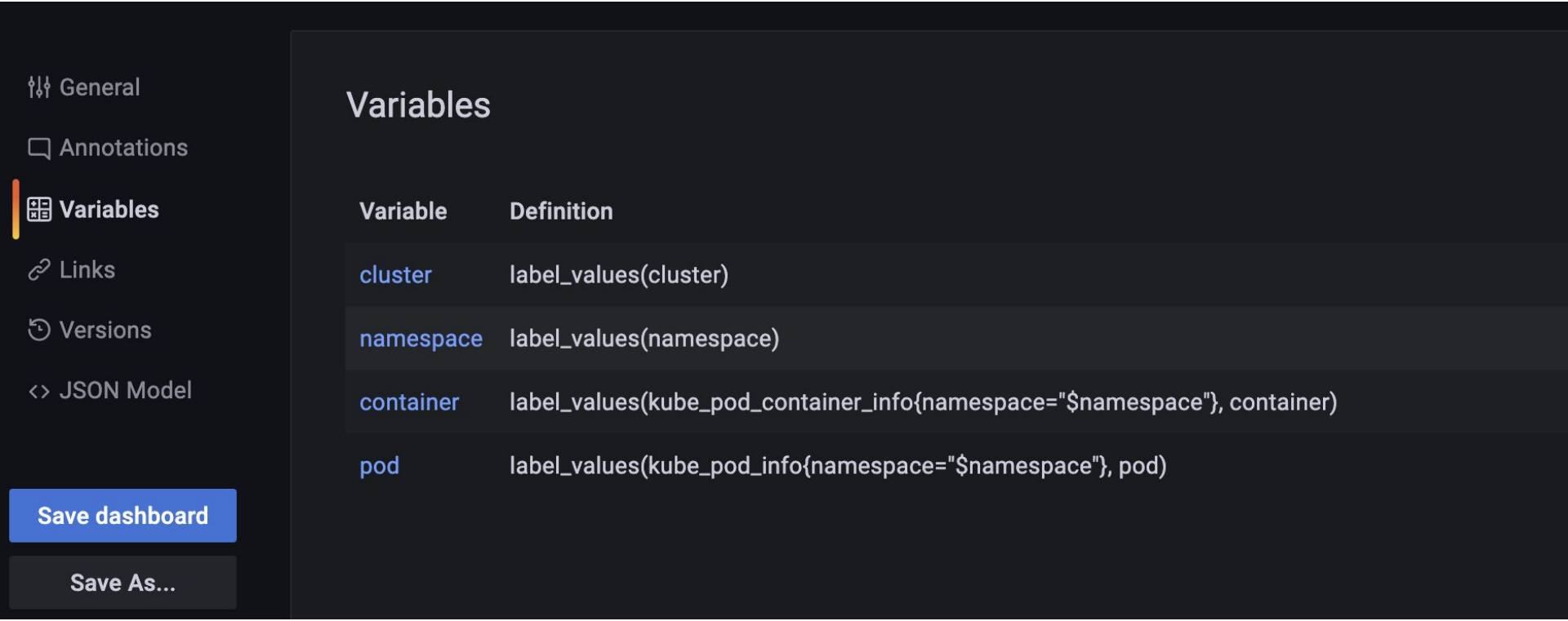
Trino

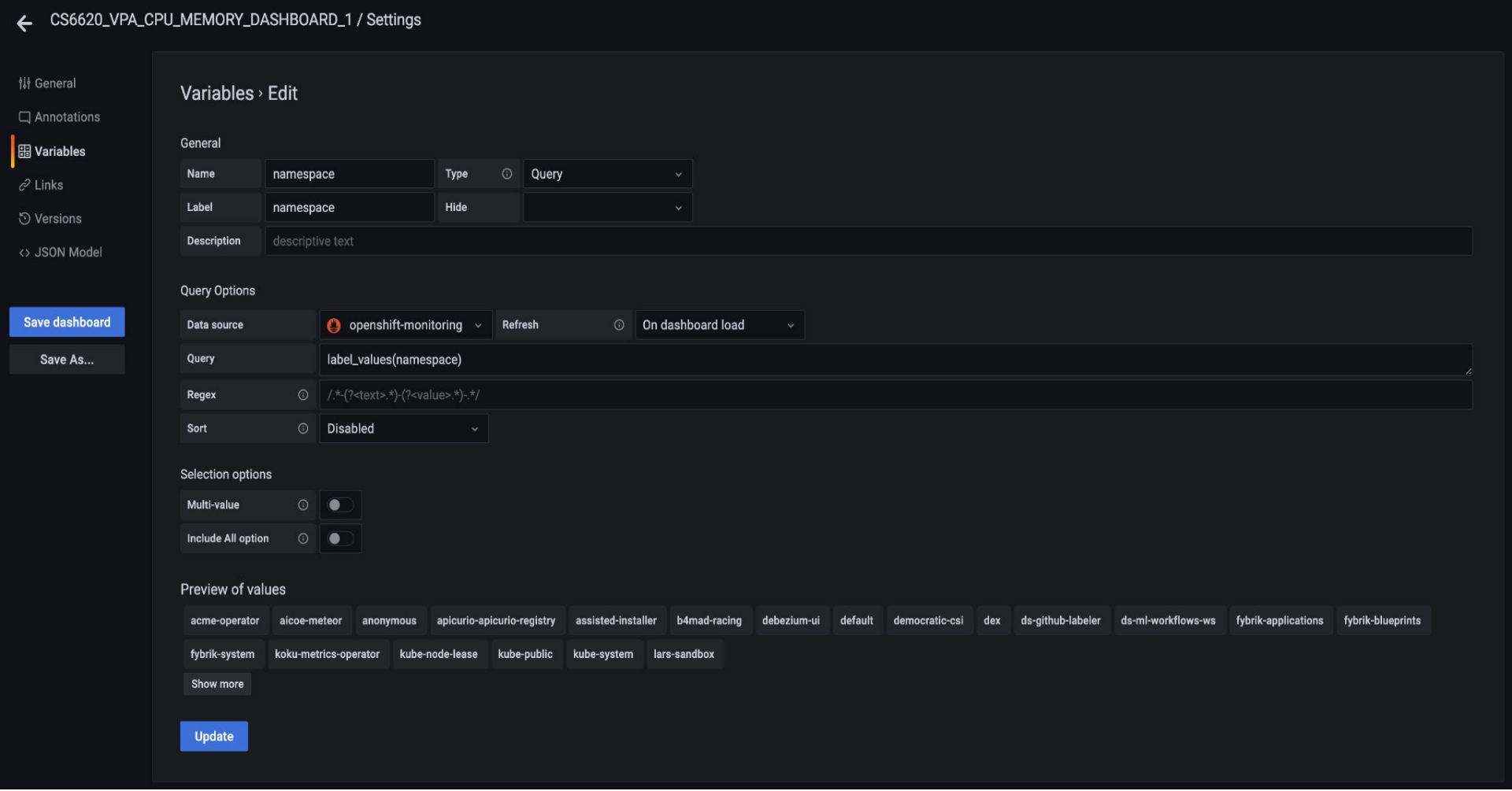
1. Trino tour
2. Initial VPA configuration and recommendations
3. Cron Job and SQL query
4. How VPA reacted Cron Job

PREVIOUSLY ON GRAFANA

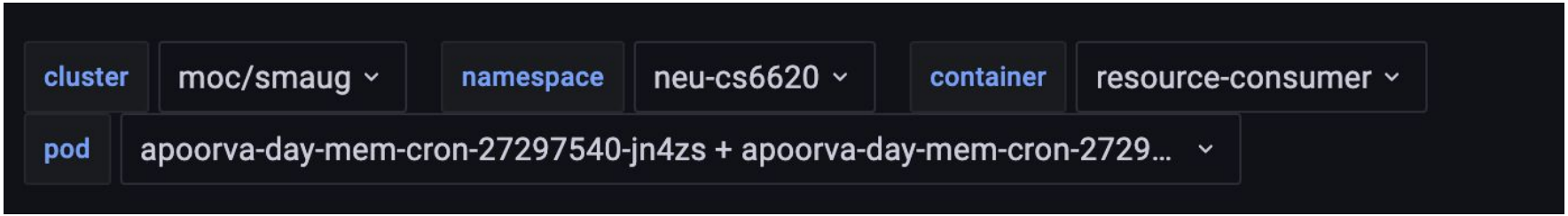


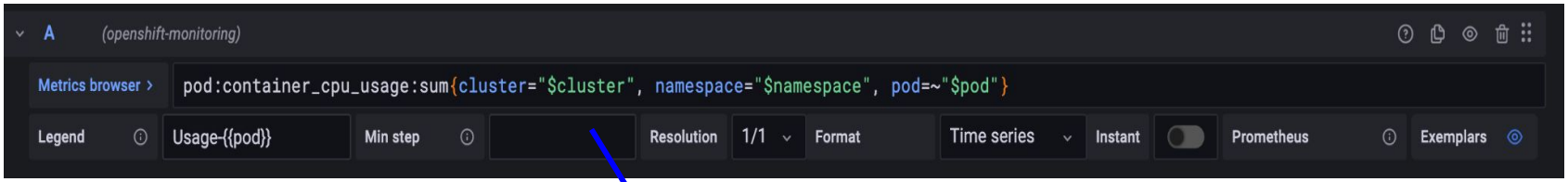

```
rate(container_cpu_usage_seconds_total {
container = "resource-consumer",
cluster = "moc/smaug",
pod = "resource-consumer-1-f5dcbc7bf-9g9zm"
}[5m])
```


General
Annotations
Variables
Links
Versions
JSON Model
Save dashboard
Save As...
Variables
Variable
Definition
cluster
label_values(cluster)
namespace
label_values(namespace)
container
label_values(kube_pod_container_info{namespace="$namespace"}, container)
pod
label_values(kube_pod_info{namespace="$namespace"}, pod)

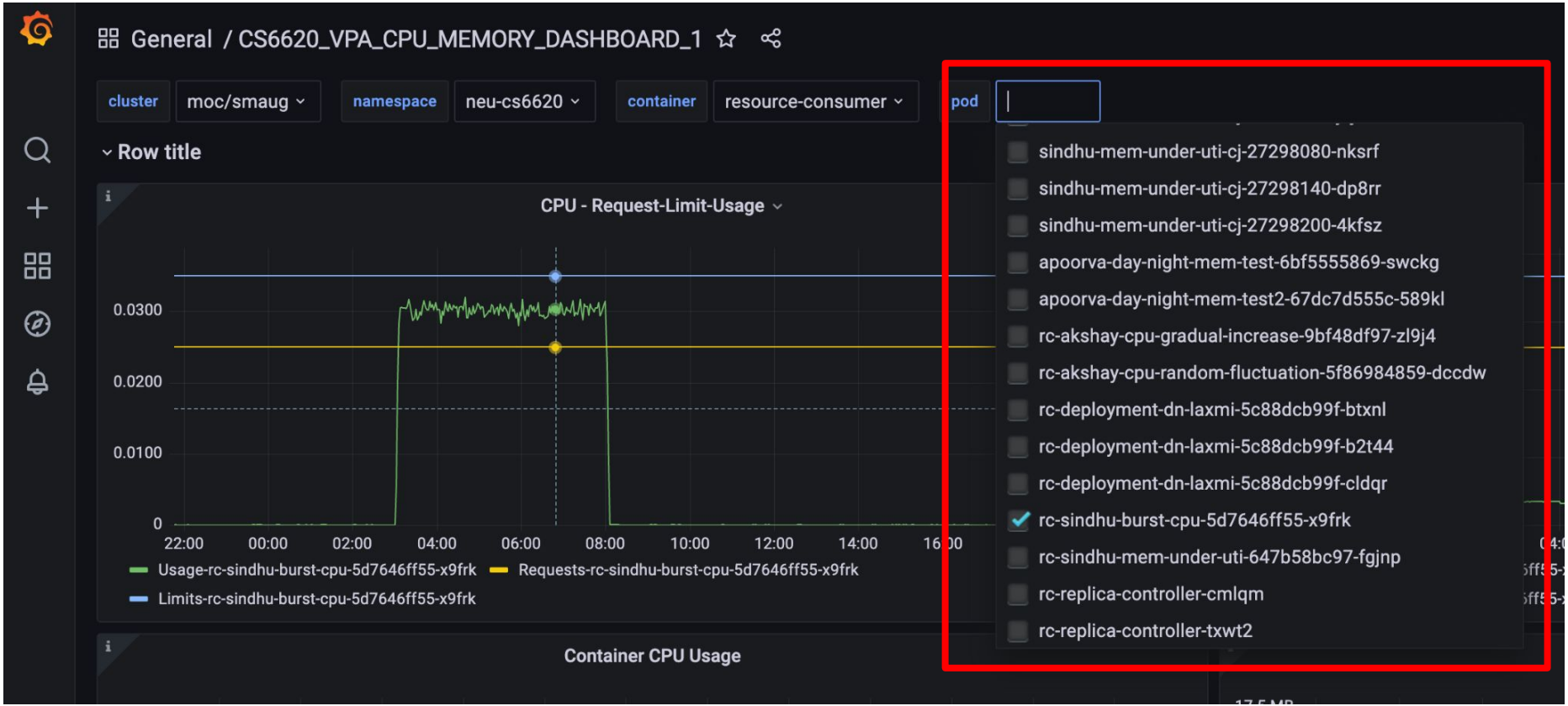

CS6620_VPA_CPU_MEMORY_DASHBOARD_1 / Settings
General
Annotations
Variables
Links
Versions
JSON Model
Save dashboard
Save As...
Variables › Edit
General
Name
namespace
Type
Query
Label
namespace
Hide
Description
descriptive text
Query Options
Data source
openshift-monitoring
Refresh
On dashboard load
Query
label_values(namespace)
Regex
/.*-(?<text>.*)-(?<value>.*)-.*/
Sort
Disabled
Selection options
Multi-value
Include All option
Preview of values
acme-operator
aicoe-meteor
anonymous
apicurio-apicurio-registry
assisted-installer
b4mad-racing
debezium-ui
default
democratic-csi
dex
ds-github-labeler
ds-ml-workflows-ws
fybrik-applications
fybrik-blueprints
fybrik-system
koku-metrics-operator
kube-node-lease
kube-public
kube-system
lars-sandbox
Show more
Update

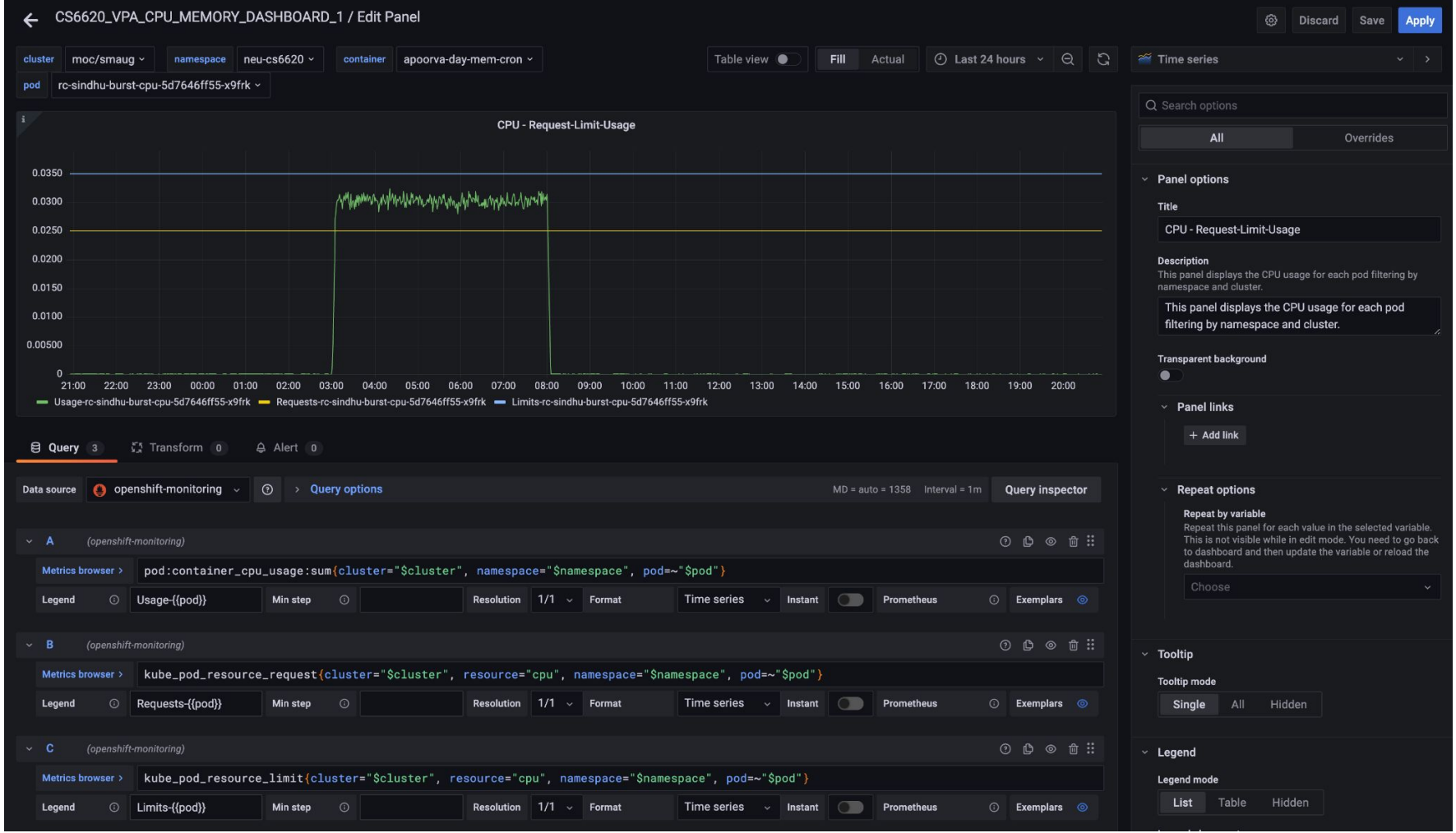
GRAFANA - PROMQL





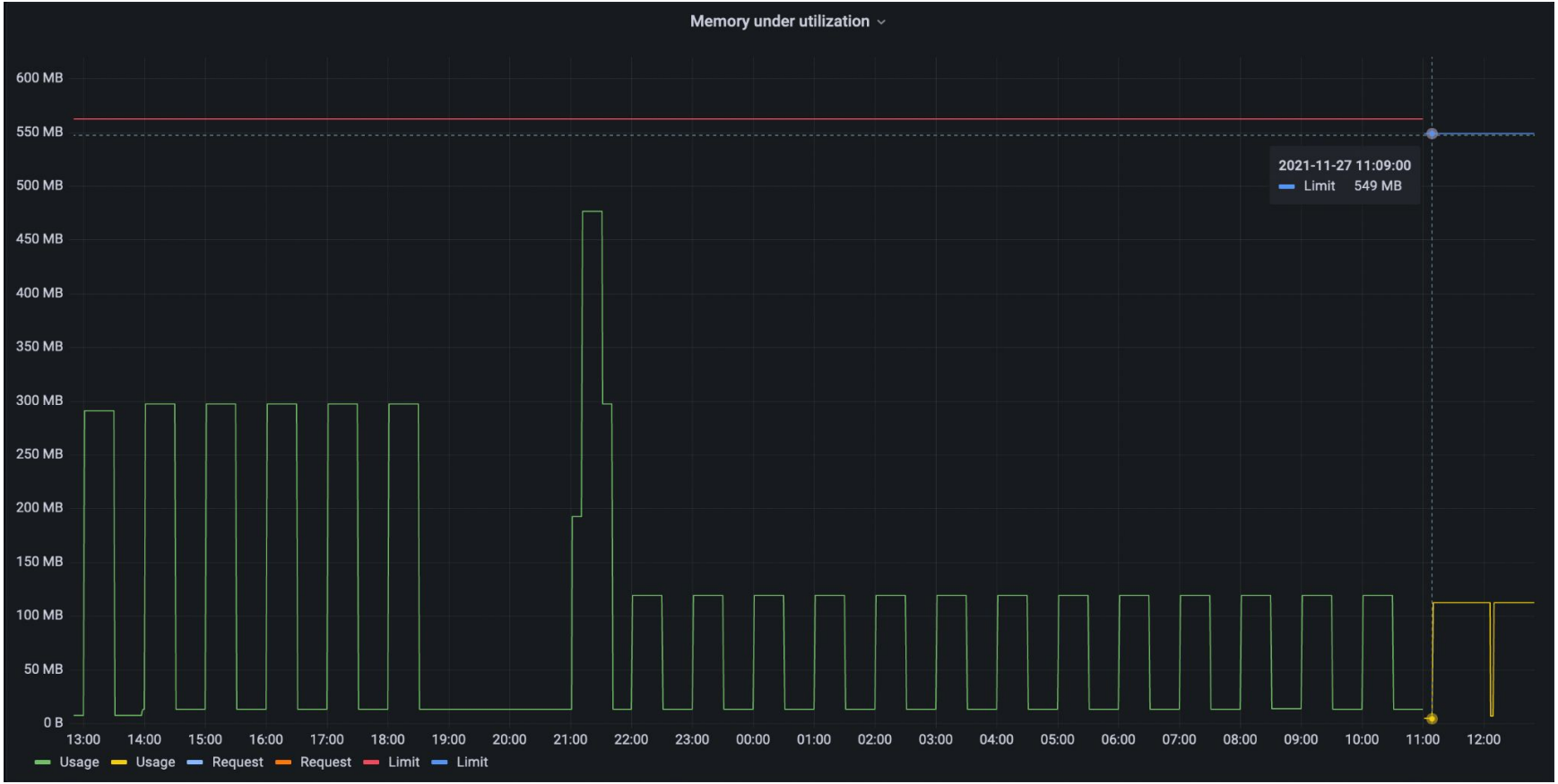

```
pod:container_cpu_usage:sum {
cluster = "$cluster",
namespace = "$namespace",
pod ~= "$pod" }
```


General / CS6620_VPA_CPU_MEMORY_DASHBOARD_1
cluster
moc/smaug
namespace
neu-cs6620
container
resource-consumer
pod
Row title
CPU - Request-Limit-Usage
0.0300
0.0200
0.0100
0
22:00
00:00
02:00
04:00
06:00
08:00
10:00
12:00
14:00
Usage-rc-sindhu-burst-cpu-5d7646ff55-x9frk
Requests-rc-sindhu-burst-cpu-5d7646ff55-x9frk
Limits-rc-sindhu-burst-cpu-5d7646ff55-x9frk
sindhu-mem-under-uti-cj-27298080-nksrf
sindhu-mem-under-uti-cj-27298140-dp8rr
sindhu-mem-under-uti-cj-27298200-4kfsz
apoorva-day-night-mem-test-6bf5555869-swckg
apoorva-day-night-mem-test2-67dc7d555c-589kl
rc-akshay-cpu-gradual-increase-9bf48df97-zl9j4
rc-akshay-cpu-random-fluctuation-5f86984859-dccdw
rc-deployment-dn-laxmi-5c88dcb99f-btxnl
rc-deployment-dn-laxmi-5c88dcb99f-b2t44
rc-deployment-dn-laxmi-5c88dcb99f-cldqr
rc-sindhu-burst-cpu-5d7646ff55-x9frk
rc-sindhu-mem-under-uti-647b58bc97-fgjnp
rc-replica-controller-cmlqm
rc-replica-controller-txwt2
Container CPU Usage

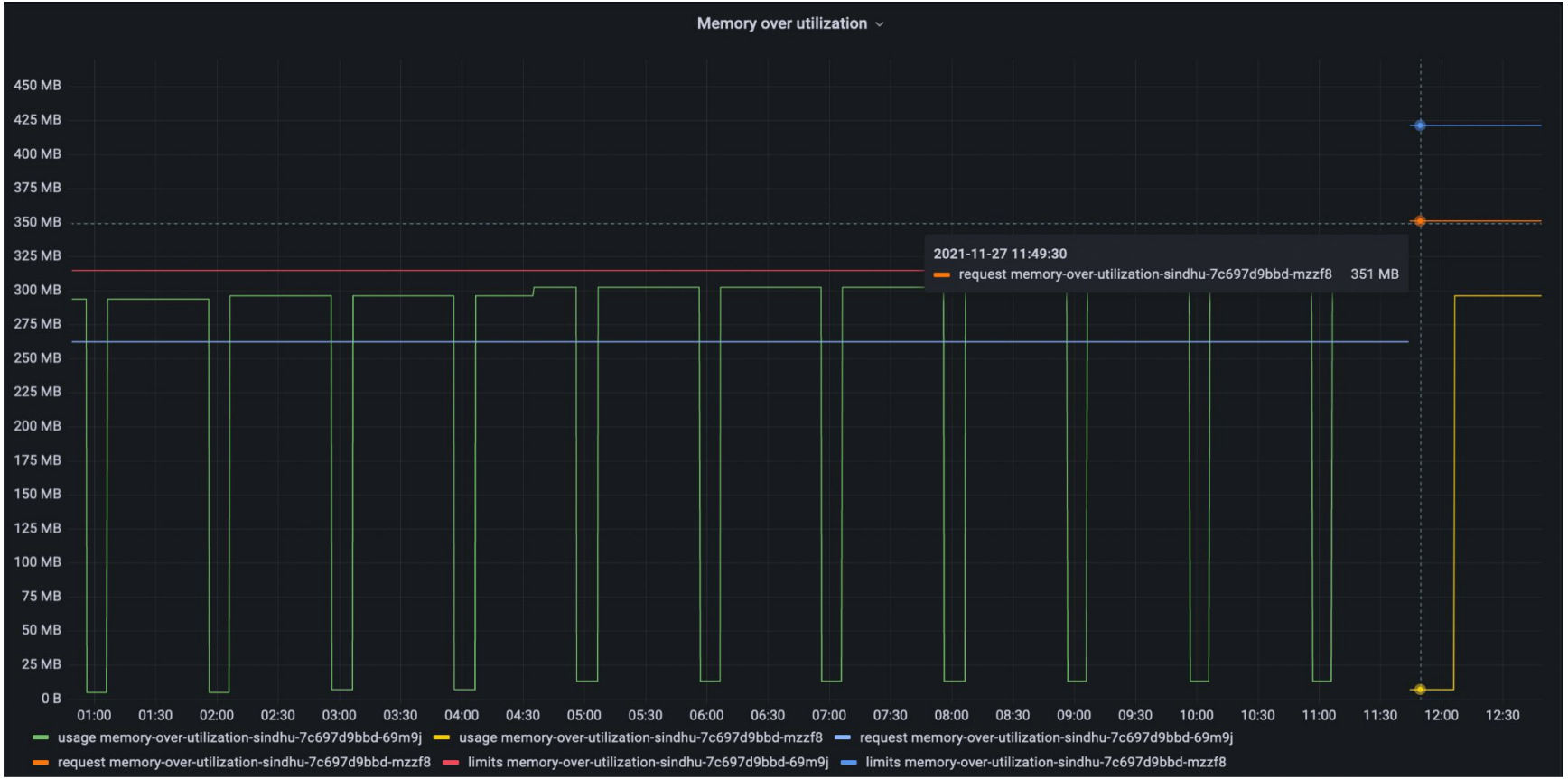

CS6620_VPA_CPU_MEMORY_DASHBOARD_1 / Edit Panel
Discard
Save
Apply
cluster
moc/smaug
namespace
neu-cs6620
container
apoorva-day-mem-cron
pod
rc-sindhu-burst-cpu-5d7646ff55-x9frk
Table view
Fill
Actual
Last 24 hours
Time series
CPU - Request-Limit-Usage
0.0350
0.0300
0.0250
0.0200
0.0150
0.0100
0.00500
0
21:00 22:00 23:00 00:00 01:00 02:00 03:00 04:00 05:00 06:00 07:00 08:00 09:00 10:00 11:00 12:00 13:00 14:00 15:00 16:00 17:00 18:00 19:00 20:00
Usage-rc-sindhu-burst-cpu-5d7646ff55-x9frk
Requests-rc-sindhu-burst-cpu-5d7646ff55-x9frk
Limits-rc-sindhu-burst-cpu-5d7646ff55-x9frk
Search options
All
Overrides
Panel options
Title
CPU - Request-Limit-Usage
Description
This panel displays the CPU usage for each pod filtering by namespace and cluster.
This panel displays the CPU usage for each pod filtering by namespace and cluster.
Transparent background
Panel links
+ Add link
Repeat options
Repeat by variable
Repeat this panel for each value in the selected variable. This is not visible while in edit mode. You need to go back to dashboard and then update the variable or reload the dashboard.
Choose
Tooltip
Tooltip mode
Single
All
Hidden
Legend
Legend mode
List
Table
Hidden
Query 3
Transform 0
Alert 0
Data source
openshift-monitoring
Query options
MD = auto = 1358
Interval = 1m
Query inspector
A (openshift-monitoring)
Metrics browser
pod:container_cpu_usage:sum{cluster="$cluster", namespace="$namespace", pod=~"$pod"}
Legend
Usage-{{pod}}
Min step
Resolution
1/1
Format
Time series
Instant
Prometheus
Exemplars
B (openshift-monitoring)
Metrics browser
kube_pod_resource_request{cluster="$cluster", resource="cpu", namespace="$namespace", pod=~"$pod"}
Legend
Requests-{{pod}}
Min step
Resolution
1/1
Format
Time series
Instant
Prometheus
Exemplars
C (openshift-monitoring)
Metrics browser
kube_pod_resource_limit{cluster="$cluster", resource="cpu", namespace="$namespace", pod=~"$pod"}
Legend
Limits-{{pod}}
Min step
Resolution
1/1
Format
Time series
Instant
Prometheus
Exemplars

Usage patterns and VPA Observations

VPA - When memory is under utilized



VPA - When memory is over utilized



VPA - Multiple requests

- 3 Cron Jobs at the same time
  1. Requests 100 millicores of CPU
  2. Requests 10 millicores of CPU
  3. Requests 120 millicores of CPU

VPA - Multiple requests

CPU Multiple requests

VPA - ON & OFF Load on CPU

TEST:

- Cron Job requests 150 milli cores of CPU every hour for 10 hours.

OBSERVATION:

- VPA gradually increased the requests and limits of CPU

Initial:

- **Requests: 25m**
- **Limits: 50m**

Updated:

- **Requests: 203m**
- **Limits: 360m**

VPA - DAY TIME

```
target:
  cpu: 25m
  memory: 262144k
```

```
target:
  cpu: 182m
  memory: 262144k
```

```
target:
  cpu: 203m
  memory: 262144k
```

Gradual increase in CPU

VPA - NIGHT TIME

TEST:

- No Load on CPU for more than 20 hours

OBSERVED:

- Reduced requests from 203 mcores to 25 mcores

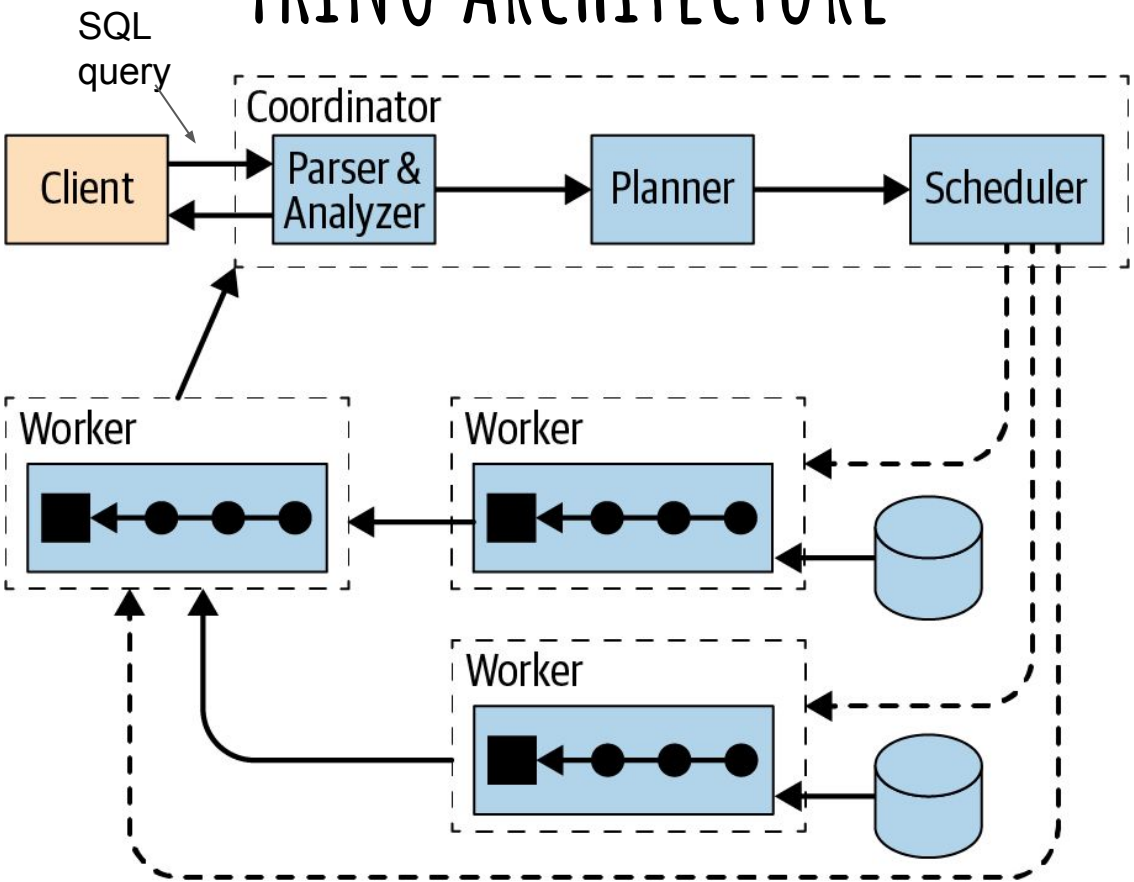
VPA target candidate: Trino

TRINO

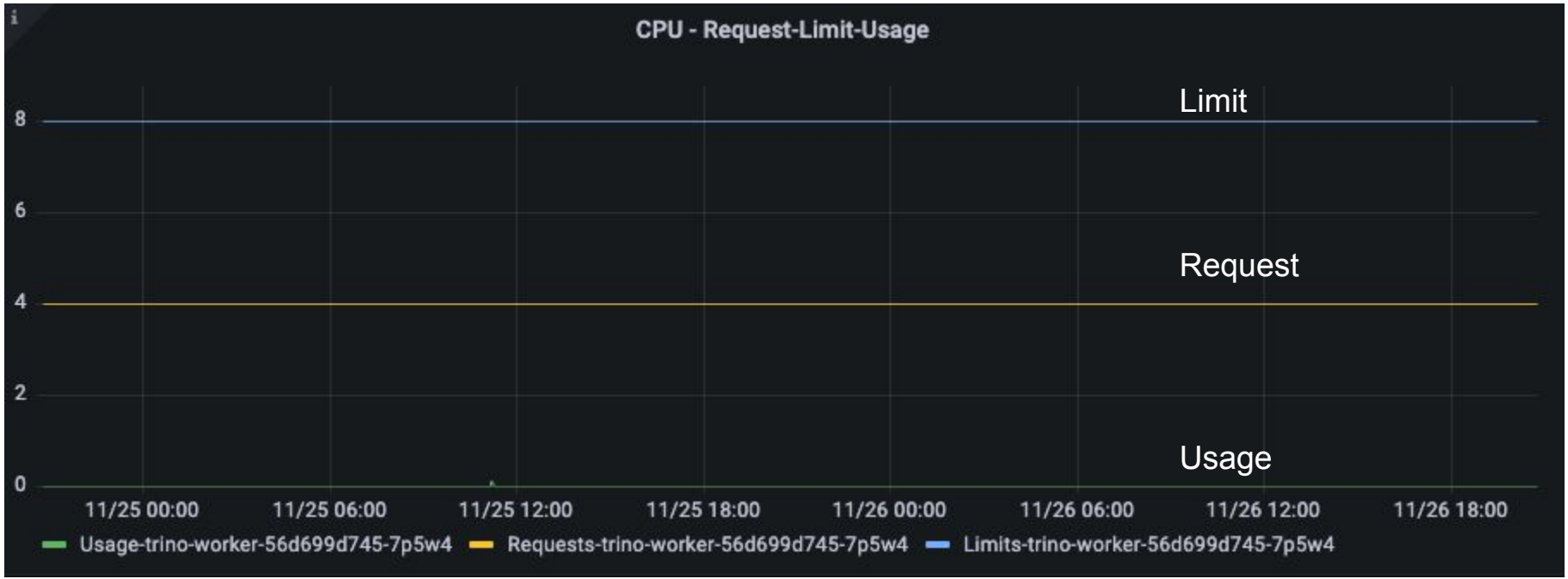
Trino is a distributed SQL query engine designed to query large data sets distributed over one or more heterogeneous data sources. (like AWS S3, HDFS, Postgres etc)

It’s not a relational database, but a tool to efficiently query vast amount of data using distributed queries.

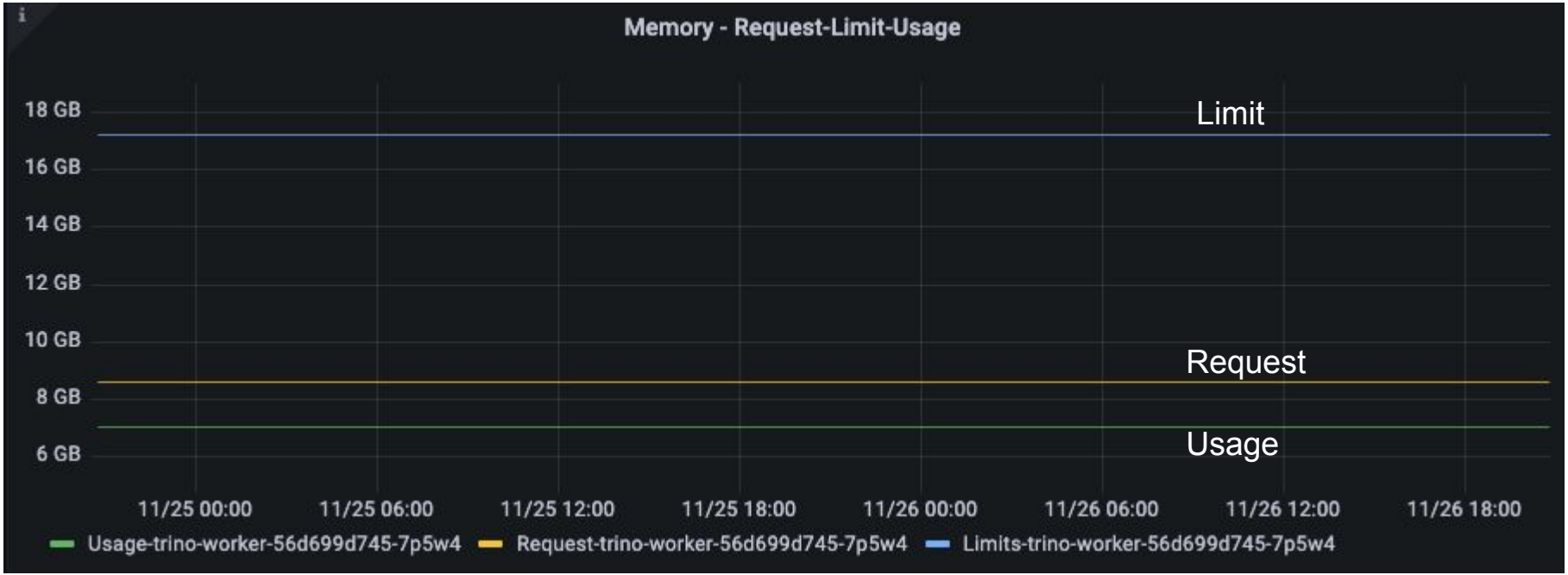
TRINO ARCHITECTURE



Trino worker Prod Cluster Usage: CPU



Trino worker Prod Cluster Usage: MEMORY



TRINO-VPA

```yaml
apiVersion: "autoscaling.k8s.io/v1"
kind: VerticalPodAutoscaler
metadata:
  name: trino-stage-v1
  namespace: opf-trino-stage
  labels:
    app: trino-stage-v1
spec:
  targetRef:
    apiVersion: "apps/v1"
    kind: Deployment
    name: trino-worker
  updatePolicy:
    updateMode: "Auto"
  resourcePolicy:
    containerPolicies:
    - containerName: "trino-worker"
      maxAllowed:
        cpu: "3000m"
        memory: "3Gi"
```

VPA Auto Update mode

TRINO-VPA

```yaml
apiVersion: "autoscaling.k8s.io/v1"
kind: VerticalPodAutoscaler
metadata:
  name: trino-stage-v1
  namespace: opf-trino-stage
  labels:
    app: trino-stage-v1
spec:
  targetRef:
    apiVersion: "apps/v1"
    kind: Deployment
    name: trino-worker
  updatePolicy:
    updateMode: "Auto"
  resourcePolicy:
    containerPolicies:
    - containerName: "trino-worker"
      maxAllowed:
        cpu: "3000m"
        memory: "3Gi"
```

VPA Auto Update mode

Initial Recommendations

Pod Memory Request

Initial Recommendations

Pod CPU Request

SOLUTION

- To fix the problem with extremely low CPU allocation, we configured “minAllowed” value of CPU/Memory resources to set boundaries.

- We recommend to use VPA in “OFF” mode update policy (recommendation only mode) for pods without much historical usage information. Once VPA recommends reasonably good recommendations, it can use “AUTO” update mode.

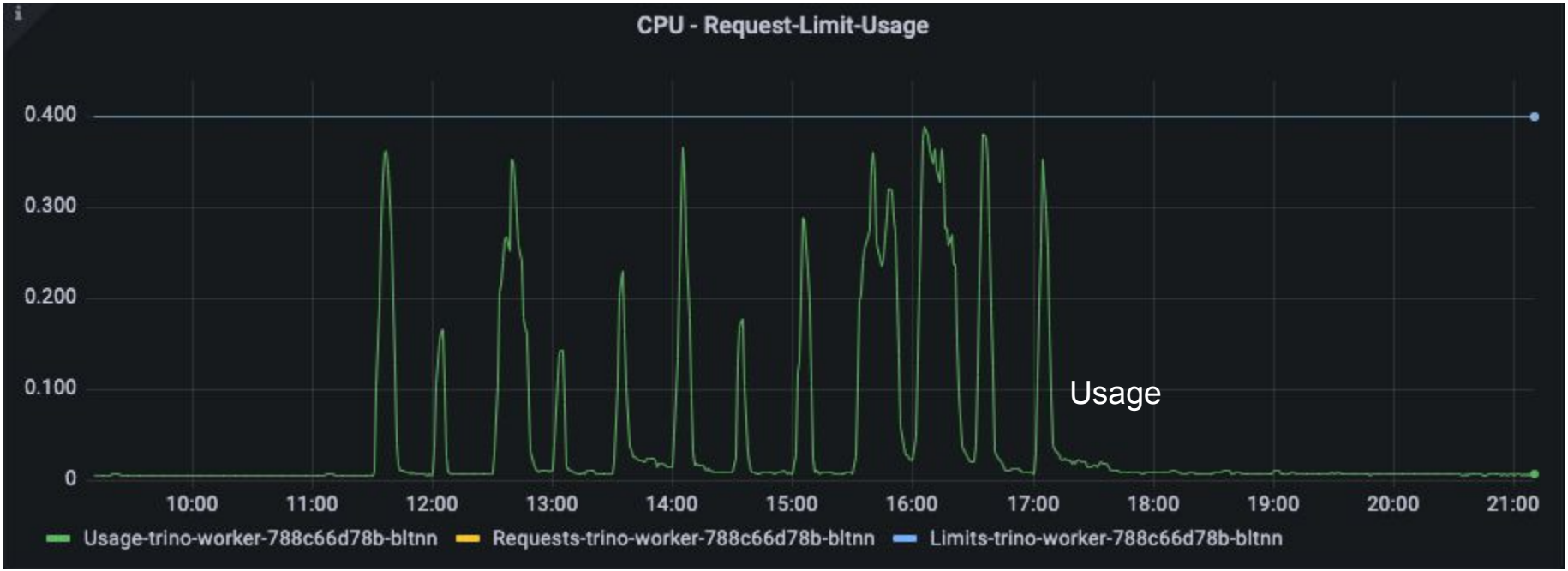
Cron job to simulate workload

```yaml
apiVersion: batch/v1
kind: CronJob
metadata:
  name: trino-stage-v1-cron
spec:
  schedule: "0/30 8-18 * * *"
  jobTemplate:
    spec:
      template:
        spec:
          containers:
          - name: trino-stage-v1-cron
            image: quay.io/opendatahub/trino:362
            command:
            - /bin/sh
            - -ec
            - trino --user=admin --password --server=https://
            trino-route-opf-trino-stage.apps.smaug.na.operate-first.cloud
            --catalog=operate_first_general --schema=test_schema --execute="select
            * from test_pass_failures where "timestamp" IN (select b.timestamp
            from test_pass_failures b where b.test like '%test%' and "timestamp"
            IN (select "timestamp" from test_pass_failures where tab
            LIKE'%redhat-assisted-installer%')) limit 2000;"
            - trino --version
            - echo "DONE!"
          restartPolicy: OnFailure
```

Run every half an hour in day time

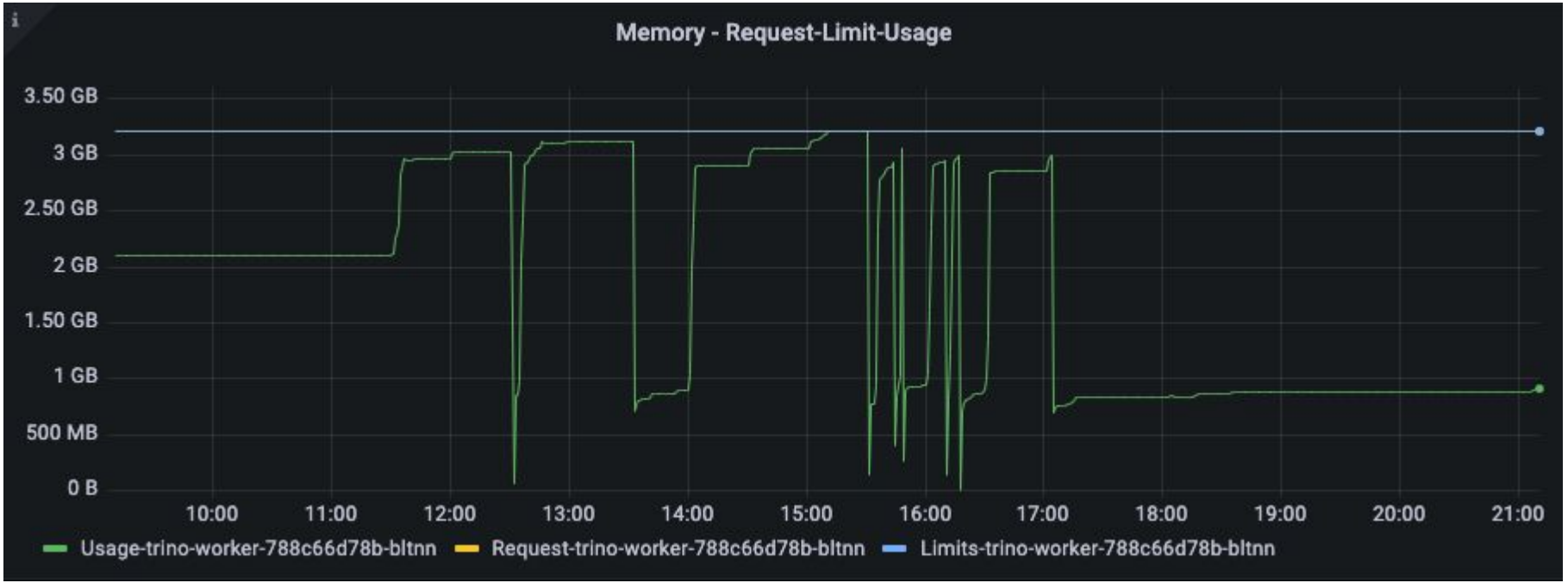
Distributed SQL Query to run on Trino

Cron job to simulate workload



CPU

Cron job to simulate workload



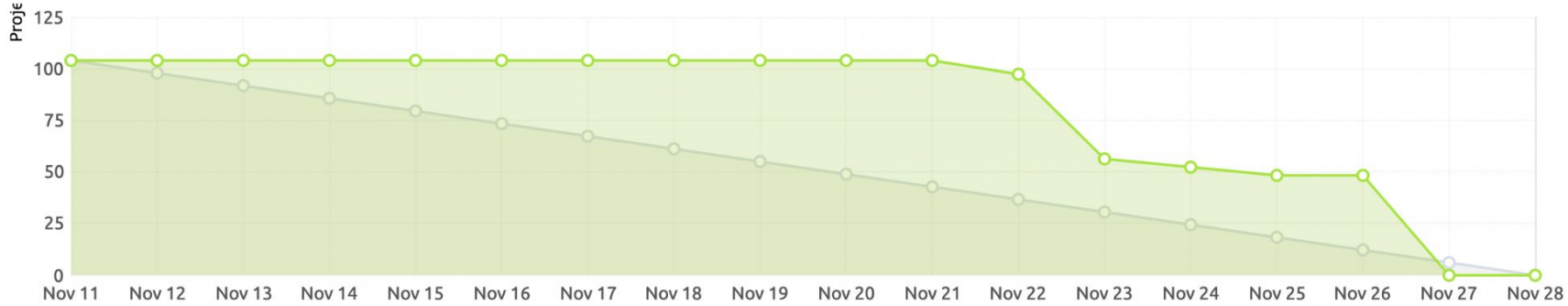
Memory

Burndown chart

Sprint Edinburgh 11 Nov 2021 to 28 Nov 2021

100% 104 total points 104 completed points 0 open tasks 22 closed tasks 0 iocaine doses

How this chart works



What we have left

- Deploying VPA on Trino production
- Document limitations of VPA
- Write a blog about various test cases we did on VPA
- Project final report