That's not a lot of data

Cloud-native solutions for FinOps

Richard “RichiH” Hartmann

Grafana Labs

Why this matters to you

Or: Who is that person?

My Background

- Director of Community @ Grafana Labs
- Prometheus team member (CNCF graduated project)
  - PromCon lead, Prometheus dev summit chair
- Founded OpenMetrics (CNCF incubating project)
- CNCF TAG Observability chair
- snmp_exporter & modbus_exporter
- Built Europe's most modern datacenter, monitored through SNMP and Modbus/TCP only
- Extensive paid and pro bono consulting on how to instrument existing and new applications

My Background

- Ran the backbone of an ISP for eleven years
  - Was the only person on call
  - My life & sanity depended on on-point monitoring & alerting
- Active in RIPE, IETF, DENOG, #networker etc
  - RFC to my name, changed RIPE NCC's IPv4 PI policies, etc.
- Prometheus transition for Germany's oldest ISP, ~5k devices
- Staffed world's largest IRC network for more than a decade
- Run conferences & monitoring from 100s to 18k attendees
  - DENOG, DebConf, FOSDEM, CCC, GrafanaCon, PromCon

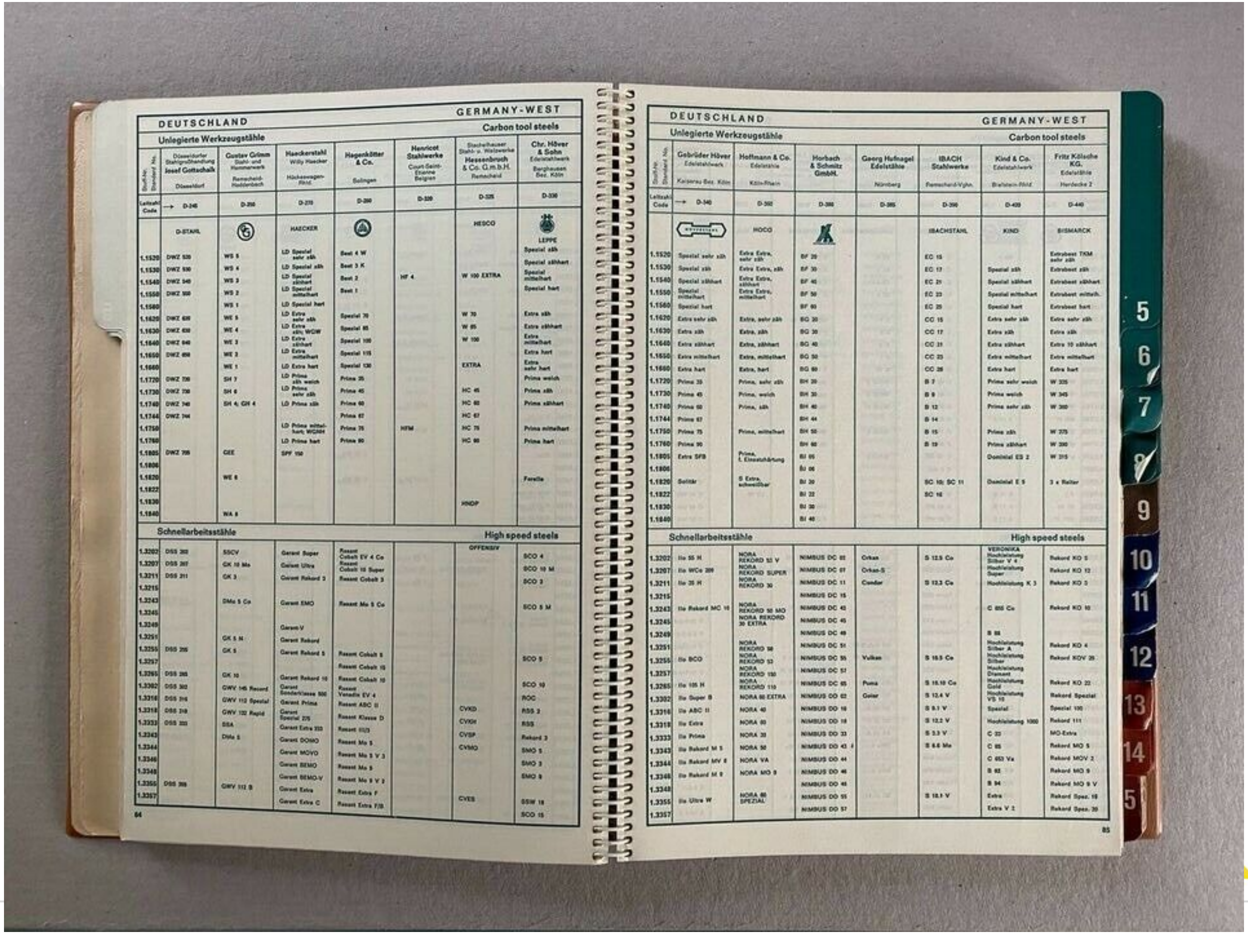

GOLD to go
The gold ATM
www.gold-to-go.com
EX ORIENTE LUX

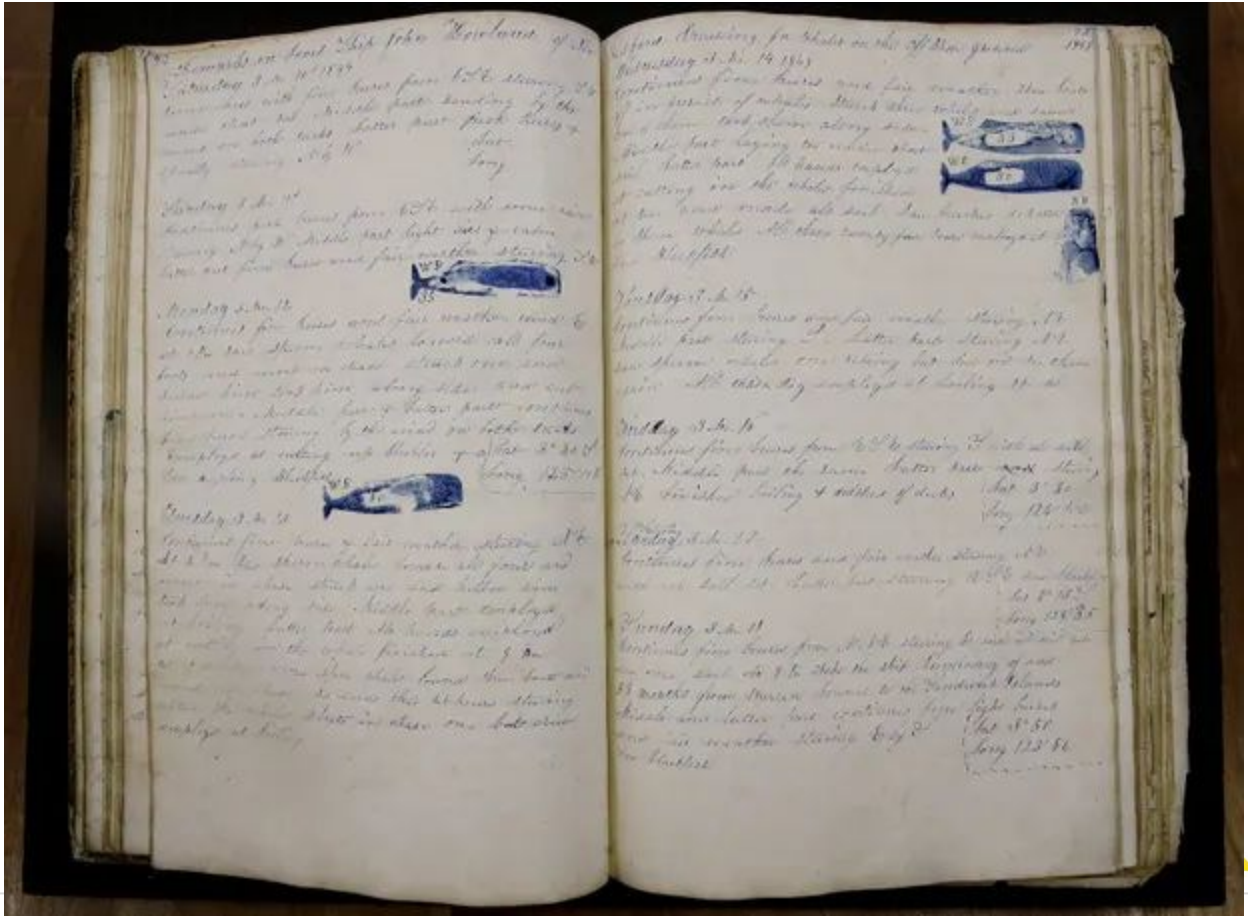

General / Stonks
2021-08-23 04:31:26 to 2021-09-15 06:25:20
symbol
AAPL
HODL
Grafanaberg Termin
HACK ATHON
Price Update: BTC/USD: $43,282.40 +0.0116%
$AAPL
Volume
High Price
$2.23K
Open Price
$2.26K
Stock price
$4.50K
$4.00K
$2.50K
$2.00K
$1.50K
$1.00K
$500.00
$0
08/24
08/26
08/28
08/30
09/01
09/03
09/05
09/07
09/09
09/11
09/13
09/15
High Price Min: $148.05 Max: $4.56K Mean: $3.52K
Top /r/wallstreetbets Stock
No data
Buy/sell volume
100000000
10000000
1000000
100000
10000
1000
100
Avg [10m]
Volume

Sneak Peak

Back to the basics

The dirty secret


DEUTSCHLAND
GERMANY-WEST
Unlegierte Werkzeugstähle
Carbon tool steels
Schnellarbeitsstähle
High speed steels
DEUTSCHLAND
GERMANY-WEST
Unlegierte Werkzeugstähle
Carbon tool steels
Schnellarbeitsstähle
High speed steels



Won’t need to convince you that numbers & events are important

Humanity runs on them, and they’re widely applicable

All this data is fundamentally the same

Observability

Observability

- Buzzword; all the rage in tech
- Control theory: Deduct the internal state of a system by looking at its inputs and outputs only
- “Monitoring” has taken on a meaning of collecting, not using data
  - One extreme: Full text indexing
  - Other extreme: Data lake

Observability is about enabling humans, and machines, to understand and predict complex systems

Inform, Optimize, Operate

Prometheus

Prometheus 101

- Time series database
- Inspired by Google's Borgmon
- Standard in cloud-native deployments and for Kubernetes
- PromQL: for processing, graphing, alerting, and export
- No hierarchical model, n-dimensional label set
  - Not equivalent to, but a great basis for, AI/ML labels

Time series

- Time series are recorded values which change over time
- Individual events can be merged into time series
- Access rates to a webserver (counter)
  - Temperatures in a datacenter (gauge)
  - Service latency (histograms)

Super easy to emit, parse & read

```
http_requests_total{env="prod",method="post",code="200"} 1027
http_requests_total{env="prod",method="post",code="400"} 3
http_requests_total{env="prod",method="post",code="500"} 12
http_requests_total{env="prod",method="get",code="200"} 20
http_requests_total{env="test",method="post",code="200"} 372
http_requests_total{env="test",method="post",code="400"} 75
```

Prometheus scale

- 1,000,000+ samples/second no problem on current hardware
- ~200,000 samples/second/core
- 16 bytes/sample compressed to 1.36 bytes/sample

The highest we saw in production on a single Prometheus instance were 100 million active times series at once!

Long term storage

- Two long term storage solutions have Prometheus-team members working on them
  - Thanos
    - Historically easier to run, but slower
    - Scales storage horizontally
  - Cortex
    - Easy to run these days
    - Scales both storage, ingester, and querier horizontally

Cortex @ Grafana (largest cluster, 2021-09)

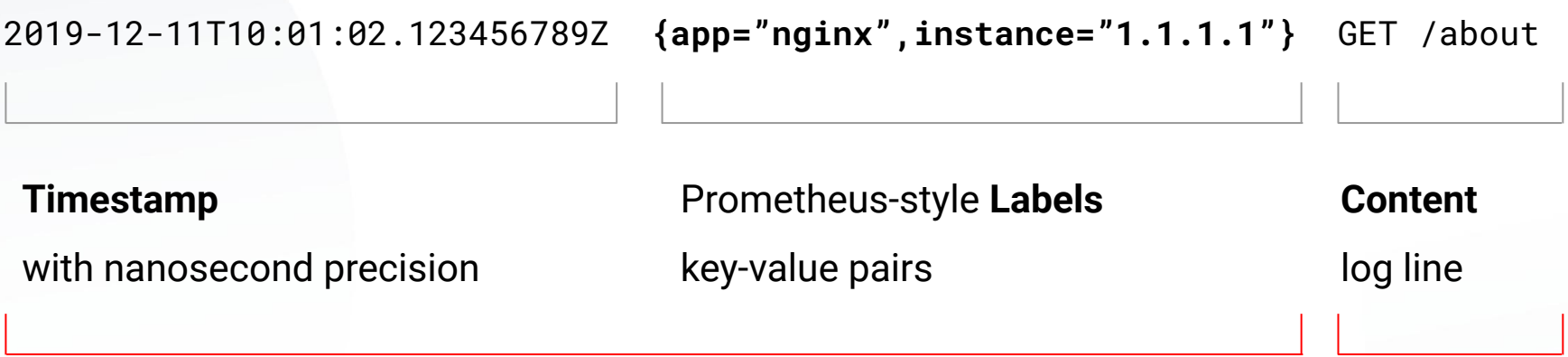
- ~65 million active series
- 668 CPU cores
- 3,349 GiB RAM

One customer running at 3 billion active series

Grafana Loki

Loki 101

- Following the same label-based system like Prometheus
- No full text index needed, incredible speed
- Work with logs at scale, without the massive cost
- Access logs with the same label sets as metrics
- Turn logs into metrics, to make it easier to work with them
- Make direct use of syslog data, via promtail


2019-12-11T10:01:02.123456789Z
{app="nginx",instance="1.1.1.1"}
GET /about
Timestamp
with nanosecond precision
Prometheus-style Labels
key-value pairs
Content
log line
indexed
unindexed

Loki @ Grafana (largest cluster, 2021-09)

- Queries regularly see 40 GiB/s
- Query terabytes of event data in under a minute
  - Including complex processing of result sets
- Undisclosed high number of Terabytes of data
- 268 CPU cores
- 1,453 GiB RAM

Grafana Tempo

(mainly relevant for FinOps itself, not finance data)

Tempo

- Exemplars: Jump from relevant logs & metrics
  - Native to Prometheus, Cortex, Thanos, and Loki
  - Exemplars work at Google scale, with the ease of Grafana
- Index and search by labelsets available for those who need it
- Object store only: No Cassandra, Elastic, etc.
- 100% compatible with OpenTelemetry Tracing, Zipkin, Jaeger
- 100% of your traces, no sampling

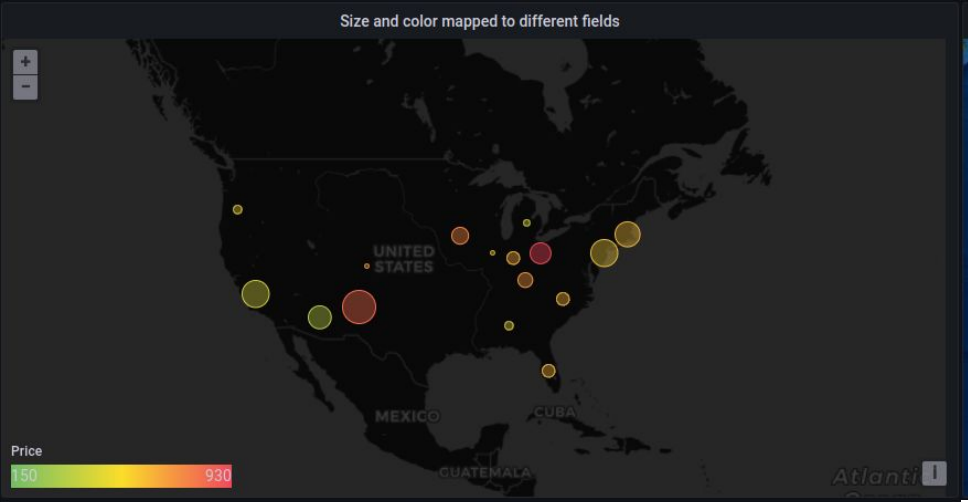
Tempo @ Grafana (largest cluster, 2021-07)

- 2,200,000 samples per second @ 350 MiB/s
- 14-day retention @ 3 copies stored
- ~240 CPU cores (includes compression cost)
- ~450 GiB RAM
- 132 TiB object storage
- Latencies:
  - p99 - 2.5s
  - p90 - 2.3s
  - p50 - 1.6s

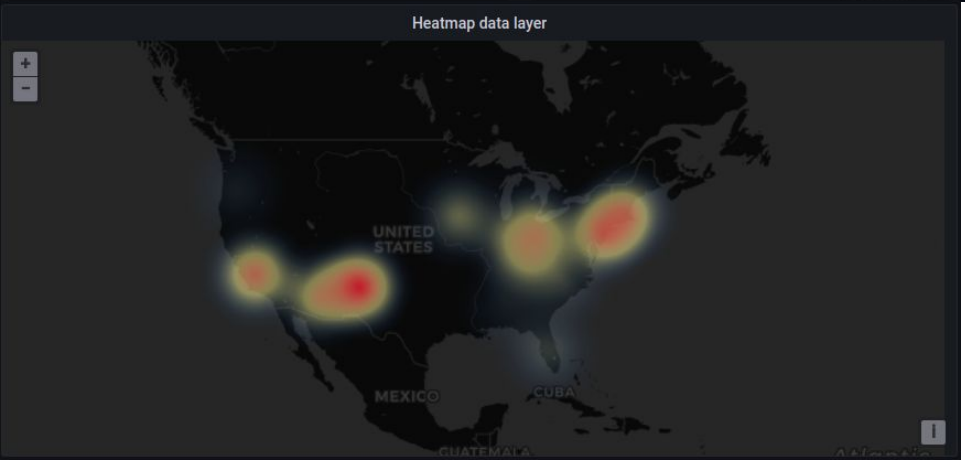
Grafana

The software


server requests
150
100
50
0
10:30
10:35
10:40
10:45
10:50
10:55
11:00
11:05
11:10
11:15
11:20
11:25
web_server_01
web_server_02
web_server_03
web_server_04
Memory / CPU
25%
20%
15%
10%
5%
0%
20 B
15 B
10 B
5 B
10:30
10:40
10:50
11:00
11:10
11:20
cpu
memory
client side full page load
5 s
4 s
3 s
2 s
1 s
0 ms
10:30
10:35
10:40
10:45
10:50
10:55
11:00
11:05
11:10
11:15
11:20
11:25
Mean
upper_25 2.20 ms
upper_50 219 ms
upper_75 740 ms
upper_90 1.39 s
upper_95 1.92 s
logins
60
40
20
0
10:30
10:35
10:40
10:45
10:50
10:55
11:00
11:05
11:10
11:15
11:20
11:25
logins
logins (-1 hour)
Traffic In/Out
4 kB
2 kB
0 B
-2 kB
-4 kB
10:30
10:35
10:40
10:45
10:50
10:55
11:00
11:05
11:10
11:15
11:20
11:25
avg
current
upper_25 2.27 B 1 B
upper_50 221 B 202 B
upper_75 736 B 821 B
upper_90 1.38 kB 1.57 kB
upper_95 1.93 kB 1.96 kB
cpu1 -1.9 kB -2.0 kB
cpu2 -736 B -821 B


Size and color mapped to different fields
UNITED STATES
MEXICO
CUBA
GUATEMALA
Price
150
930


Base layer ArcGIS wold imagery


Heatmap data layer
UNITED STATES
MEXICO
CUBA
GUATEMALA


Color points by color scale
50 °
45 °
40 °
35 °
30 °
25 °
20 °
15 °
10 °
5 °
0 °
05:30
06:00
06:30
07:00
07:30
08:00
08:30
09:00
09:30
10:00
10:30
11:00
A-series
B-series


State timeline strings
Level A
LOW
HIGH
NORMAL
LOW
NORMAL
HIGH
Level B
NORMAL
LOW
CRITICAL
LOW
NORMAL
HIGH
Level C
NORMAL
CRITICAL
LOW
NORMAL
05:45 06:00 06:15 06:30 06:45 07:00 07:15 07:30 07:45 08:00 08:15 08:30 08:45 09:00 09:15 09:30 09:45 10:00 10:15 10:30 10:45 11:00 11:15 11:3
State timeline with time series + thresholds
backend_01
backend_02
backend_03
backend_04
05:45 06:00 06:15 06:30 06:45 07:00 07:15 07:30 07:45 08:00 08:15 08:30 08:45 09:00 09:15 09:30 09:45 10:00 10:15 10:30 10:45 11:00 11:15 11:3
< 50 °
50 °+
300 °+
Same query/data as above
400 °
300 °
200 °
100 °
0 °
05:45 06:00 06:15 06:30 06:45 07:00 07:15 07:30 07:45 08:00 08:15 08:30 08:45 09:00 09:15 09:30 09:45 10:00 10:15 10:30 10:45 11:00 11:15 11:3
backend_01
backend_02
backend_03
backend_04
Status history with time series + thresholds
Cellar 12.7 23.3 27.3 23.3 21.7 21.0 18.7 24.1 23.3 26.7 26.7 29.7
Living room 12.1 6.20 11.6 20.2 27.3 23.4 24.3 15.5 16.7 20.2 16.7 23.9
Porch 14.0 2.95 4.80 11.5 9.70 3.49 12.7 14.4 17.9 11.8 3.22 -2.8
Bedroom 14.3 21.3 26.0 33.3 24.9 27.7 29.6 40 34.9 39.5 40 40
05:30 06:00 06:30 07:00 07:30 08:00 08:30 09:00 09:30 10:00 10:30 11:00

https://play.grafana.org/

Grafana

For business

Fortune 50

- Paying customers
  - 40% of Fortune 10
  - 50% of Fortune 20
  - 40% of Fortune 50
- 92% of Fortune 50 use our technology

Awards & recognition

- Enterprise Tech 30 2021 (medium size Top 10 and corp dev Top 10)
- Inc Best Work Places 2020, 2021
- Forbes America's Best Startup Employers 2021
- Gartner Cool Vendor 2021
- DevOps Dozen 2020 - Loki. Most innovative open source project
- Forbes Cloud 100 2021 (this was before our Series C)
- DevOps Dozen 2020 - Loki: Most Innovative DevOps Open Source Project
- 50 Must Have SecOps Tools - #1 Dashboards: Grafana

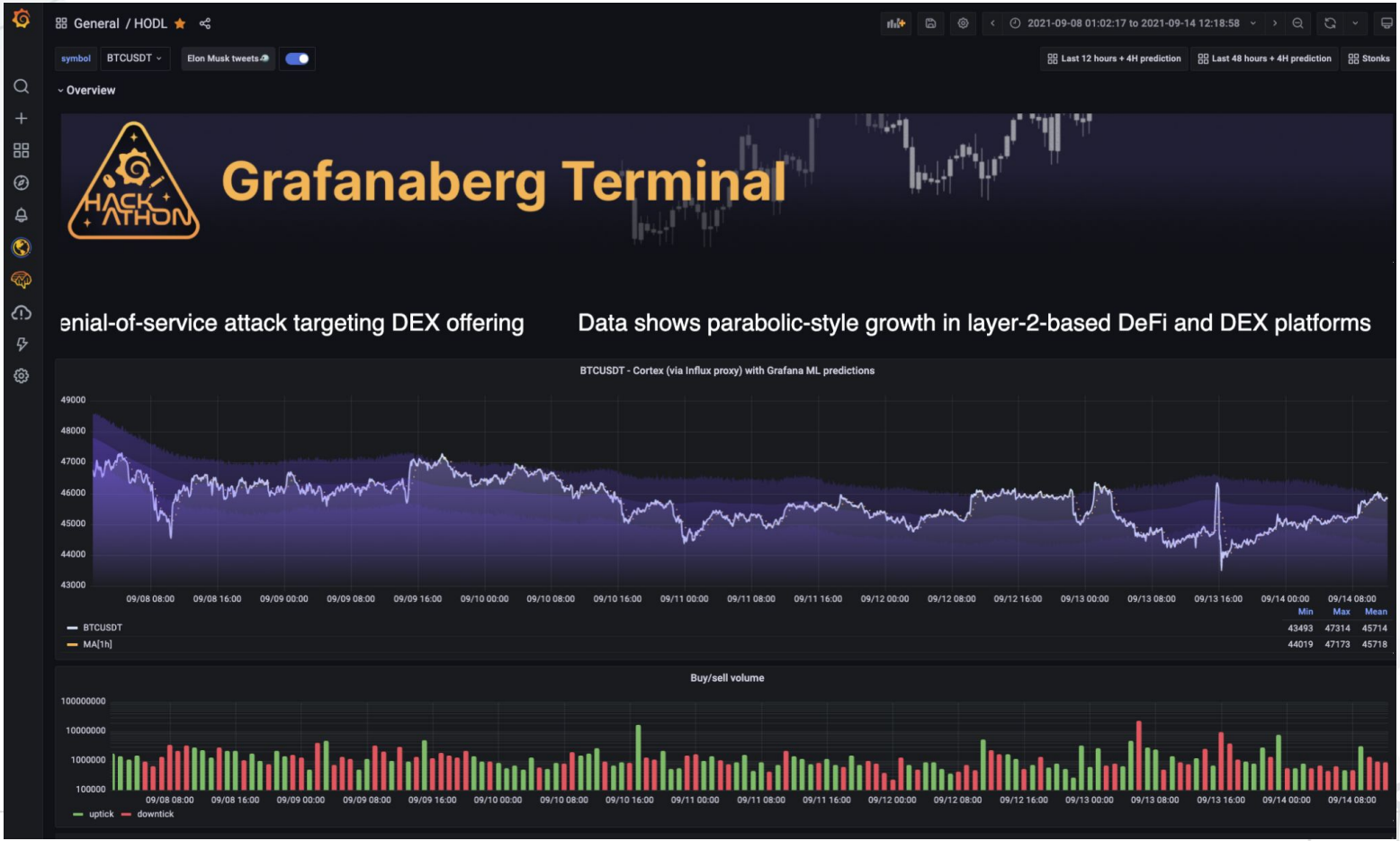

JPMorgan Chase & Co.
Hall of Innovation
Grafana Labs
Technology Innovation Symposium 2020

GrafanaCONline 2021

How Robinhood scaled from 100M to 700M time series with Grafana Enterprise Metrics

Wednesday, Jun 9, 14:15 ET, 18:15 UTC

Also known as: When GME (GameStop) led to GEM (Grafana Enterprise Metrics). After Robinhood, the stock-trading app, was inundated with GameStop trades in January, its two-person observability team turned to Grafana Enterprise Metrics to scale from 100M time series on Prometheus to 700M on Cortex. This session will share the lessons they learned along the way.


General / HODL
2021-09-08 01:02:17 to 2021-09-14 12:18:58
symbol
BTCUSDT
Elon Musk tweets
Last 12 hours + 4H prediction
Last 48 hours + 4H prediction
Stonks
Overview
Grafanaberg Terminal
HACK ATHON
enial-of-service attack targeting DEX offering
Data shows parabolic-style growth in layer-2-based DeFi and DEX platforms
BTCUSDT - Cortex (via Influx proxy) with Grafana ML predictions
49000
48000
47000
46000
45000
44000
43000
09/08 08:00
09/08 16:00
09/09 00:00
09/09 08:00
09/09 16:00
09/10 00:00
09/10 08:00
09/10 16:00
09/11 00:00
09/11 08:00
09/11 16:00
09/12 00:00
09/12 08:00
09/12 16:00
09/13 00:00
09/13 08:00
09/13 16:00
09/14 00:00
09/14 08:00
Min
Max
Mean
BTCUSDT
43493
47314
45714
MA[1h]
44019
47173
45718
Buy/sell volume
100000000
10000000
1000000
100000
09/08 08:00
09/08 16:00
09/09 00:00
09/09 08:00
09/09 16:00
09/10 00:00
09/10 08:00
09/10 16:00
09/11 00:00
09/11 08:00
09/11 16:00
09/12 00:00
09/12 08:00
09/12 16:00
09/13 00:00
09/13 08:00
09/13 16:00
09/14 00:00
09/14 08:00
uptick
downtick

All of this is Open Source and you can run it yourself

It’s also available On-Prem and Cloud aaS

Thank You!

richih@richih.org
twitter.com/TwitchiH
github.com/RichiH/talks