



Exporting Kubernetes Event Objects for Better Observability

KubeCon/CloudNativeCon North America 2019 San Diego



MUSTAFA AKIN | AHMET ŞEKER | SRE @ ATLASSIAN OPSGENIE

Agenda

K8s Event API

Interesting Kubernetes Events

Proper Monitoring & Alerting

Event Exporter Tool

Kubernetes Event API

What is an Event and why they are thrown?

**“Event is a report of an event
somewhere in the cluster”**

Kubernetes Source Code

**“Event is a report of an event
somewhere in the cluster”**

Kubernetes Source Code

AN EVENT EXAMPLE

```
apiVersion: v1
kind: Event
metadata:
  name: xxx.15d3018d822b6959
  namespace: default
count: 1
eventTime: null
type: Normal
message: pulling image "my-cool-app:0.1"
reason: Pulling
firstTimestamp: "2019-11-01T10:00:02Z"
involvedObject:
  apiVersion: v1
  fieldPath: spec.containers{container}
  kind: Pod
  name: my-app
  namespace: default
source:
  component: kubelet
  host: ip-10-35-44-212.us-west-2.compute.internal
```

Event Components



Message

A human-readable description of the status of this operation



Reason

Short, machine understandable string, in other words: Enum



Type

Currently holds only Normal & Warning, but custom type can be given if desired.



Involved Object

The object that this event is about, like Pod, Deployment, Node etc.



Source

The component reporting this event, short machine understandable string. i.e kube-scheduler



Count

The number of times the event has occurred

When are Events published?

Informational

Pod scheduled, images pulled, Node healthy,
Deployment is updated, ReplicaSet is scaled,
Container is killed

Warnings

Pods have errors, persistent volumes are not
bound yet

Errors

Node is down, Persistent Volume is not found,
Cannot create a LoadBalancer in the Cloud
Provider

How can you publish custom events?

Directly

Use the REST API directly, or with a SDK (i.e. client-go) to create the Event Object with required fields.

Event Recorder

A helper for K8s controllers create events easier.

Also see another talk for more information:

Emitting, Consuming, and Presenting: The Event Lifecycle - Jesse Dearing, VMware
<https://kccncna19.sched.com/event/UaY1>

KUBECTL DESCRIBE POD

```
makin@makin-OptiPlex-5090: ~ (zsh) [281]
```

```
Host Port: <none>
State: Running
  Started: Thu, 21 Nov 2019 09:07:18 -0800
Ready: True
Restart Count: 0
Environment: <none>
Mounts:
  /tmp from tmp-dir (rw)
  /var/run/secrets/kubernetes.io/serviceaccount from metrics-server-token-xfhj6 (ro)

Conditions:
  Type        Status
  Initialized  True
  Ready       True
  ContainersReady  True
  PodScheduled  True

Volumes:
  tmp-dir:
    Type:      EmptyDir (a temporary directory that shares a pod's lifetime)
    Medium:
    SizeLimit: <unset>
  metrics-server-token-xfhj6:
    Type:      Secret (a volume populated by a Secret)
    SecretName: metrics-server-token-xfhj6
    Optional:   false
  QoS Class:  BestEffort
  Node-Selectors: <none>
  Tolerations: node.kubernetes.io/not-ready:NoExecute for 300s
                node.kubernetes.io/unreachable:NoExecute for 300s

Events:
  Type  Reason  Age   From           Message
  ----  -----  --   --  -----
  Normal Scheduled  13s  default-scheduler  Successfully assigned kube-system/metrics-server-6c6c6c6457-xtljb to ip-10-35-13-60.us-west-2.compute.internal
  Normal Pulling   12s  kubelet, ip-10-35-13-60.us-west-2.compute.internal  pulling image "k8s.gcr.io/metrics-server-amd64:v0.3.3"
  Normal Pulled   11s  kubelet, ip-10-35-13-60.us-west-2.compute.internal  Successfully pulled image "k8s.gcr.io/metrics-server-amd64:v0.3.3"
  Normal Created   11s  kubelet, ip-10-35-13-60.us-west-2.compute.internal  Created container
  Normal Started   11s  kubelet, ip-10-35-13-60.us-west-2.compute.internal  Started container
```

KUBECTL GET EVENTS

```
makin@C02YJ0YRJGH7: ~ (zsh) 2%1

3m21s    Normal   Scheduled          Pod      Successfully assigned kube-system/filebeat-5xkpr to ip-10-35-96-45.us-west-2.compute.internal
3m20s    Warning  FailedCreatePodSandBox  Pod      Failed create pod sandbox: rpc error: code = Unknown desc = [failed to set up sandbox container "dd378d5616aa5fe15bfe4e86e
46ca8dcbbcb2b86b877b1732ce8884d5ce5cbf" network for pod "filebeat-5xkpr": NetworkPlugin_cni failed to set up pod "filebeat-5xkpr_kube-system" network: rpc error: code = Unavailable desc = al
l SubConns are in TransientFailure, latest connection error: connection error: desc = "transport: Error while dialing dial tcp 127.0.0.1:50051: connect: connection refused", failed to clean u
p sandbox container "dd378d5616aa5fe15bfe4e86e46ca8dcbbcb2b86b877b1732ce8884d5ce5cbf" network for pod "filebeat-5xkpr": NetworkPlugin_cni failed to teardown pod "filebeat-5xkpr_kube-system"
network: rpc error: code = Unavailable desc = all SubConns are in TransientFailure, latest connection error: connection error: desc = "transport: Error while dialing dial tcp 127.0.0.1:50051:
connect: connection refused"]
2m41s    Normal   SandboxChanged        Pod      Pod sandbox changed, it will be killed and re-created.
2m40s    Normal   Pulling             Pod      pulling image "416306766477.dkr.ecr.us-west-2.amazonaws.com/delivery/filebeat:1.0"
2m38s    Normal   Pulled              Pod      Successfully pulled image "416306766477.dkr.ecr.us-west-2.amazonaws.com/delivery/filebeat:1.0"
2m38s    Normal   Created              Pod      Created container
2m38s    Normal   Started              Pod      Started container
8m26s    Normal   TaintManagerEviction Pod      Marking for deletion Pod kube-system/filebeat-kg4vs
8m26s    Normal   Killing              Pod      Killing container with id docker://filebeat:Need to kill Pod
8m26s    Normal   SuccessfulDelete    DaemonSet  Deleted pod: filebeat-kg4vs
3m22s    Normal   SuccessfulCreate   DaemonSet  Created pod: filebeat-5xkpr
3m41s    Normal   Scheduled          Pod      Successfully assigned kube-system/kube-proxy-2jl22 to ip-10-35-96-45.us-west-2.compute.internal
3m40s    Normal   Pulling             Pod      pulling image "602401143452.dkr.ecr.us-west-2.amazonaws.com/eks/kube-proxy:v1.11.5"
3m38s    Normal   Pulled              Pod      Successfully pulled image "602401143452.dkr.ecr.us-west-2.amazonaws.com/eks/kube-proxy:v1.11.5"
3m28s    Normal   Created              Pod      Created container
3m28s    Normal   Started              Pod      Started container
3m42s    Normal   SuccessfulCreate   DaemonSet  Created pod: kube-proxy-2jl22
3m41s    Normal   Scheduled          Pod      Successfully assigned kube-system/metricbeat-d7dvw to ip-10-35-96-45.us-west-2.compute.internal
3m40s    Normal   Pulling             Pod      pulling image "416306766477.dkr.ecr.us-west-2.amazonaws.com/metricbeat:latest"
3m33s    Normal   Pulled              Pod      Successfully pulled image "416306766477.dkr.ecr.us-west-2.amazonaws.com/metricbeat:latest"
3m28s    Normal   Created              Pod      Created container
3m28s    Normal   Started              Pod      Started container
3m42s    Normal   SuccessfulCreate   DaemonSet  Created pod: metricbeat-d7dvw
81s     Normal   Killing              Pod      Killing container with id docker://metrics-server:Need to kill Pod
81s     Normal   Scheduled          Pod      Successfully assigned kube-system/metrics-server-6c6c6c6457-xtljb to ip-10-35-13-60.us-west-2.compute.internal
80s     Normal   Pulling             Pod      pulling image "k8s.gcr.io/metrics-server-amd64:v0.3.3"
79s     Normal   Pulled              Pod      Successfully pulled image "k8s.gcr.io/metrics-server-amd64:v0.3.3"
79s     Normal   Created              Pod      Created container
79s     Normal   Started              Pod      Started container
81s     Normal   SuccessfulCreate   ReplicaSet  Created pod: metrics-server-6c6c6c6457-xtljb
4m4s    Normal   NoPods              PodDisruptionBudget  No matching pods found

→ ~
```

**For scalability issues on
etcd, events are stored only
for 1-hour by default.**

**Usually a separate etcd
cluster is deployed to reduce
the load.**

Interesting Kubernetes Events

If a tree falls in a forest, and no one is around to hear it, does it make a sound?

THE ONES YOU PROBABLY KNOW

Unhealthy
Pulled
Started
Scaled
Preempted
Starting
Failed
SuccessfulDelete

INFREQUENT EVENTS

FailedCreatePodSandBox
NetworkNotReady
LeaderElection
FailedAttachVolume
ScaleDownFailed
ImageGCFailed
FailedToUpdateEndpoint
TaintManagerEviction

SOME EVENTS FROM THE KUBERNETES SOURCE CODE

ProvisioningFailed

FailedToCreateEndpoint

ClusterIPNotValid

UpdateLoadBalancerFailed

CheckLimitsForResolvConf

PortRangeFull

InvalidDiskCapacity

ClusterIPAlreadyAllocated

InvalidEnvironmentVariableNames

WaitForFirstConsumer

FailedNeedsStart

FailedPlacementReason

MissingClusterDNS

CalculateExpectedPodCountFailed

TLSConfigChanged

2K

Events hourly for a 10-node not-so-busy stable cluster

700k

Events hourly for a 200-node for busy dev cluster

70

Unique Reasons

DISTRIBUTION OF EVENTS PER INVOLVED OBJECT

Mostly Pods

They are the unit of computation and there are probably lots of replicas

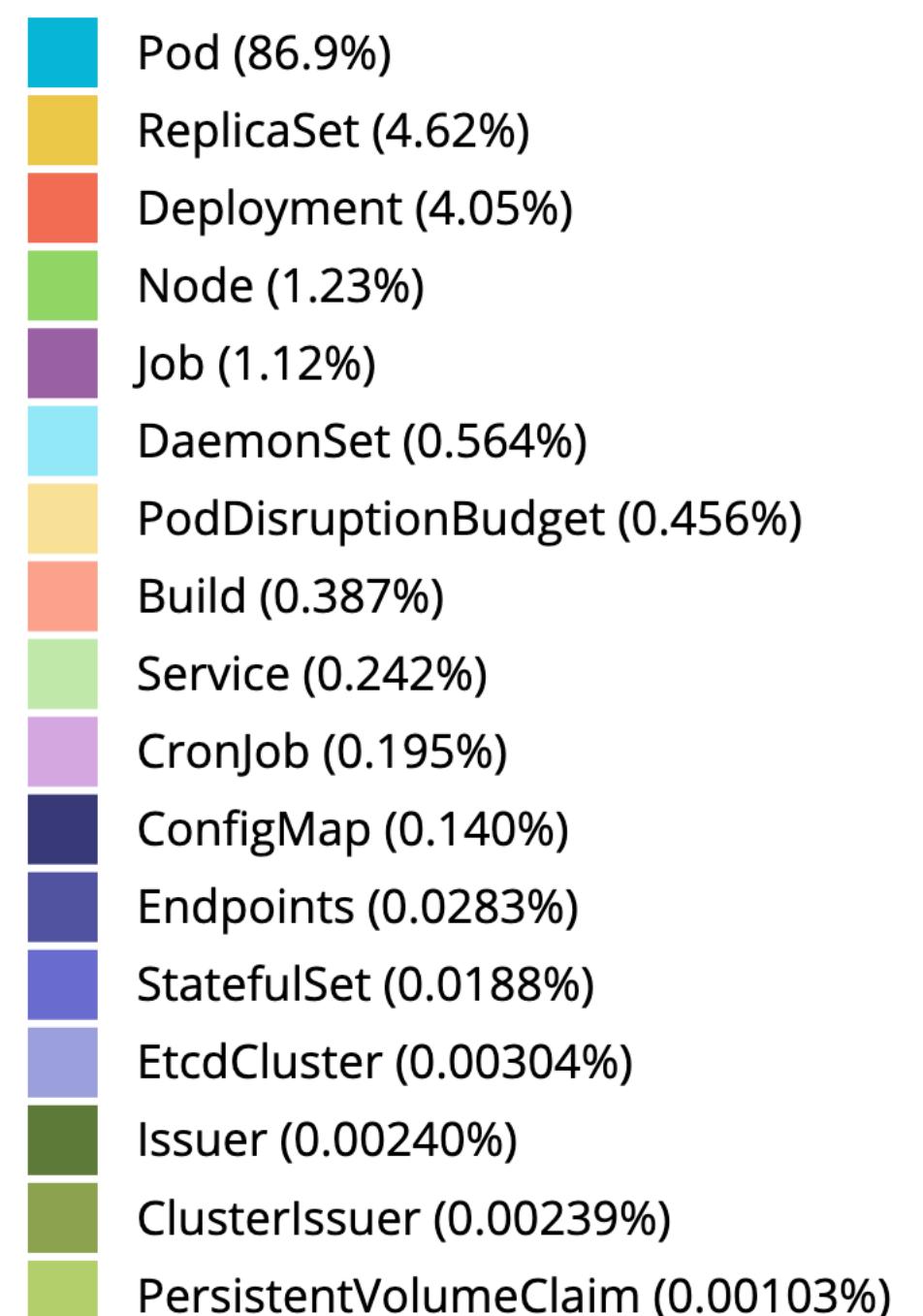
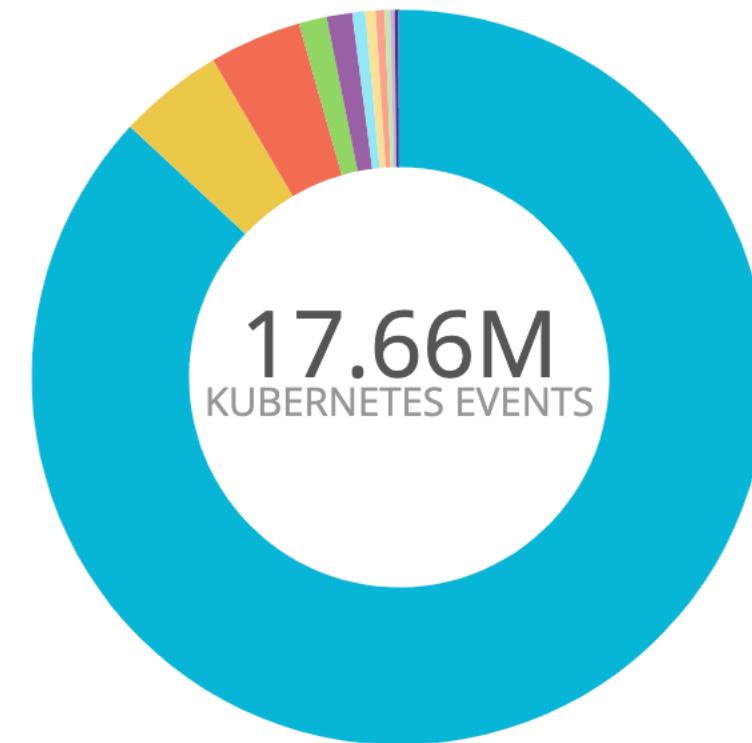
ReplicaSet & Deployments

Since they are main Pod controllers, they publish a lot of events

Node

They come and go with *cluster-autoscaler* and their health might fluctuate if you are not careful enough.

Since 1 day ago | Lab Shared



DISTRIBUTION OF EVENTS PER REASON / STABLE CLUSTER

Pod Creation

SuccessfulCreate, Started, Scheduled, Created, Pulled is all related to new Pods

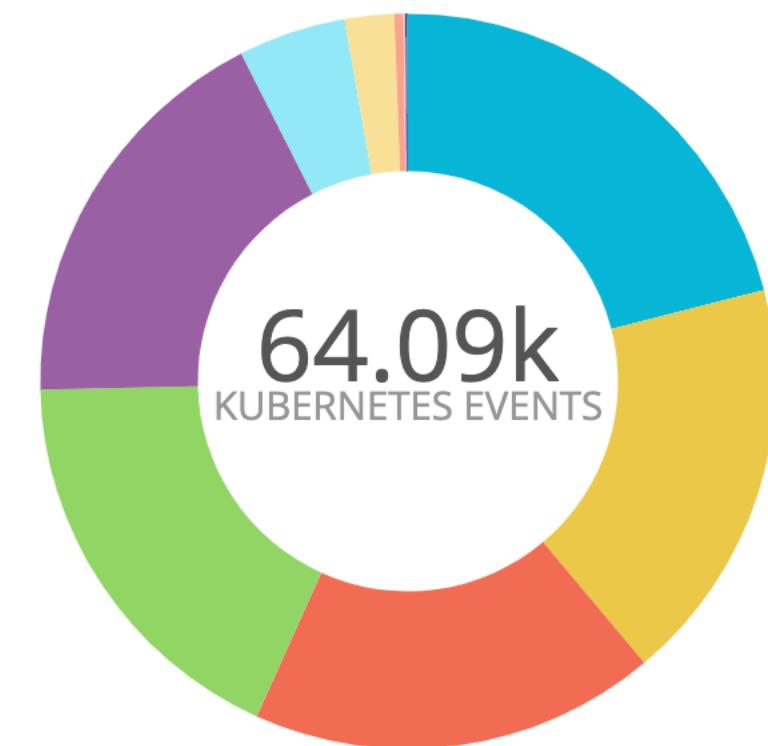
Jobs

CronJob & Jobs also states are published as events.

Node

They come and go with *cluster-autoscaler* and their health might fluctuate if you are not careful enough.

Since 1 day ago



SuccessfulCreate (21.0%)
Started (17.9%)
Scheduled (17.9%)
Created (17.9%)
Pulled (17.9%)
SawCompletedJob (4.73%)
NoControllers (2.17%)
SuccessfulDelete (0.403%)
SuccessfulRescale (0.0468%)
ScalingReplicaSet (0.0468%)
CalculateExpectedPodCountFailed (0.0421%)
Killing (0.0265%)
UnexpectedJob (0.0203%)
Pulling (0.00624%)
DeadlineExceeded (0.00312%)

DISTRIBUTION OF EVENTS PER REASON / DEV CLUSTER

Unhealthy & Readiness

Readiness probes might need tweaking in dev environments.

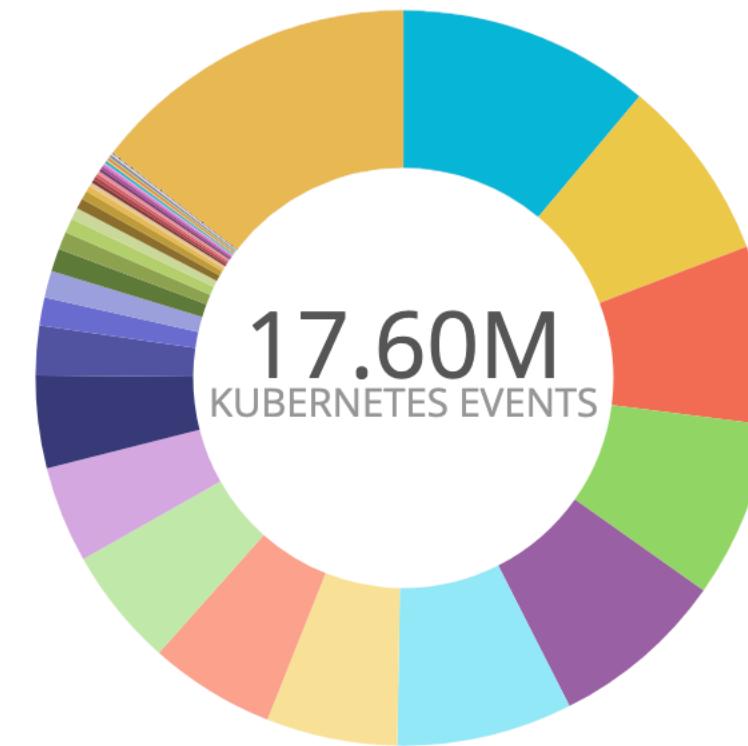
Schedule Fails & Pre-emption

High-volume pod schedule and keeping costs

Sandbox

CNI, Docker & Kubelet bugs happen at scale and you can miss them.

Since 1 day ago | Lab Shared



Unhealthy (11.1%)
BackOff (8.08%)
Pulling (7.85%)
Pulled (7.75%)
Created (7.74%)
Started (7.72%)
Killing (5.77%)
FailedScheduling (5.54%)
Scheduled (5.24%)
SuccessfulCreate (4.25%)
ScalingReplicaSet (4.06%)
SuccessfulDelete (2.18%)
FailedCreatePodSandBox (1.24%)
SandboxChanged (1.17%)
NotTriggerScaleUp (1.01%)
Preempted (0.775%)
TriggeredScaleUp (0.640%)

Proper Monitoring and Alerting

What should be an alert and notification from those K8s Events?

EVENTS -> ALERT

Alert should be
designed for
human
consumption.



EVENTS -> ALERT

They should be
structured,
precise,
actionable and
noise-free.



“Back-off Restarting Failed Container”

Involved Object: Pod name

This event has happened for a Pod and should be included in the message.

Namespace

We utilize namespaces for organization and segregation and it can specify importance.

Labels

Events normally do not have labels, but we fetch them in our tool for embedding more information, so they can be routed correctly.

MONITORING & OBSERVABILITY

We can also extract information and metrics from the events for extra observability.





Alerts

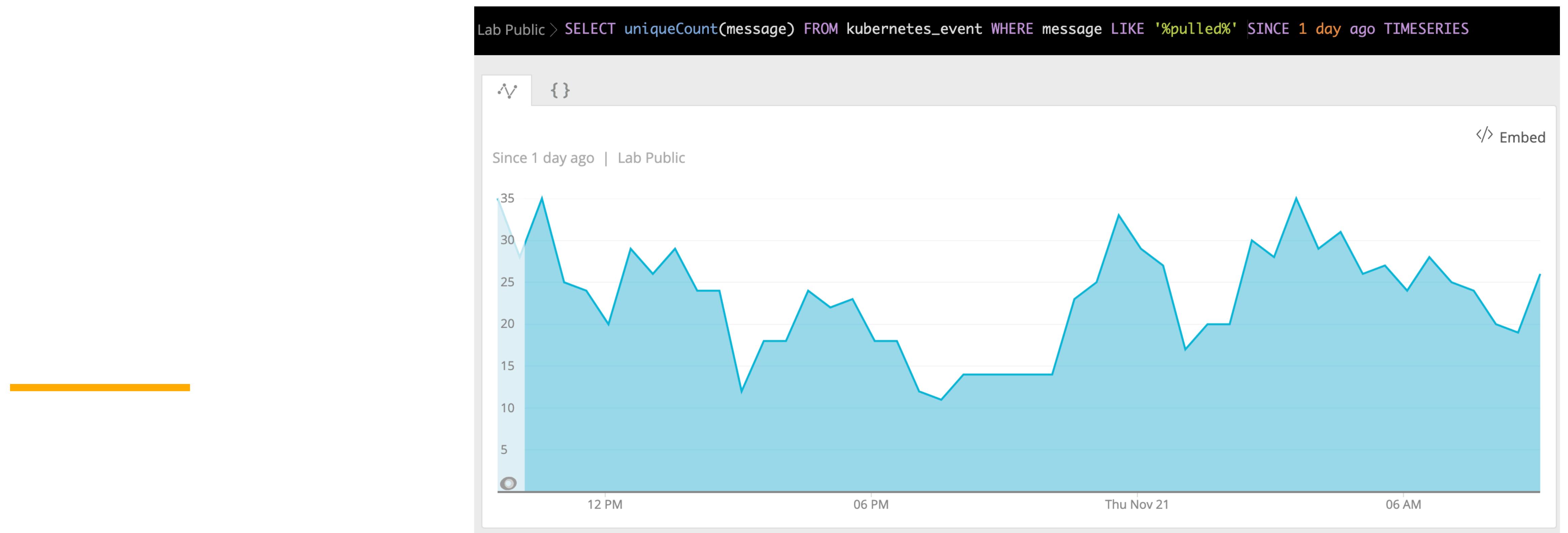
Usually Warnings and some Information can be transformed into alerts for human consumption.



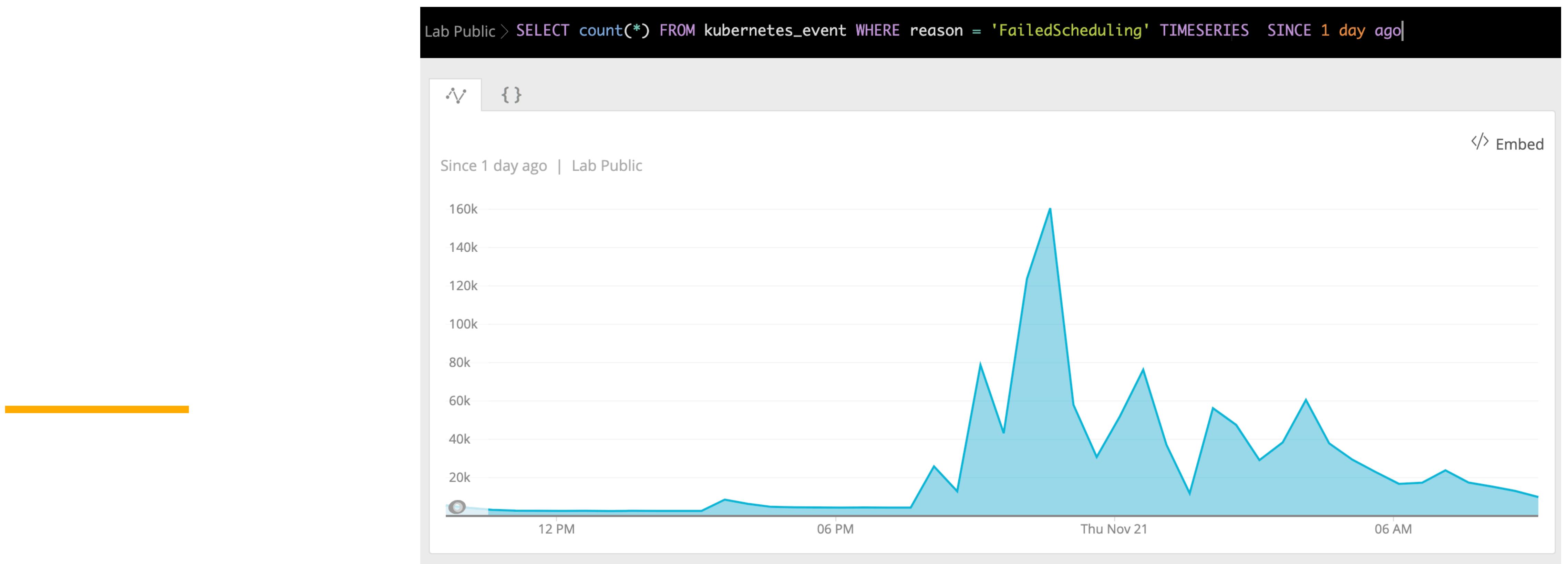
Monitoring

Aggregation & filtering of many events over a time-series can give you uncharted information about the state of cluster.

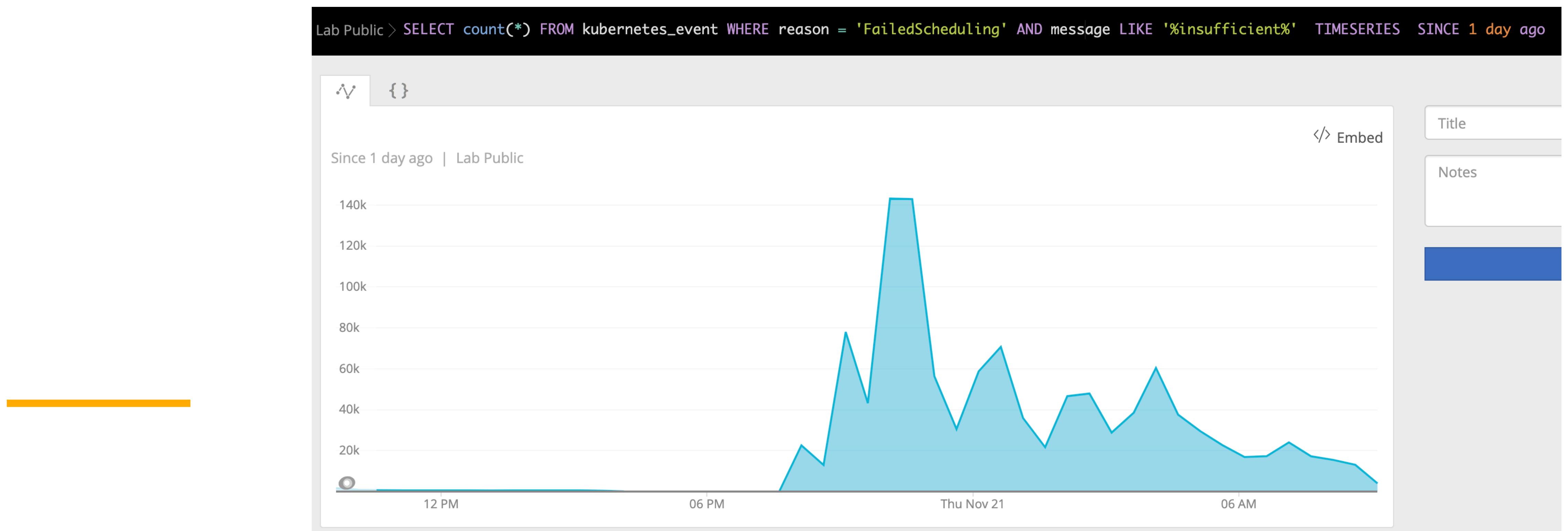
How many different images are pulled hourly?



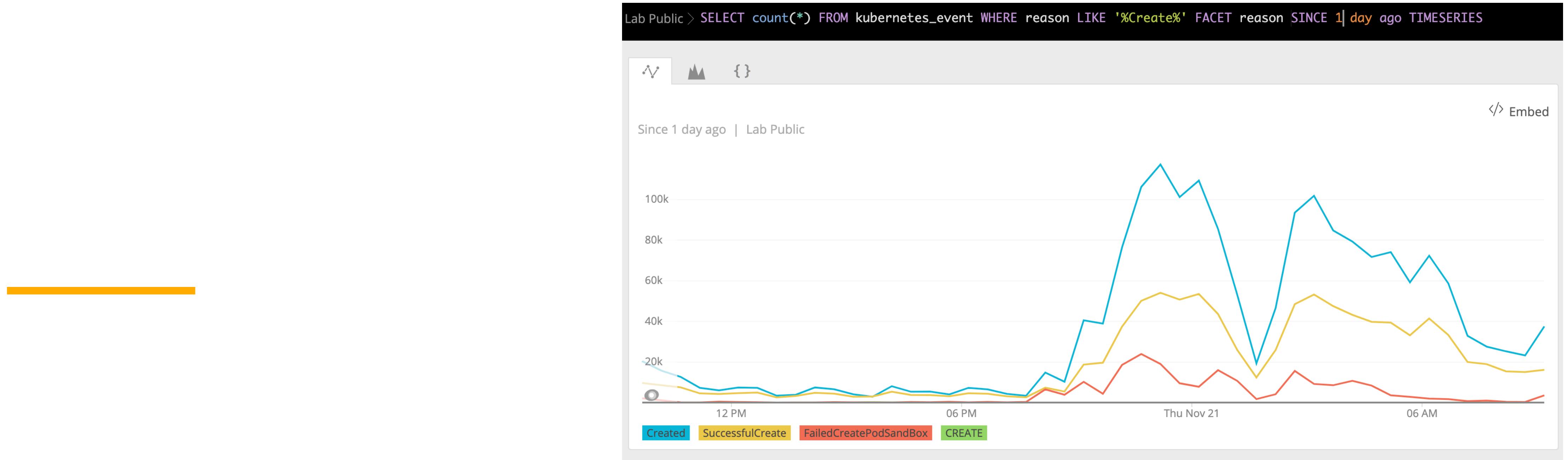
How often pods cannot be scheduled?



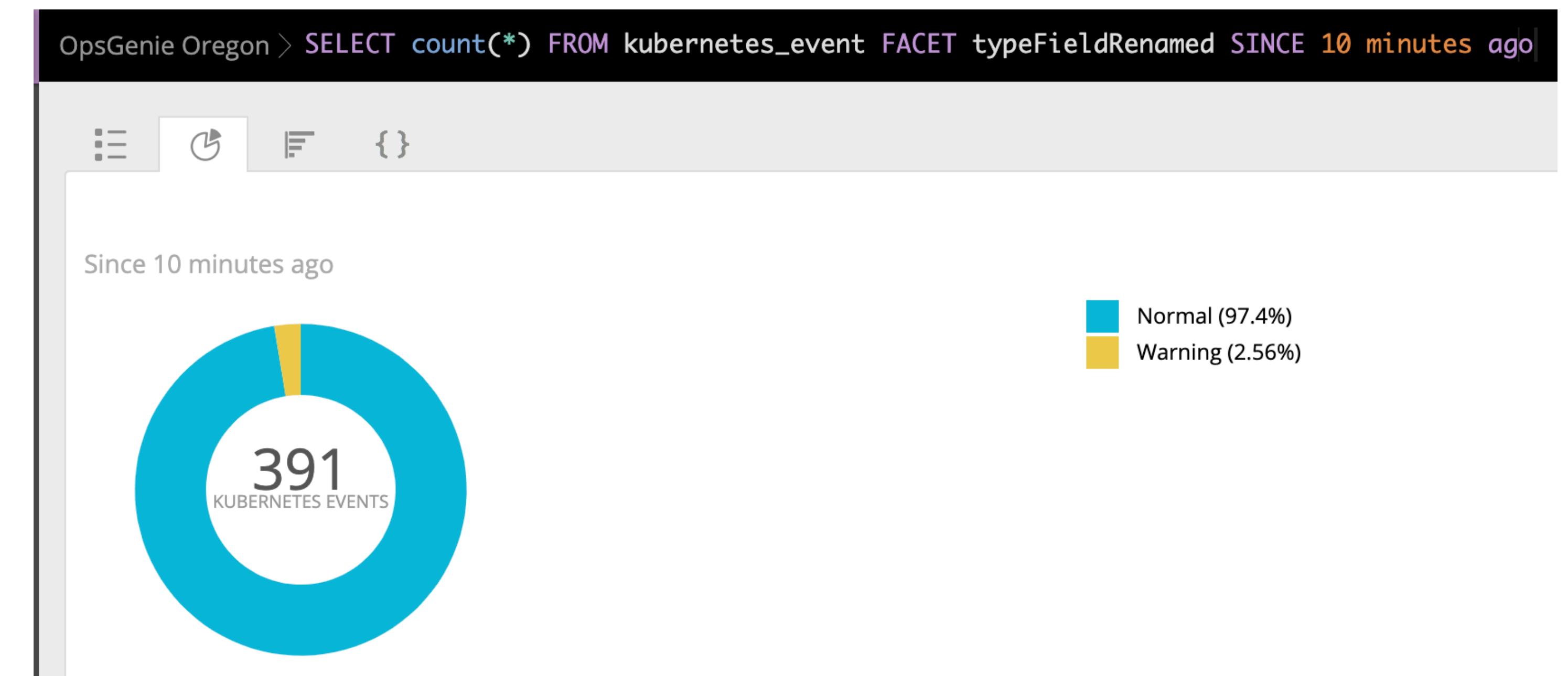
Are they related to capacity errors?



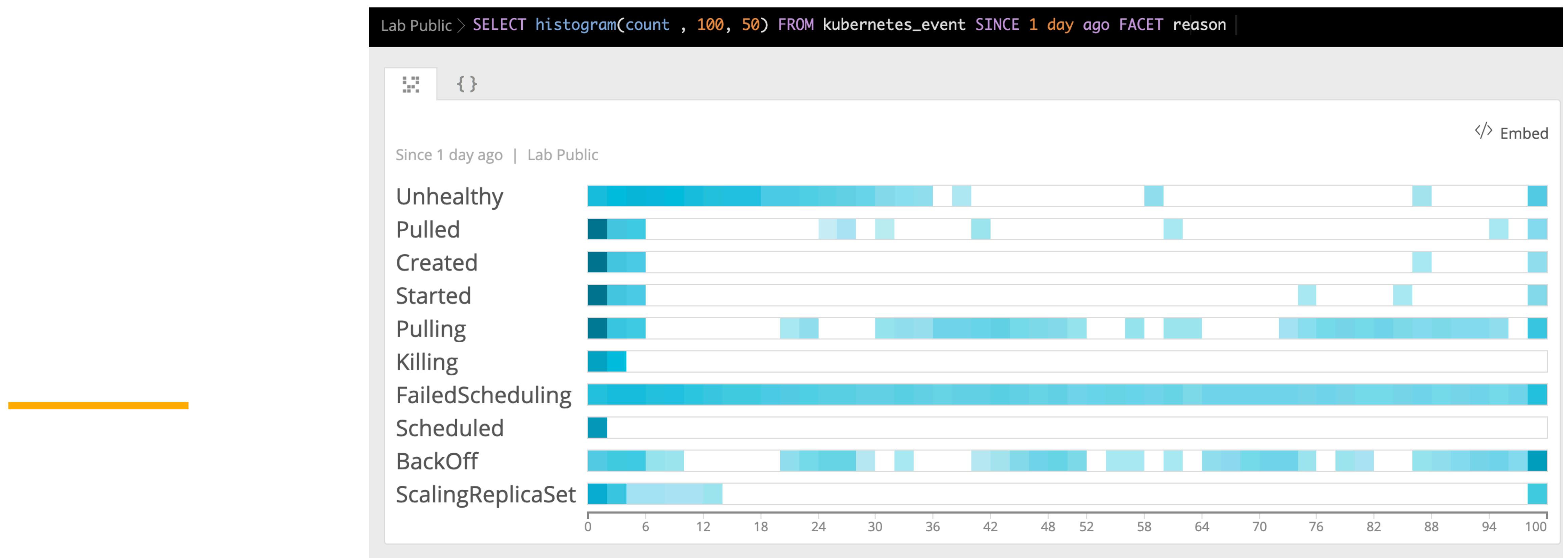
What is the distribution of Pods, Deployments Created/ Updated through out the day?



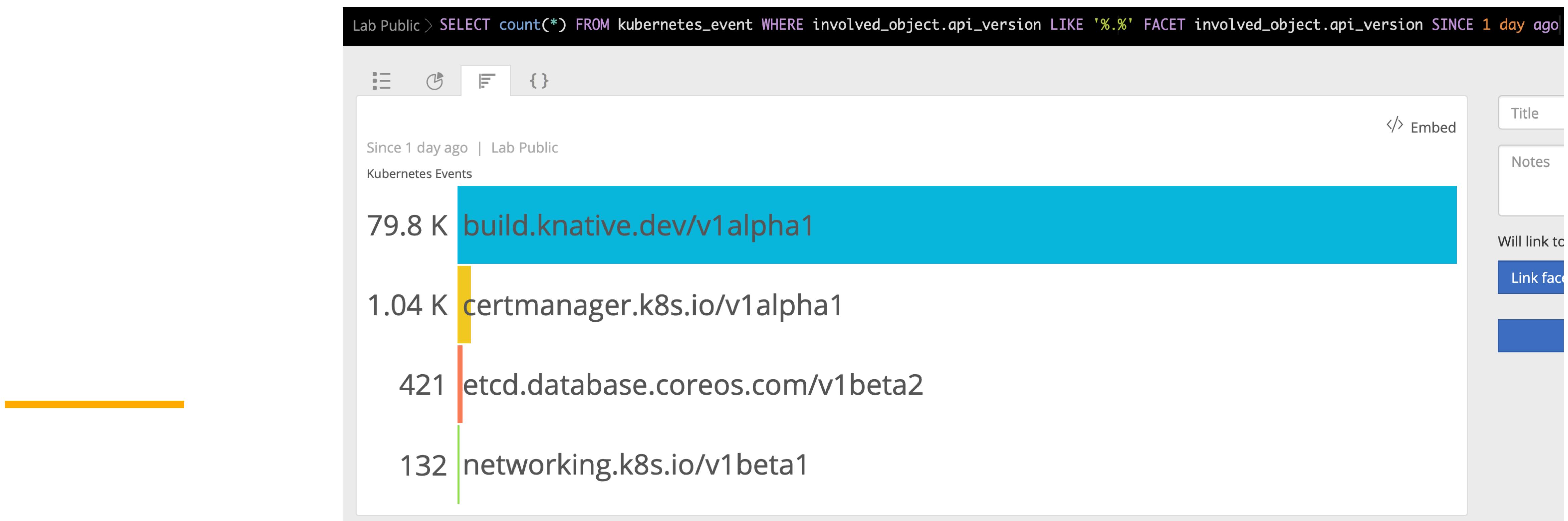
Are there significant amount of 'Warning' events right now?



What does the distribution of event counts per reason?



How many times did your *Custom Event* occur?



Is Cluster Autoscaler publishing interesting events?

Lab Public > `SELECT message FROM kubernetes_event WHERE involved_object.name='cluster-autoscaler-status' AND metadata.namespace = 'kube-system'`

TIMESTAMP	MESSAGE
21 Nov 09:41:34	(combined from similar events): Scale-down: removing node ip-10-35-121-132.us-west-2.compute.in...
21 Nov 09:41:34	Scale-down: removing node ip-10-35-106-248.us-west-2.compute.internal, utilization: {0.090875 0.07...
21 Nov 09:41:34	Scale-down: removing node ip-10-35-5-61.us-west-2.compute.internal, utilization: {0.090875 0.07797...
21 Nov 09:41:34	Scale-down: removing node ip-10-35-123-77.us-west-2.compute.internal, utilization: {0.090875 0.076...
21 Nov 09:41:34	Scale-down: removing node ip-10-35-48-125.us-west-2.compute.internal, utilization: {0.240875 0.279...
21 Nov 09:41:34	Scale-down: removing node ip-10-35-67-59.us-west-2.compute.internal, utilization: {0.278375 0.3392...
21 Nov 09:41:34	Scale-down: empty node ip-10-35-87-129.us-west-2.compute.internal removed
21 Nov 09:41:34	Scale-down: removing node ip-10-35-11-123.us-west-2.compute.internal, utilization: {0.090875 0.076...
21 Nov 09:41:34	Scale-down: empty node ip-10-35-13-139.us-west-2.compute.internal removed
21 Nov 09:41:34	Scale-down: removing empty node ip-10-35-46-73.us-west-2.compute.internal

Since 60 minutes ago | Lab Public

Embed CSV

Title Notes

Is cert-manager able to renew certificates properly?

```
Lab Public > SELECT uniques(message) FROM kubernetes_event WHERE involved_object.api_version = 'certmanager.k8s.io/v1alpha1' SINCE 7 day ago
```

Since 7 days ago | Lab Public

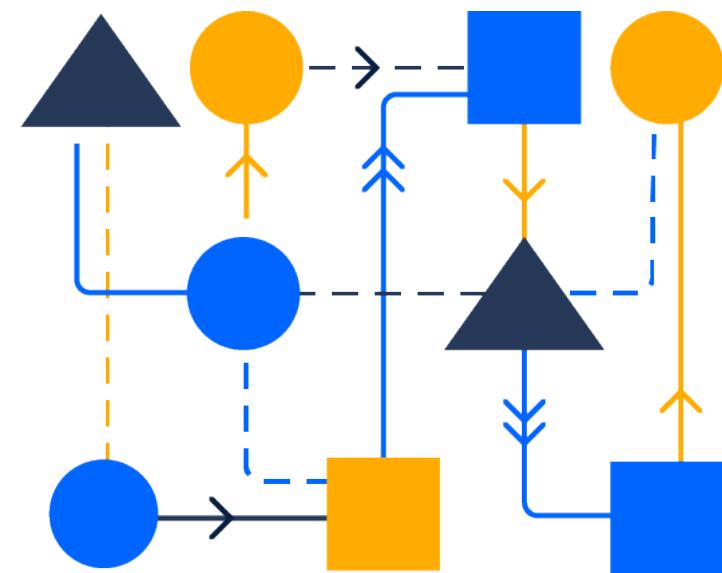
Signing CA verified

Failed to update ACME account:acme: urn:ietf:params:acme:error:invalidEmail: Unable to update account :: email domain name does not end in a IANA suffix

Event Exporter Tool

The implementation details and output types

Features



Event Routing & Filtering

Events are received from a single endpoint and filtered and routed based on their fields with regexes to route relevant events.



Multiple Outputs

Each output has different use cases, so tool allows using all of them with the routing rules to avoid deploying multiple instances.



Payload Customization

The pushed data can be customized to fit custom needs so that it can be easily embedded in the monitoring stack of many users.

Outputs



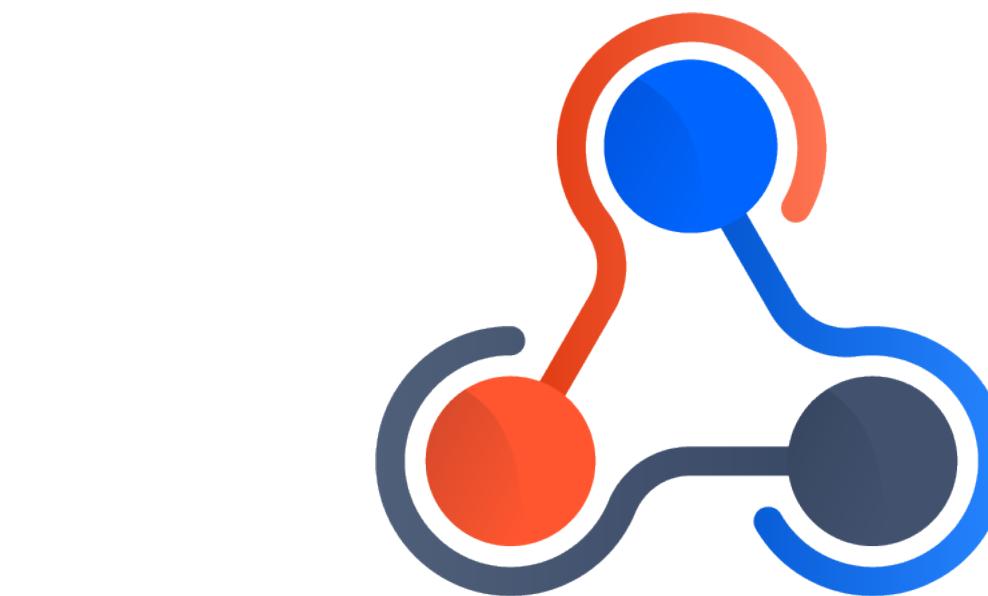
Reporting

When you want to export more data



Alerting

The critical events for the eyes of the on-call



Notification

Push some of events for notification, extra processing



An example Alert in Opsgenie

Alerts

P3 Nov 21, 2019 10:25 AM - No Owner Event FailedScheduling for mcatal/neo4j-core-0 on K8s cluster x1 #7 event FailedScheduling neo4j-core-0 Pod + Add tag OPEN

Details Activity log Responder states

Source 104.192.139.233 ELAPSED TIME 0h 0m 10s

Integration test_API (API)

Responders test

Owner Team test

Alias 3e7de5c2-0c8c-11ea-9332-06c74e6d2df8

Last Updated At Nov 21, 2019 10:25 AM

Description

```
{ "metadata": { "name": "neo4j-core-0.15d940a822ba9de4", "namespace": "mcatal", "selfLink": "/api/v1/namespaces/mcatal/events/neo4j-core-0.15d940a822ba9de4", "uid": "3e7de5c2-0c8c-11ea-9332-06c74e6d2df8", "resourceVersion": "310696000", "creationTimestamp": "2019-11-21T18:25:07Z" }, "reason": "FailedScheduling", "message": "0/42 nodes are available: 42 node(s) didn't match node selector.", "source": { "component": "default-scheduler" }, "firstTimestamp": "2019-11-21T18:23:55Z", "lastTimestamp": "2019-11-21T18:25:07Z", "count": 13, "type": "Warning", "eventTime": null, "reportingComponent": "", "reportingInstance": "", "involvedObject": { "kind": "Pod", "namespace": "mcatal", "name": "neo4j-core-0", "uid": "e0773cd7-0bf1-11ea-9332-06c74e6d2df8", "apiVersion": "v1", "resourceVersion": "310695948", "labels": { "app": "neo4j", "component": "core", "controller-revision-hash": "neo4j-core-6bd969d6b8", "opsgenie.com/role": "devopspoc", "statefulset.kubernetes.io/pod-name": "neo4j-core-0" } } }
```

Priority P3 - Moderate

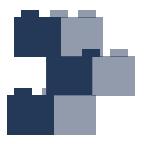
Extra Properties + Add extra property

Implementation



Watcher

Writing a Kubernetes resource watcher requires some care



Generic Client

Events are enriched with objects labels to be used in routing and filtering



Output buffering

Many types of outputs, we've tried to utilized goroutines efficiently.

Configuration

Routing

Outputs

Payload
Customization

```
route:  
  match:  
    - receiver: dump  
routes:  
  - drop:  
    - namespace: test*  
    - type: Normal  
  - match:  
    - receiver: slack  
      kind: Pod  
    - receiver: alert  
      kind: Pod  
      namespace: prod  
      reason: "Failed*"
```

Configuration

Routing

Outputs

Payload

Customization

```
- name: personal-message
slack:
  apiKey: "xoxo-12345"
  channel: "{{ .InvolvedObject.Labels.Owner }}"
  message: "Your pod has a msg {{ .InvolvedObject.Name }}"

- name: dump
elasticsearch:
  addresses:
    - http://localhost:9200
index: kubernetes-events

- name: high-priority-alert
opsgenie:
  apikey: xxx
  priority: "P3"
  message: "Event {{ .Reason }} for {{ .InvolvedObject.Namespace }}/{{ .InvolvedObject.Name }} on K8s cluster"
  alias: "{{ .UID }}"
  description: "<pre>{{ toPrettyJson . }}</pre>"
tags:
  - "event"
  - "{{ .Reason }}"
  - "{{ .InvolvedObject.Kind }}"
  - "{{ .InvolvedObject.Name }}"
```

Configuration

Routing

Outputs

Payload
Customization

```
- name: appMetric
  kinesis:
    region: us-west-2
    streamname: applicationMetric
    layout:
      region: "us-west-2"
      eventType: "kubeevent"
      createdAt: "{{ .GetTimestampMs }}"
      details:
        message: "{{ .Message }}"
        reason: "{{ .Reason }}"
        type: "{{ .Type }}"
        count: "{{ .Count }}"
        kind: "{{ .InvolvedObject.Kind }}"
        name: "{{ .InvolvedObject.Name }}"
        namespace: "{{ .Namespace }}"
        component: "{{ .Source.Component }}"
        host: "{{ .Source.Host }}"
        labels: "{{ toJson .InvolvedObject.Labels}}"
```

Where did this project come from?

Attended KubeCon '19 Barcelona

We loved everyone sharing experiences and their tooling in an open and welcoming environment

Open Source an In-House Project

We already have many tools to improve our own observability and wanted to share our experience with the whole world as a generic tool.

More to Come

We loved open-sourcing our stuff to share with the community, and we are working on sharing our more internal projects.

Next: Alternative Kubernetes Dashboard,
Golang Batching Library



Thanks for joining us!
Any questions, comments?

<https://github.com/opsgenie/kubernetes-event-exporter>

Meet us at the booth!

 **ATLASSIAN BOOTH G20**