Grafana Loki: *Like Prometheus, but for logs.*

Tom Wilkie, Feb 2019

Grafana Labs

Demo

Tom Wilkie *VP Product, Grafana Labs*

Previously*: Kausal, Weaveworks, Google, Acunu, Xensource*

Prometheus & Cortex maintainer, mixins authors etc

Twitter: *@tom_wilkie*   Email: *tom@grafana.com*

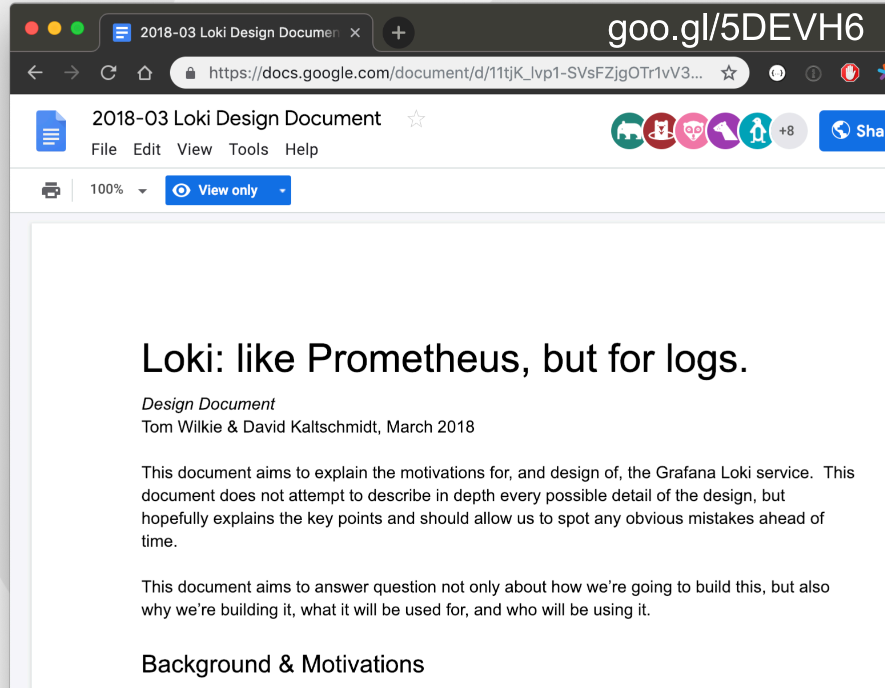
Grafana Labs

Loki is a horizontally-scalable, highly-available, multi-tenant log aggregation system inspired by Prometheus.

03/18	Project started
12/18	Launched at KubeCon
12/18	#1 on HN for ~12hrs!
01/19	~5k GitHub stars

https://github.com/grafana/loki

goo.gl/5DEVH6


2018-03 Loki Design Documen
https://docs.google.com/document/d/11tjK_lvp1-SVsFZjgOTr1vV3...
2018-03 Loki Design Document
File
Edit
View
Tools
Help
+8
Sha
100%
View only

Loki: like Prometheus, but for logs.

Design Document
Tom Wilkie & David Kaltschmidt, March 2018

This document aims to explain the motivations for, and design of, the Grafana Loki service. This document does not attempt to describe in depth every possible detail of the design, but hopefully explains the key points and should allow us to spot any obvious mistakes ahead of time.

This document aims to answer question not only about how we're going to build this, but also why we're building it, what it will be used for, and who will be using it.

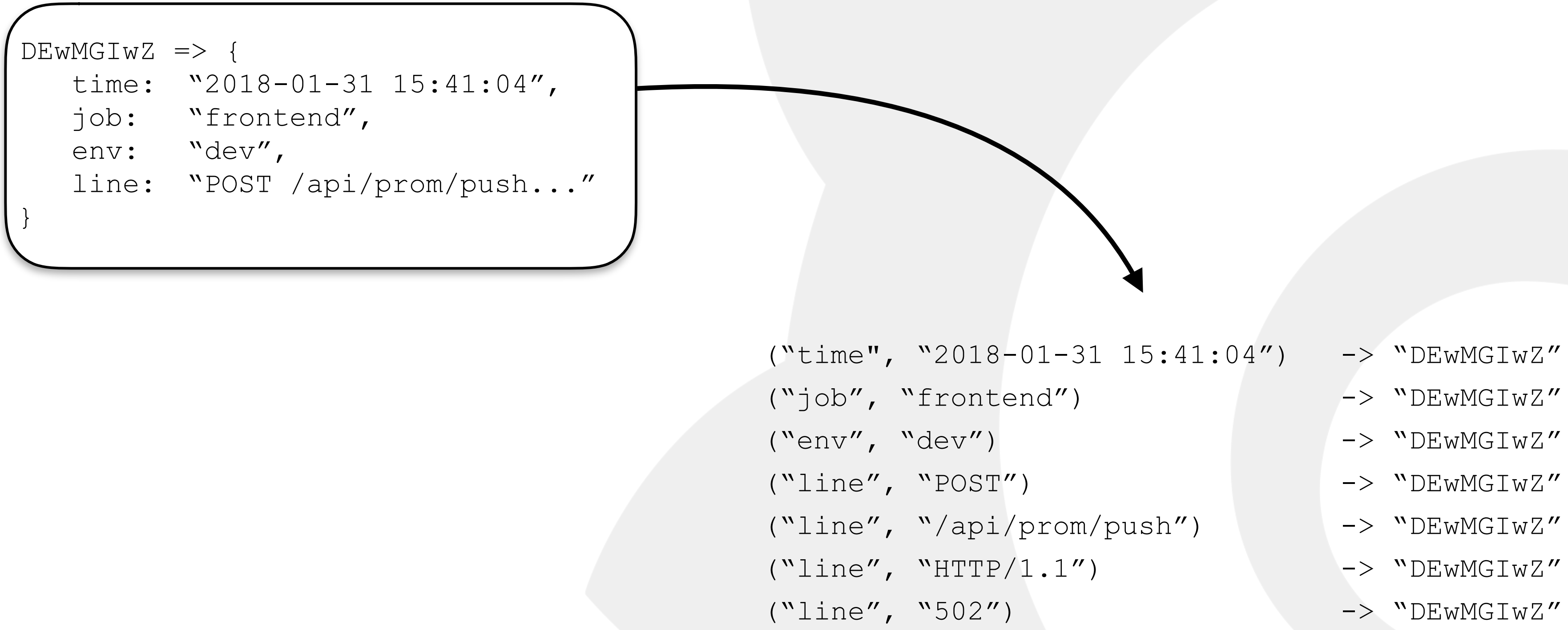
Background & Motivations

#0 Simple and cost effective to operate

#1 Integrated with existing observability tools

#2 Cloud Native and Airplane Friendly

#0 Simple to scale

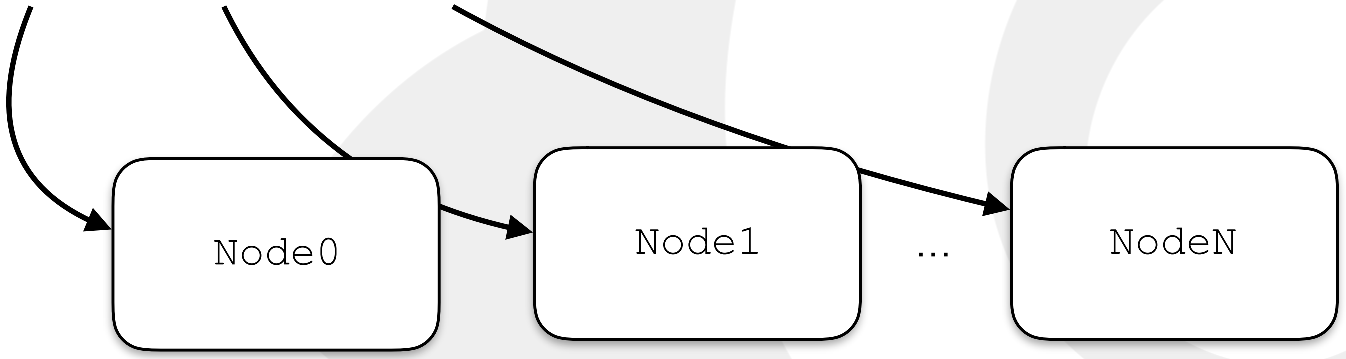


```
DEwMGIwZ => {
   time:  “2018-01-31 15:41:04”,
   job:   “frontend”,
   env:   “dev”,
   line:  “POST /api/prom/push...”
}
```

```
(“time", “2018-01-31 15:41:04”)   -> “DEwMGIwZ”
(“job”, “frontend”)               -> “DEwMGIwZ”
(“env”, “dev”)                    -> “DEwMGIwZ”
(“line”, “POST”)                  -> “DEwMGIwZ”
(“line”, “/api/prom/push”)        -> “DEwMGIwZ”
(“line”, “HTTP/1.1”)              -> “DEwMGIwZ”
(“line”, “502”)                   -> “DEwMGIwZ”
```

Existing log aggregation systems do full text indexing and support complex queries

```
("time", "2018-01-31 15:41:04")   -> "DEwMGIwZ"
("job", "frontend")               -> "DEwMGIwZ"
("env", "dev")                    -> "DEwMGIwZ"
("line", "POST")                  -> "DEwMGIwZ"
("line", "/api/prom/push")        -> "DEwMGIwZ"
("line", "HTTP/1.1")              -> "DEwMGIwZ"
("line", "502")                   -> "DEwMGIwZ"
```



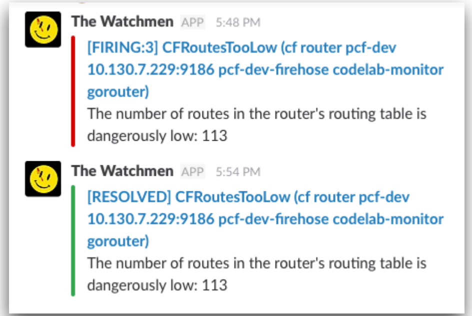
Existing log aggregation systems do full text indexing and support complex queries

```
{job=“frontend”, env=“dev”} => {
  time:  “2018-01-31 15:41:04”,
  line:  “POST /api/prom/push HTTP/1.1 502 0"
}
```

Loki doesn’t index the text of the logs, instead grouping entries into “streams” and indexing those with labels.

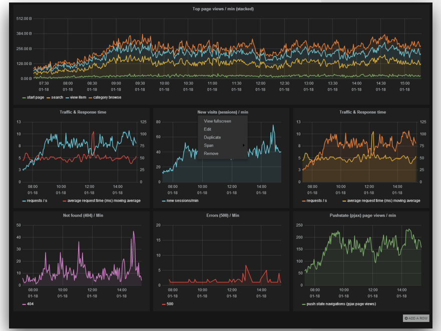
#1 Integrated with existing tools

1. Alert

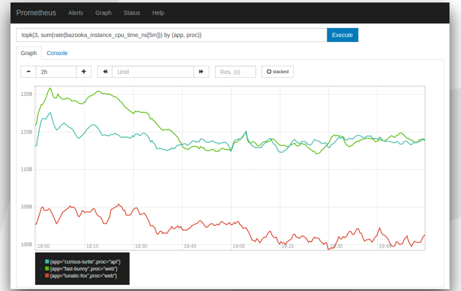

The Watchmen APP 5:48 PM
[FIRING:3] CFRoutesTooLow (cf router pcf-dev 10.130.7.229:9186 pcf-dev-firehose codelab-monitor gorouter)
The number of routes in the router's routing table is dangerously low: 113

The Watchmen APP 5:54 PM
[RESOLVED] CFRoutesTooLow (cf router pcf-dev 10.130.7.229:9186 pcf-dev-firehose codelab-monitor gorouter)
The number of routes in the router's routing table is dangerously low: 113

2. Dashboard


Top page views / min (stacked)
Traffic & Response time
New visits (sessions) / min
Traffic & Response time
Not found (404) / Min
Errors (500) / Min
Pushstate (pjax) page views / min

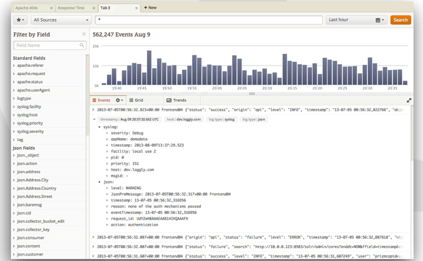
3. Adhoc Query



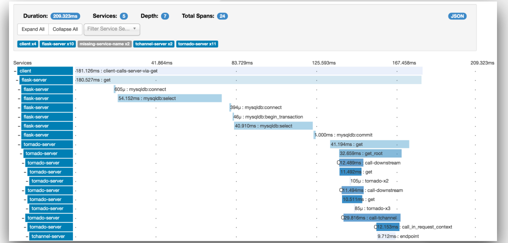
```
topk(3, sum(rate(bazooka_instance_cpu_time_ns[5m])) by (app, proc))
```

🟦 {app="curious-turtle",proc="api"}
🟩 {app="fast-bunny",proc="web"}
🟥 {app="lunatic-fox",proc="web"}

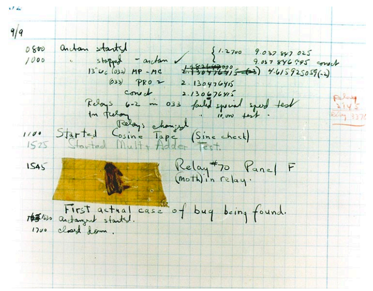
4. Log Aggregation



5. Distributed Tracing



Fix!


9/9

0800 antan started
1000 " stopped - antan ✓ {1.2700 9.037 847 025
9.037 846 995 conect
13"uc (032) MP - MC ~~1.982147000~~ ~~2.130476415~~ ~~(-3)~~ 4.615925059(-2)
(033) PRO 2 2.130476415
conect 2.130676415
Relays 6-2 in 033 failed special speed test
in relay " 10,000 test .
Relays changed
1100 Started Cosine Tape (Sine check)
1525 Started Mult+ Adder Test.
1545 Relay #70 Panel F
(moth) in relay.
First actual case of bug being found.
~~1630~~ 1630 antangent started.
1700 closed down.

Relay 2145
Relay 3370

Prometheus' data model is very simple:

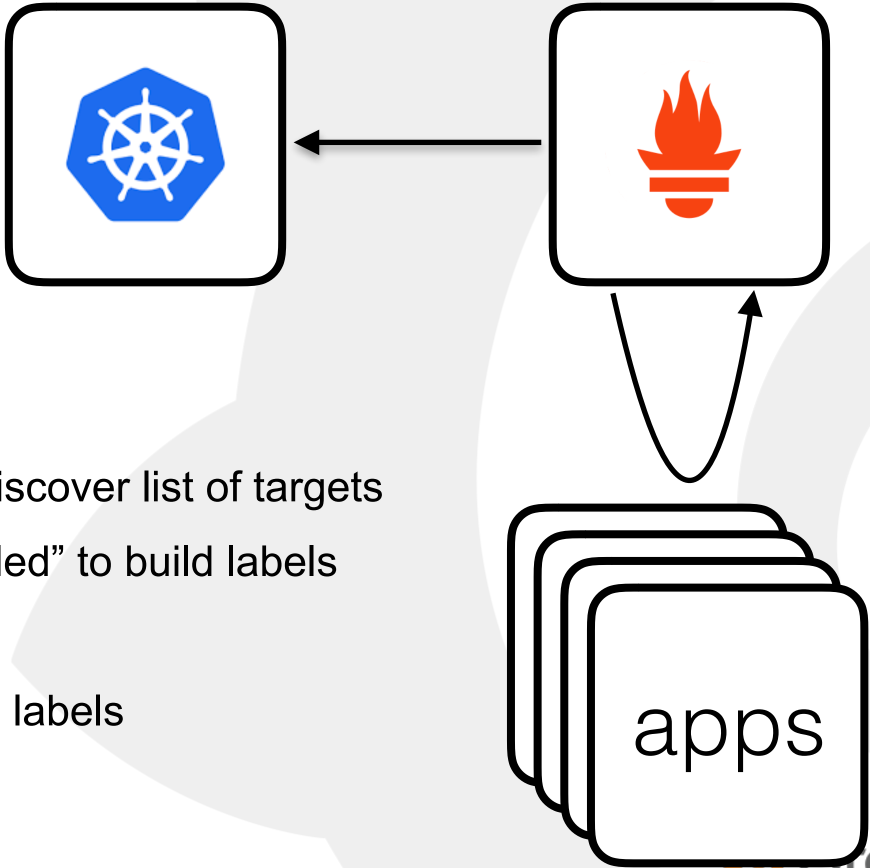
```
<identifier> → [ (t0, v0), (t1, v1), ... ]
```

Timestamps are millisecond int64, values are float64

Identifiers are bags of (label, value) pairs:

```
{job="foo", instance="bar", ... }
```

https://www.slideshare.net/Docker/monitoring-the-prometheus-way-julius-voltz-prometheus



#0 Prometheus talks to k8s to discover list of targets

#1 Target information is “relabelled” to build labels

#2 Metrics are pulled from apps

#3 Target labels added to series labels

What is Relabelling?

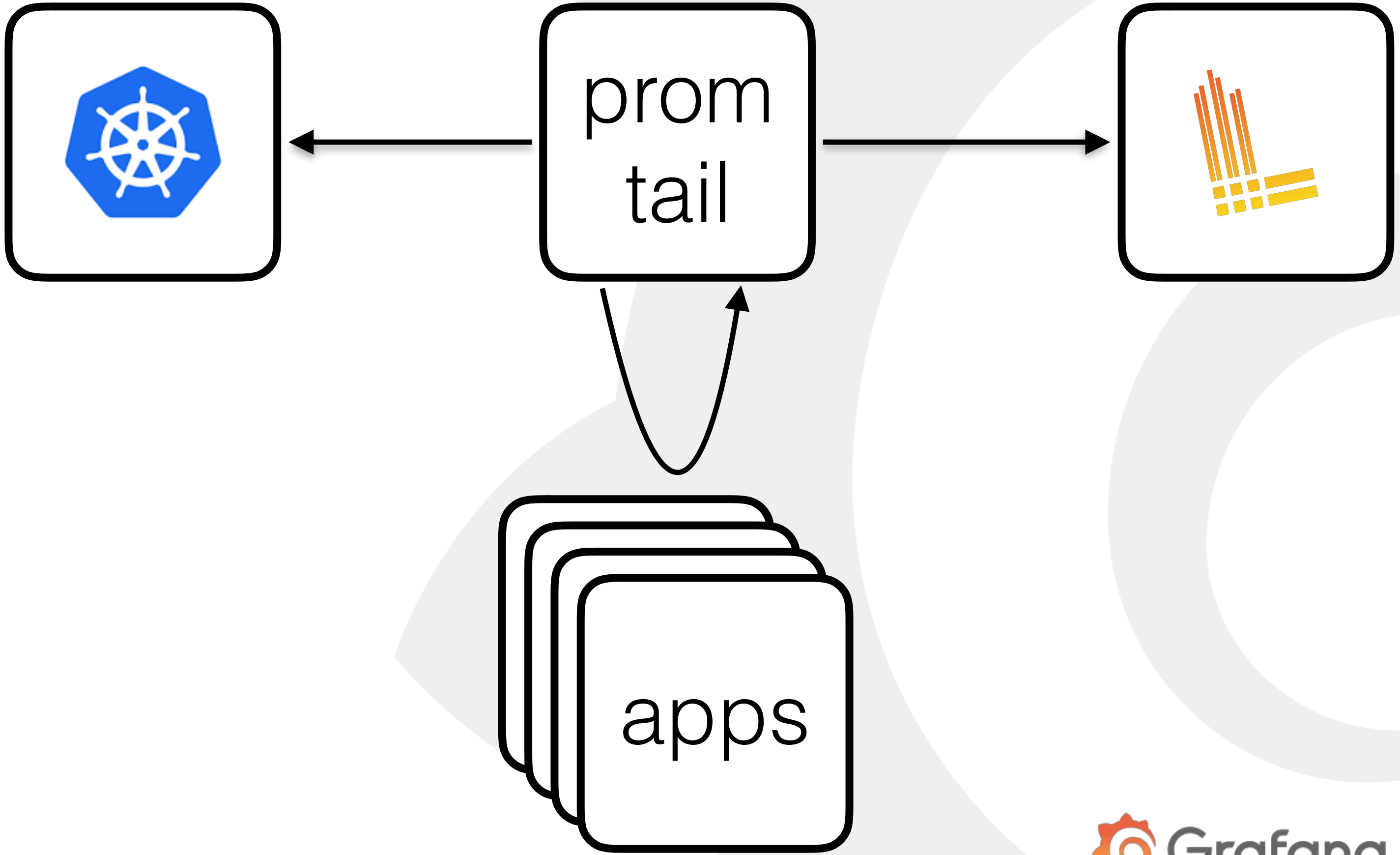
Loki’s data model is very similar:

```
<identifier> → [ (t0, v0), (t1, v1), ... ]
```

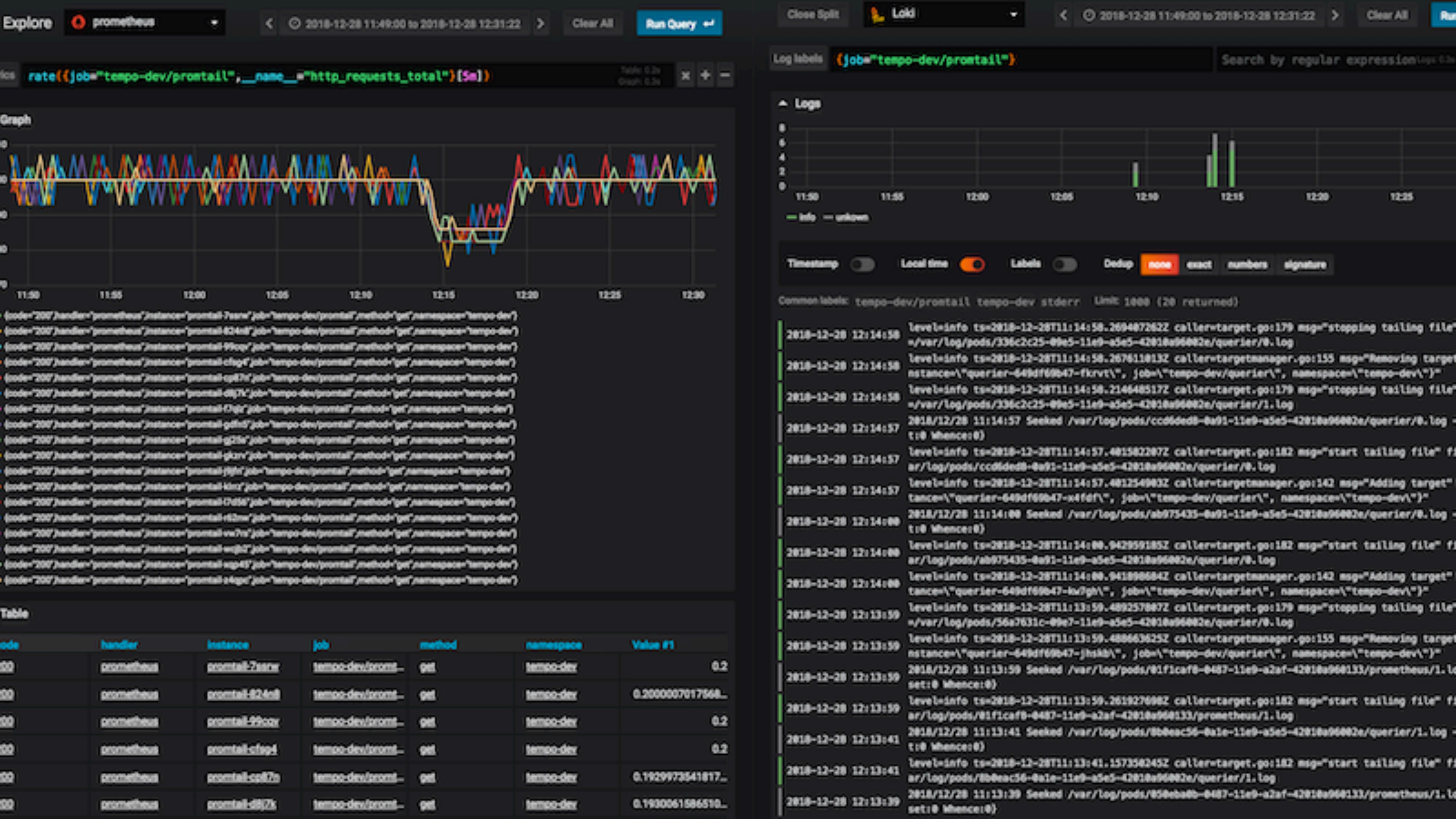
Timestamps are nanosecond floats, values are byte arrays.

Identifiers are the same - label sets.

https://www.slideshare.net/Docker/monitoring-the-prometheus-way-julius-voltz-prometheus



Grafana Labs



```
rate({job="tempo-dev/promtail",__name__="http_requests_total"}[5m])
```

```
{job="tempo-dev/promtail"}
```

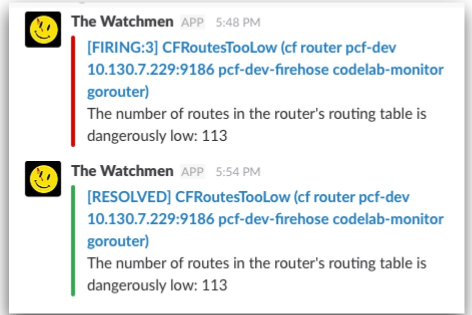
```
{code="200",handler="prometheus",instance="promtail-7ssrw",job="tempo-dev/promtail",method="get",namespace="tempo-dev"}
{code="200",handler="prometheus",instance="promtail-824n8",job="tempo-dev/promtail",method="get",namespace="tempo-dev"}
{code="200",handler="prometheus",instance="promtail-99cqv",job="tempo-dev/promtail",method="get",namespace="tempo-dev"}
{code="200",handler="prometheus",instance="promtail-cfsg4",job="tempo-dev/promtail",method="get",namespace="tempo-dev"}
{code="200",handler="prometheus",instance="promtail-cp87n",job="tempo-dev/promtail",method="get",namespace="tempo-dev"}
{code="200",handler="prometheus",instance="promtail-d8j7k",job="tempo-dev/promtail",method="get",namespace="tempo-dev"}
{code="200",handler="prometheus",instance="promtail-f7qjz",job="tempo-dev/promtail",method="get",namespace="tempo-dev"}
{code="200",handler="prometheus",instance="promtail-gdfn5",job="tempo-dev/promtail",method="get",namespace="tempo-dev"}
{code="200",handler="prometheus",instance="promtail-gj25e",job="tempo-dev/promtail",method="get",namespace="tempo-dev"}
{code="200",handler="prometheus",instance="promtail-gkzrv",job="tempo-dev/promtail",method="get",namespace="tempo-dev"}
{code="200",handler="prometheus",instance="promtail-j9jfn",job="tempo-dev/promtail",method="get",namespace="tempo-dev"}
{code="200",handler="prometheus",instance="promtail-kkrz",job="tempo-dev/promtail",method="get",namespace="tempo-dev"}
{code="200",handler="prometheus",instance="promtail-l7d56",job="tempo-dev/promtail",method="get",namespace="tempo-dev"}
{code="200",handler="prometheus",instance="promtail-r62nw",job="tempo-dev/promtail",method="get",namespace="tempo-dev"}
{code="200",handler="prometheus",instance="promtail-vw7rs",job="tempo-dev/promtail",method="get",namespace="tempo-dev"}
{code="200",handler="prometheus",instance="promtail-wcjb2",job="tempo-dev/promtail",method="get",namespace="tempo-dev"}
{code="200",handler="prometheus",instance="promtail-wsp45",job="tempo-dev/promtail",method="get",namespace="tempo-dev"}
{code="200",handler="prometheus",instance="promtail-z4gpx",job="tempo-dev/promtail",method="get",namespace="tempo-dev"}
```

ode	handler	instance	job	method	namespace	Value #1
00	prometheus	promtail-7ssrw	tempo-dev/promt...	get	tempo-dev	0.2
00	prometheus	promtail-824n8	tempo-dev/promt...	get	tempo-dev	0.2000007017568...
00	prometheus	promtail-99cqv	tempo-dev/promt...	get	tempo-dev	0.2
00	prometheus	promtail-cfsg4	tempo-dev/promt...	get	tempo-dev	0.2
00	prometheus	promtail-cp87n	tempo-dev/promt...	get	tempo-dev	0.1929973541817...
00	prometheus	promtail-d8j7k	tempo-dev/promt...	get	tempo-dev	0.1930061586510...

Common labels: tempo-dev/promtail tempo-dev stderr Limit: 1000 (20 returned)

```
2018-12-28 12:14:58  level=info ts=2018-12-28T11:14:58.269407262Z caller=target.go:179 msg="stopping tailing file"
                     =/var/log/pods/336c2c25-09e5-11e9-a5e5-42010a96002e/querier/0.log
2018-12-28 12:14:58  level=info ts=2018-12-28T11:14:58.267611013Z caller=targetmanager.go:155 msg="Removing target
                     nstance=\"querier-649df69b47-fkrvt\", job=\"tempo-dev/querier\", namespace=\"tempo-dev\"}"
2018-12-28 12:14:58  level=info ts=2018-12-28T11:14:58.214648517Z caller=target.go:179 msg="stopping tailing file"
                     =/var/log/pods/336c2c25-09e5-11e9-a5e5-42010a96002e/querier/1.log
2018-12-28 12:14:57  2018/12/28 11:14:57 Seeked /var/log/pods/ccd6ded8-0a91-11e9-a5e5-42010a96002e/querier/0.log -
                     t:0 Whence:0}
2018-12-28 12:14:57  level=info ts=2018-12-28T11:14:57.401582207Z caller=target.go:182 msg="start tailing file" fi
                     ar/log/pods/ccd6ded8-0a91-11e9-a5e5-42010a96002e/querier/0.log
2018-12-28 12:14:57  level=info ts=2018-12-28T11:14:57.401254903Z caller=targetmanager.go:142 msg="Adding target"
                     tance=\"querier-649df69b47-x4fdf\", job=\"tempo-dev/querier\", namespace=\"tempo-dev\"}"
2018-12-28 12:14:00  2018/12/28 11:14:00 Seeked /var/log/pods/ab975435-0a91-11e9-a5e5-42010a96002e/querier/0.log -
                     t:0 Whence:0}
2018-12-28 12:14:00  level=info ts=2018-12-28T11:14:00.942959185Z caller=target.go:182 msg="start tailing file" fi
                     ar/log/pods/ab975435-0a91-11e9-a5e5-42010a96002e/querier/0.log
2018-12-28 12:14:00  level=info ts=2018-12-28T11:14:00.941890684Z caller=targetmanager.go:142 msg="Adding target"
                     tance=\"querier-649df69b47-kw7gh\", job=\"tempo-dev/querier\", namespace=\"tempo-dev\"}"
2018-12-28 12:13:59  level=info ts=2018-12-28T11:13:59.489257807Z caller=target.go:179 msg="stopping tailing file"
                     =/var/log/pods/56a7631c-09e7-11e9-a5e5-42010a96002e/querier/0.log
2018-12-28 12:13:59  level=info ts=2018-12-28T11:13:59.488663625Z caller=targetmanager.go:155 msg="Removing target
                     nstance=\"querier-649df69b47-jhskb\", job=\"tempo-dev/querier\", namespace=\"tempo-dev\"}"
2018-12-28 12:13:59  2018/12/28 11:13:59 Seeked /var/log/pods/01f1caf0-0487-11e9-a2af-42010a960133/prometheus/1.lo
                     set:0 Whence:0}
2018-12-28 12:13:59  level=info ts=2018-12-28T11:13:59.261927698Z caller=target.go:182 msg="start tailing file" fi
                     ar/log/pods/01f1caf0-0487-11e9-a2af-42010a960133/prometheus/1.log
2018-12-28 12:13:41  2018/12/28 11:13:41 Seeked /var/log/pods/0b0eac56-0a1e-11e9-a5e5-42010a96002e/querier/1.log -
                     t:0 Whence:0}
2018-12-28 12:13:41  level=info ts=2018-12-28T11:13:41.157350245Z caller=target.go:182 msg="start tailing file" fi
                     ar/log/pods/0b0eac56-0a1e-11e9-a5e5-42010a96002e/querier/1.log
2018-12-28 12:13:39  2018/12/28 11:13:39 Seeked /var/log/pods/050eba0b-0487-11e9-a2af-42010a960133/prometheus/1.lo
                     set:0 Whence:0}
```

1. Alert



The Watchmen APP 5:48 PM

[FIRING:3] CFRoutesTooLow (cf router pcf-dev 10.130.7.229:9186 pcf-dev-firehose codelab-monitor gorouter)

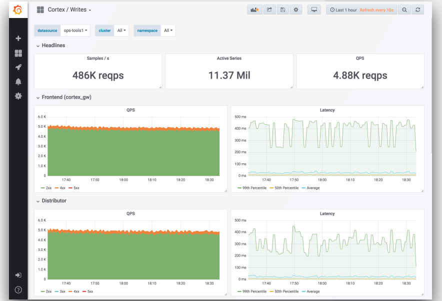
The number of routes in the router's routing table is dangerously low: 113

The Watchmen APP 5:54 PM

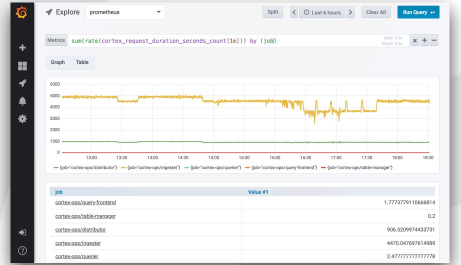
[RESOLVED] CFRoutesTooLow (cf router pcf-dev 10.130.7.229:9186 pcf-dev-firehose codelab-monitor gorouter)

The number of routes in the router's routing table is dangerously low: 113

2. Dashboard



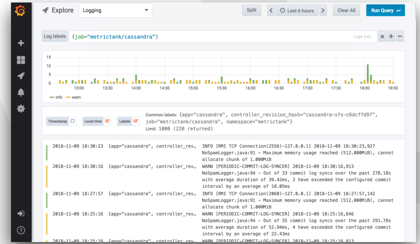
3. Adhoc Query


Explore
prometheus
Split
Last 6 hours
Clear All
Run Query
Metrics
Graph
Table
{job="cortex-ops/distributor"}
{job="cortex-ops/ingester"}
{job="cortex-ops/querier"}
{job="cortex-ops/query-frontend"}
{job="cortex-ops/table-manager"}

```
sum(rate(cortex_request_duration_seconds_count[1m])) by (job)
```

job	Value #1
cortex-ops/query-frontend	1.7773779110666814
cortex-ops/table-manager	0.2
cortex-ops/distributor	906.5209974433731
cortex-ops/ingester	4470.047697614989
cortex-ops/querier	2.4777777777777778

4. Log Aggregation

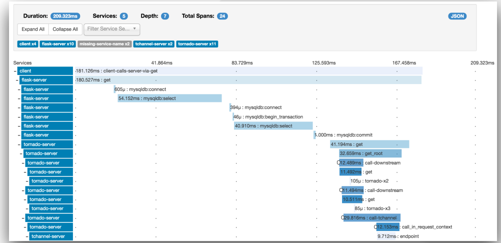


```
{job="metrictank/cassandra"}
```

```
Common labels: {app="cassandra", controller_revision_hash="cassandra-sfs-c6dcffd97",
job="metrictank/cassandra", namespace="metrictank"}
Limit: 1000 (228 returned)

2018-11-09 18:30:23  {app="cassandra", controller_rev…  INFO [RMI TCP Connection(2556)-127.0.0.1] 2018-11-09 18:30:23,927
                                                        NoSpamLogger.java:91 - Maximum memory usage reached (512.000MiB), cannot
                                                        allocate chunk of 1.000MiB
2018-11-09 18:30:16  {app="cassandra", controller_rev…  WARN [PERIODIC-COMMIT-LOG-SYNCER] 2018-11-09 18:30:16,913
                                                        NoSpamLogger.java:94 - Out of 33 commit log syncs over the past 278.18s
                                                        with average duration of 39.42ms, 2 have exceeded the configured commit
                                                        interval by an average of 18.05ms
2018-11-09 18:27:57  {app="cassandra", controller_rev…  INFO [RMI TCP Connection(2060)-127.0.0.1] 2018-11-09 18:27:57,142
                                                        NoSpamLogger.java:91 - Maximum memory usage reached (512.000MiB), cannot
                                                        allocate chunk of 1.000MiB
2018-11-09 18:25:16  {app="cassandra", controller_rev…  WARN [PERIODIC-COMMIT-LOG-SYNCER] 2018-11-09 18:25:16,846
                                                        NoSpamLogger.java:94 - Out of 35 commit log syncs over the past 291.78s
                                                        with average duration of 52.94ms, 4 have exceeded the configured commit
                                                        interval by an average of 22.43ms
2018-11-09 18:25:16  {app="cassandra", controller_rev…  WARN [PERIODIC-COMMIT-LOG-SYNCER] 2018-11-09 18:25:16,813
```

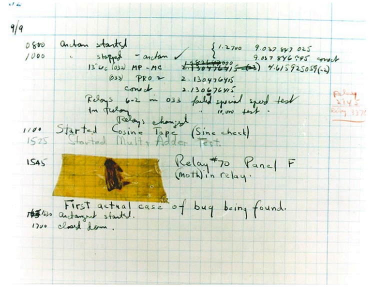
5. Distributed Tracing



Duration: 209.323ms Services: 5 Depth: 7 Total Spans: 24

```
client            181.126ms : client-calls-server-via-get
flask-server      180.527ms : get
flask-server      605μ : mysqldb:connect
flask-server      54.152ms : mysqldb:select
flask-server      394μ : mysqldb:connect
flask-server      46μ : mysqldb:begin_transaction
flask-server      40.910ms : mysqldb:select
flask-server      5.000ms : mysqldb:commit
tornado-server    41.194ms : get
tornado-server    32.659ms : get_root
tornado-server    12.489ms : call-downstream
tornado-server    11.492ms : get
tornado-server    105μ : tornado-x2
tornado-server    11.494ms : call-downstream
tornado-server    10.511ms : get
tornado-server    85μ : tornado-x3
tornado-server    29.816ms : call-tchannel
tornado-server    12.153ms : call_in_request_context
tchannel-server   9.712ms : endpoint
```

Fix!

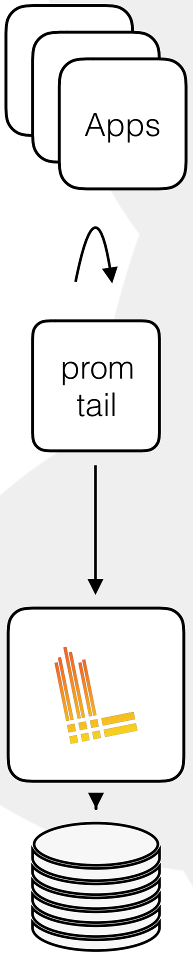


9/9

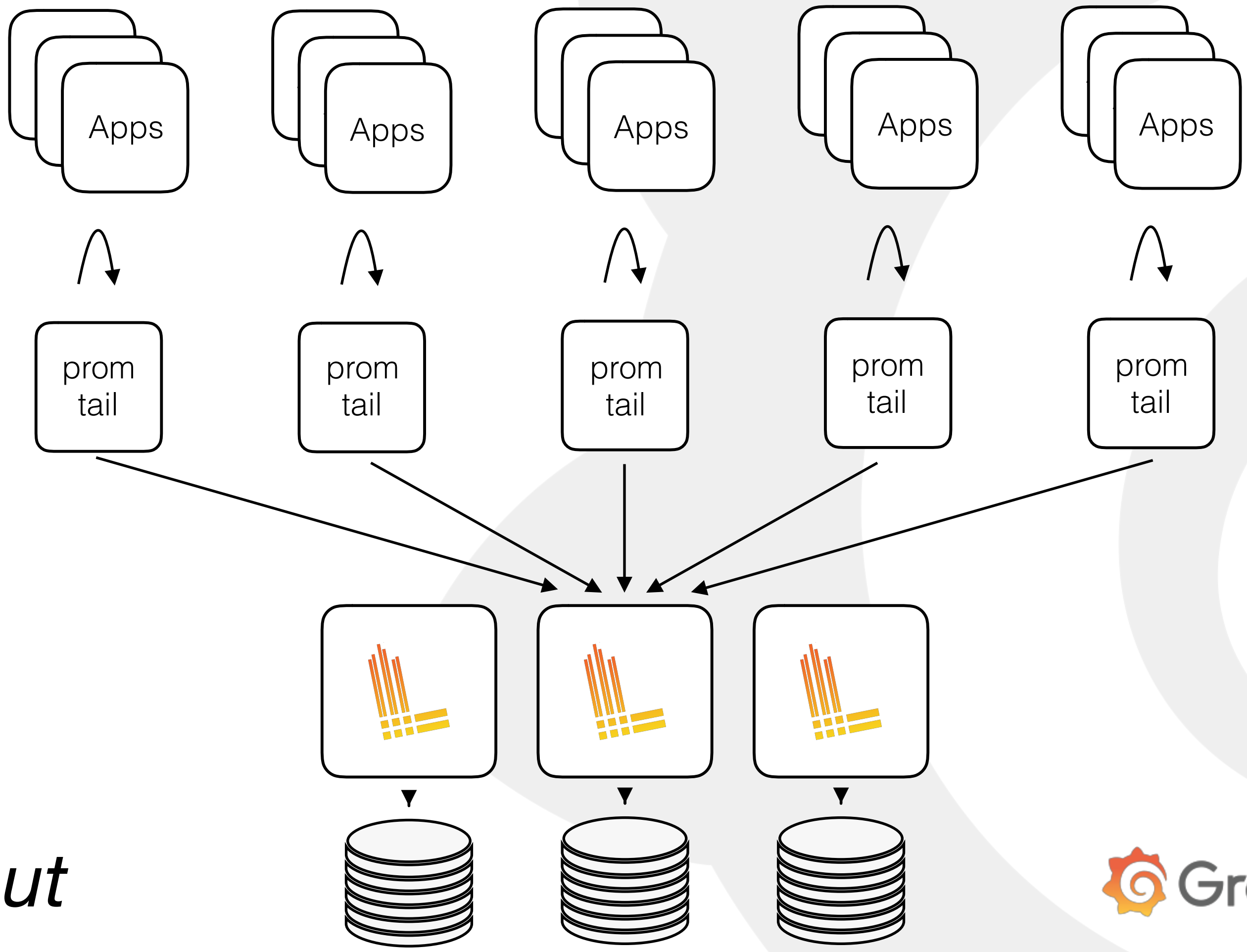
0800 Antan started
1000 " stopped - antan ✓ {1.2700 9.037 847 025 / 9.037 846 995 correct
13" uc (032) MP - MC ~~1.982147000~~ ~~2.130476415~~ ~~(-3)~~ 4.615925059(-2)
(033) PRO 2 2.130476415
correct 2.130676415
Relays 6-2 in 033 failed special speed test
in relay " 11,000 test.
Relays changed
1100 Started Cosine Tape (Sine check)
1525 Started Mult + Adder Test.
1545 Relay #70 Panel F
(moth) in relay.
First actual case of bug being found.
1630 Antangent started.
1700 closed down.

Relay 2145
Relay 3370

#2 Cloud Native and Airplane Friendly

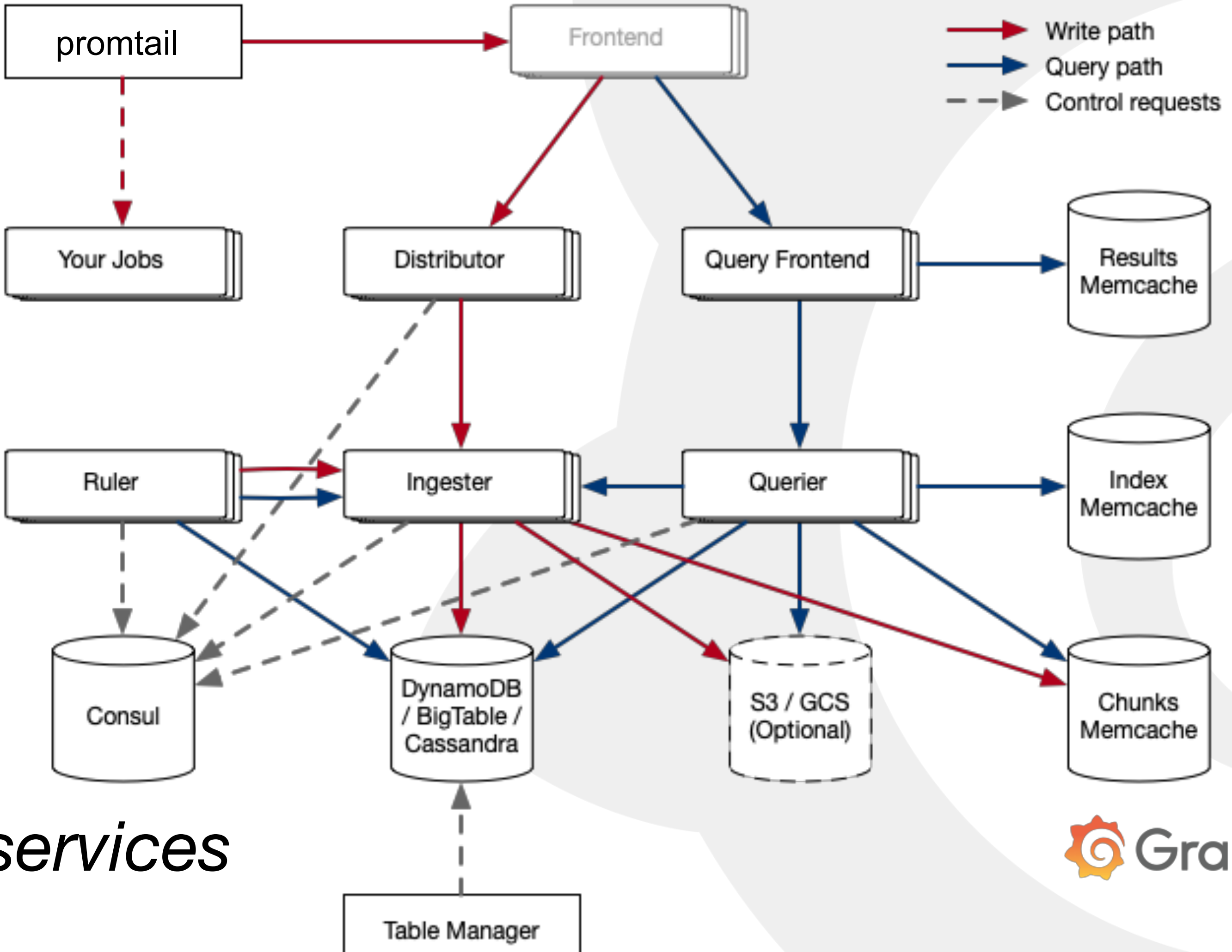


Airplane Friendly



Scale out

Grafana Labs



Microservices

Grafana Labs

Containerised

Kubernetes Native

Cloud Storage

#0 Simple and cost effective to operate

#1 Integrated with existing observability tools

#2 Cloud Native and Airplane Friendly

Demo

Whats next?

Grafana Labs

https://goo.gl/8h58zu

```
rate(({job=”app”} | “/foo” ! “/foo/bar”)[1m])

extract({job=”default/nginx”}, “code=(\d+)”, “co
    |> {code >= 500}

sum(extract({job=”app”}, “code=(\d+)”))
```

LogQL

Grafana Labs

Improve clustering & durability

Add Alerts & Rules off logs

Make it easier to get context, ad hoc filtering

Launch first beta in ~April

Thanks! Questions?

(we're hiring)