

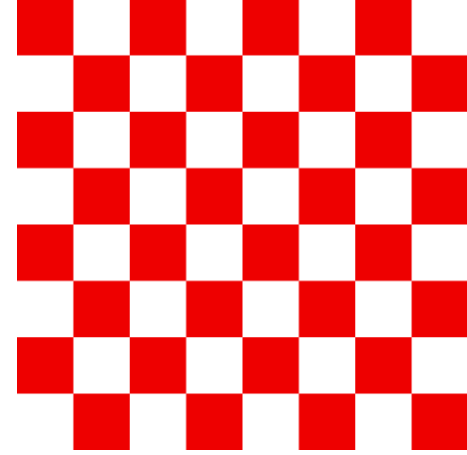
Monitoring Kubernetes Cluster with an External Zabbix Server

The ZABBIX logo, consisting of the word "ZABBIX" in white capital letters on a red rectangular background.

Dario Sindičić

Who are we?

- Government IT company with over than 50 years of existence
- Ministry of Finance, Tax administration, Custom Administration
- In the current transition into the mostly containerized environment
- And those newly stuff also needs monitoring



ZABBIX

























NPIS IT

Pods

Filter

Name

Search by name... /

Name ↑	Status ↓	Ready ↓	Restarts ↓	Owner ↓	Memory ↓	Created ↓
 ansible-postgres-13-0	 Running	1/1	0	 ansible-postgres-13	69.7 MiB	 28 Jul 202
 ansible-task-7b5d489ccd-5fn59	 Running	4/4	0	 ansible-task-7b5d489ccd	1,224.7 MiB	 28 Jul 202
 ansible-web-58b56d48b6-9q24f	 Running	3/3	0	 ansible-web-58b56d48b6	1,431.1 MiB	 28 Jul 202
 awx-0	 Running	4/4	0	 awx	2,454.9 MiB	 23 Jul 202
 awx-operator-controller-manager-6f6fd689b4-8x4h2	 Running	2/2	0	 awx-operator-controller-manager-6f6fd689b4	99.8 MiB	 28 Jul 202
 postgresql-696fd9cf4d-92wb4	 Running	1/1	0	 postgresql-696fd9cf4d	182.7 MiB	 23 Jul 202

ZABBIX

NPIS IT

Traditional approach



ZABBIX

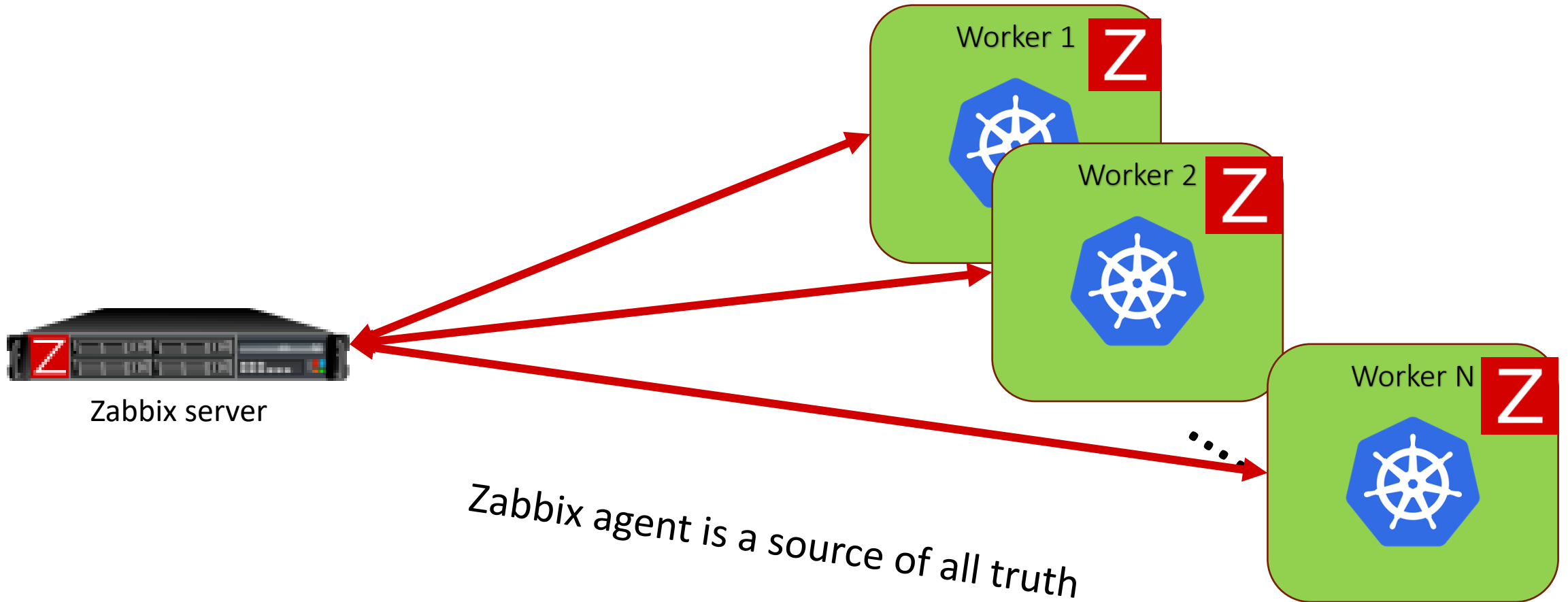
Zabbix is a **Universal** Open Source
enterprise-level monitoring solution

Alexei Vladishev

ZABBIX

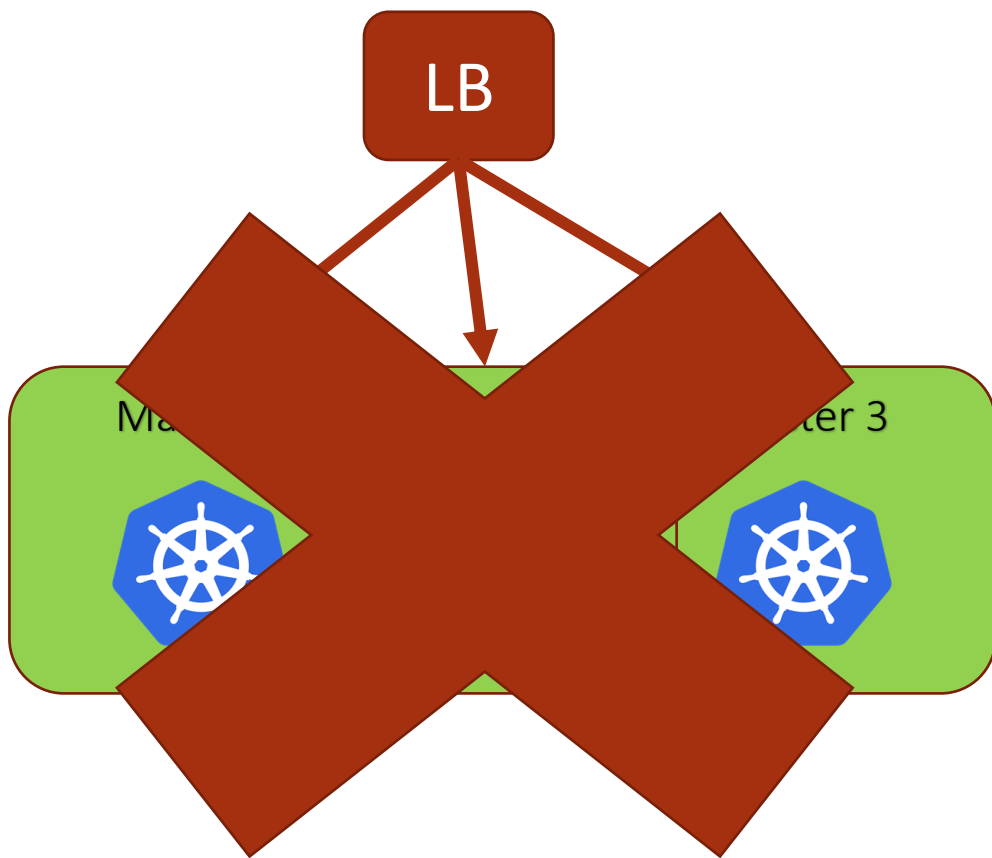
NPIS IT

Our solution



ZABBIX

api.kubernetes.local



ZABBIX

What do we want?

CrashLoop is detected api-88666df67-pbxjq - [REDACTED]

Liveness probe is failing api-64fc944776-5vxpd - [REDACTED]

Memory usage is approaching its limit - [REDACTED]-is-58b589cbdd-tf95z
cp4i-acc-prod

OOM pod mongodb-0 is killed - ibm-backup-restore

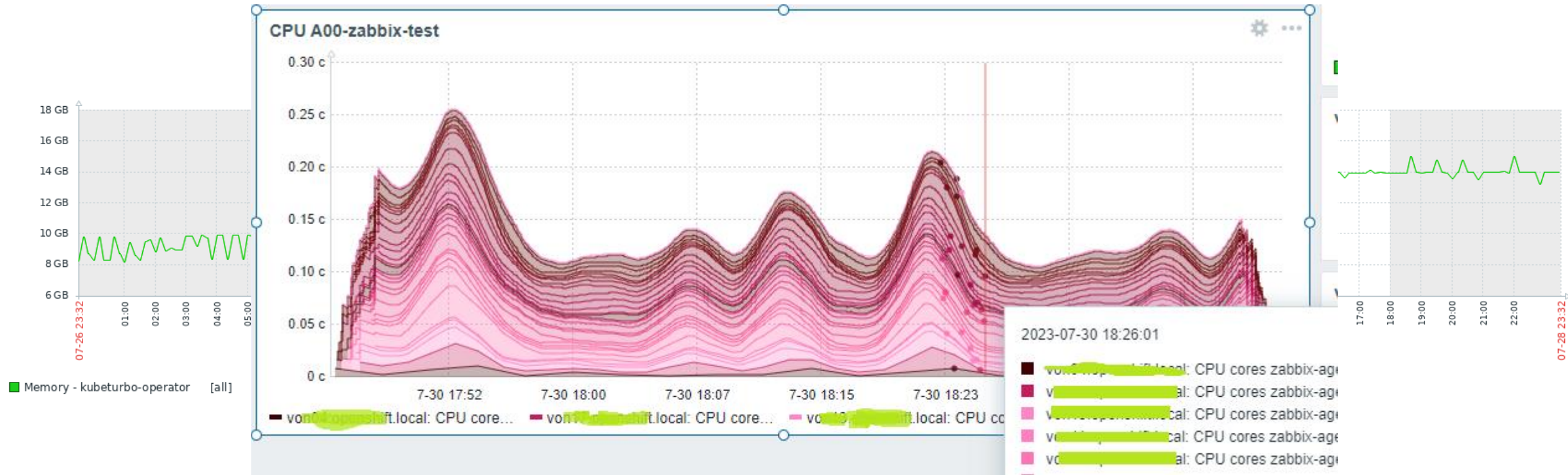
....

ZABBIX



**Emergency
Response**

What do we want?



ZABBIX

NPIS IT

```
graph LR; A[Worker registration] --> B[Pod discovery]; B --> C[Pod lifecycle and data];
```

Worker
registration

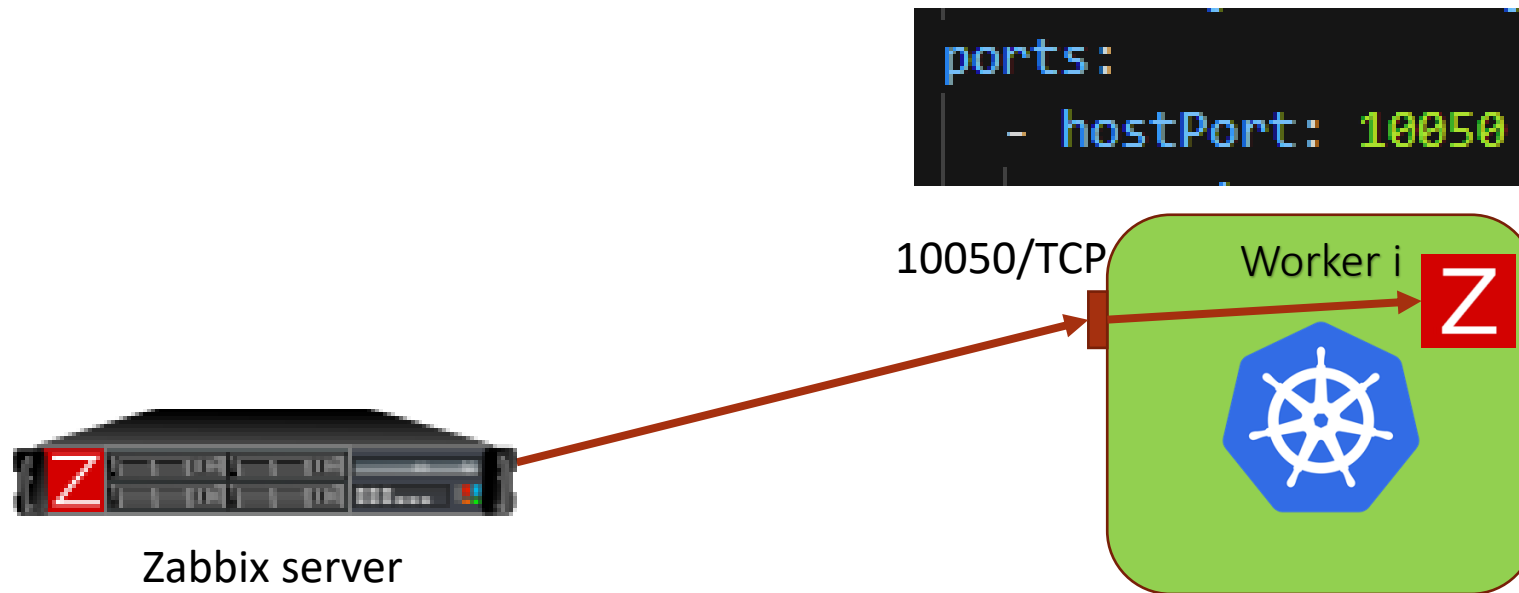
Pod
discovery

Pod lifecycle
and data

ZABBIX

NPIS IT

Worker registration



```
env:  
- name: NODE_NAME  
  valueFrom:  
    fieldRef:  
      apiVersion: v1  
      fieldPath: spec.nodeName
```

HostMetadataItem=system.run[echo \$NODE_NAME]

[worker0.poc.openshift.local](#)

Enabled

ZBX

ZABBIX

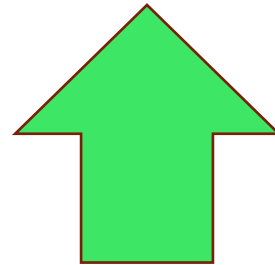
Worker
registration



Pod
discovery



Pod lifecycle
and data



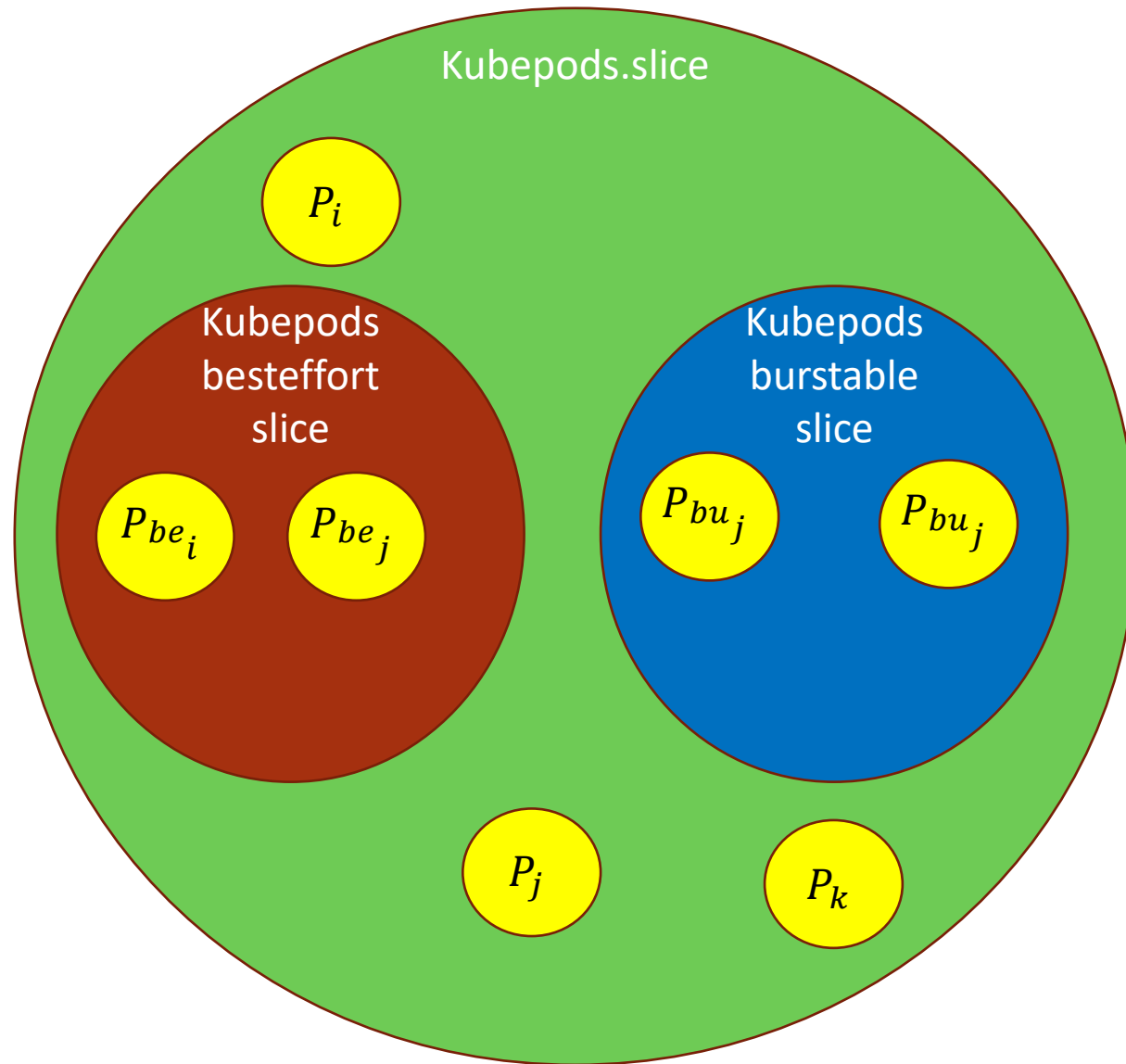
ZABBIX

Pod discovery

- What are pods actually?
- Pod == bunch of processes which are bundled together in slice Systemd structure (or Cgroup for non SystemD systems)
- So why wouldn't we discover all slices and then match them with other useful information

Pod discovery

- Pod = Systemd slice
- 3 main classes of pods



```
CGroup: /kubepods.slice/kubepods-besteffort.slice
kubepods-besteffort-pod0e37594e_3cb2_4dc9_a334_042158589113.slice
kubepods-besteffort-pod14a840fc_a1f3_42d7_b405_ec123ad0a706.slice
kubepods-besteffort-pod33cb8a69_2039_419b_970f_5589754249e0.slice
kubepods-besteffort-pod5087f6e0_4af0_4716_884c_a61022427e3b.slice
kubepods-besteffort-pod6a53b255_27f4_446a_8483_e1ea7924c120.slice
kubepods-besteffort-pod6a841de2_b32b_4e8e_b859_375d631522e0.slice
kubepods-besteffort-pod6da43013_28ae_497b_b20c_dd2b4450593a.slice
kubepods-besteffort-pod8e73981f_70d0_469d_b915_5016522f394d.slice
kubepods-besteffort-pod99c628c7_5c22_41e1_9978_87ce11958277.slice
kubepods-besteffort-podb3b13744_ae5a_4b3a_9cff_55f062270e7f.slice
kubepods-besteffort-podecedefff_a01c_4670_9ee9_42d5fda3bf96.slice
```

systemd.unit.discovery[slice]

/var/log/pods

project-d00-dev_web-ts-6594bf8bf8-7k5cz_14a840fc-a1f3-42d7-b405-ec123ad0a706

Namespace

Pod name

Slice suffix

ZABBIX

NPIS IT

Discovery result

```
{
  "data":[
    {
      "{#SLICE_NAME}": "6b1241e9_8a94_4b87_9ab2_5bc68aad2718",
      "{#POD_NAME}": „httpd-jlprs-5886b7c5f6-lzjx2",
      "{#POD_NAMESPACE}": „p00-project-dev”
    }
    ,{
      "{#SLICE_NAME}": "347a3ab8_1807_4579_97aa_98e35bd39111",
      "{#POD_NAME}": „zabbix-rsyslog-agent-79b6f878f8-vmn44",
      "{#POD_NAMESPACE}": „a00-zabbix-test”
    }
    ,{
      "{#SLICE_NAME}": "19ae9143_d351_40a4_aa95_196949c0029e",
      "{#POD_NAME}": "el-helm-pipeline-5b58867866-h5c88",
      "{#POD_NAMESPACE}": "a00-pipelines”
    }
  ].....
}
```



```
graph LR; A[Worker registration] --> B[Pod discovery]; B --> C[Pod lifecycle and data]; D[ZABBIX] --- E[NPIS IT];
```

Worker
registration

Pod
discovery

Pod lifecycle
and data

ZABBIX

NPIS IT

CPU/Memory

```
[Slice]  
MemoryAccounting=yes  
CPUAccounting=yes  
BlockIOAccounting=yes
```

Kubernetes default behaviour

CPU is more than 30% in use by pod maven-java11-wjf78 from namespace b00-jenkins on host

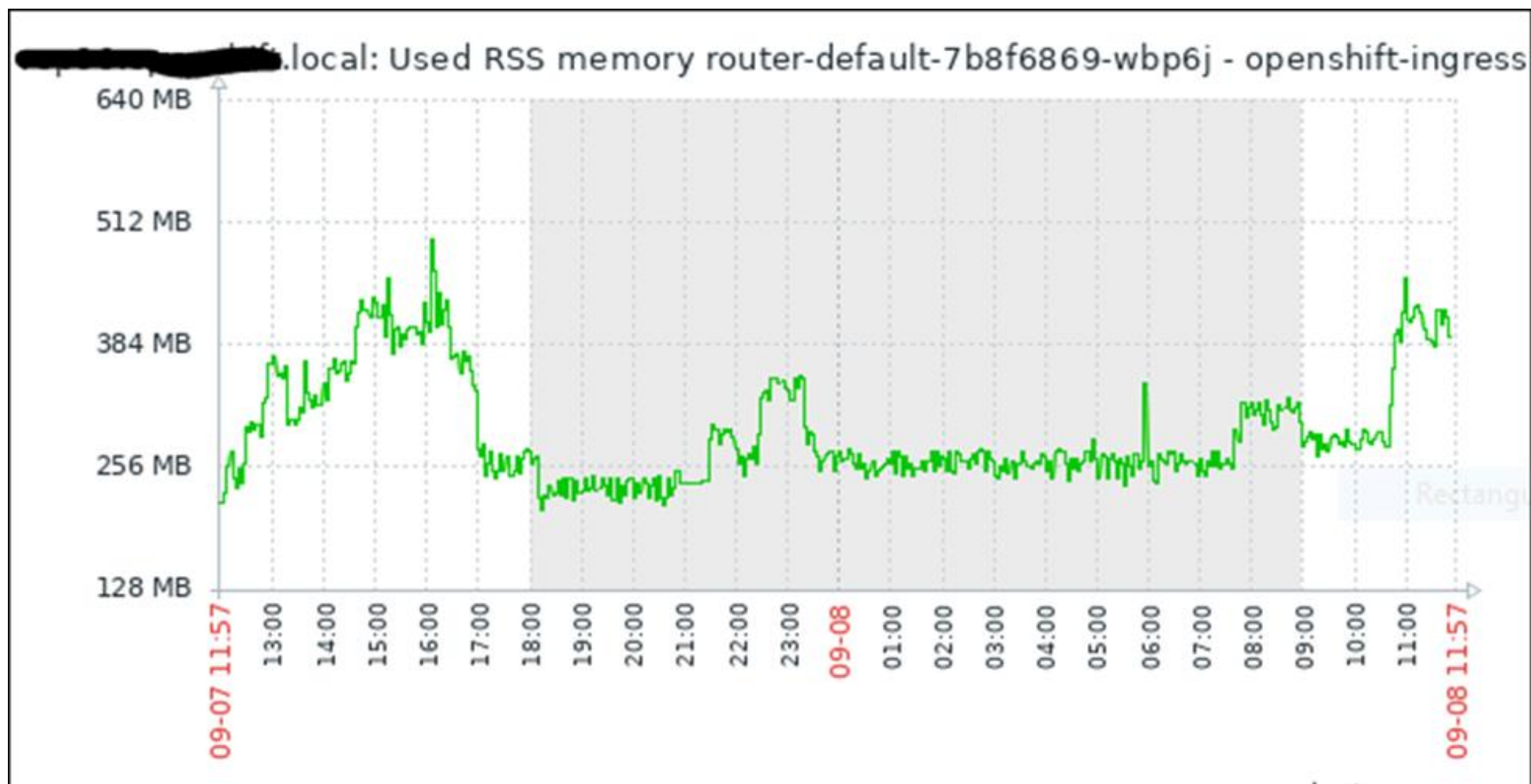
CPU

```
systemd.cgroup.cpu[  
  ../kubepods.slice/kubepods-burstable.slice/kubepods-burstable-pod{#SLICE_NAME}.slice  
  ,total  
]
```

Memory

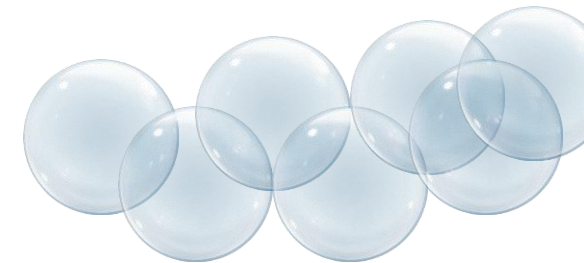
```
systemd.cgroup.mem[  
  ../kubepods.slice/kubepods-burstable.slice/kubepods-burstable-pod{#SLICE_NAME}.slice  
  ,total_rss  
]
```

ZABBIX



ZABBIX

NPIS IT



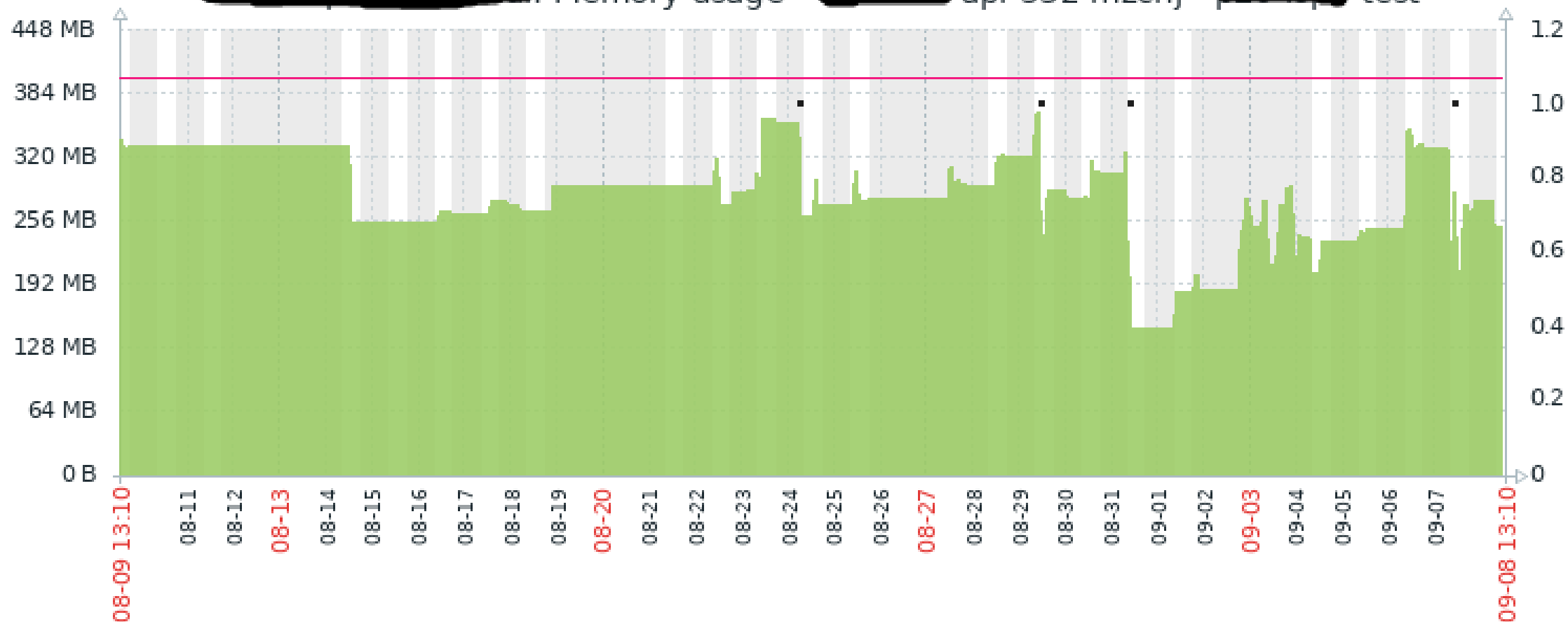
Memory limit

```
- resources:
  limits:
    cpu: 697m
    memory: 205Mi
```

- Pods can be limited = Cgroup
- If they overreach over the limit, they are killed OOM (burst)
- Memory limit can be read from a file

```
vfs.file.contents[
  _kubepods.slice/kubepods-burstable.slice/kubepods-burstable-pod{#SLICE_NAME}.slice/memory.limit_in_bytes
]
```

Shift 1: Memory usage - api-332-mzcnj - test



	last	min	avg	max
Used RSS memory api-332-mzcnj - test	249.25 MB	46.7 MB	278.41 MB	381.37 MB
Memory limit api-332-mzcnj - test	400 MB	400 MB	400 MB	400 MB
OOM pod killed - api-332-mzcnj - test	1	1	1	1

Trigger

- If memory is getting close to the memory limit, trigger will activate, warning that application might be killed by the OS
- `last(systemd.cgroup.mem[SLICE, total_rss]) > last(vfs.file.contents(mem_limit) * 0.9)`

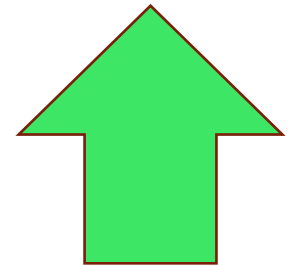
Worker
registration



Pod
discovery



Pod lifecycle
and data

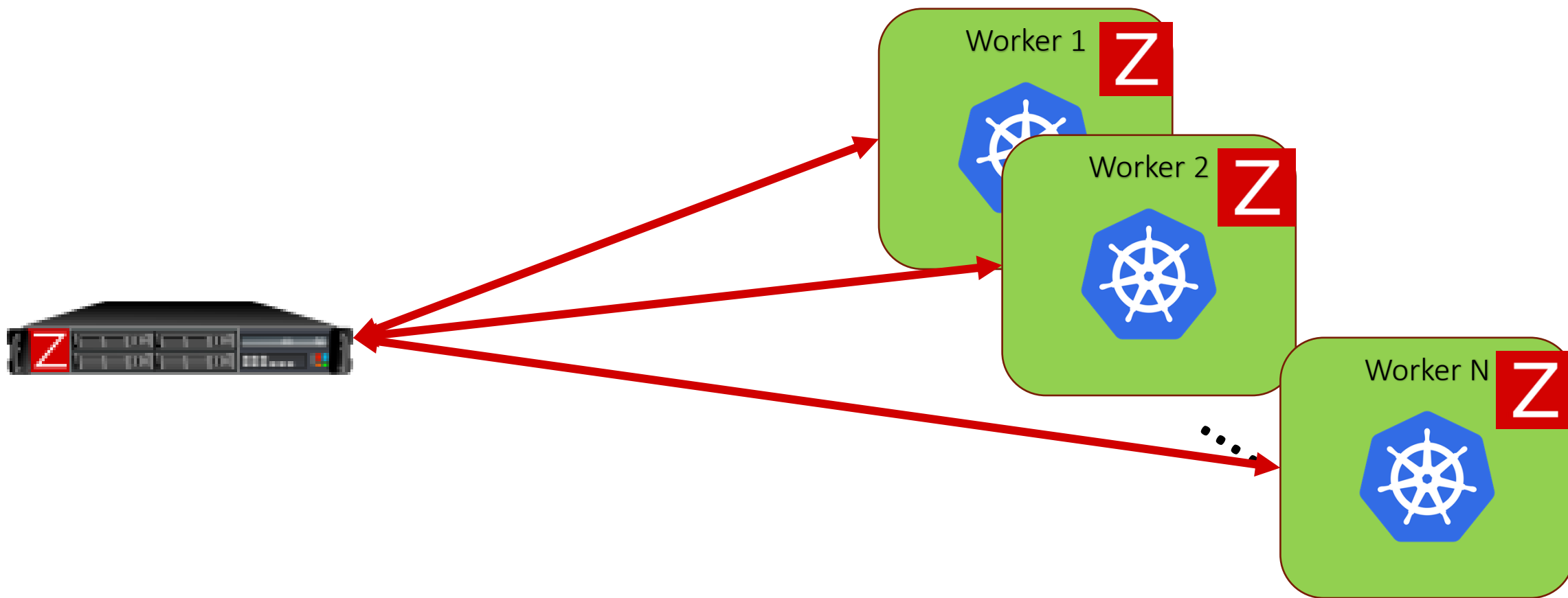


ZABBIX

Pod lifecycle

- Crashloop, liveness, readiness probe, OOM etc

P	web-6f88ffc758-s6lrz	! CrashLoopBackOff	0/1	5116	⋮
<hr/>					
P	api-view-59ff48f85f-c94h9				
Generated from kubelet on va-13-pensmrt .local					
Readiness probe warning: Probe terminated redirects, Response body:					



ZABBIX

Pod lifecycle

- Crashloop, liveness, readiness probe, OOM etc
- How does perform liveness probe, crashloop, OOM ...?

kubelet.service
kernel logs

"Probe failed" probeType="Liveness" podUID=f441e218-083d-42f3-be2b-708cbf2727f7

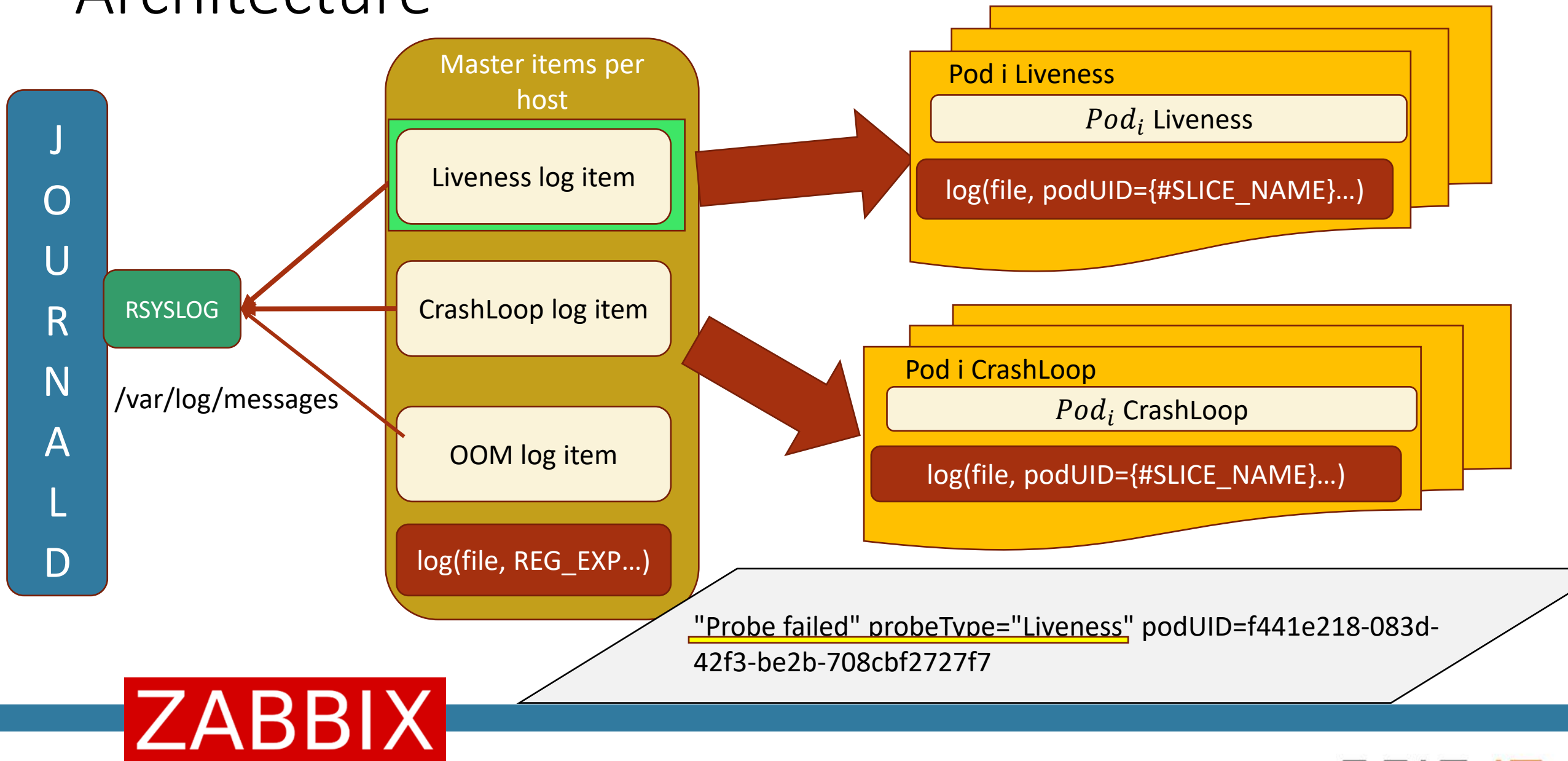
Error syncing pod, skipping" err="failed to "StartContainer" for c1 with CrashLoopBackOff:
back-off 2m40s restarting failed container=c1 podUID=f08b29c7-1fef-4472-97b5-
3dfc37e48569

oom-kill:constraint=CONSTRAINT_MEMCG,nodemask=(null),
oom_memcg=/kubepods.slice/kubepods-burstable.slice/kubepods-burstable-
pod7890bf4c_22ab_4900_988b_e7528d65e4b8.slice
,task=mongod,pid=3795709,uid=1004020000

Architecture

Pod discovery rule

INSTANT NOTIFICATION

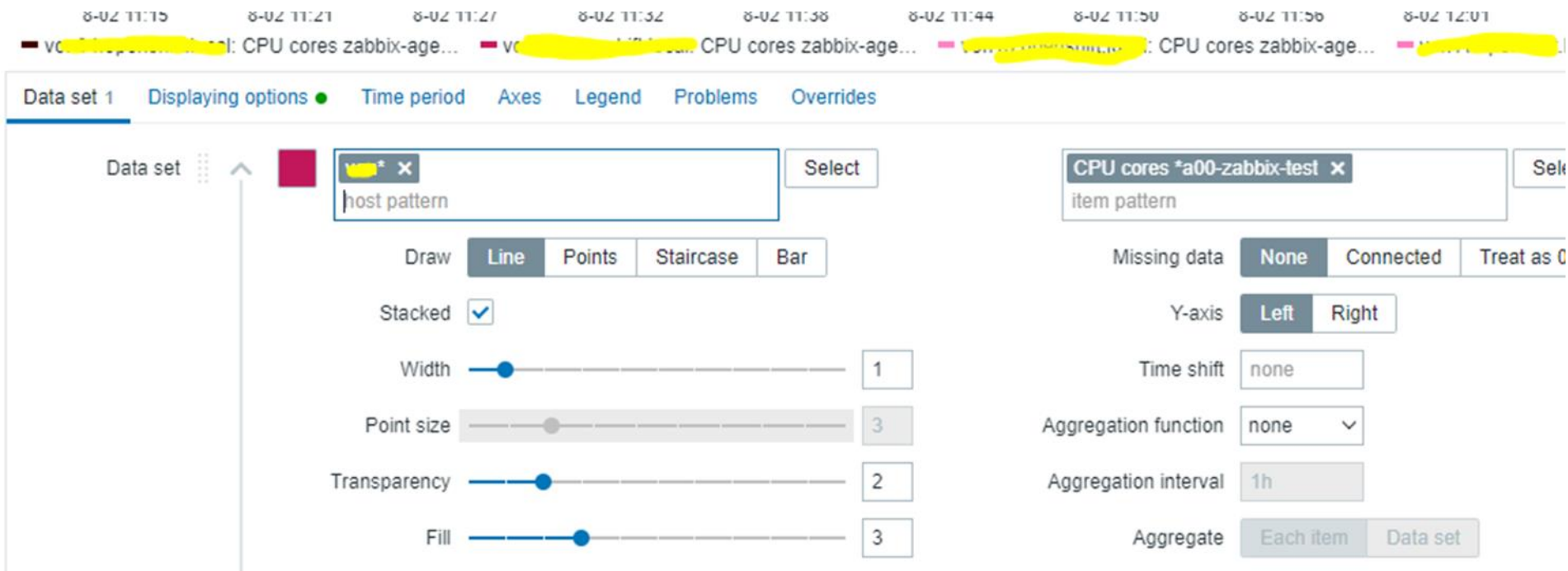


Graphs

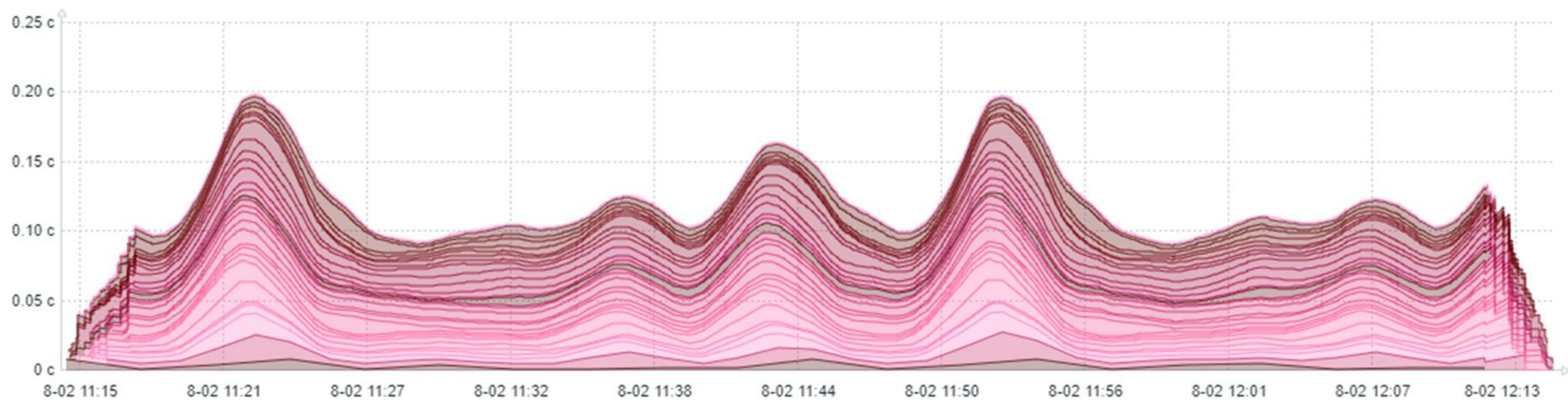
Item prototypes

All templates / Openshift nodes - OCP4 Discovery list / Openshift besteffort pods

<input type="checkbox"/>	Name ▲
<input type="checkbox"/>	... CPU cores {#POD_NAME} - {#POD_NAMESPACE}
<input type="checkbox"/>	... CPU utilization {#POD_NAME} - {#POD_NAMESPACE}
<input type="checkbox"/>	... <u>CrashLoopBackOff</u> : CrashLoopBackOff detected- {#POD_NAME} - {#POD_NAMESPACE}
<input type="checkbox"/>	... <u>Liveness probe failed</u> : Liveness probe failed - {#POD_NAME} - {#POD_NAMESPACE}
<input type="checkbox"/>	... <u>OOM kill</u> : OOM pod killed - {#POD_NAME} - {#POD_NAMESPACE}
<input type="checkbox"/>	... Used RSS memory {#POD_NAME} - {#POD_NAMESPACE}



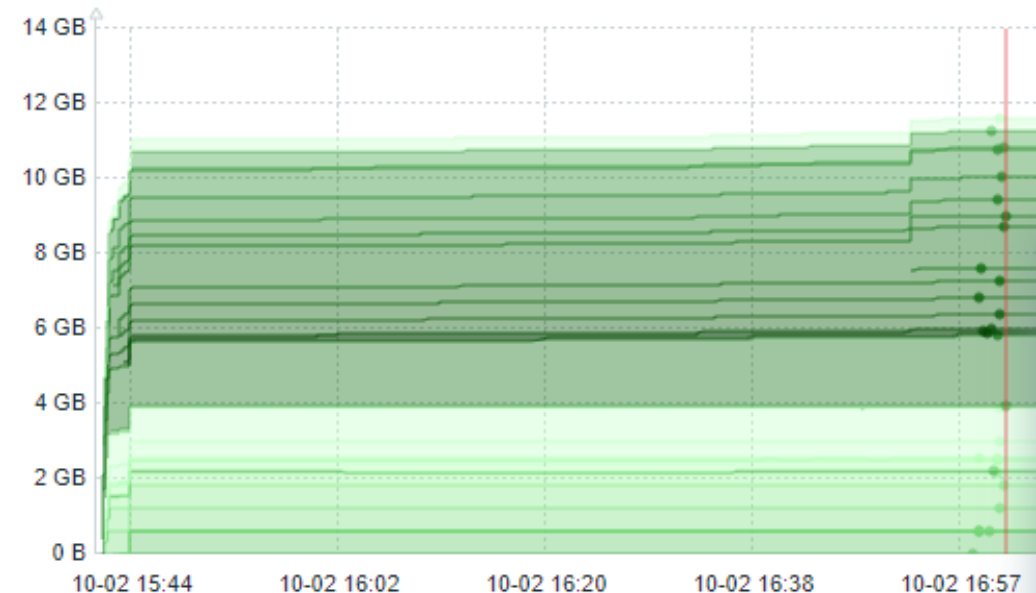
ZABBIX



ZABBIX

NPIS IT

Graph



von16-openshift.local: Used RS... von16-openshift.local: Used RS... von16-open

2023-10-02 17:01:00

- Used RSS memory business-67f67566f4-rcnrf - 184-184-184-184-test: 1.88 GB
- Used RSS memory business-0 - b52-184-184-test: 1.1 GB
- Used RSS memory k8sensor-586bdb4c84-gx2rt - instana-agent: 870.64 MB
- Used RSS memory cci-view-5dbcbf5fb5-jz458 - b49-184-184-test: 762.74 MB
- Used RSS memory h15i-view-cd85bb6d9-x86hs - b46-184-184-test: 646.22 MB
- Used RSS memory evjencanja-gradjani-7959658ccc-b27pk - m48-184-184-test: 635.19 MB
- Used RSS memory h15-business-0 - b46-184-184-test: 631.82 MB
- Used RSS memory cci-validator-65567b4699-q9qwf - b46-184-184-test: 609.11 MB
- Used RSS memory job-manager-787dcff984-757lr - ibm-backup-restore: 477.32 MB
- Used RSS memory core-6b857d697b-9tmd2 - 477-184-184-test: 456.41 MB
- Used RSS memory backup-location-deployment-649f7c7dd7-bdspm - ibm-backup-restore: 456.4 MB
- Used RSS memory backend-cms-76b6d4658b-gs6m5 - 424-184-184-test: 453.67 MB
- Used RSS memory business-portal-7bf855f89c-4mvv8 - 477-184-184-test: 424.27 MB
- Used RSS memory baas-transaction-manager-7d65b45c4b-fhgjr - baas: 387.32 MB
- Used RSS memory batch-265-ksnqf - b54-184-184-test: 352.31 MB
- Used RSS memory korisnik-admin-public-55cbd7bc5-grk6t - m88-184-184-test: 344.07 MB
- Used RSS memory business-5b9d69d9c7-nw8hw - f20-184-184-test: 307.65 MB
- Used RSS memory h6-view-99-kvjzh - b46-184-184-test: 305.03 MB
- Used RSS memory h7-view-98-gmbdp - b46-184-184-test: 297.28 MB
- Used RSS memory api-eusluga-77-4lncz - m37-184-184-test: 91.03 MB

Displaying 20 of 31 found

ZABBIX

NPIS IT

Sum CPU/Memory per Namespace

ZBXNEXT-6452

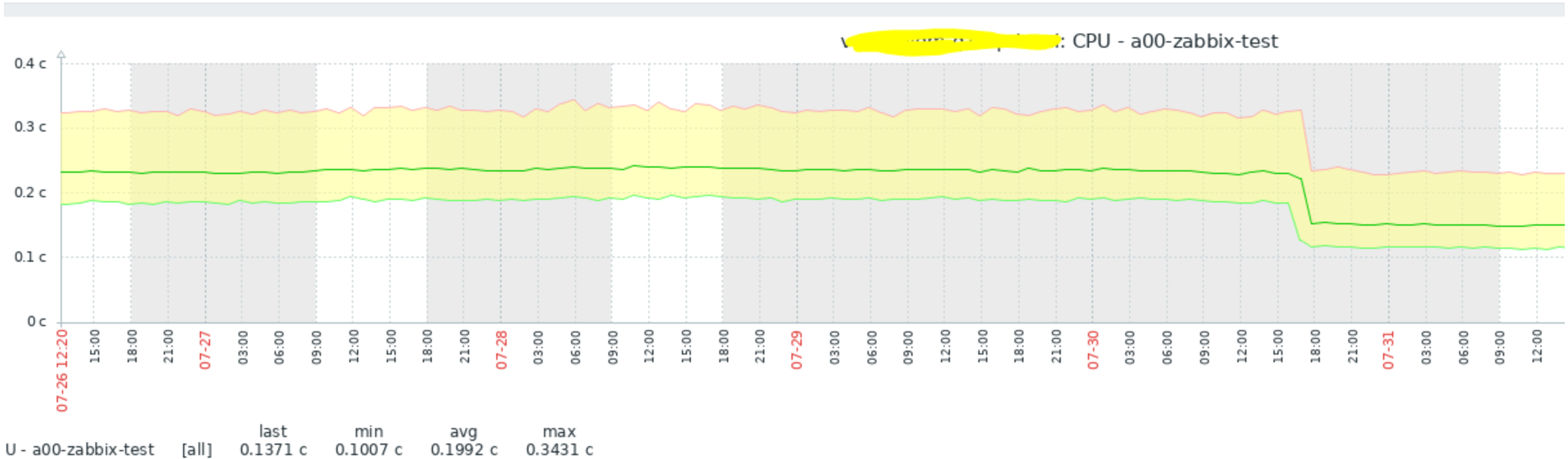
Item prototypes

All templates / Openshift nodes - OCP4		Discovery list / Openshift besteffort pods
<input type="checkbox"/>	Name ▲	
<input type="checkbox"/>	...	CPU cores {#POD_NAME} - {#POD_NAMESPACE}
<input type="checkbox"/>	...	CPU utilization {#POD_NAME} - {#POD_NAMESPACE}
<input type="checkbox"/>	...	CrashLoopBackOff : CrashLoopBackOff detected- {#POD_NAME} - {#POD_NAMESPACE}
<input type="checkbox"/>	...	Liveness probe failed : Liveness probe failed - {#POD_NAME} - {#POD_NAMESPACE}
<input type="checkbox"/>	...	OOM kill : OOM pod killed - {#POD_NAME} - {#POD_NAMESPACE}
<input type="checkbox"/>	...	Used RSS memory {#POD_NAME} - {#POD_NAMESPACE}

Itemprototype tag

Application: {#POD_NAMESPACE}

```
sum(
  last_foreach(
    /*/openshift.cores[*,total]?[(
      group={$KUBERNETES_NONPROD_GROUP}
      and tag="Application:CPU"
      and tag="Application:a00-zabbix-test")])
  )
)
```





THANK YOU

ZABBIX

ZABBIX

NPIS IT

Appendix

```
UserParameter=openshift.pods.discovery.besteffort, zabbix_agentd -t "systemd.unit.discovery[slice]" 2>/dev/null |  
tail -n1 | sed 's/^systemd.unit.discovery\[slice\]\s*\[s|\(.*\)\]/\1/g' | jq -r '.[0].data[] | .{"#UNIT.NAME}" | grep  
kubepods-besteffort-p | sed "s/^kubepods-besteffort-pod\(.*\).slice/\1/" | while read slice; do echo $slice,$(echo  
$slice | sed "s/_/-/g" | while read dir; do ls /host/var/log/pods | grep "$dir" | head -n1 | sed  
"s/^\/host\/var\/log\/pods\////g" | sed "s/^\[A-Za-z0-9-\]*\)_\[A-Za-z0-9-\]*\).*\/2,\1/g"; done ); done | awk -F','  
'BEGIN {print "{ \"data\":["} NR>1{printf ","} {printf "{ \"#SLICE_NAME\": \"%s\", \"#POD_NAME\":  
\"%s\", \"#POD_NAMESPACE\": \"%s\"}"\n", $1, $2, $3 } END {print "]}"}
```

```
UserParameter=openshift.pods.discovery.burstable, zabbix_agentd -t "systemd.unit.discovery[slice]" 2>/dev/null | tail  
-n1 | sed 's/^systemd.unit.discovery\[slice\]\s*\[s|\(.*\)\]/\1/g' | jq -r '.[0].data[] | .{"#UNIT.NAME}" | grep  
kubepods-burstable-p | sed "s/^kubepods-burstable-pod\(.*\).slice/\1/" | while read slice; do echo $slice,$(echo  
$slice | sed "s/_/-/g" | while read dir; do ls /host/var/log/pods | grep "$dir" | head -n1 | sed  
"s/^\/host\/var\/log\/pods\////g" | sed "s/^\[A-Za-z0-9-\]*\)_\[A-Za-z0-9-\]*\).*\/2,\1/g"; done ); done | awk -F','  
'BEGIN {print "{ \"data\":["} NR>1{printf ","} {printf "{ \"#SLICE_NAME\": \"%s\", \"#POD_NAME\":  
\"%s\", \"#POD_NAMESPACE\": \"%s\"}"\n", $1, $2, $3 } END {print "]}"}
```