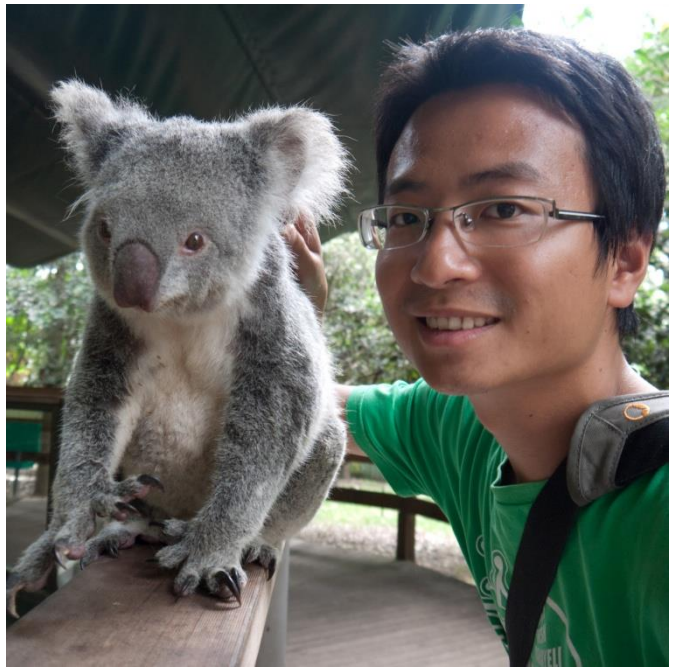
The Art of Container Monitoring

Derek Chen

2016.9.22

About me



- DevOps Engineer at Trend Micro
  - Agile transformation
  - Micro service and cloud service
  - Docker integration
  - Monitoring system development
- Automate all the things
- Make everything smoother
- Find me at derekhound@gmail.com

Why monitoring?

- We want to know when things go wrong
- We want to know when things aren't quite right
- We want to know in advance of problems

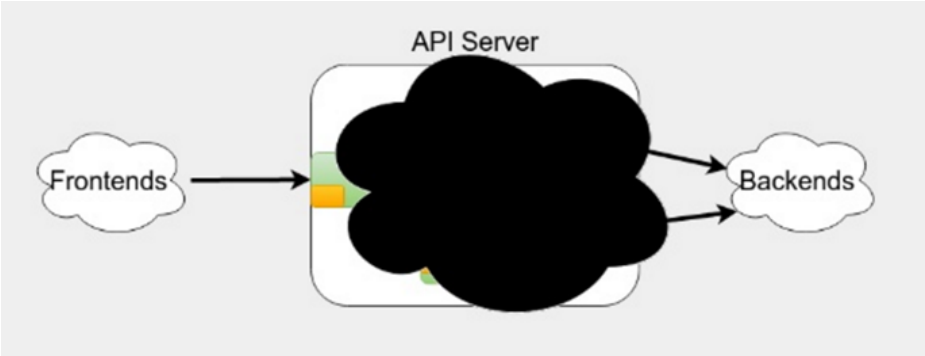
Microscope?
Magnifier?
Telescope?

Blackbox vs Whitebox

- Blackbox
  - Requires no participation of the monitored system
  - Observes external functionality, “what the user see”
  - End to end test
- Whitebox
  - Collects data internally provided by the target system
  - Has more granular information about the system
  - Can provide warning of problems before they occur

Blackbox tests measure…

- Can you ping the server
- Can you fetch a page from the server
- Does it have the correct contents

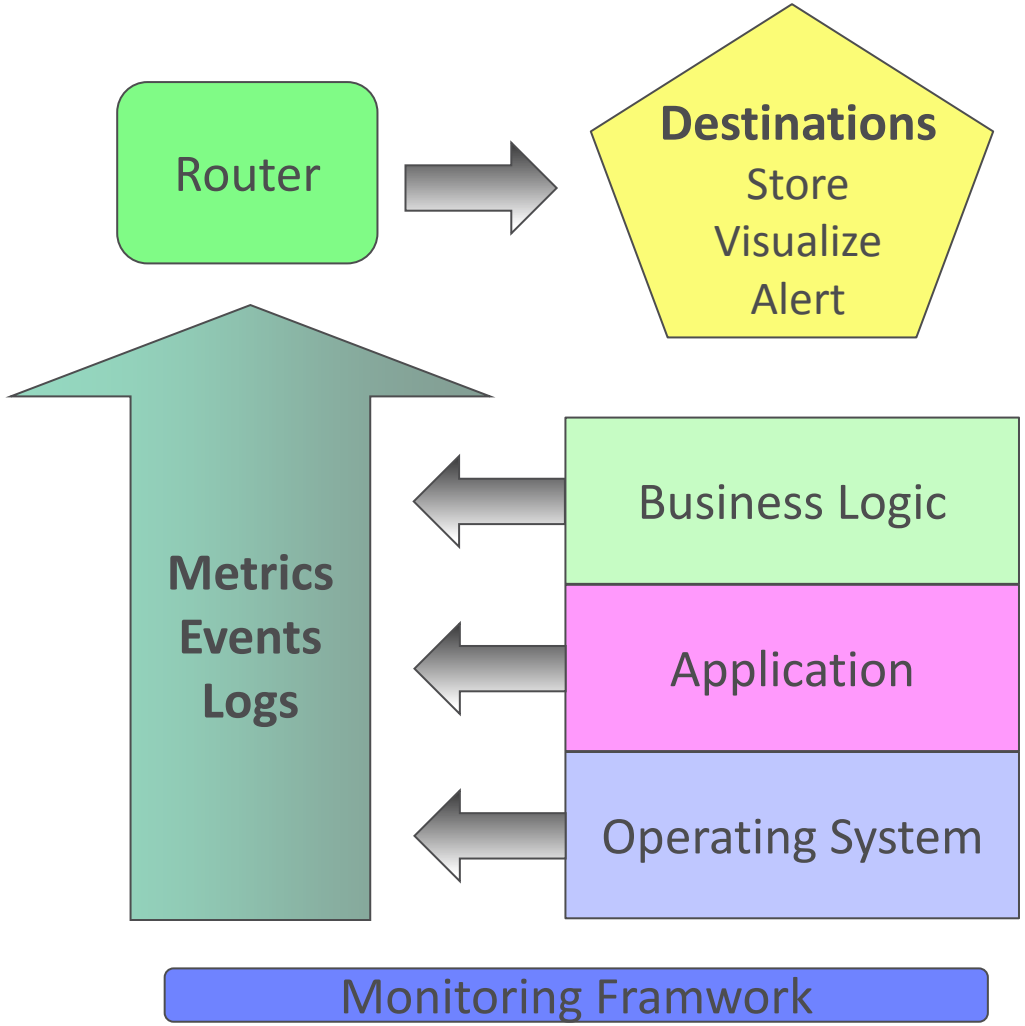


Whitebox monitoring collects…

- Network statistics (packets/bytes sent/received)
- System load (cpu, memory, disk usage)
- Application statistics (connection count, query per second)

Goal

- Be metric, event, and log
- Allow us to easily visualize the stat of our environment
- Provide contextual and useful notification.
- Focus on whitebox monitoring instead of blackbox monitoring



[image] https://www.artofmonitoring.com

Why Open Source Software Monitoring?

- The practice of system admin is changing quickly
  - DevOps
  - Infrastructure as Code
  - Visualization
  - Cloud Platform

- Commercial solution can’t keep up!

[image] http://devops.com/2014/05/05/meet-infrastructure-code

Monolithic

- If all-in-one gets the job done, then great
- Good for smaller scale, non-tech-focused companies

Modular

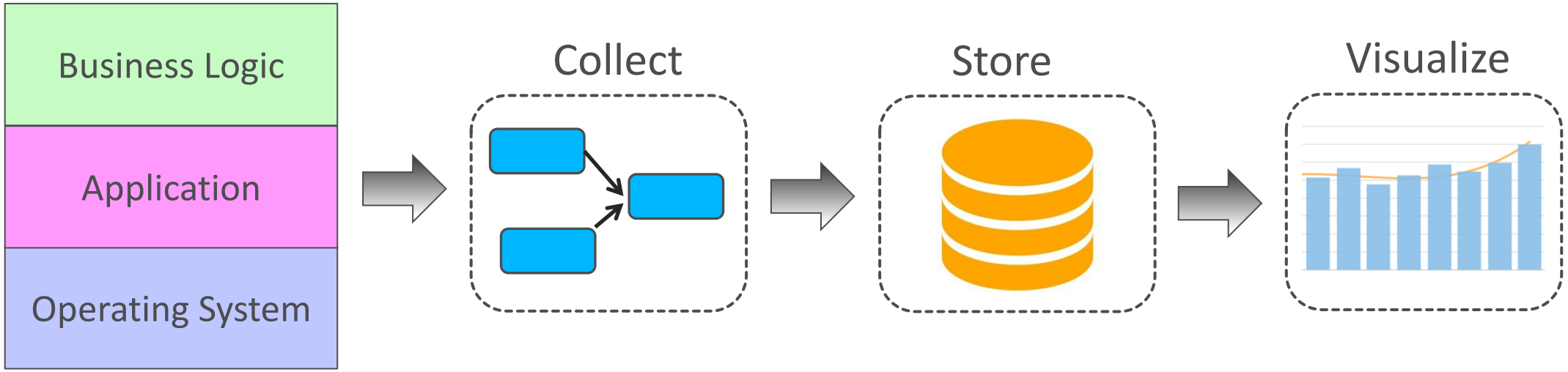
- DevOps requires flexibility and innovation
- Good for tech driven and ops-focused companies

Metric

We’ll rely most heavily on metrics to help us understand what’s going on in our environment.

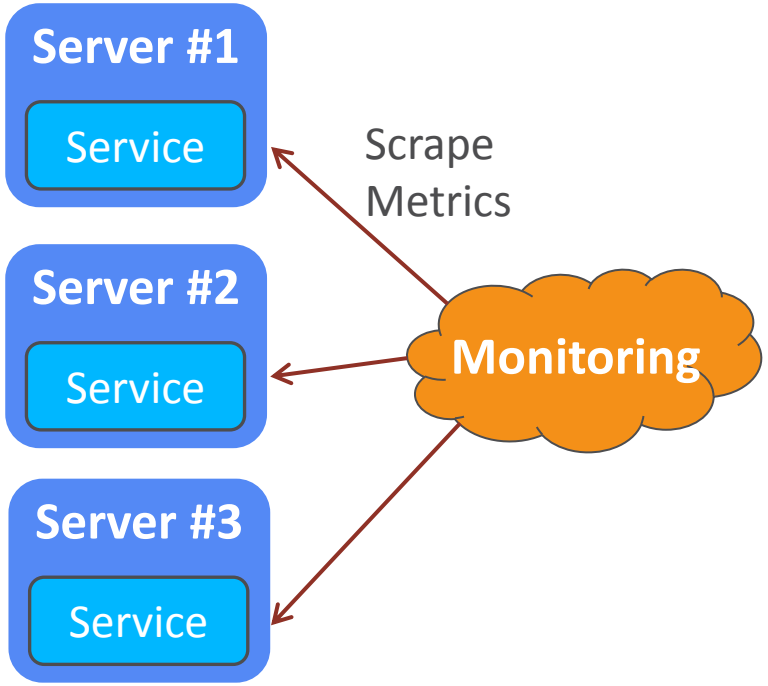
Metric

- Provide a dynamic, real-time picture of the state of your infrastructure



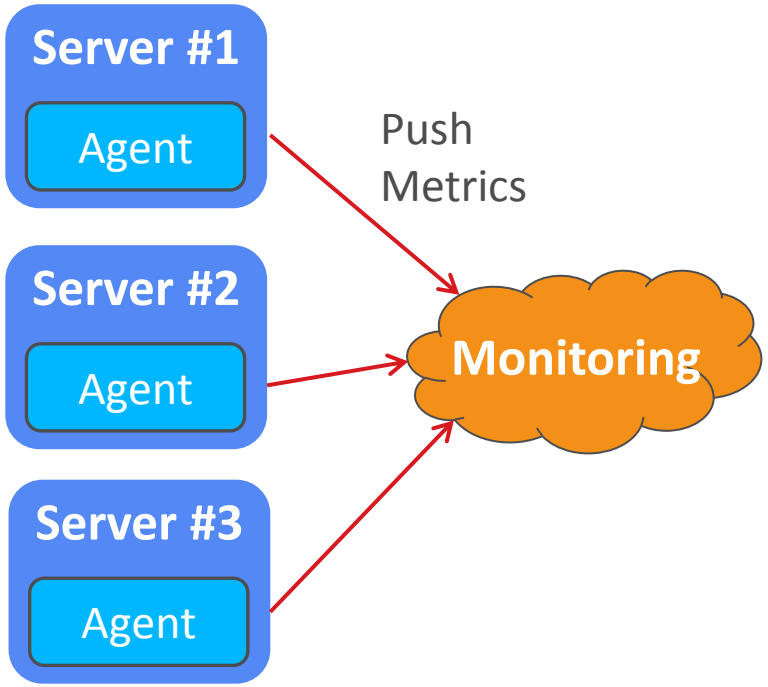
Pull Mode

A central collector periodically requests metrics from each monitored system



Push Mode

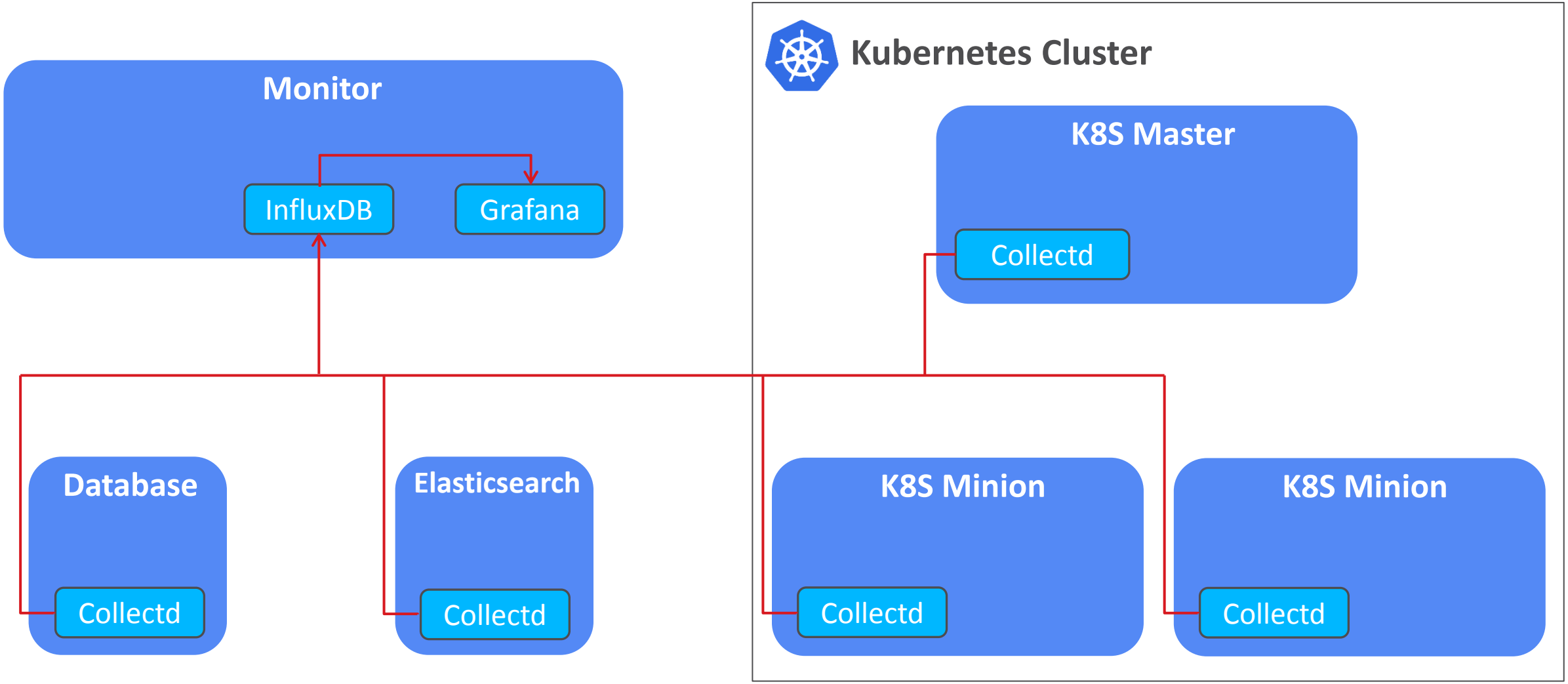
Metrics are periodically sent by each monitored system to a central collector



Collectd

The system statistics collection daemon

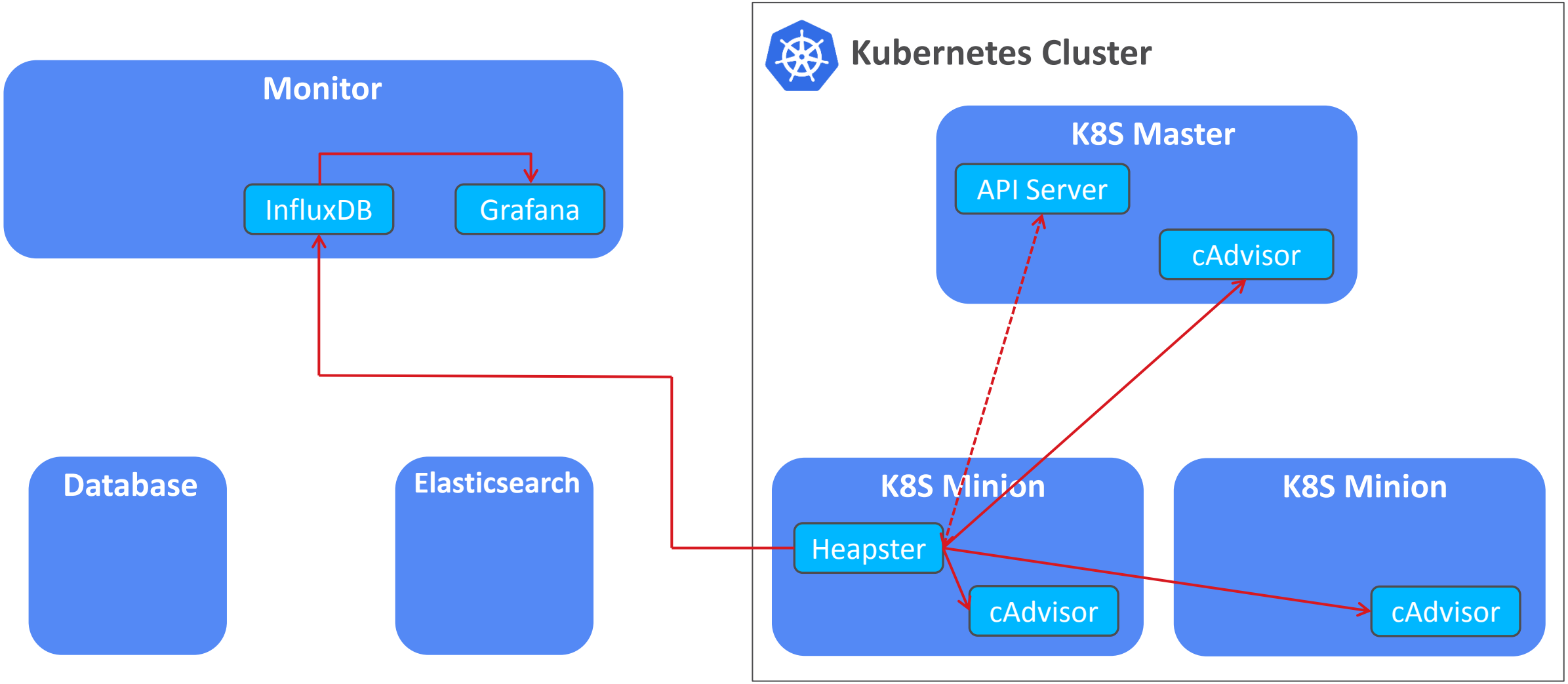
Host Machine OS Metric



cAdvisor

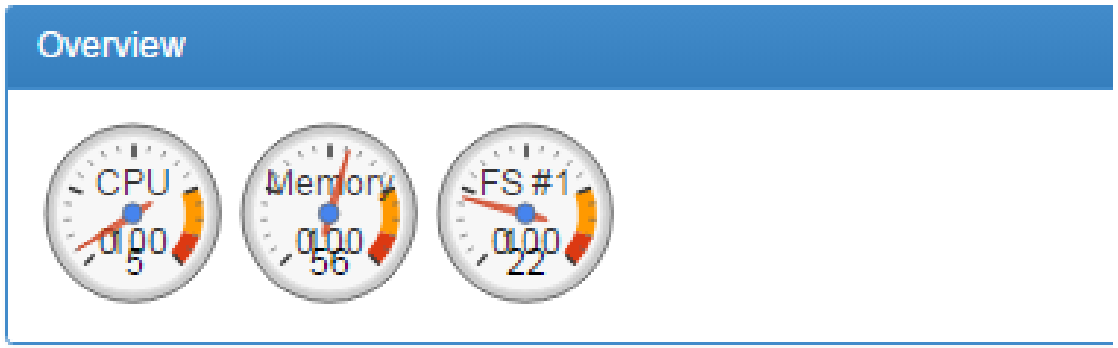
Collects, aggregates, processes, and exports information about running containers.

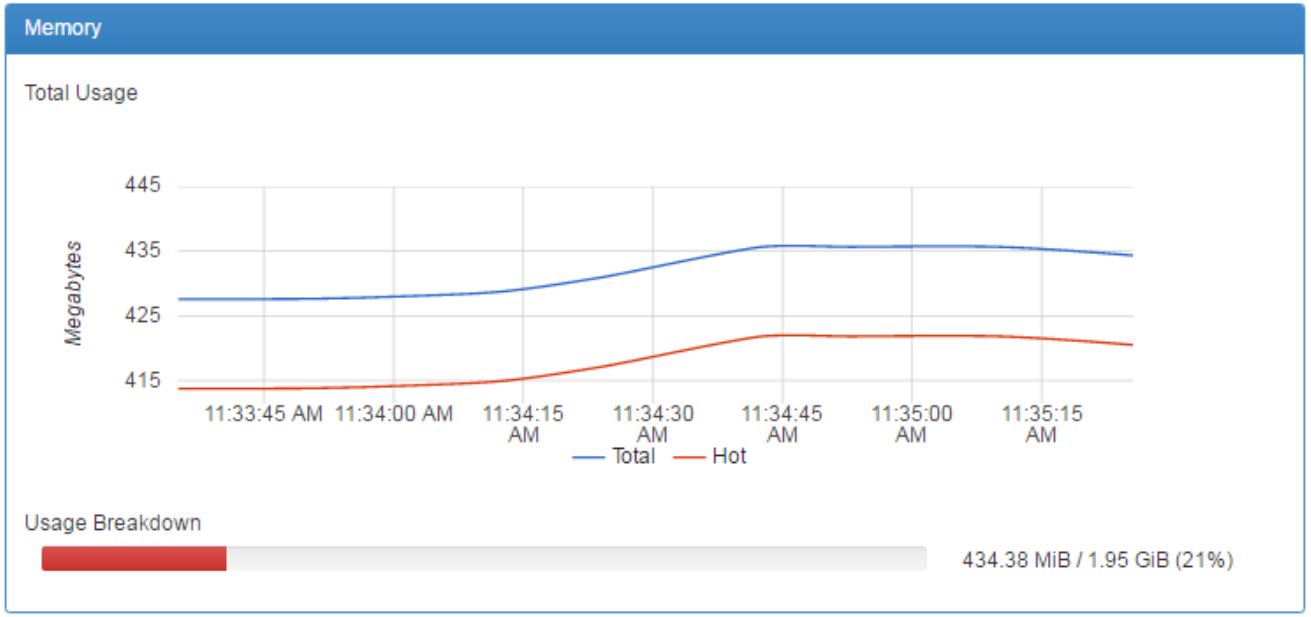
Container OS Metric

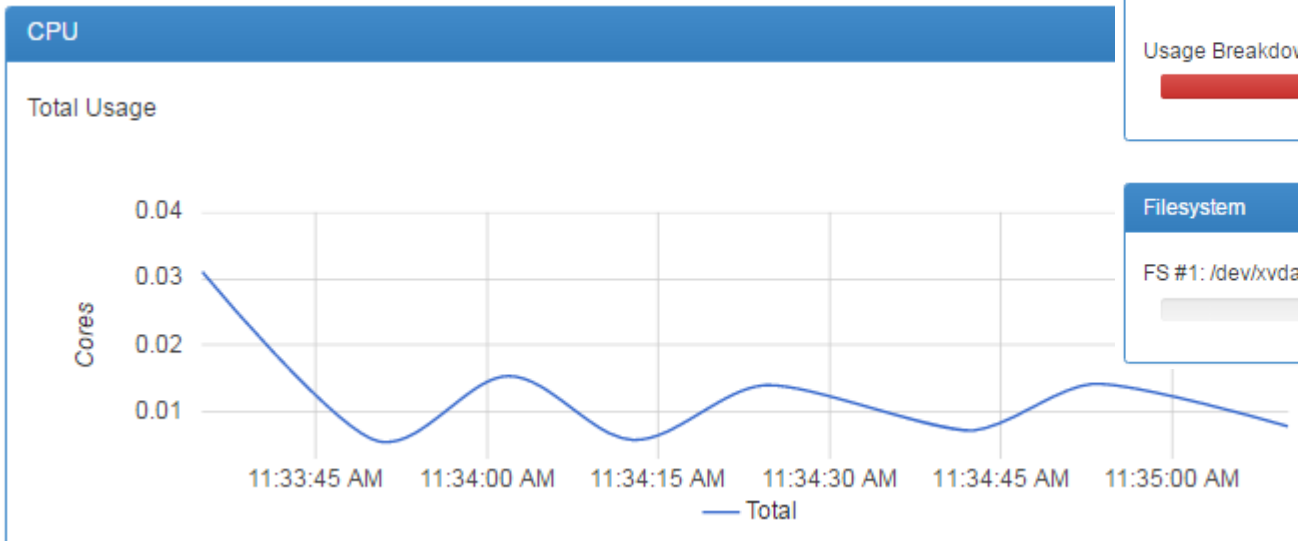


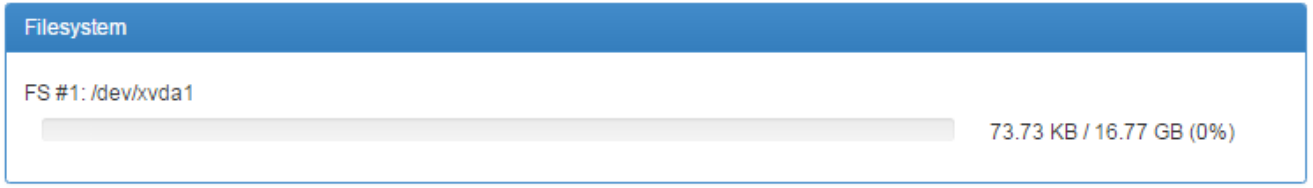
cAdvisor

http://127.0.0.1:4194
http://127.0.0.1:4194/metrics





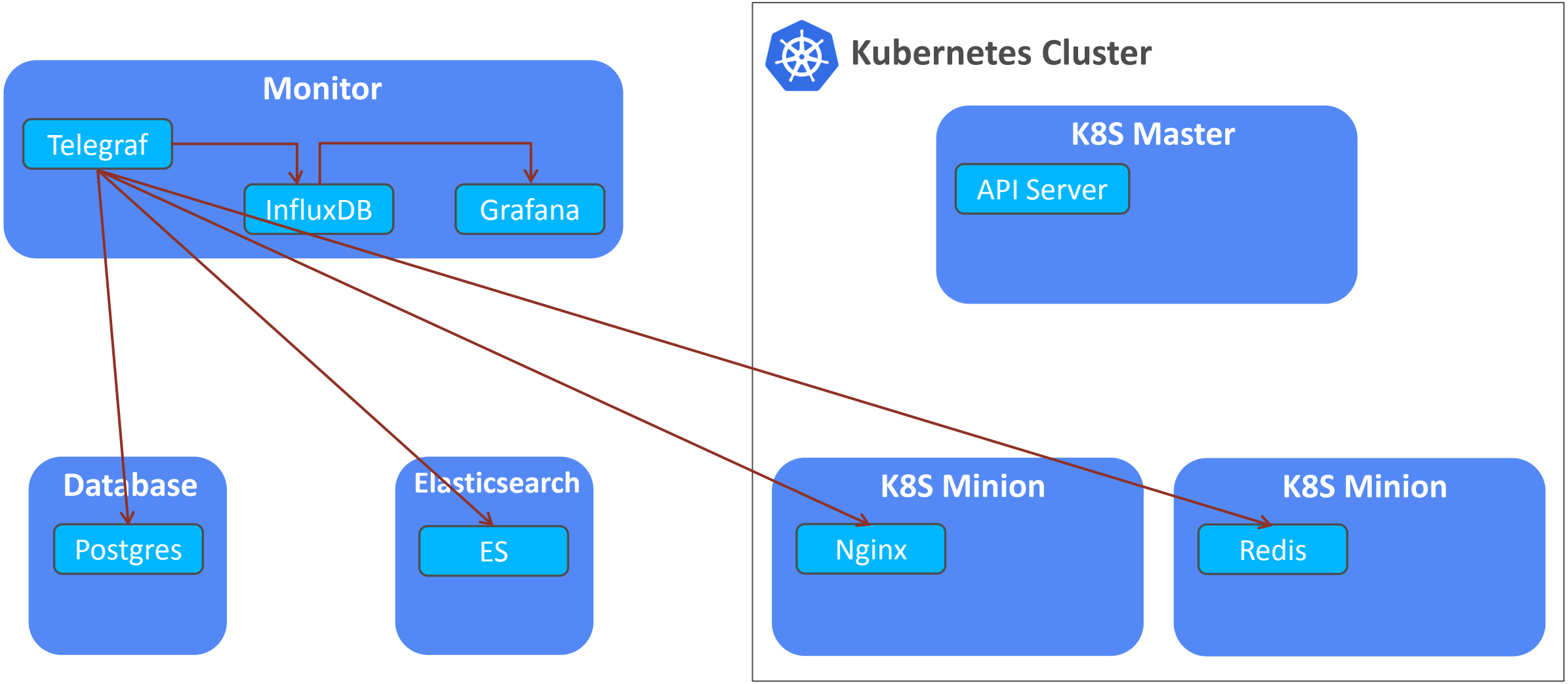




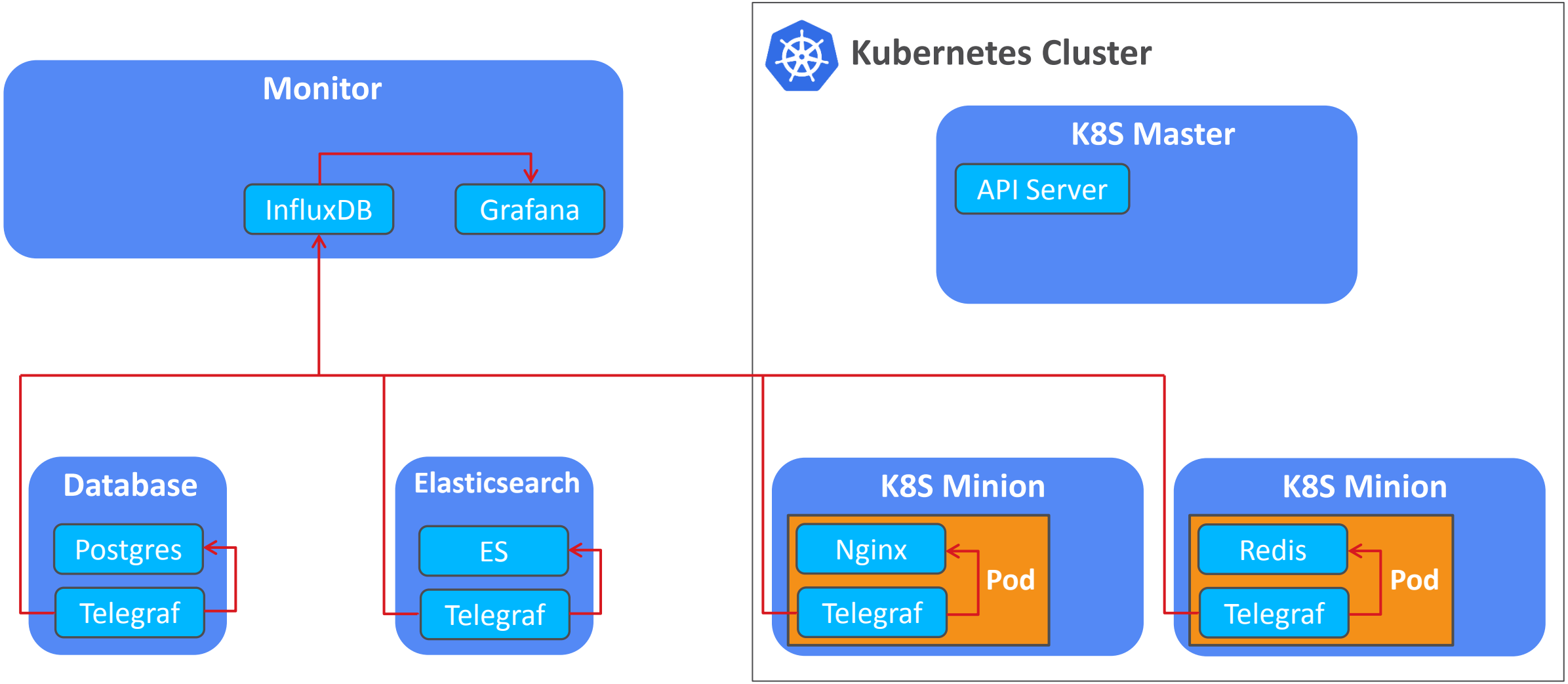
Telegraf

Collects time series data from a variety of sources

Application Metric – Pull Mode



Application Metric – Push Mode

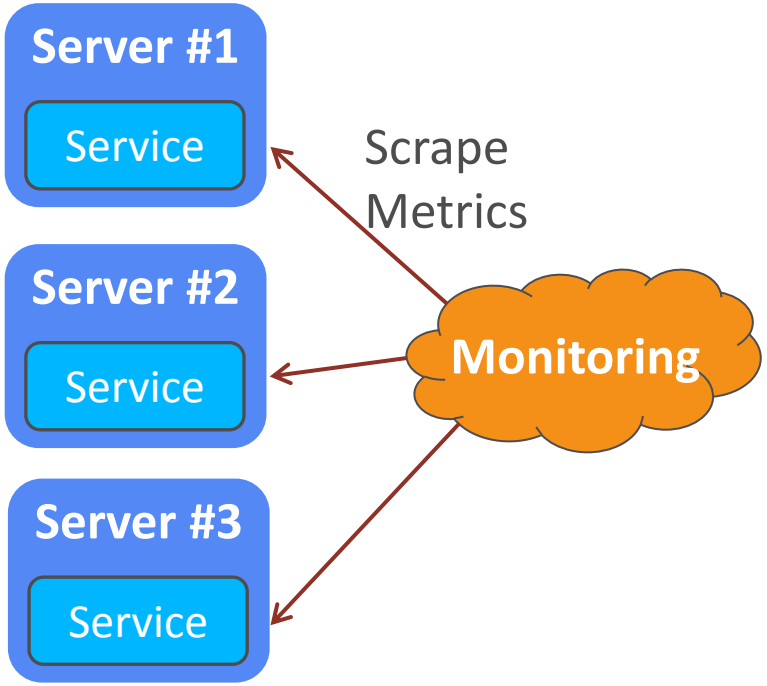


Telegraf

- An agent written in Go
- Easy to contribute or develop your plugins
- Supported Input Plugins
  - System: cpu, mem, net, disk, processes and etc.
  - Application: docker, nginx, postgresql, redis, and etc.
  - Third party api: aws cloudwatch
- Supported output plugins
  - influxdb, graphite, cloudwatch, datadog and etc.

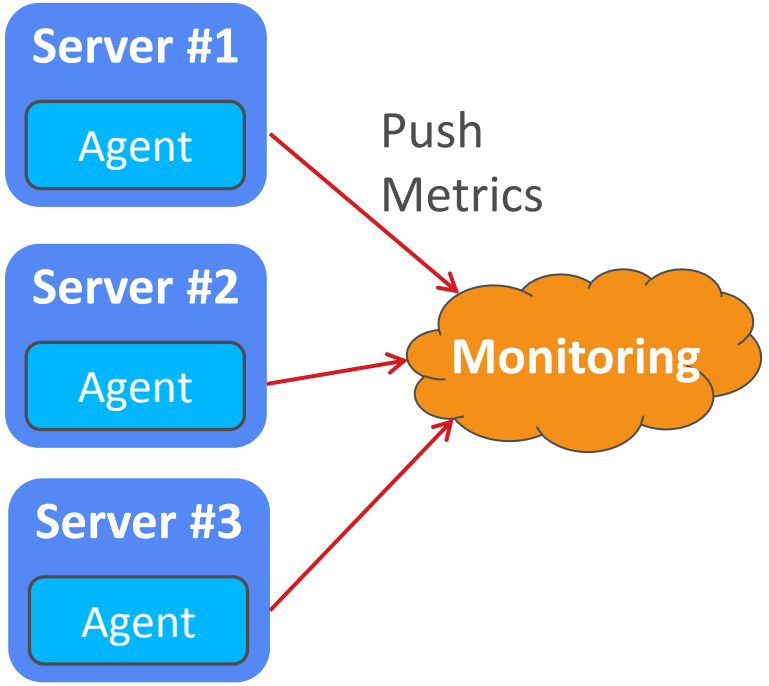
Pull Mode

- Collector discoveries services periodically
- Collector needs to talk to target services
  - Overlay network. Ex: Flannel, Weave, etc.
  - Forward proxy



Push Mode

- Agent sends metrics as soon as it starts up
- Suitable for short-lived containers or dynamic environments



InfluxDB

Time series database stores all the metrics

InfluxDB

- Similar scope to Graphite
- Written in Go
- No external dependency
- Ready for billions row
- Several client libraries
- SQL style queries

InfluxDB Schema

InfluxDB Write Data Documentation Database: telegraf

Query: select * from nginx order by time desc limit 3;

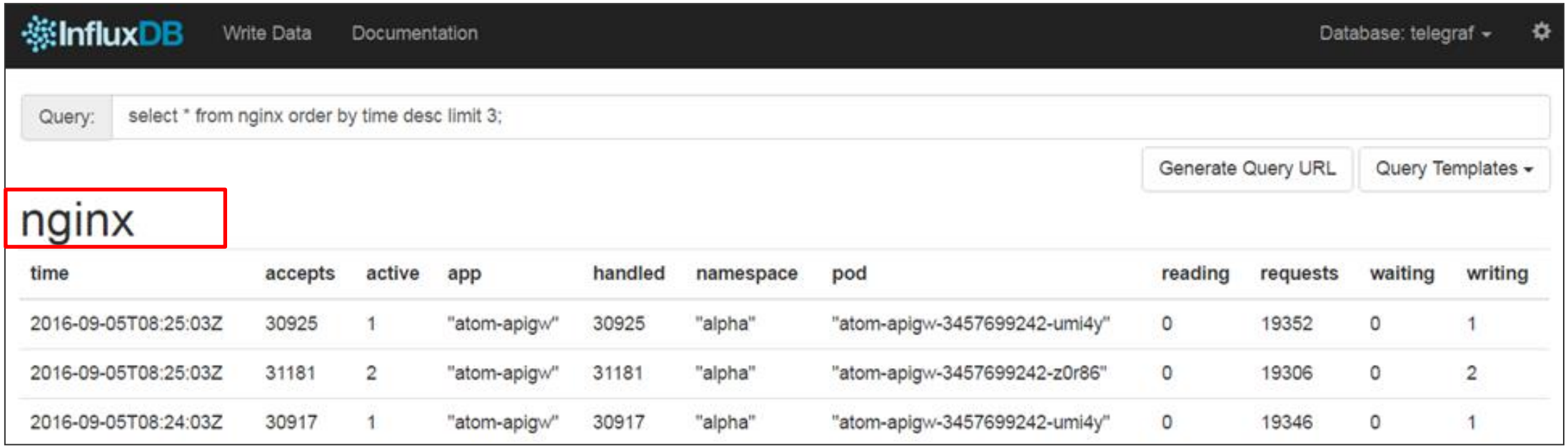
Generate Query URL Query Templates

nginx

time	accepts	active	app	handled	namespace	pod	reading	requests	waiting	writing
2016-09-05T08:25:03Z	30925	1	"atom-apigw"	30925	"alpha"	"atom-apigw-3457699242-umi4y"	0	19352	0	1
2016-09-05T08:25:03Z	31181	2	"atom-apigw"	31181	"alpha"	"atom-apigw-3457699242-z0r86"	0	19306	0	2
2016-09-05T08:24:03Z	30917	1	"atom-apigw"	30917	"alpha"	"atom-apigw-3457699242-umi4y"	0	19346	0	1

- Measurements (e.g. cpu, mem, disk, net, nginx)
- Timestamp (nano second)
- Tags (e.g. app=atom-apigw namespace=alpha)
- Fields (e.g. accepts=30925, active=1, handled=30925)

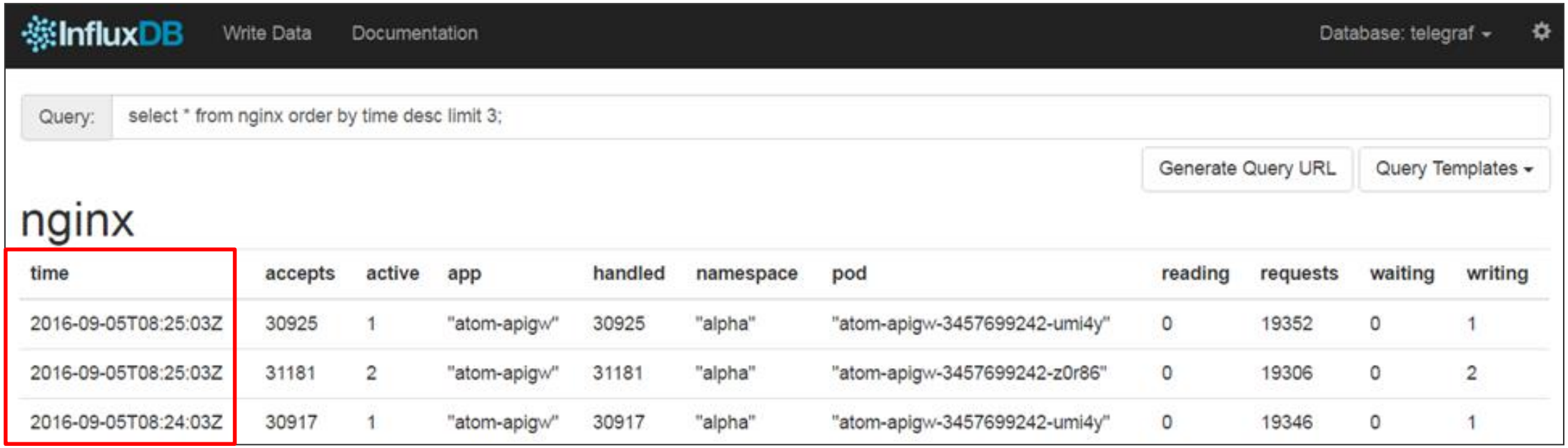
InfluxDB Schema



time	accepts	active	app	handled	namespace	pod	reading	requests	waiting	writing
2016-09-05T08:25:03Z	30925	1	"atom-apigw"	30925	"alpha"	"atom-apigw-3457699242-umi4y"	0	19352	0	1
2016-09-05T08:25:03Z	31181	2	"atom-apigw"	31181	"alpha"	"atom-apigw-3457699242-z0r86"	0	19306	0	2
2016-09-05T08:24:03Z	30917	1	"atom-apigw"	30917	"alpha"	"atom-apigw-3457699242-umi4y"	0	19346	0	1

- Measurements (e.g. cpu, mem, disk, net, nginx)
- Timestamp (nano second)
- Tags (e.g. app=atom-apigw namespace=alpha)
- Fields (e.g. accepts=30925, active=1, handled=30925)

InfluxDB Schema



time	accepts	active	app	handled	namespace	pod	reading	requests	waiting	writing
2016-09-05T08:25:03Z	30925	1	"atom-apigw"	30925	"alpha"	"atom-apigw-3457699242-umi4y"	0	19352	0	1
2016-09-05T08:25:03Z	31181	2	"atom-apigw"	31181	"alpha"	"atom-apigw-3457699242-z0r86"	0	19306	0	2
2016-09-05T08:24:03Z	30917	1	"atom-apigw"	30917	"alpha"	"atom-apigw-3457699242-umi4y"	0	19346	0	1

- Measurements (e.g. cpu, mem, disk, net, nginx)
- Timestamp (nano second)
- Tags (e.g. app=atom-apigw namespace=alpha)
- Fields (e.g. accepts=30925, active=1, handled=30925)

InfluxDB Schema

InfluxDB Write Data Documentation Database: telegraf

Query: select * from nginx order by time desc limit 3;

Generate Query URL Query Templates

nginx

time	accepts	active	app	handled	namespace	pod	reading	requests	waiting	writing
2016-09-05T08:25:03Z	30925	1	"atom-apigw"	30925	"alpha"	"atom-apigw-3457699242-umi4y"	0	19352	0	1
2016-09-05T08:25:03Z	31181	2	"atom-apigw"	31181	"alpha"	"atom-apigw-3457699242-z0r86"	0	19306	0	2
2016-09-05T08:24:03Z	30917	1	"atom-apigw"	30917	"alpha"	"atom-apigw-3457699242-umi4y"	0	19346	0	1

- Measurements (e.g. cpu, mem, disk, net, nginx)
- Timestamp (nano second)
- Tags (e.g. app=atom-apigw namespace=alpha)
- Fields (e.g. accepts=30925, active=1, handled=30925)

InfluxDB Schema

InfluxDB Write Data Documentation Database: telegraf

Query: select * from nginx order by time desc limit 3;

Generate Query URL Query Templates

nginx

time	accepts	active	app	handled	namespace	pod	reading	requests	waiting	writing
2016-09-05T08:25:03Z	30925	1	"atom-apigw"	30925	"alpha"	"atom-apigw-3457699242-umi4y"	0	19352	0	1
2016-09-05T08:25:03Z	31181	2	"atom-apigw"	31181	"alpha"	"atom-apigw-3457699242-z0r86"	0	19306	0	2
2016-09-05T08:24:03Z	30917	1	"atom-apigw"	30917	"alpha"	"atom-apigw-3457699242-umi4y"	0	19346	0	1

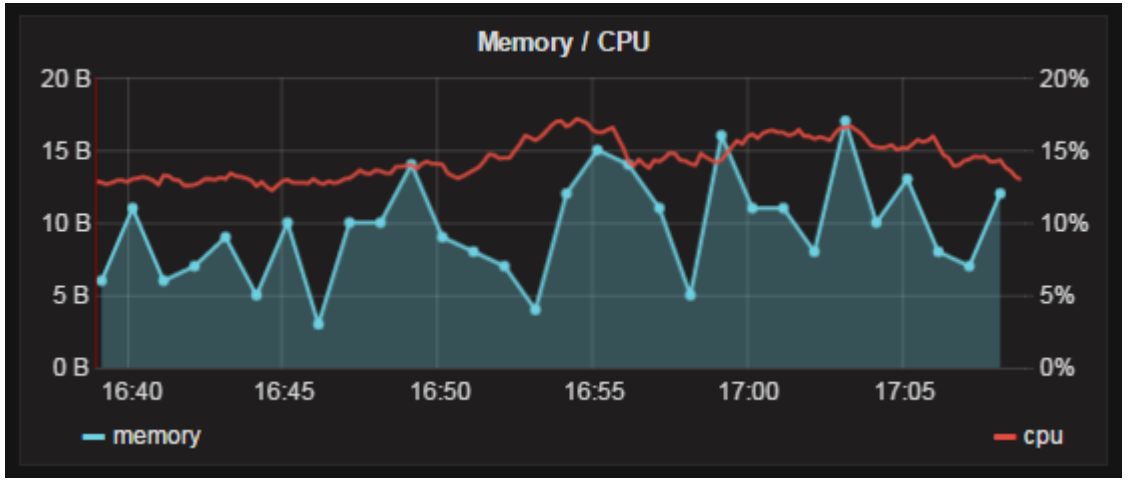
- Measurements (e.g. cpu, mem, disk, net, nginx)
- Timestamp (nano second)
- Tags (e.g. app=atom-apigw namespace=alpha)
- Fields (e.g. accepts=30925, active=1, handled=30925)

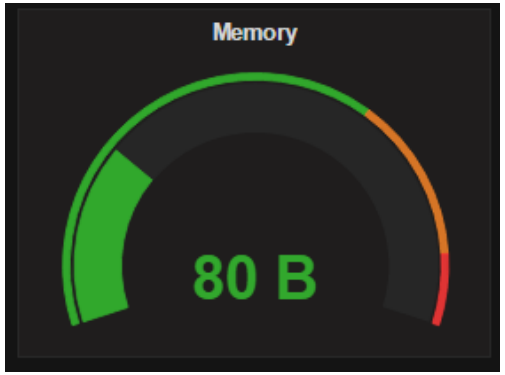
Grafana

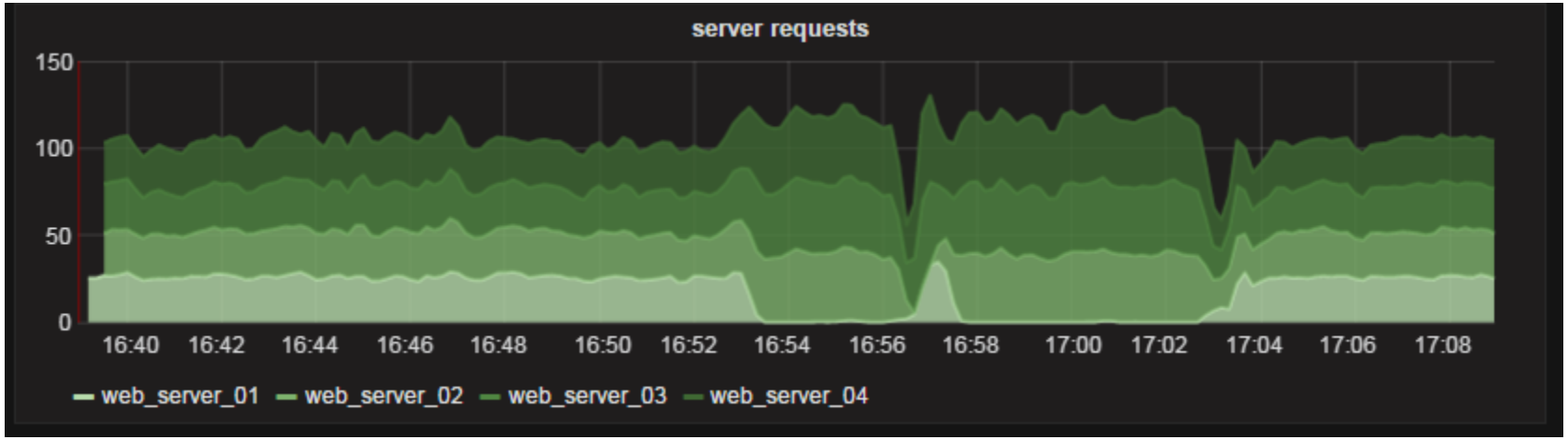
Querying and visualizing time series and metrics

Grafana

Finally! What we can to see!
Less talk more demo…







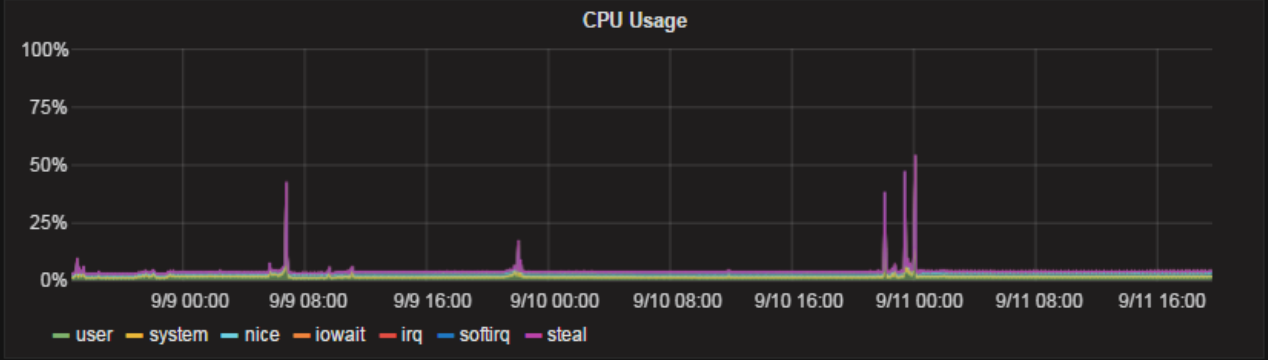
System - Overview

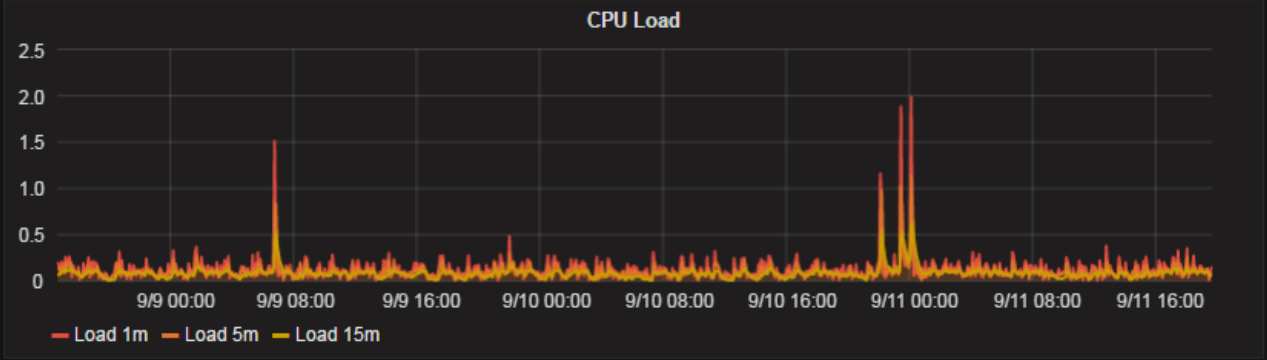
Zoom Out

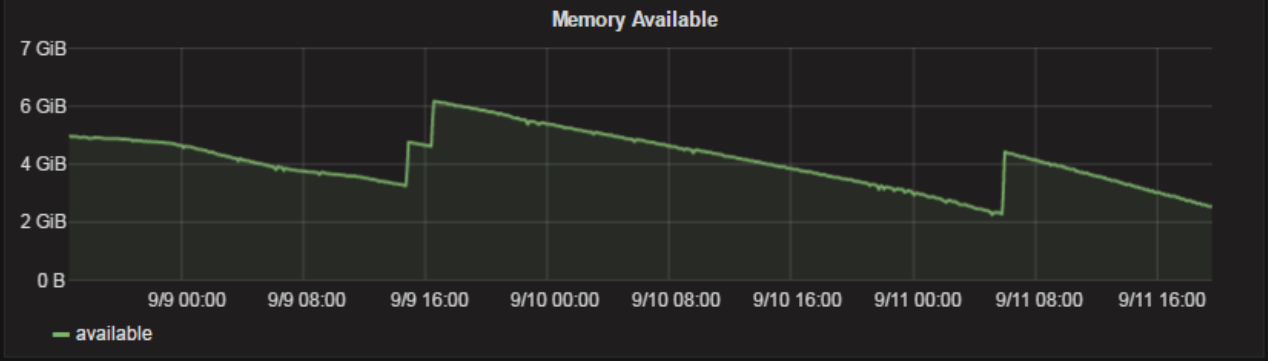
Sep 8, 2016 16:39:01 to Sep 11, 2016 19:35:50

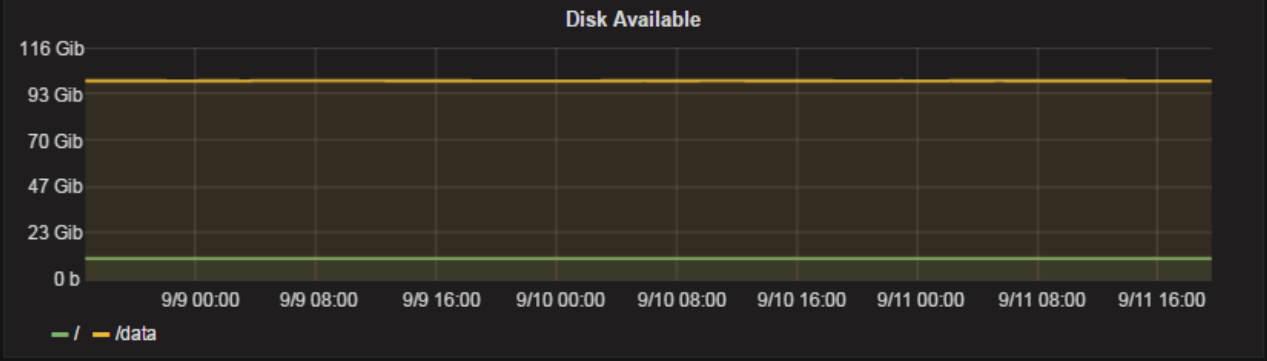
Host: ops-monitor101.ap-northeast-1

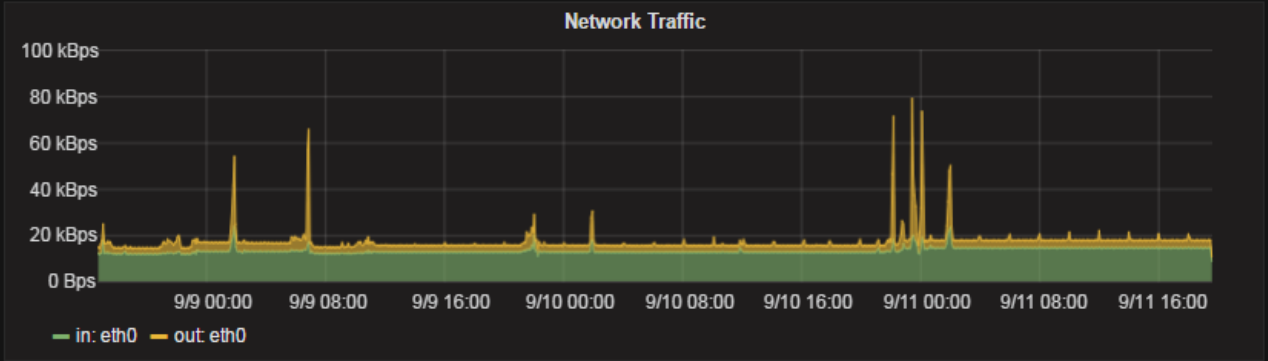
System Uptime	Virtual CPUs	Memory Available	Memory Total	Disk Available	Disk Total
1.3 weeks	2	30%	7.80 GiB	70%	15.62 GiB

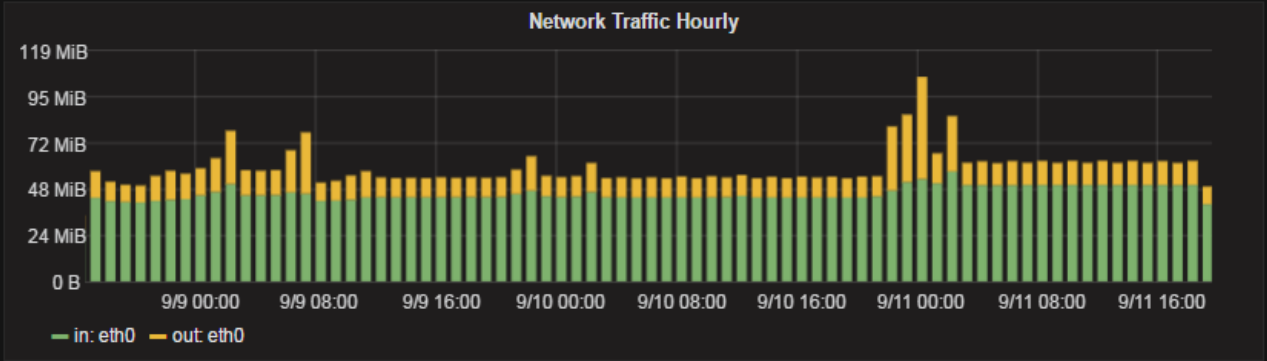












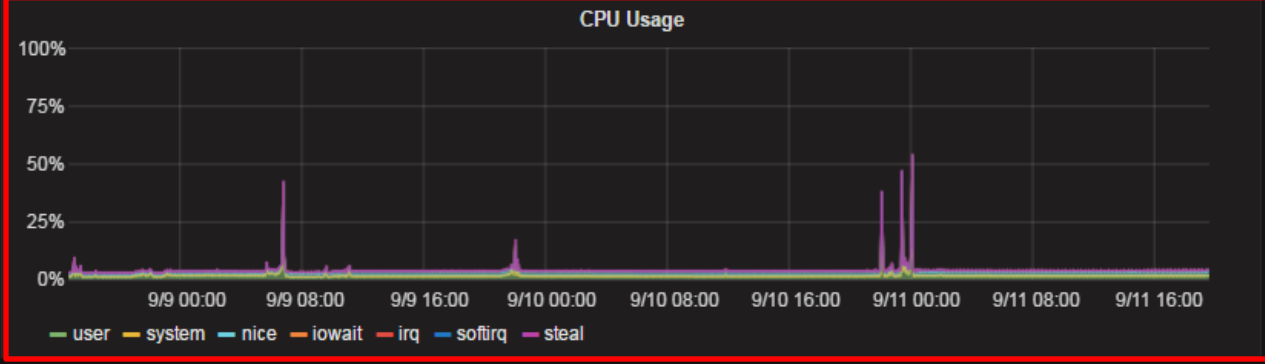
System - Overview

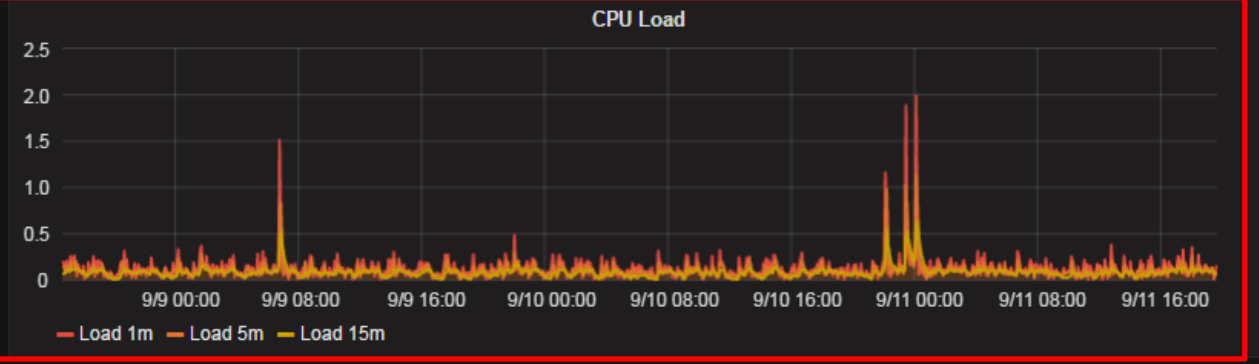
Zoom Out

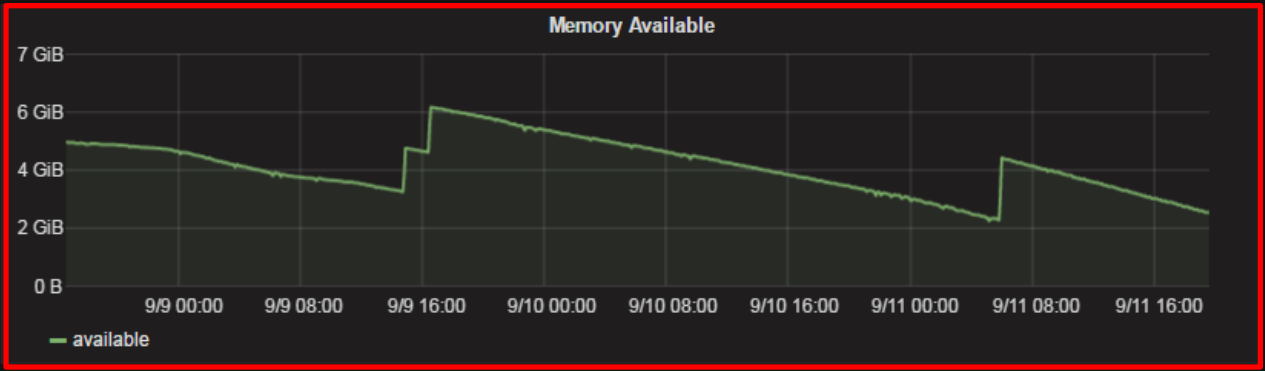
Sep 8, 2016 16:39:01 to Sep 11, 2016 19:35:50

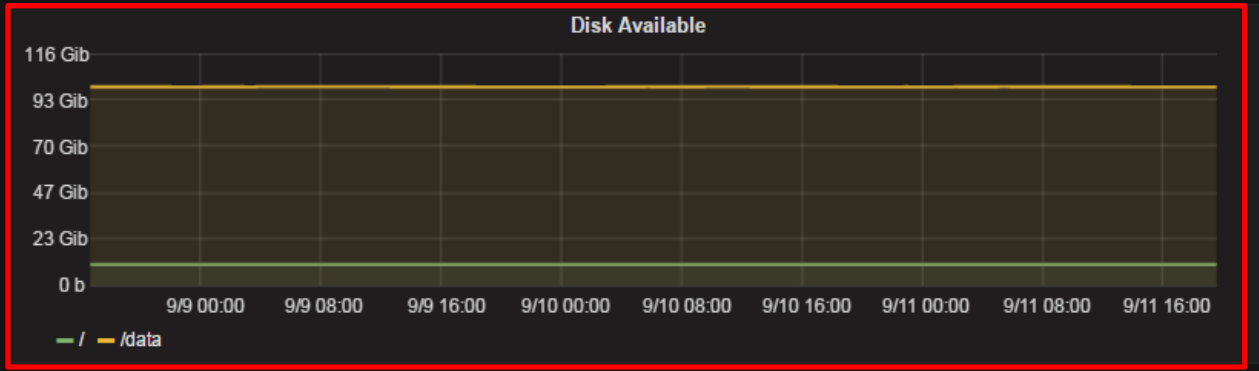
Host: ops-monitor101.ap-northeast-1

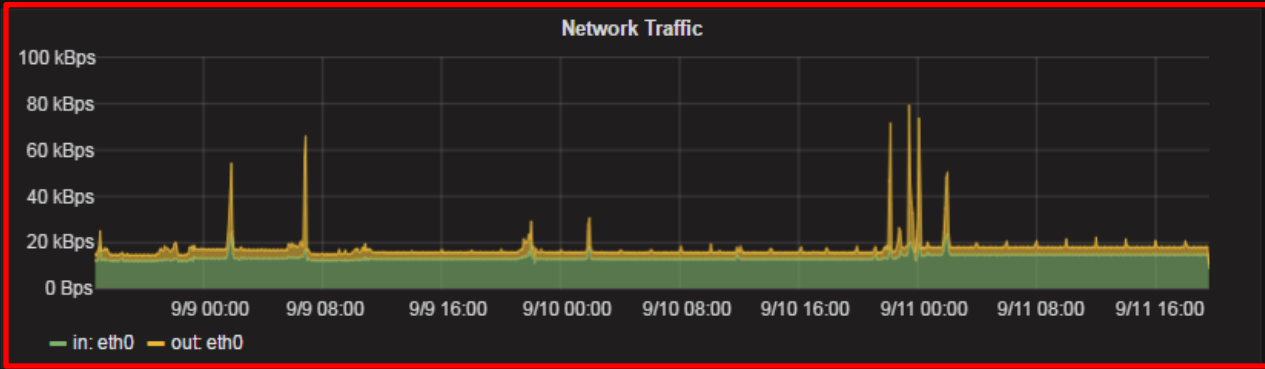
System Uptime	Virtual CPUs	Memory Available	Memory Total	Disk Available	Disk Total
1.3 weeks	2	30%	7.80 GiB	70%	15.62 GiB

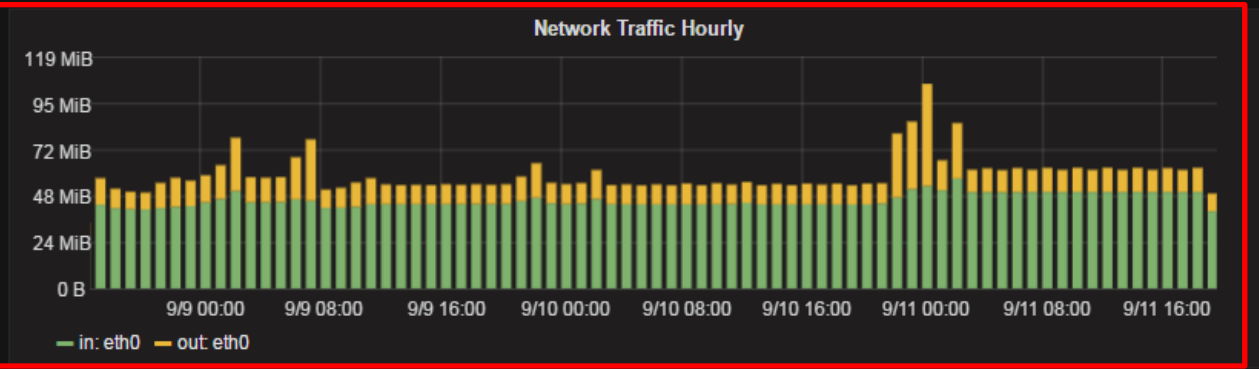












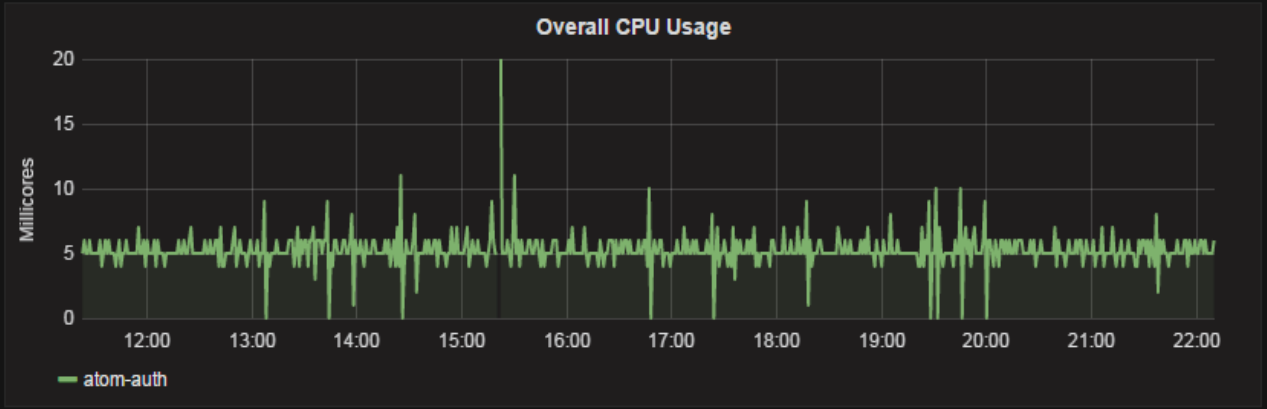
Kubernetes - Containers

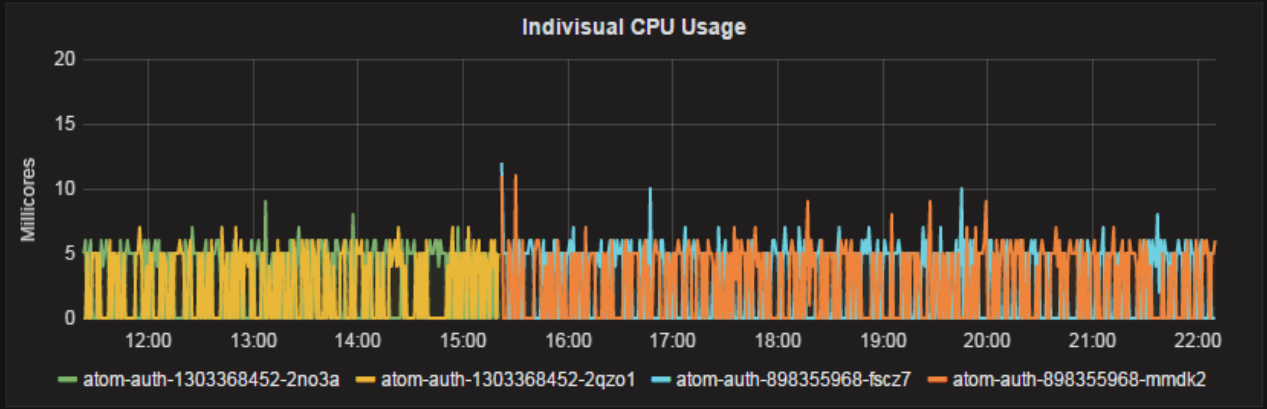
Zoom Out

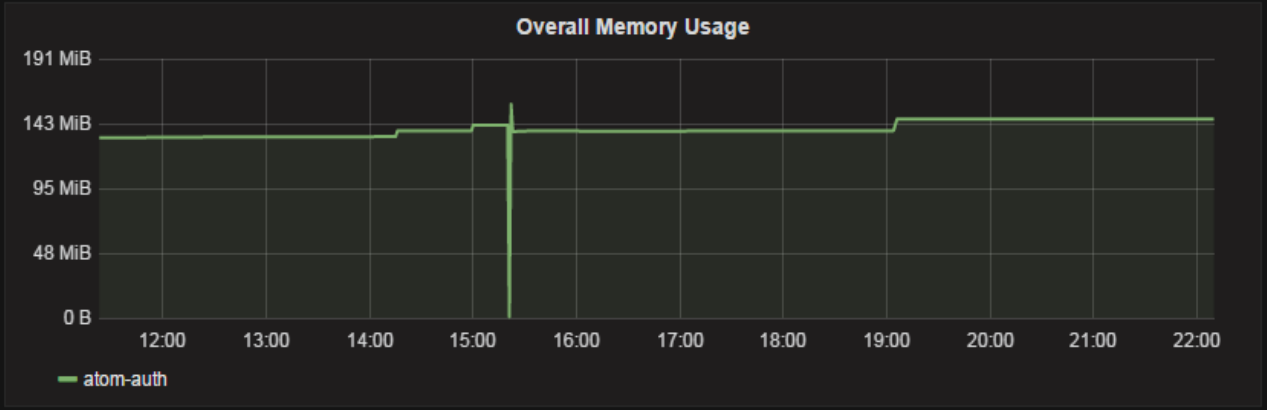
Sep 7, 2016 11:22:46 to Sep 7, 2016 22:10:20

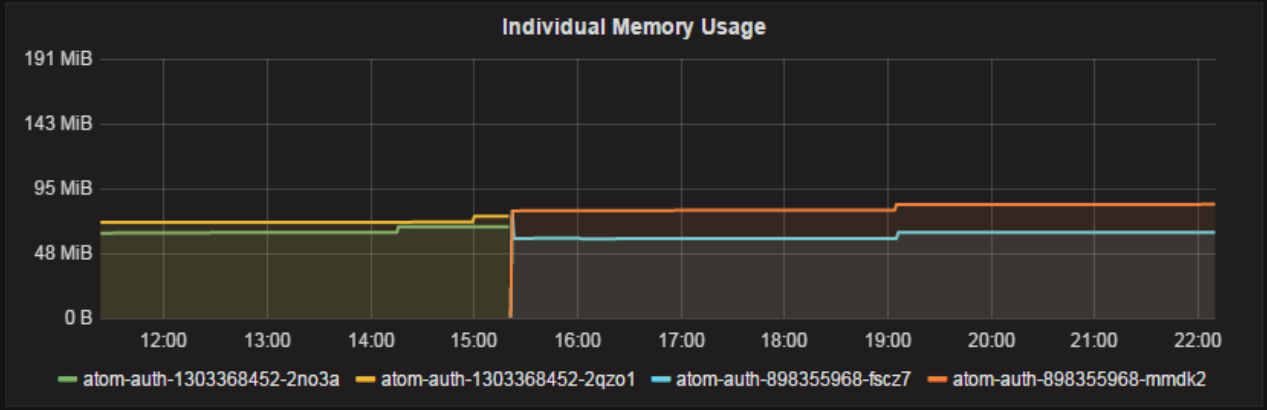
namespace: alpha

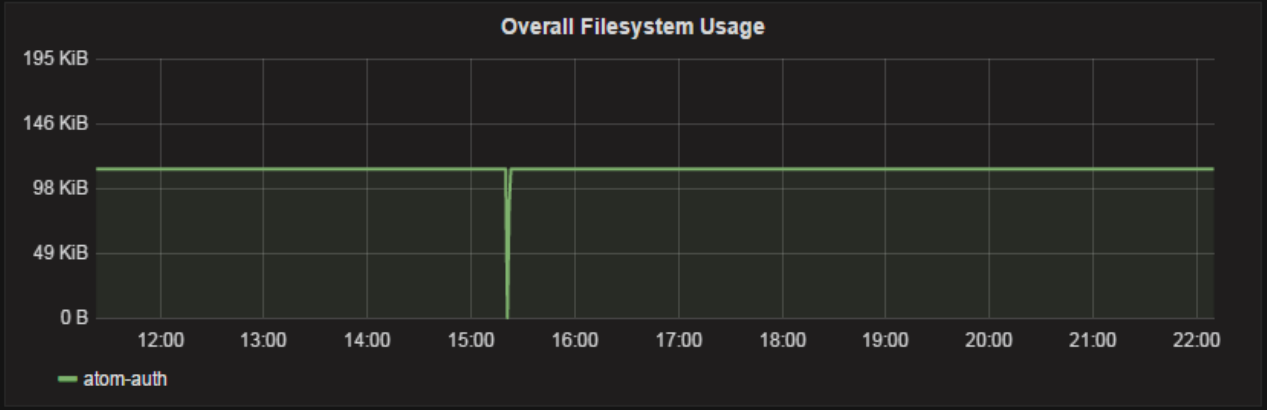
container: atom-auth

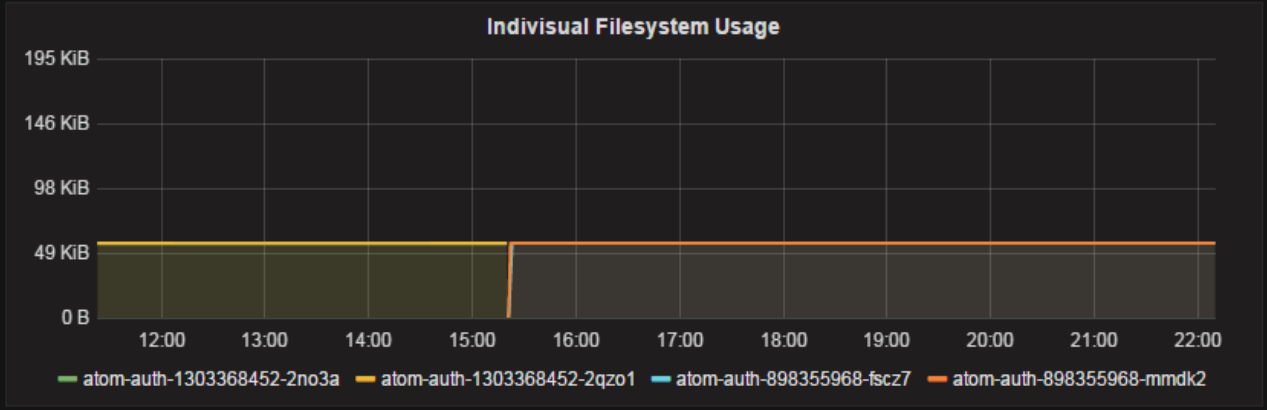












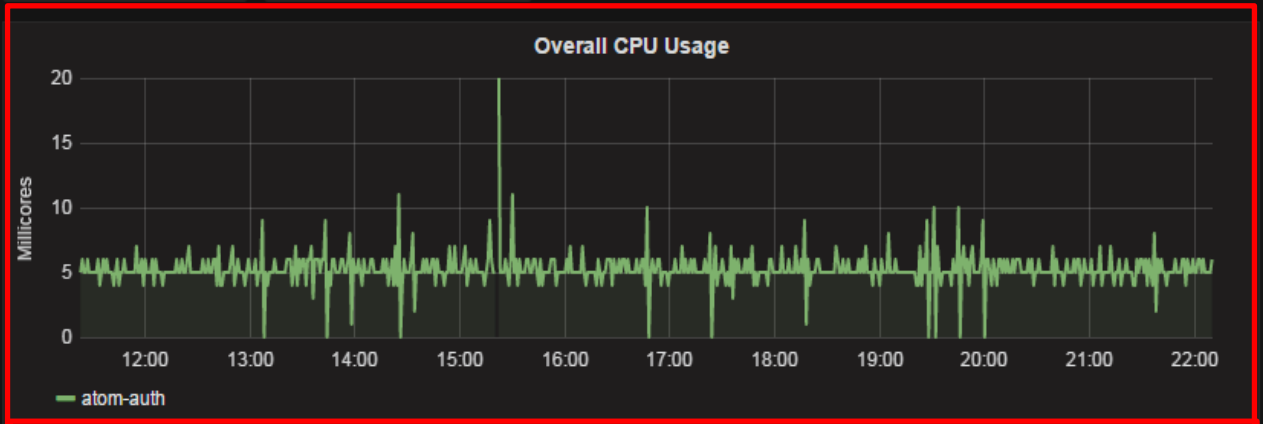
Kubernetes - Containers

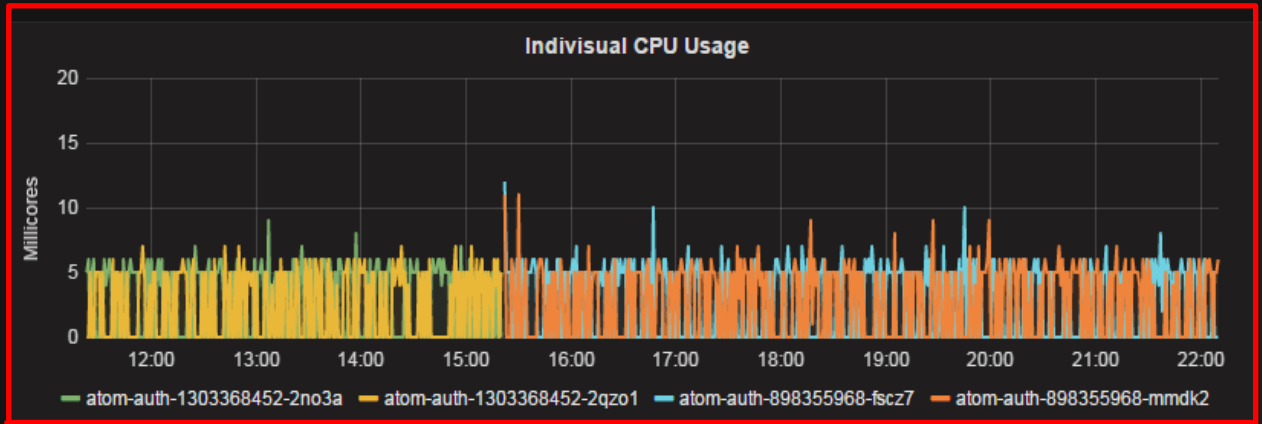
Zoom Out

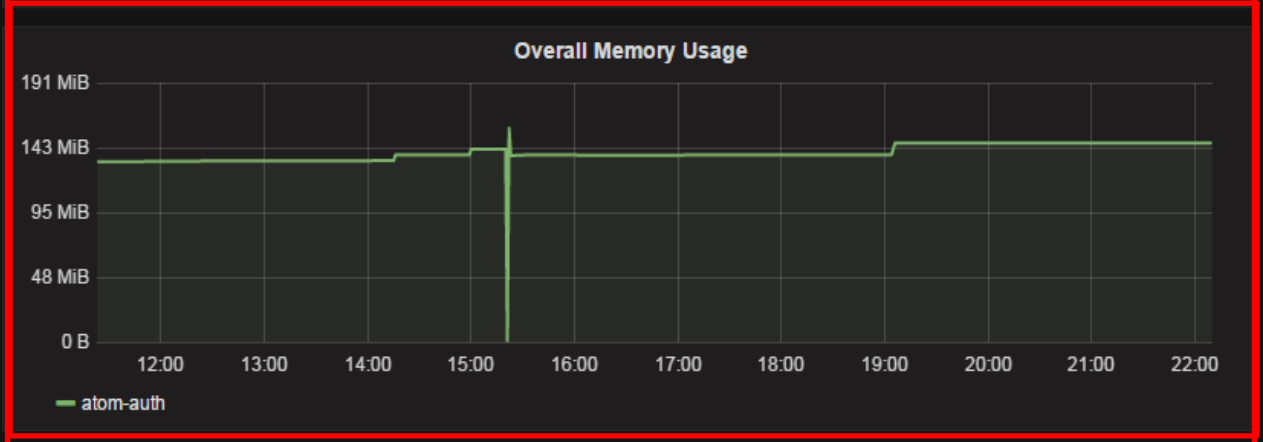
Sep 7, 2016 11:22:46 to Sep 7, 2016 22:10:20

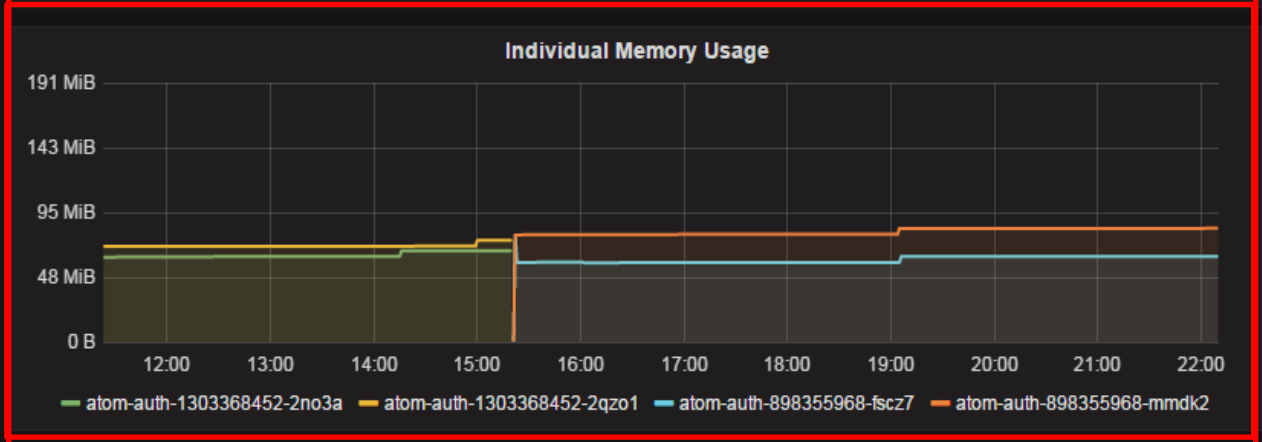
namespace: alpha

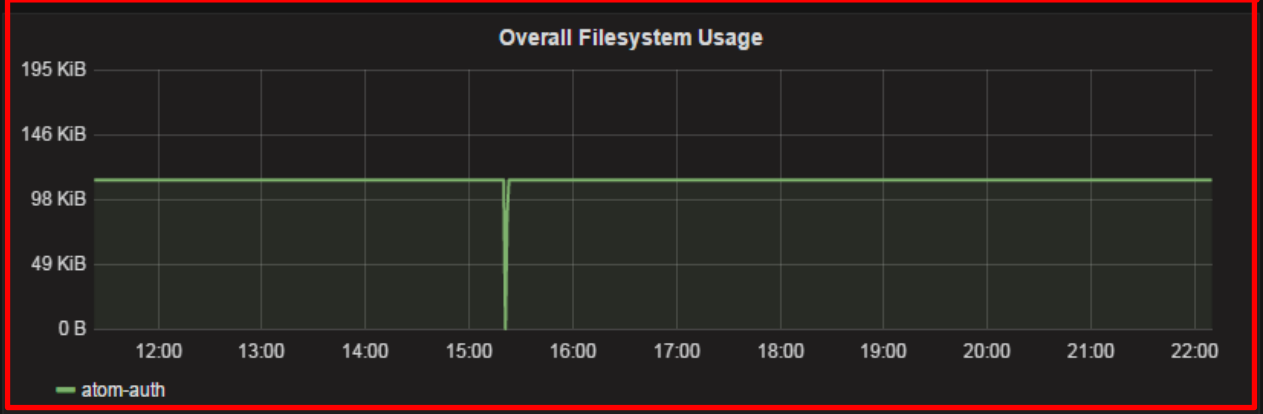
container: atom-auth

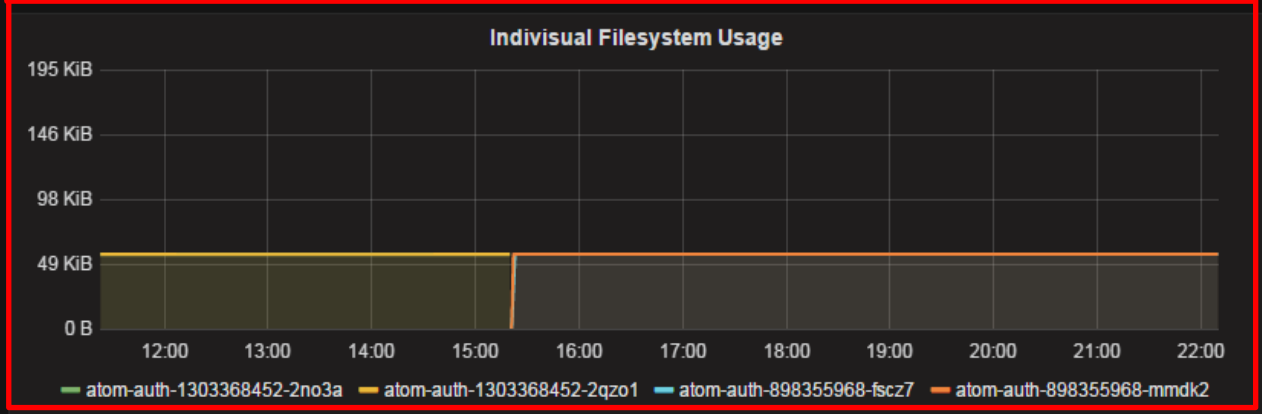










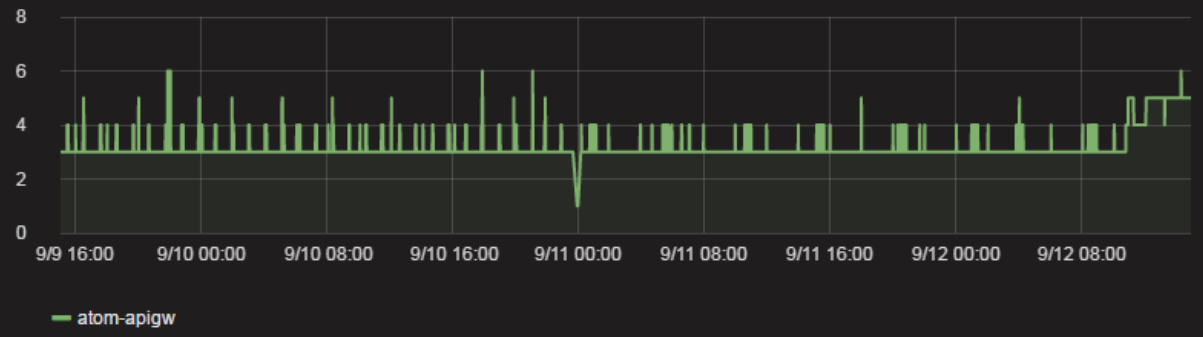


Application - Nginx

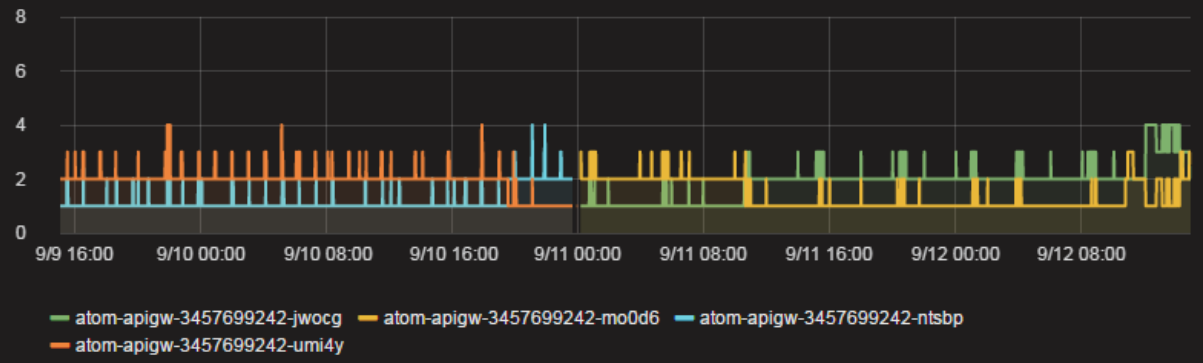
Zoom Out | Last 3 days

namespace: alpha | app: atom-apigw

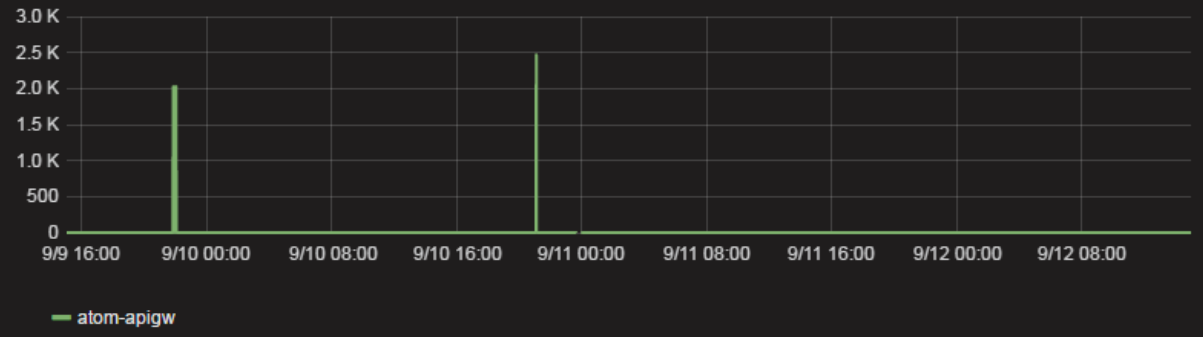
Overall Active Connections



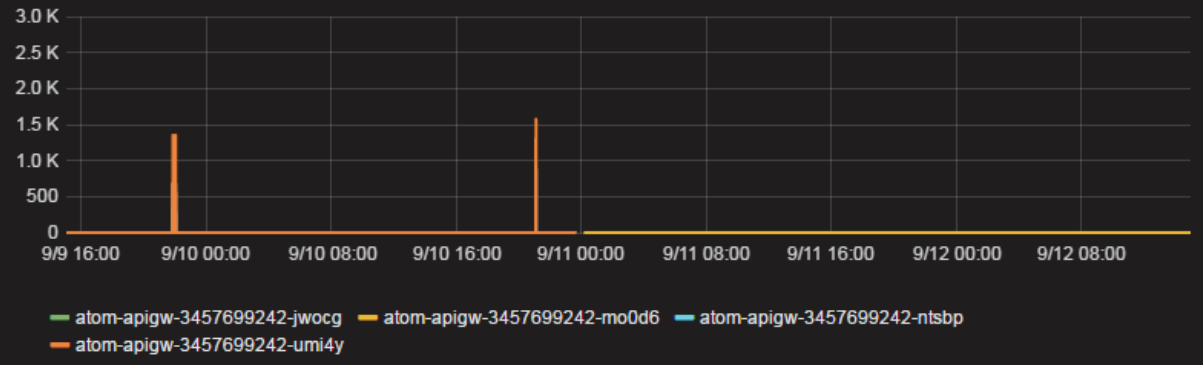
Indivisual Active Connections



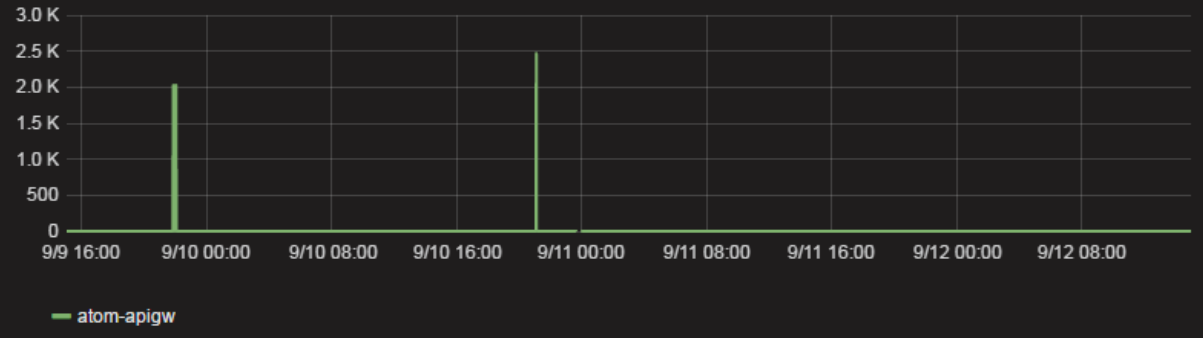
Overall Accepts / s



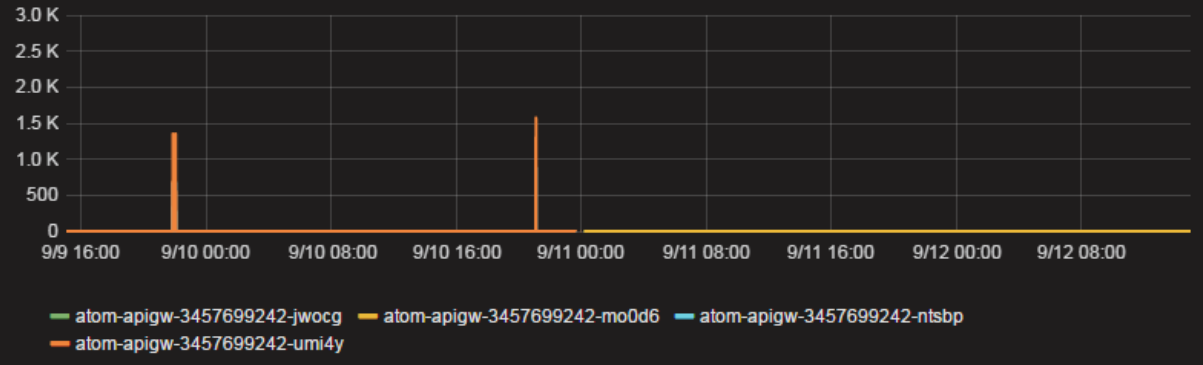
Indivisual Accepts / s



Overall Handled / s



Indivisual Handled / s

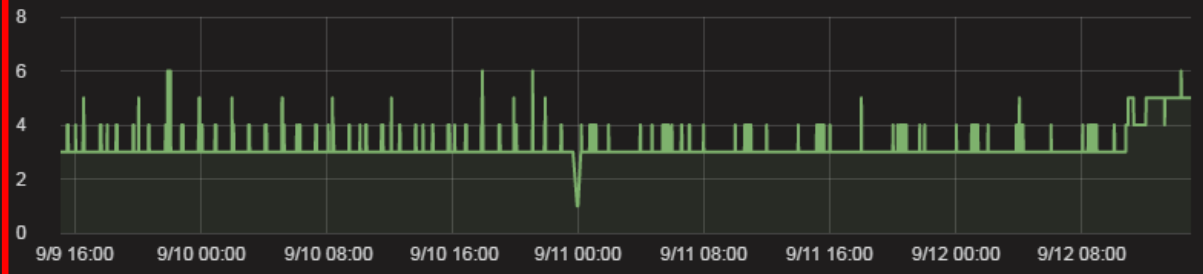


Application - Nginx

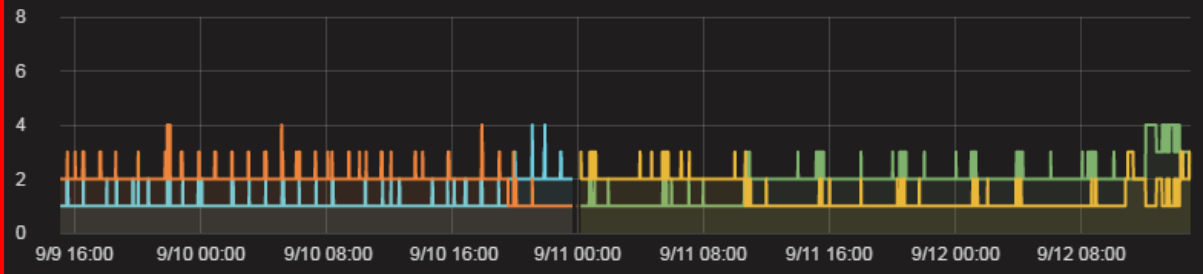
Zoom Out | Last 3 days

namespace: alpha | app: atom-apigw

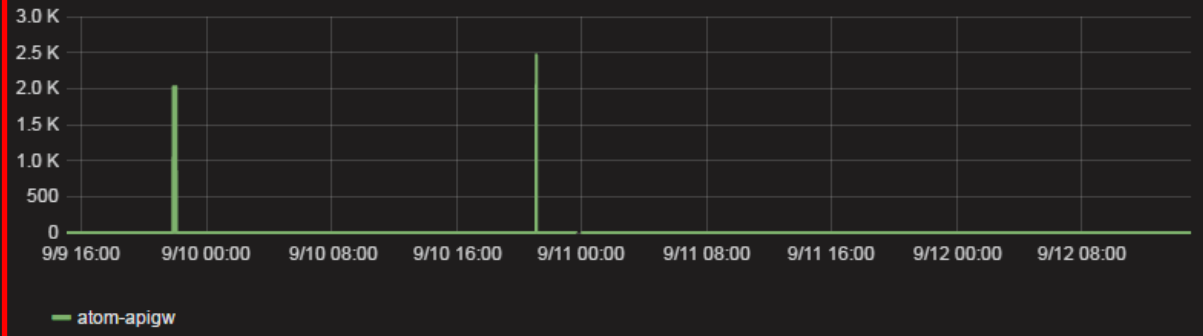
Overall Active Connections



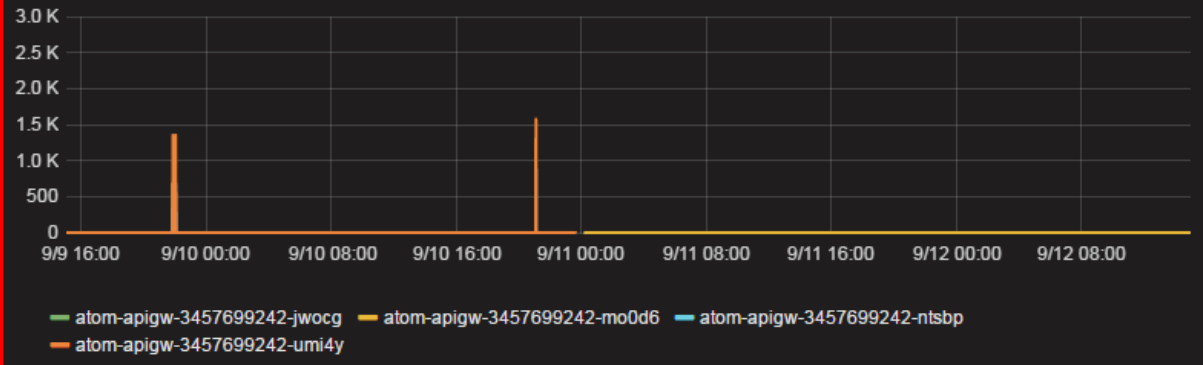
Indivisual Active Connections



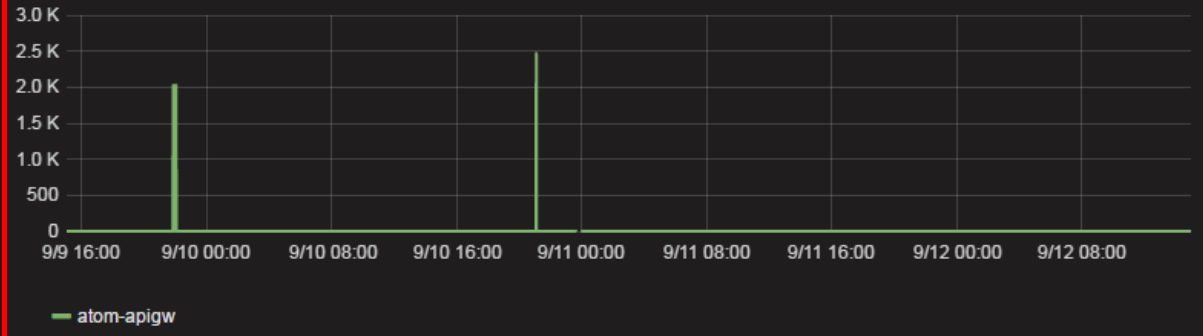
Overall Accepts / s



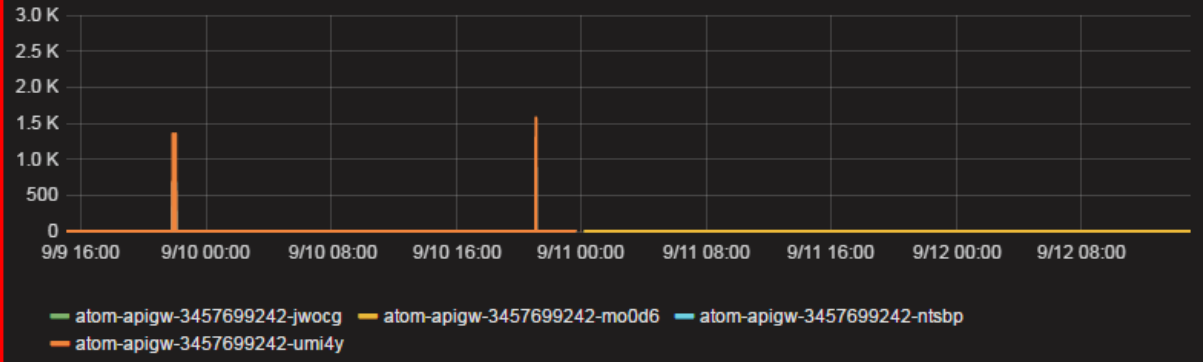
Indivisual Accepts / s



Overall Handled / s



Indivisual Handled / s



Event

We'll generally use events to let us know about changes and occurrences in our environment.

Icinga2

A monitoring system which checks the availability of your resources, notifies users of outages

icinga2

- Originally forked from **Nagios** in 2009
- Independent version **Icinga2** since 2014

- Monitors **everything**
- Gathering **status**
- Collect **performance** data

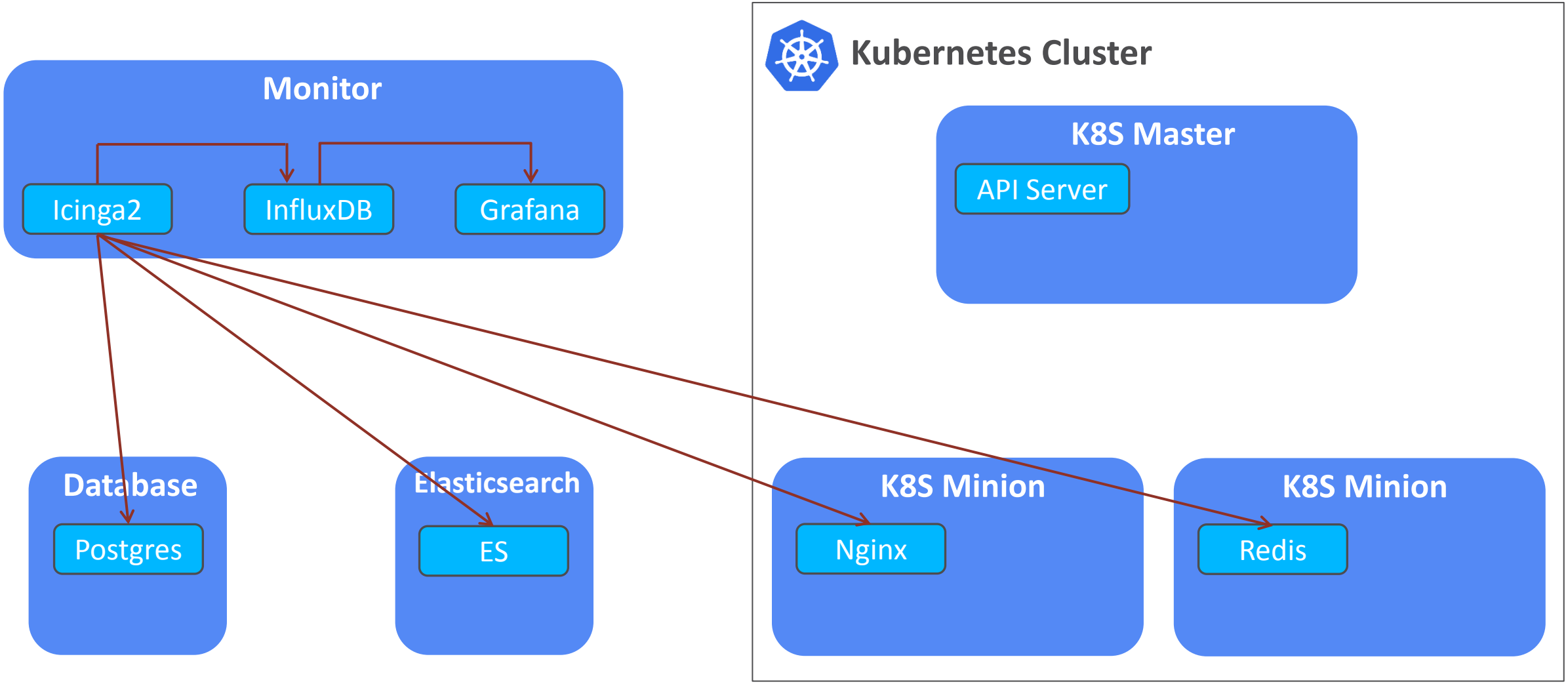
/use/lib/nagios/plugins/plugin_name

Any program which returns

- 0 – OK
- 1 – WARNING
- 2 – CRITICAL

Message to STDOUT

Event Monitoring



icinga

Search ...

Dashboard

Problems 5

Overview

Tactical Overview

Hosts

Services

Hostgroups

Servicegroups

Contactgroups

Contacts

Comments

Downtimes

History

Reporting

System

Configuration

dchen

Hosts

9 Hosts: 9

1 row(s) selected

25 Sort by Hostname

Search...

UP since Aug 29	atom-a-share101.ap-northeast-1 PING OK - Packet loss = 0%, RTA = 0.45 ms
UP since Aug 29	atom-alpha OK: Check was successful.
UP since 02:32	k8s.alpha.atom-apigw-3457699242-ntsbp (1 unhandled service) PING OK - Packet loss = 0%, RTA = 0.67 ms
UP since Sep 2	k8s.alpha.atom-apigw-3457699242-umi4y PING OK - Packet loss = 0%, RTA = 0.65 ms
UP since Aug 29	ops-k8smaster101.ap-northeast-1 PING OK - Packet loss = 0%, RTA = 0.47 ms
UP since Aug 29	ops-k8sminion101.ap-northeast-1 PING OK - Packet loss = 0%, RTA = 0.53 ms
UP since Aug 29	ops-k8sminion102.ap-northeast-1 PING OK - Packet loss = 0%, RTA = 0.67 ms
UP since Sep 2	ops-k8sminion103.ap-northeast-1 PING OK - Packet loss = 0%, RTA = 0.57 ms
UP since Aug 29	ops-monitor101.ap-northeast-1 (4 unhandled services) PING OK - Packet loss = 0%, RTA = 0.04 ms

Host | Services | History

UP since Aug 29 — ops-monitor101.ap-northeast-1 172.31.24.105

22 Services: 4 18

CRITICAL 33m 4s	DOCKER_PROCESS PROCS CRITICAL: 0 processes with args '/usr/bin/docker daemon', UID = 0 (root)
OK Sep 1	FLANNEL_PROCESS PROCS OK: 1 process with command name 'flanneld', UID = 0 (root)
OK Sep 1	GRAFANA_PROCESS PROCS OK: 1 process with command name 'grafana-server', UID = 116 (grafana)
OK 2d 2h	ICINGA2_PROCESS PROCS OK: 1 process with command name 'icinga2', UID = 114 (nagios)
OK Sep 1	INFLUXDB_PROCESS PROCS OK: 1 process with command name 'influxd', UID = 998 (influxdb)
OK Sep 1	INODE_USAGE_ROOT DISK OK - free space: / 10905 MB (71% inode=78%):
OK Sep 1	MEMORY_USAGE OK: Free memory percentage is 66% (5297 MiB)
OK Sep 1	MYSQL_PROCESS PROCS OK: 1 process with args '/usr/sbin/mysqld', UID = 115 (mysql)
OK Aug 30	NGINX_STATUS NGINX OK - 0.047 sec. response time, Active: 2 (Writing: 1 Reading: 0 Waiting: 1) ReqPerSec: 0.186 ConnPerSec: 0.068 ReqPerConn: 2.352
OK Aug 29	NRPE_SERVICE NRPE Working
CRITICAL 35m 30s	NTP_PEER NTP CRITICAL: No response from NTP server
CRITICAL 35m 34s	NTP_PROCESS PROCS CRITICAL: 0 processes with command name 'ntpd', UID = 112 (ntp)

icinga

Hosts | Host | Services | History

- Search ...
- Dashboard
- Problems 5
- Overview
  - Tactical Overview
  - Hosts
  - Services
  - Hostgroups
  - Servicegroups
  - Contactgroups
  - Contacts
  - Comments
  - Downtimes
- History
- Reporting
- System
- Configuration
- dchen

9 Hosts: 9 | 1 row(s) selected

25 | Sort by Hostname

Search...

UP since Aug 29	atom-a-share101.ap-northeast-1 PING OK - Packet loss = 0%, RTA = 0.45 ms
UP since Aug 29	atom-alpha OK: Check was successful.
UP since 02:32	k8s.alpha.atom-apigw-3457699242-ntsbp (1 unhandled service) PING OK - Packet loss = 0%, RTA = 0.67 ms
UP since Sep 2	k8s.alpha.atom-apigw-3457699242-umi4y PING OK - Packet loss = 0%, RTA = 0.65 ms
UP since Aug 29	ops-k8smaster101.ap-northeast-1 PING OK - Packet loss = 0%, RTA = 0.47 ms
UP since Aug 29	ops-k8sminion101.ap-northeast-1 PING OK - Packet loss = 0%, RTA = 0.53 ms
UP since Aug 29	ops-k8sminion102.ap-northeast-1 PING OK - Packet loss = 0%, RTA = 0.67 ms
UP since Sep 2	ops-k8sminion103.ap-northeast-1 PING OK - Packet loss = 0%, RTA = 0.57 ms
UP since Aug 29	ops-monitor101.ap-northeast-1 (4 unhandled services) PING OK - Packet loss = 0%, RTA = 0.04 ms

UP since Aug 29 — ops-monitor101.ap-northeast-1 — 172.31.24.105

22 Services: 4 | 18

CRITICAL 33m 4s	DOCKER_PROCESS PROCS CRITICAL: 0 processes with args '/usr/bin/docker daemon', UID = 0 (root)
OK Sep 1	FLANNEL_PROCESS PROCS OK: 1 process with command name 'flanneld', UID = 0 (root)
OK Sep 1	GRAFANA_PROCESS PROCS OK: 1 process with command name 'grafana-server', UID = 116 (grafana)
OK 2d 2h	ICINGA2_PROCESS PROCS OK: 1 process with command name 'icinga2', UID = 114 (nagios)
OK Sep 1	INFLUXDB_PROCESS PROCS OK: 1 process with command name 'influxd', UID = 998 (influxdb)
OK Sep 1	INODE_USAGE_ROOT DISK OK - free space: / 10905 MB (71% inode=78%):
OK Sep 1	MEMORY_USAGE OK: Free memory percentage is 66% (5297 MiB)
OK Sep 1	MYSQL_PROCESS PROCS OK: 1 process with args '/usr/sbin/mysqld', UID = 115 (mysql)
OK Aug 30	NGINX_STATUS NGINX OK - 0.047 sec. response time, Active: 2 (Writing: 1 Reading: 0 Waiting: 1) ReqPerSec: 0.186 ConnPerSec: 0.068 ReqPerConn: 2.352
OK Aug 29	NRPE_SERVICE NRPE Working
CRITICAL 35m 30s	NTP_PEER NTP CRITICAL: No response from NTP server
CRITICAL 35m 34s	NTP_PROCESS PROCS CRITICAL: 0 processes with command name 'ntpd', UID = 112 (ntp)

ICINGA

Search ...
Dashboard
Problems 5
Overview
Tactical Overview
Hosts
Services
Hostgroups
Servicegroups
Contactgroups
Contacts
Comments
Downtimes
History
Reporting
System
Configuration
dchen

Hosts

9 Hosts: 9 | 1 row(s) selected

25 | Sort by Hostname

Search...

UP since Aug 29	atom-a-share101.ap-northeast-1 PING OK - Packet loss = 0%, RTA = 0.45 ms
UP since Aug 29	atom-alpha OK: Check was successful.
UP since 02:32	k8s.alpha.atom-apigw-3457699242-ntsbp (1 unhandled service) PING OK - Packet loss = 0%, RTA = 0.67 ms
UP since Sep 2	k8s.alpha.atom-apigw-3457699242-umi4y PING OK - Packet loss = 0%, RTA = 0.65 ms
UP since Aug 29	ops-k8smaster101.ap-northeast-1 PING OK - Packet loss = 0%, RTA = 0.47 ms
UP since Aug 29	ops-k8sminion101.ap-northeast-1 PING OK - Packet loss = 0%, RTA = 0.53 ms
UP since Aug 29	ops-k8sminion102.ap-northeast-1 PING OK - Packet loss = 0%, RTA = 0.67 ms
UP since Sep 2	ops-k8sminion103.ap-northeast-1 PING OK - Packet loss = 0%, RTA = 0.57 ms
UP since Aug 29	ops-monitor101.ap-northeast-1 (4 unhandled services) PING OK - Packet loss = 0%, RTA = 0.04 ms

Pod

Host

Host | Services | History

UP since Aug 29 — ops-monitor101.ap-northeast-1 — 172.31.24.105

22 Services: 4 18

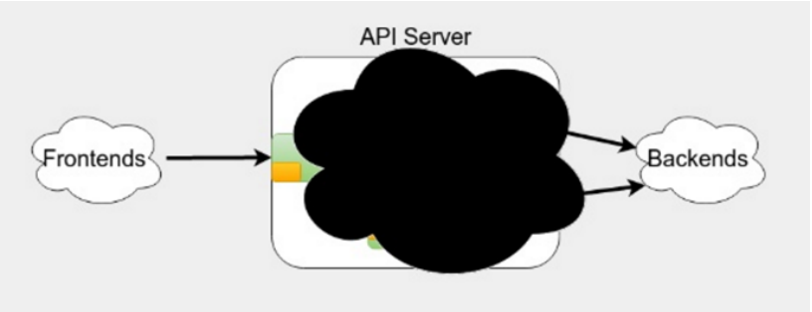
CRITICAL 33m 4s	DOCKER_PROCESS PROCS CRITICAL: 0 processes with args '/usr/bin/docker daemon', UID = 0 (root)
OK Sep 1	FLANNEL_PROCESS PROCS OK: 1 process with command name 'flanneld', UID = 0 (root)
OK Sep 1	GRAFANA_PROCESS PROCS OK: 1 process with command name 'grafana-server', UID = 116 (grafana)
OK 2d 2h	ICINGA2_PROCESS PROCS OK: 1 process with command name 'icinga2', UID = 114 (nagios)
OK Sep 1	INFLUXDB_PROCESS PROCS OK: 1 process with command name 'influxd', UID = 998 (influxdb)
OK Sep 1	INODE_USAGE_ROOT DISK OK - free space: / 10905 MB (71% inode=78%):
OK Sep 1	MEMORY_USAGE OK: Free memory percentage is 66% (5297 MiB)
OK Sep 1	MYSQL_PROCESS PROCS OK: 1 process with args '/usr/sbin/mysqld', UID = 115 (mysql)
OK Aug 30	NGINX_STATUS NGINX OK - 0.047 sec. response time, Active: 2 (Writing: 1 Reading: 0 Waiting: 1) ReqPerSec: 0.186 ConnPerSec: 0.068 ReqPerConn: 2.352
OK Aug 29	NRPE_SERVICE NRPE Working
CRITICAL 35m 30s	NTP_PEER NTP CRITICAL: No response from NTP server
CRITICAL 35m 34s	NTP_PROCESS PROCS CRITICAL: 0 processes with command name 'ntpd', UID = 112 (ntp)

Server Monitoring

- About changes and occurrences in our environment
  - Is cpu load too high?
  - Is memory not enough?
  - Is docker engine still alive?

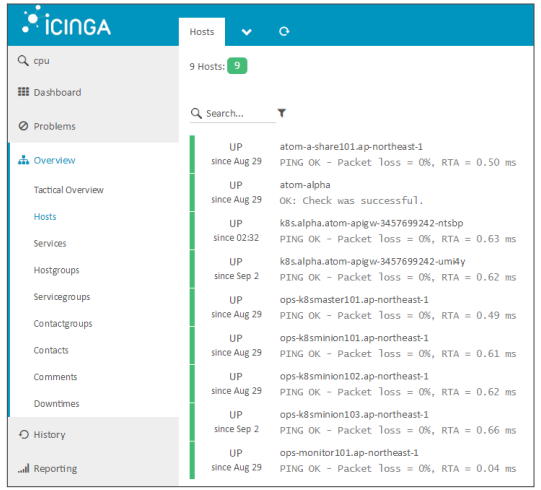
External Monitoring

- End to end test from user's perspective
  - Can you connect ssh port 22?
  - Can you browse web page?
  - Can you request RESTful API successfully?



Dashboard

- Provides a clear view for current environment

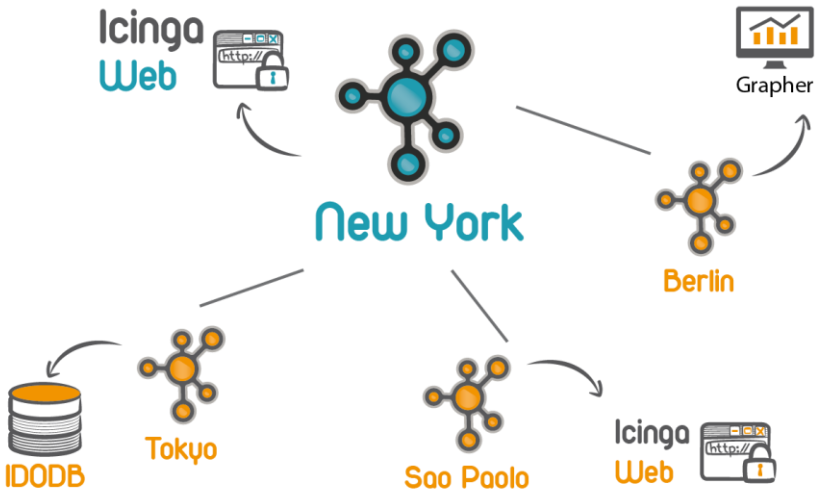


Alerting

- Notifies using email, slack, pager and etc.
- Notification escalation

Others…

- Distributed Monitoring
- Reloads config without interrupt checks



Log

Logs are a subset of events. They're often most useful for fault diagnosis and investigation

EFK

- EFK stack is a mature logging solution

Shipping all of your log to where it should go

The main part to store your data with high availability

Visualize will power your data. To know more about its value.

Summary

Monitoring Service Stack

Collect

Collectd

cAdvisor

Telegraf

Fluentd

Store

InfluxDB

Elasticserach

Visualize

Grafana

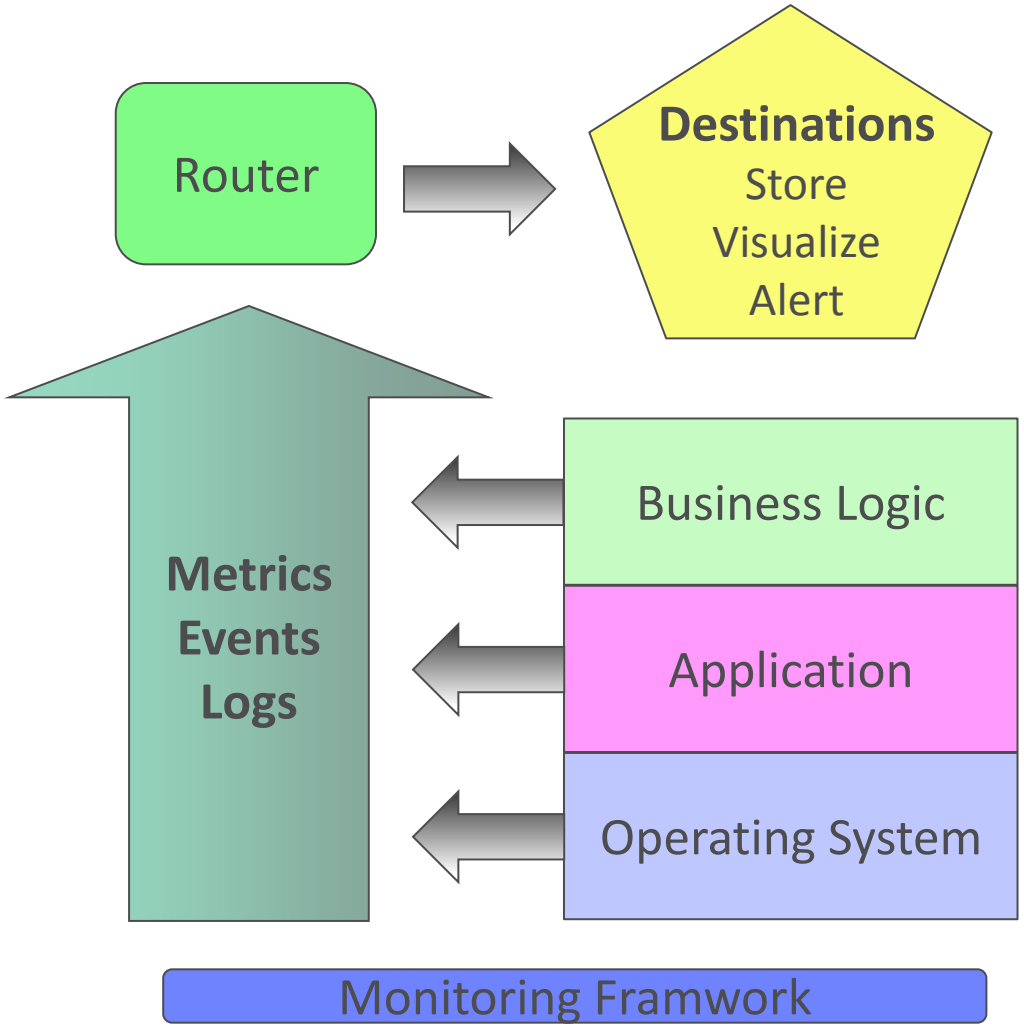
Kibana

Alert

Icinga2

Last but Not least

- Data Flow
  - Blackbox vs Whitebox
  - Pull Mode vs Push Mode
- Data Content
  - Opeating System, Application, Business Logic
- Data Type
  - Metric, Events, Logs



[image] https://www.artofmonitoring.com

THANKS!

Any Questions?