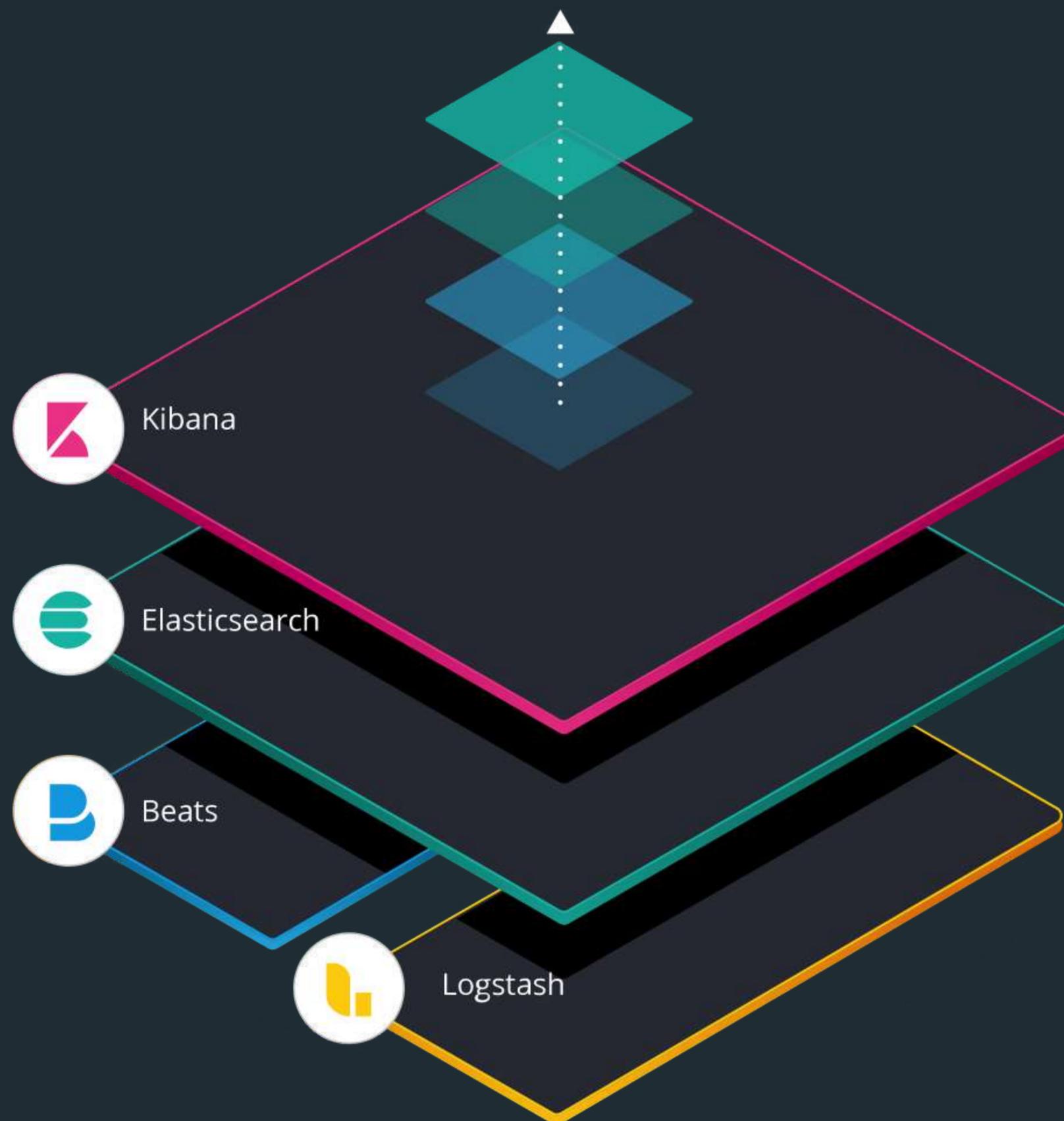




Workshop: Logging with the Elastic Stack

Alexander Reelsen
@spinscale
alex@elastic.co





Agenda

- Why use a search engine for logging?
- Log centralization
- Logging challenges
- Deployment
- Demo & workshop
- Logging patterns
- Q & A

Prerequisite

- docker
- docker-compose
- git
- java

Prerequisites

- `git clone https://github.com/xeraa/java-logging`
- `cd java-logging`
- `./gradlew assemble`
- `docker-compose up --build`



Logging?

Why use Elastic Stack for logging?



But why?

- Fundamental for debugging production issues
- Logs are decentralized
- Containers containing logs are ephemeral
- Logs are not standardized
- Correlations are hard

No standards...

```
1.2.3.4 - - [06/Nov/2014:19:10:38 +0600] "GET /news/foo.html
HTTP/1.1" 404 177 "-" "Mozilla/5.0 (iPhone; CPU iPhone OS 6_0
like Mac OS X) AppleWebKit/536.26 (KHTML, like Gecko) Version/6.0
Mobile/10A5376e Safari/8536.25 (compatible; Googlebot/2.1;
+http://www.google.com/bot.html)"
```

No standards...

```
Sep 12 10:15:08 rhincodon logd[64]: #DECODE failed to resolve  
UUID: [pc:0x7fff65ec1ac7 ns:0x06 type:0x82 flags:0x8208 main:A5  
2374C3-0F9D-3062-A636-131B737C4589 pid:945]
```

No standards...

```
[2019-09-12T10:23:45,900][INFO ][o.e.c.s.ClusterApplierService]
[rhincodon] master node changed {previous [], current
[{rhincodon}{q3Rj1oGxRdm176yLo9d9UA}{Vq4FpFk1RbCyVFAVKU7ukQ}
{127.0.0.1}{127.0.0.1:9300}{dim}{ml.machine_memory=17179869184,
xpack.installed=true, ml.max_open_jobs=20}]}, term: 6, version:
57, reason: Publication{term=6, version=57}
```

Preprocessing to the rescue

- Date normalization
- Information extraction
- Field normalization

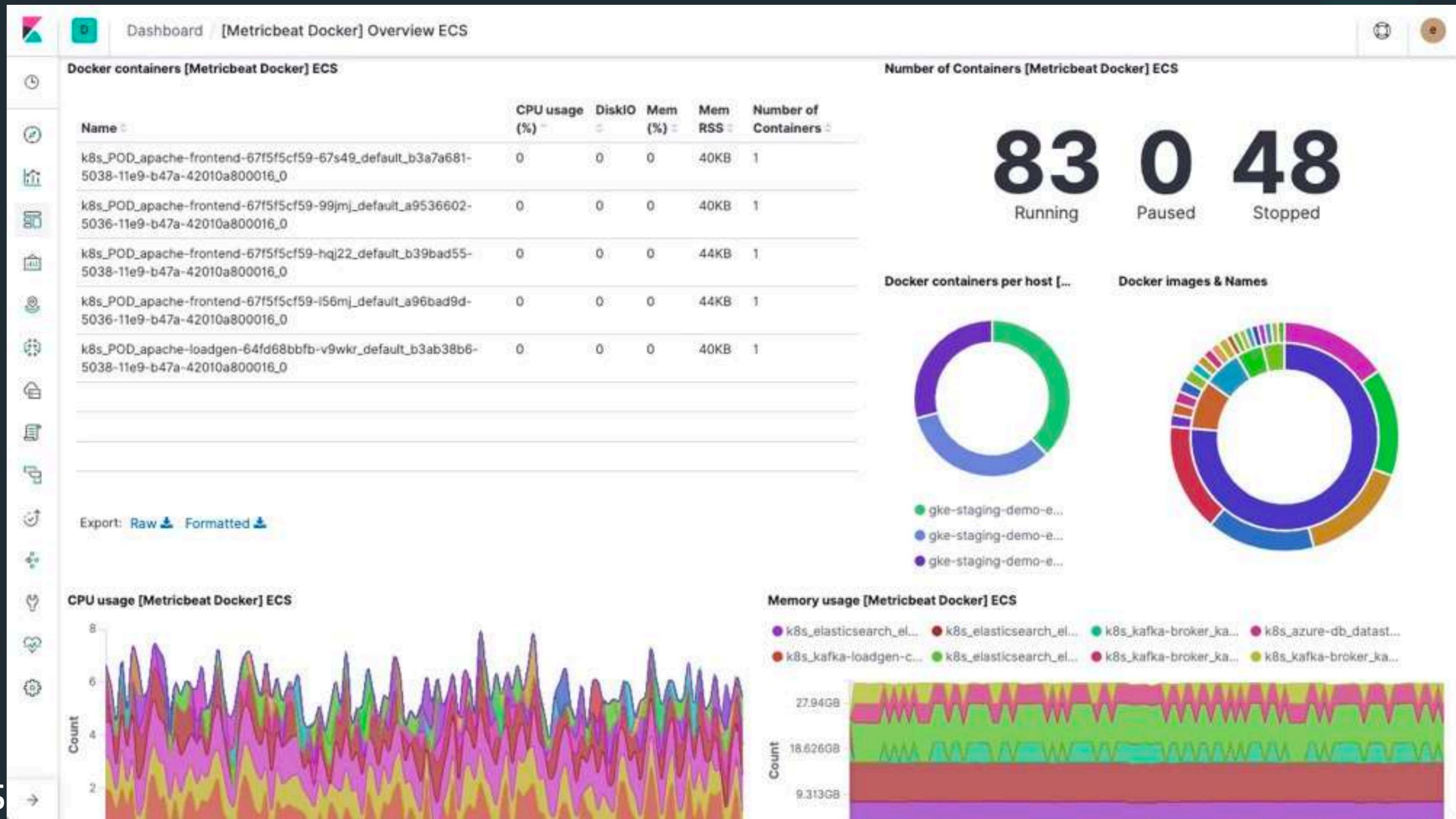
Time series have a lifecycle

- Recent data is more important
- Recent data is queried more often
- Older data less searched
- Old data may require archival due to compliance

Time series is a search

- Max response time per 10 minute window since yesterday
- Documents: All documents from yesterday till now
- Aggregate in 10 minute buckets ($6*24$)
- For each bucket, extract max value

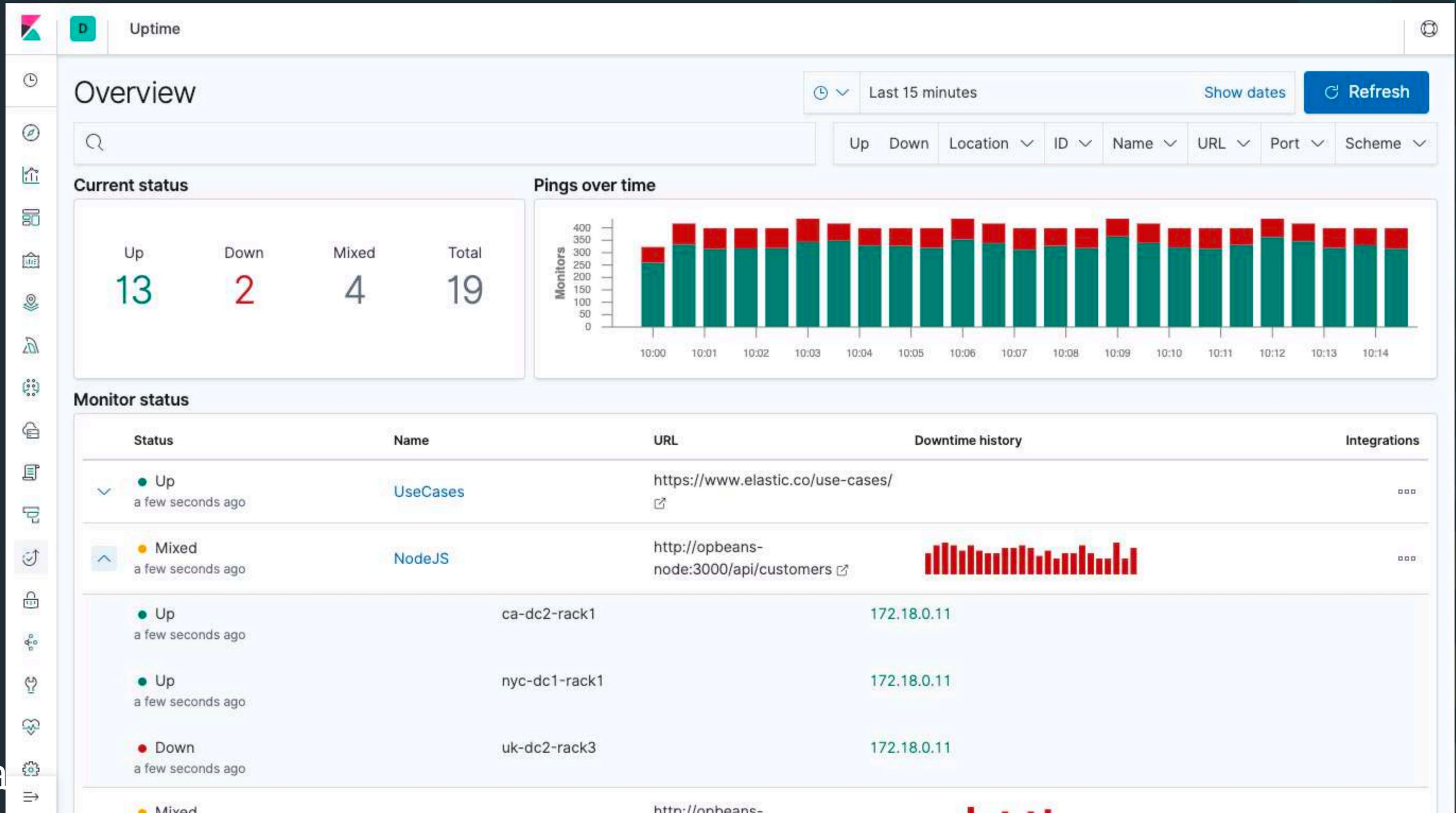
Dashboards & Time Series



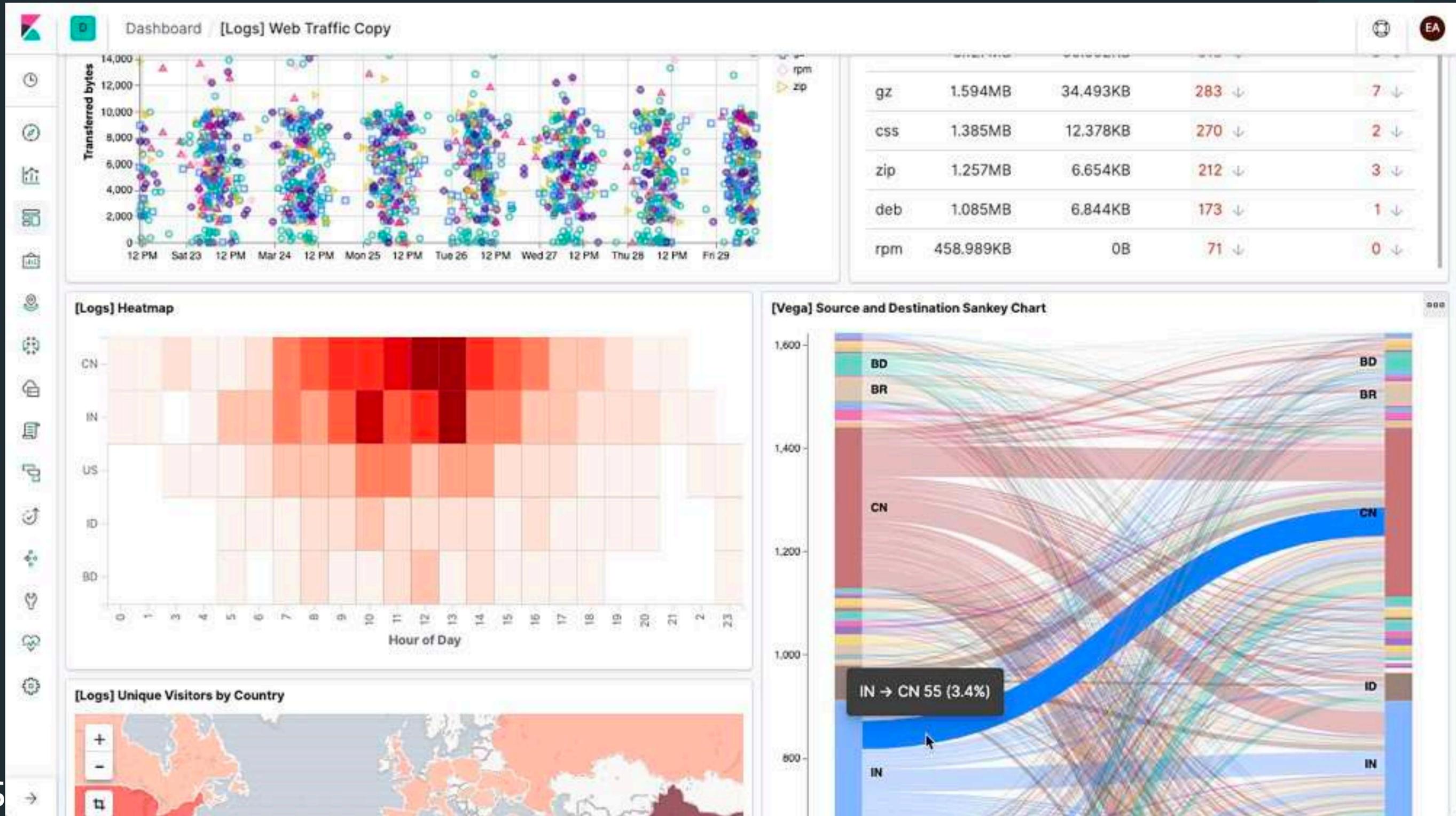
Dashboards & Time Series



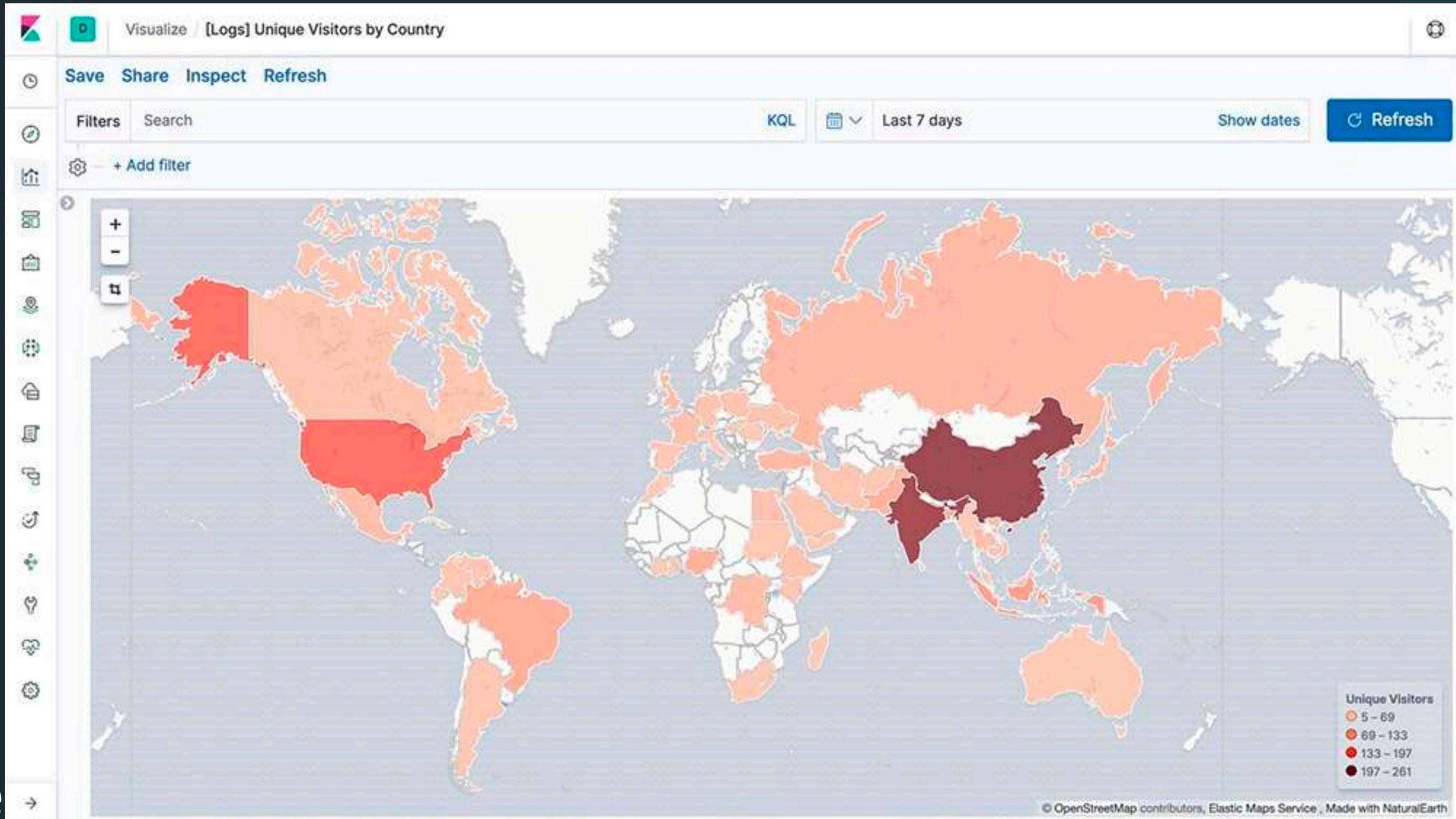
Dashboards & Time Series



Dashboards & Time Series



Dashboards & Time Series



Standardizing data

What is ECS?

The Elastic Common Schema (ECS) is an open source specification, developed with support from the Elastic user community. ECS defines a common set of fields to be used when storing event data in Elasticsearch, such as logs and metrics.

ECS specifies field names and Elasticsearch datatypes for each field, and provides descriptions and example usage. ECS also groups fields into ECS levels, which are used to signal how much a field is expected to be present. You can learn more about ECS levels in [Guidelines and Best Practices](#). Finally, ECS also provides a set of naming guidelines for adding custom fields.

The goal of ECS is to enable and encourage users of Elasticsearch to normalize their event data, so that they can better analyze, visualize, and correlate the data represented in their events.

ECS has been scoped to accommodate a wide variety of events, spanning:

- **Event sources:** whether the source of your event is an Elastic product, a third-party product, or a custom application built by your organization.
- **Ingestion architectures:** whether the ingestion path for your events includes Beats processors, Logstash, Elasticsearch ingest node, all of the above, or none of the above.
- **Consumers:** whether consumed by API, Kibana queries, dashboards, apps, or other means.



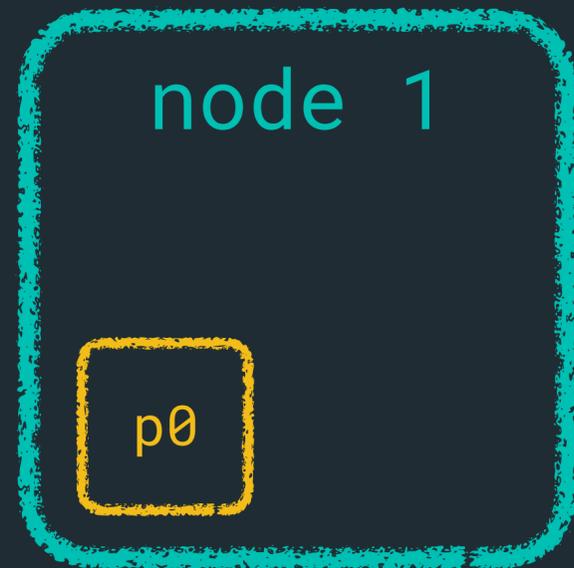
Elasticsearch overview



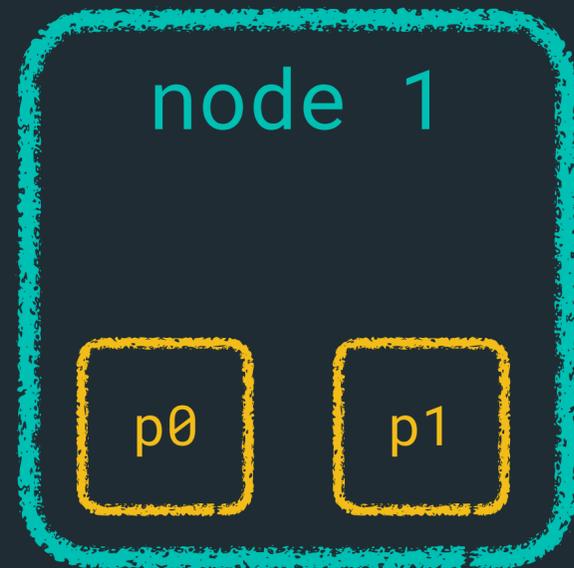
Elasticsearch in 10 seconds

- Search Engine (FTS, Analytics, Geo), real-time
- Distributed, scalable, highly available, resilient
- Interface: HTTP & JSON
- Centrepiece of the Elastic Stack

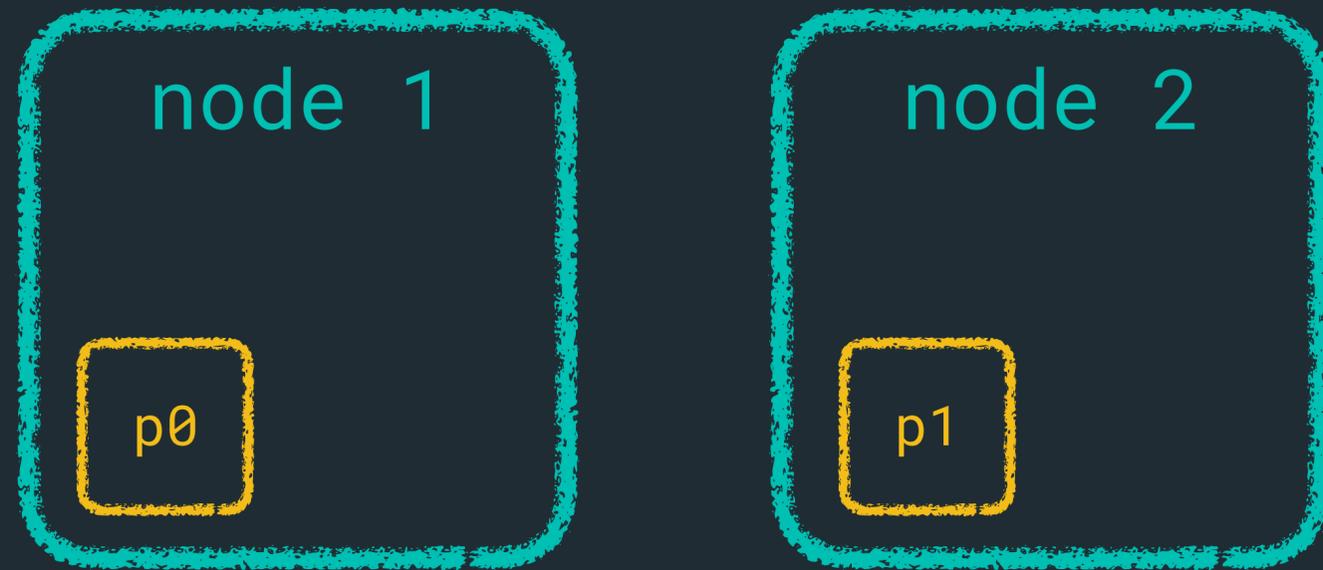
Elasticsearch - a distributed system



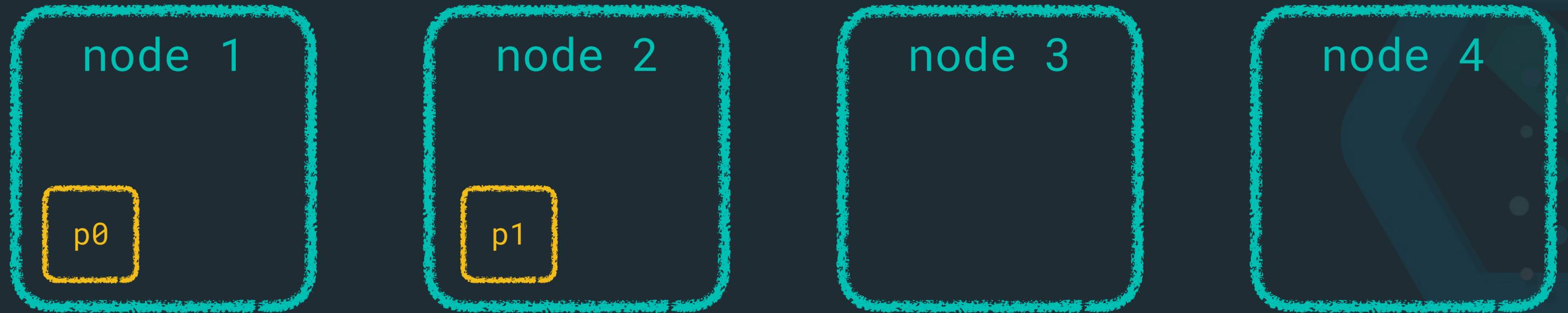
Elasticsearch - a distributed system



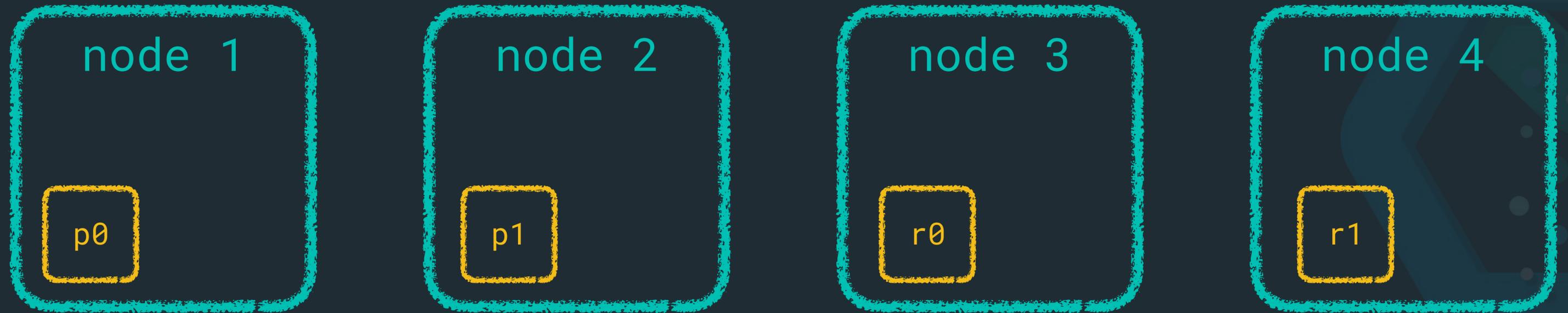
Elasticsearch - a distributed system



Elasticsearch - a distributed system



Elasticsearch - a distributed system





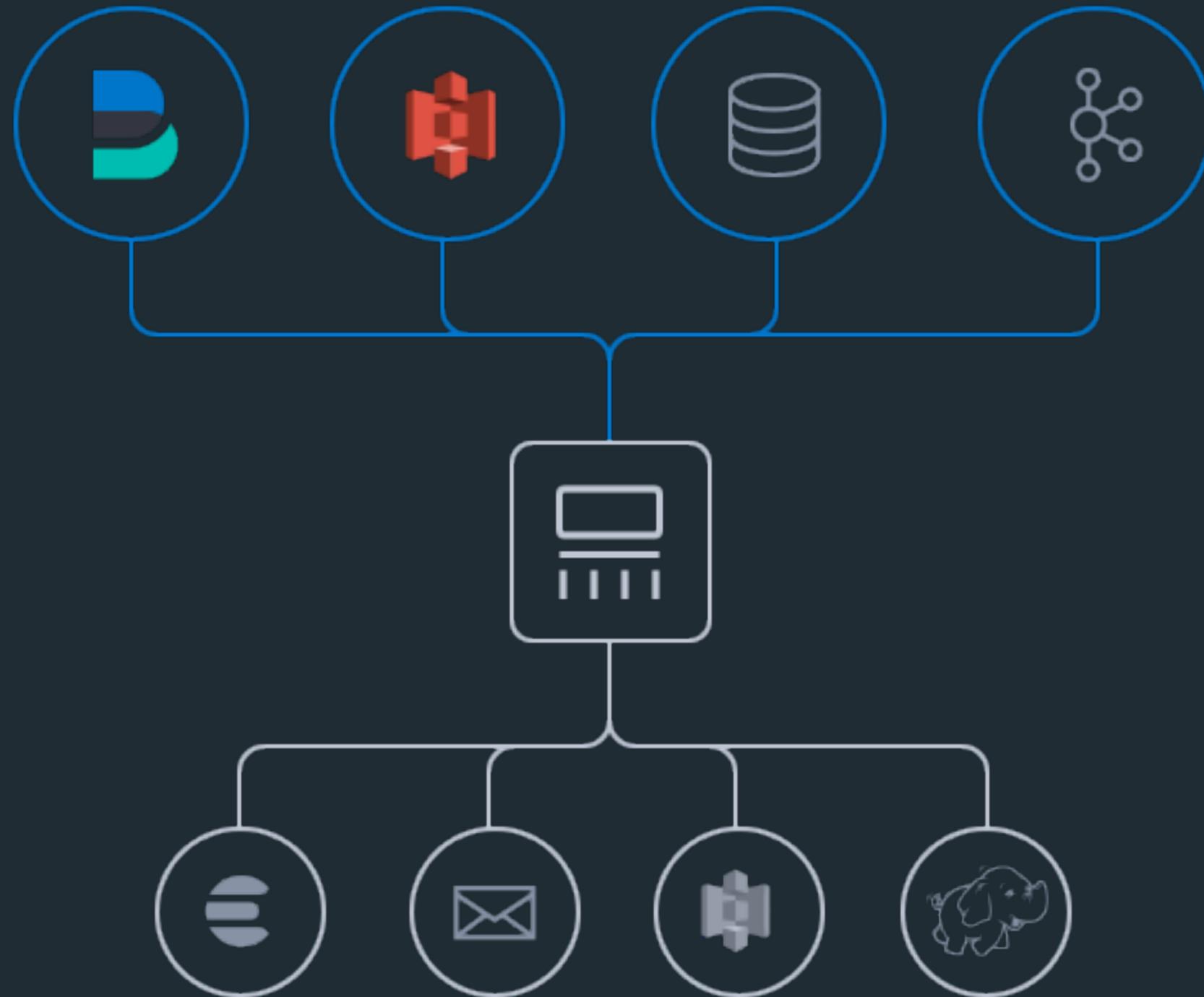
Ingest overview



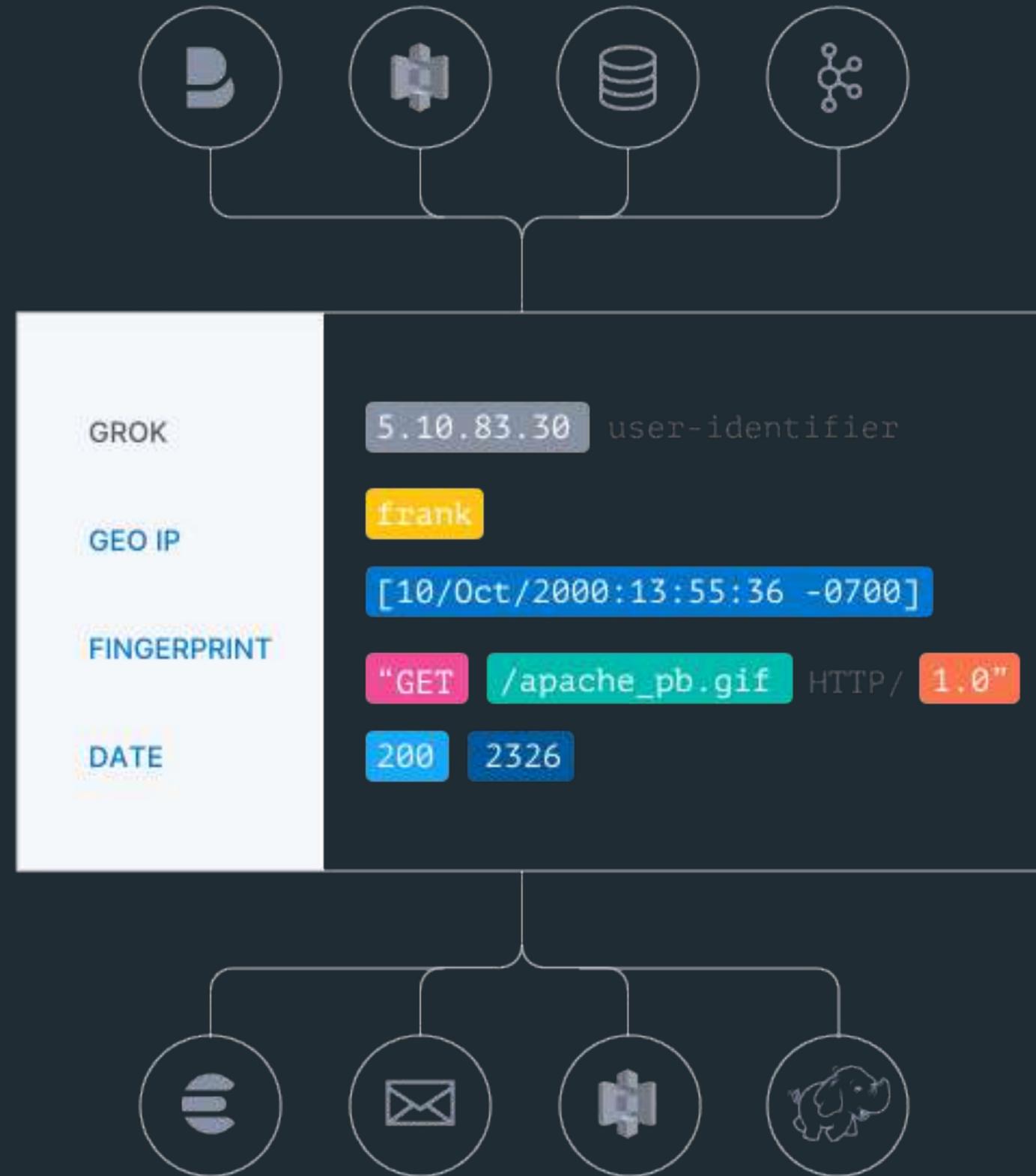
Ingestion

- Logstash: extensible dynamic data collection
- Beats: specialized single purpose data shipper
- your own rolled integration, it's all HTTP!

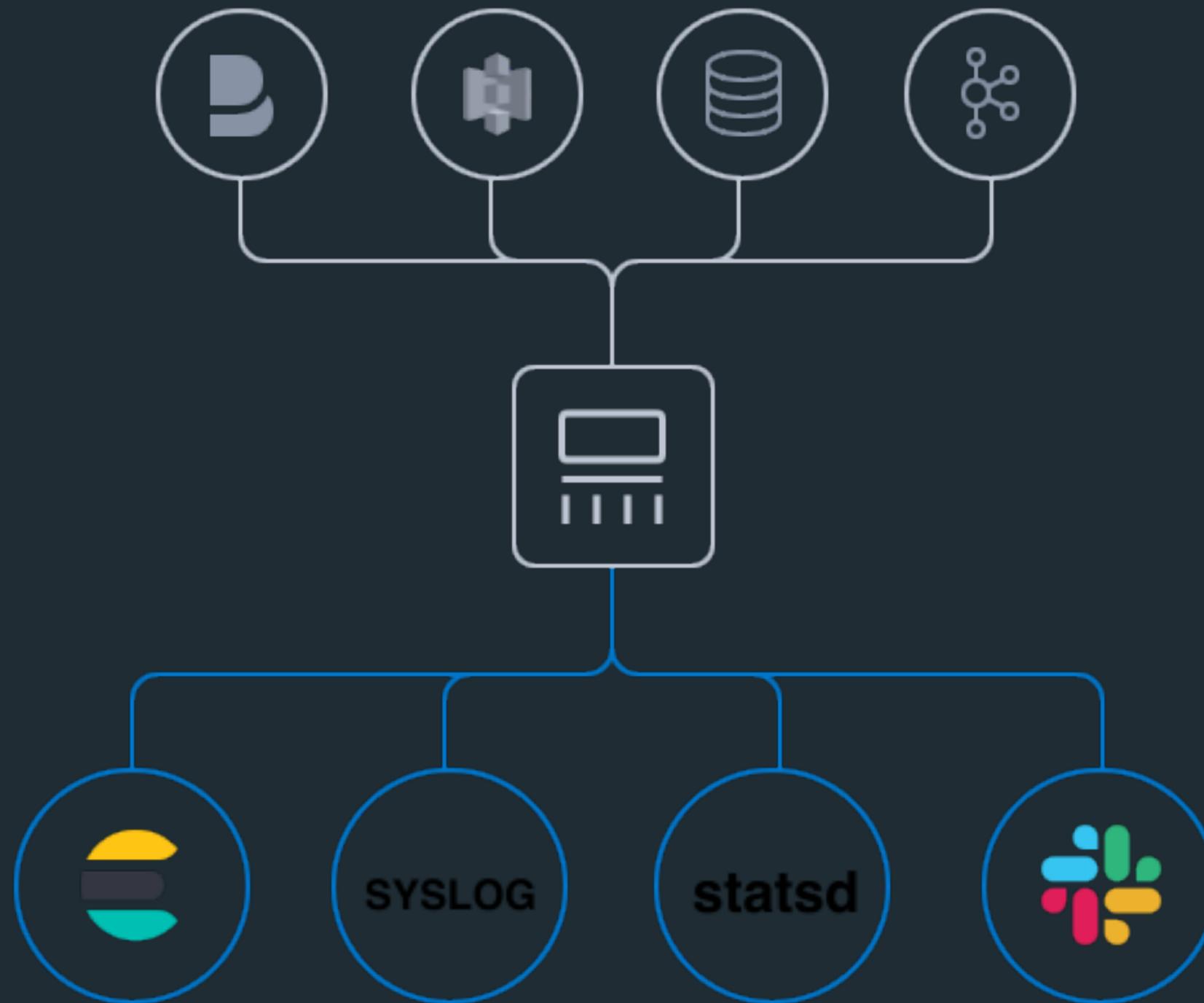
Logstash



Logstash



Logstash



Beats

- Filebeat
- Metricbeat
- Packetbeat
- Winlogbeat
- Auditbeat
- Heartbeat
- Functionbeat
- Journalbeat

Filebeat s

- Apache
- Auditd
- AWS
- CEF
- Cisco
- Coredns
- Elasticsearch
- Envoyproxy
- Google Cloud
- haproxy
- IBM MQ
- Icinga
- IIS
- Iptables
- Kafka
- Kibana
- Logstash
- MongoDB
- MSSQL
- MySQL
- nats
- NetFlow
- Nginx
- Osquery
- Palo Alto Networks
- PostgreSQL
- RabbitMQ
- Redis
- Santa
- Suricata
- Traefik
- Zeek (Bro)

Metricbeat modules

- Aerospike
- Apache
- aws
- Beat
- Ceph
- CockroachDB
- consul
- coredns
- Couchbase
- couchdb
- Docker
- Dropwizard
- Elasticsearch
- envoyproxy
- Etcd
- Golang
- Graphite
- HAProxy
- HTTP
- Jolokia
- Kafka
- Kibana
- Kubernetes
- kvm
- Logstash
- Memcached
- MongoDB
- MSSQL
- Munin
- MySQL
- Nats
- Nginx
- Oracle
- PHP_FPM
- PostgreSQL
- Prometheus
- RabbitMQ
- Redis
- Statsd
- System
- traefik
- uwsgi
- vSphere
- Windows
- ZooKeeper



Solutions



Elastic APM

- Distributed tracing
- APM server
- Kibana application
- Agents: Java, .NET, Node, Python, Ruby, RUM, Go
- Alerting & ML integration

Elastic Logs

The screenshot displays the Elastic Logs interface. At the top, a search bar contains the query `kubernetes.pod.uid: f2046b20-a4d8-11e9-8210-42010a8e0164`. Below the search bar is a table of log entries with columns for **Timestamp**, **event.dataset**, and **Message**. The messages are truncated, showing patterns like `[redis.log][warning]`, `[redis.log][notice]`, and `[redis.log][warning]`.

On the right side, a panel titled "Log event document details" shows the full document structure for a selected log entry. The fields and their values are as follows:

Field	Value
@timestamp	2019-07-12T19:55:54.287Z
_id	-x_B52sBAOoy4w2krX2H
_index	filebeat-7.2.0-2019.07.10-000001
agent.ephemeral_id	5a20c2e1-f35c-4896-8cac-4dd25cfe3a6f
agent.hostname	filebeat-dynamic-xzfm
agent.id	552da04f-2e32-440c-8ee2-51c511cd9233
agent.type	filebeat
agent.version	7.2.0
cloud.availability_zone	us-east1-b
cloud.instance.id	5328326164652724159
cloud.instance.name	gke-jamie-pmm-default-pool-625d4cbf-b8bz
cloud.machine.type	n1-standard-2

At the bottom of the log list, a message states "No additional entries found" with a "Load again" button.

Elastic SIEM

The screenshot displays the Elastic SIEM interface. The top navigation bar includes 'Overview', 'Hosts', 'Network', and 'Timeline'. The 'Hosts' tab is active, showing a search bar with the query 'e.g. host.name: "foo"'. Below the search bar, the 'Hosts' section displays 'Last Event: in 20 days' and a card for 'Hosts' with a count of 904 and a line graph. The 'All Hosts' section shows 'Showing: 21,116 Hosts' and a table with columns for 'Name' and 'beats-ci-immutable-ubuntu-1604-'. The main search results pane shows a query: 'event.action:"config_change" and event.dataset:"file"'. The results table has columns for '@timestamp', 'event.severity', 'event.category', 'event.action', and 'host.name'. The first row shows an event on 'Jun 3, 2019 @ 19:40:15.160' with severity '--', category 'audit-rule', action 'executed', and host 'siem-es'. The session details for this event are: '# unset', 'root', '@ siem-es', 'in /', 'executed', '> ip route ls table local type local scope host dev eth0 proto 66 with result success'. The interface also includes a 'Notes' section with 0 notes, a 'Refresh' button, and a 'Load More' button at the bottom. The status bar indicates '25 of 14846121 Events' and 'Updated 4 minutes ago'.

Elastic Uptime

The screenshot displays the Elastic Uptime dashboard. At the top, the 'Overview' section shows a search bar and a time range selector set to 'Last 15 minutes'. Below this, the 'Current status' section provides a summary: 13 monitors are 'Up', 2 are 'Down', and 4 are 'Mixed', for a total of 19 monitors. To the right, the 'Pings over time' section features a bar chart showing the number of monitors that are up (green) and down (red) from 10:00 to 10:14. The chart shows a relatively stable number of up monitors, with a slight dip around 10:06 and 10:11.

Current status

Status	Count
Up	13
Down	2
Mixed	4
Total	19

Pings over time

Time	Up (Green)	Down (Red)
10:00	250	50
10:01	300	50
10:02	280	50
10:03	300	50
10:04	280	50
10:05	280	50
10:06	250	50
10:07	280	50
10:08	280	50
10:09	300	50
10:10	280	50
10:11	250	50
10:12	300	50
10:13	280	50
10:14	280	50

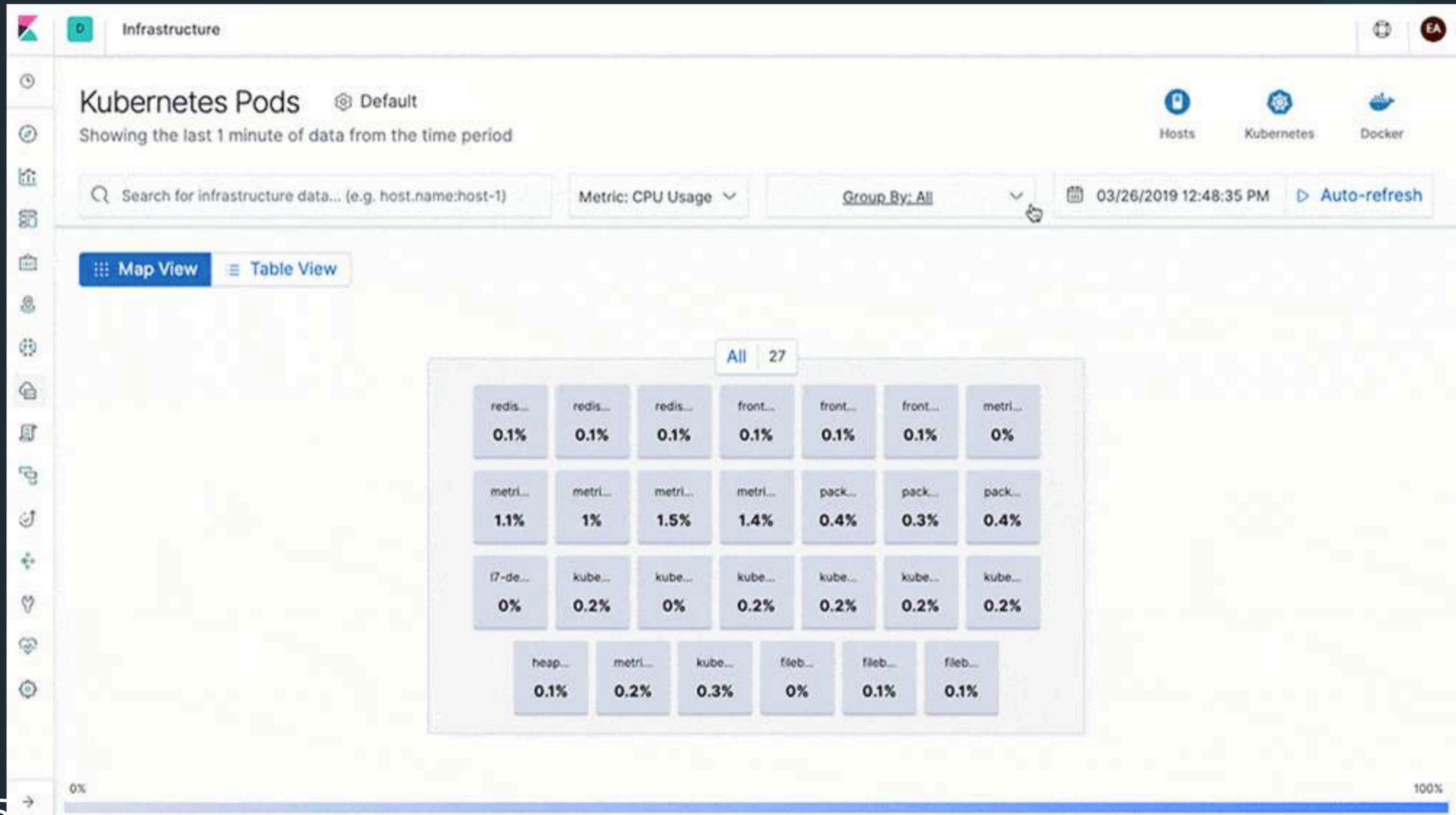
Monitor status

Status	Name	URL	Downtime history	Integrations
Up a few seconds ago	UseCases	https://www.elastic.co/use-cases/		...
Mixed a few seconds ago	NodeJS	http://opbeans-node:3000/api/customers		...
Up a few seconds ago	ca-dc2-rack1	172.18.0.11		
Up a few seconds ago	nyc-dc1-rack1	172.18.0.11		
Down a few seconds ago	uk-dc2-rack3	172.18.0.11		
Mixed		http://opbeans-		

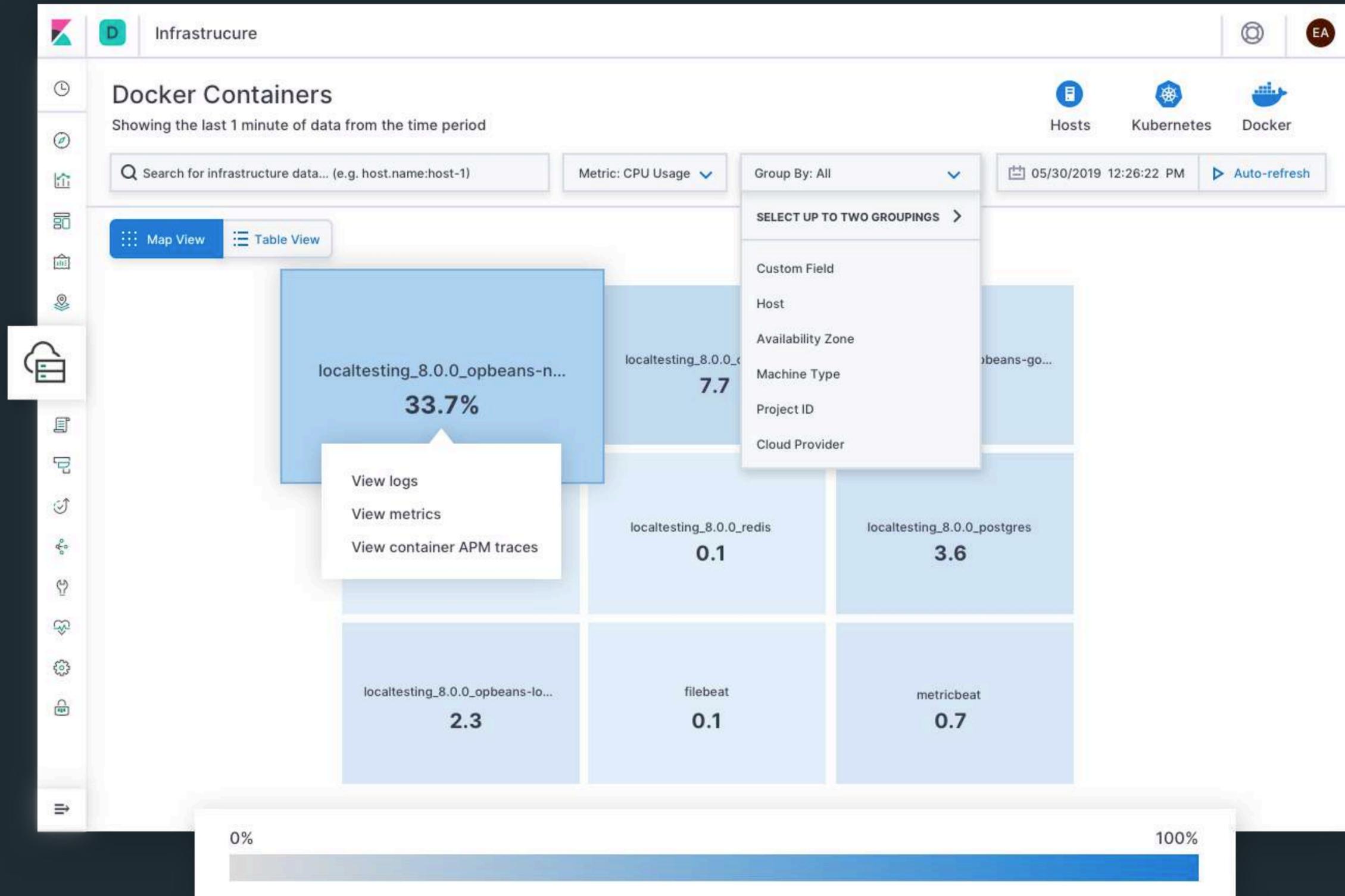
Elastic Uptime

```
1. heartbeat.monitors:
2. - type: http
3.   schedule: '@every 5s'
4.   urls:
5.     - "https://discuss.elastic.co/"
6.     - "https://www.elastic.co"
7.     - "https://demo.elastic.co"
8.   check.response.status: 200
9. - type: http
10.  schedule: '@every 5s'
11.  urls:
12.    - "http://www.elastic.co"
13.  check.response.status: 301
14. - type: http
15.  schedule: '@every 5s'
16.  urls:
17.    - "https://www.elastic.co/solutions/apm"
18.  check.response:
19.    status: 200
20.    body: "Open Source Application Performance Monitoring"
```

Elastic Infrastructure



Elastic Infrastructure





Deployment options



Distributions

- zip, tar.gz, RPM, DEB
- debian/rpm repositories, homebrew tap
- Docker, Helm chart
- K8s Operator (ECK)

Elastic Cloud



Elasticsearch Service

Easily spin up deployments on AWS, GCP or Azure with Kibana and features you can't get anywhere else.

AS LOW AS

\$16

per month

[Product Overview](#)

[Start Trial](#)

By submitting you agree to the [Elastic Cloud Standard Terms of Service](#) and to receive occasional emails from Elastic. Your personal data will be processed in accordance with Elastic's [privacy statement](#).



Elastic App Search Service

Build a fast, relevant search experience for your custom application in just a few minutes.

AS LOW AS

\$49

per month

[Product Overview](#)

[Start Trial](#)

By submitting you agree to the [Elastic Cloud Standard Terms of Service](#) and to receive occasional emails from Elastic. Your personal data will be processed in accordance with Elastic's [privacy statement](#).



Elastic Site Search Service

Everything you need to deliver a powerful search experience for your website — without the learning curve.

AS LOW AS

\$79

per month

[Product Overview](#)

[Start Trial](#)

By submitting you agree to the [Elastic Cloud Standard Terms of Service](#) and to receive occasional emails from Elastic. Your personal data will be processed in accordance with Elastic's [privacy statement](#).

Elastic Cloud Enterprise





meetup.com RSVP stream demo

Time series data...





logging workshop demo

—
start your engines...





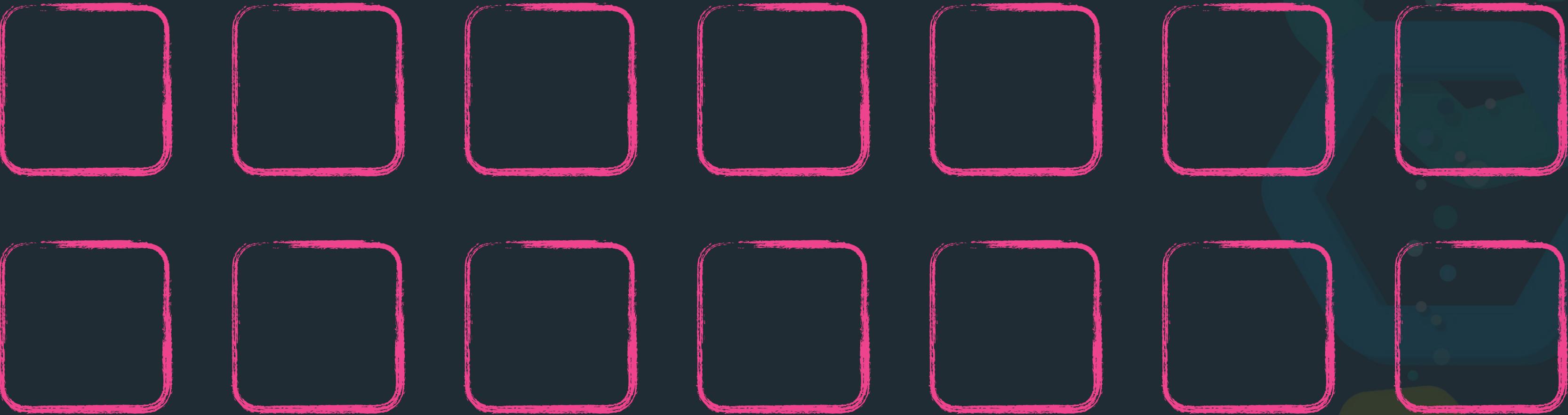
Logging patterns



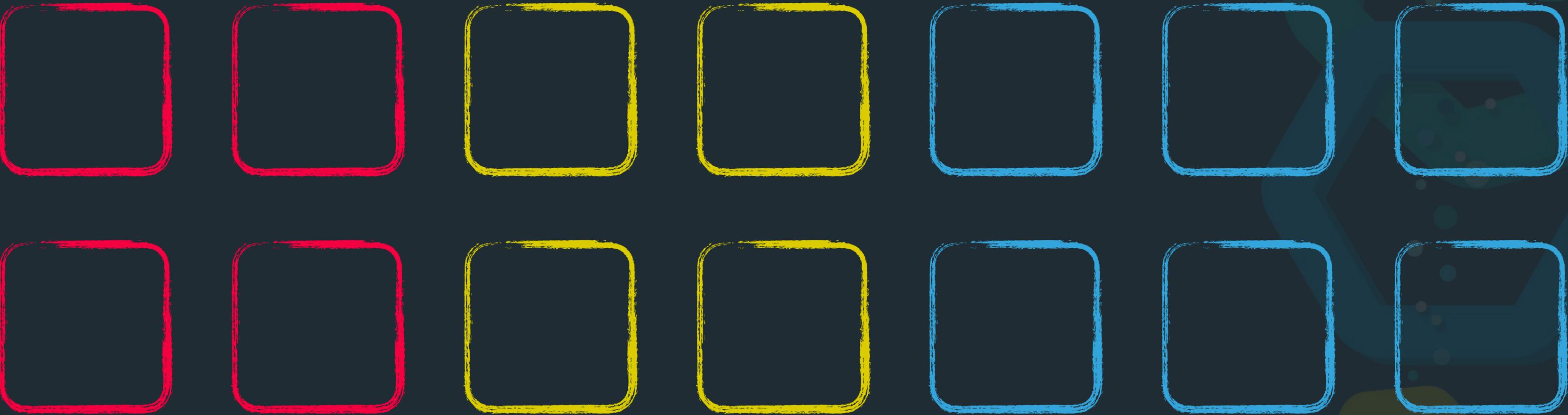
Time based data

- time based data has properties
- current data gets indexed
- more recent data gets searched more
- old data is still required 'just in case'

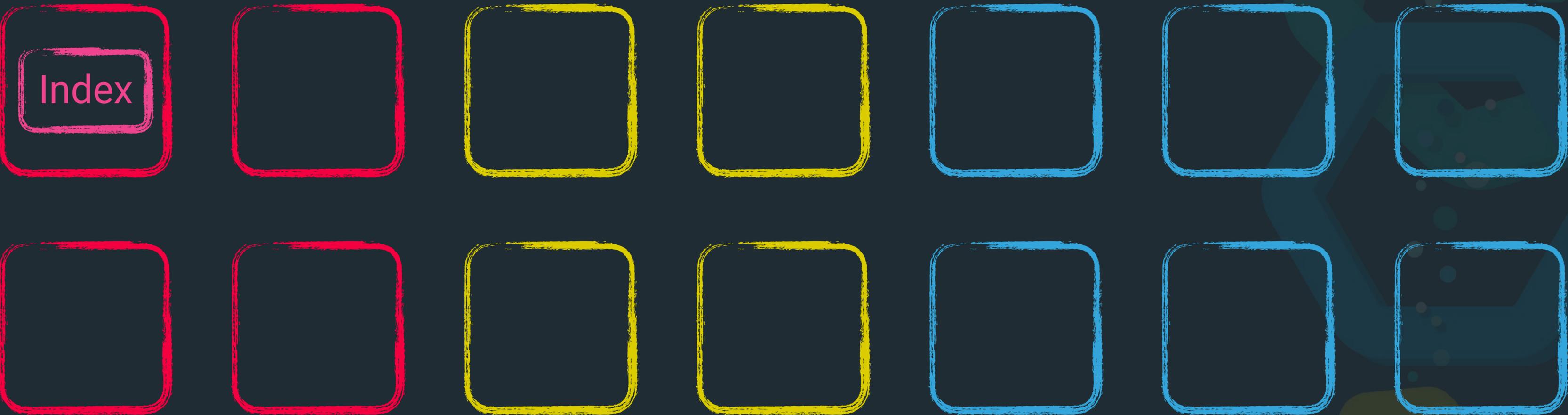
Homogeneous architecture



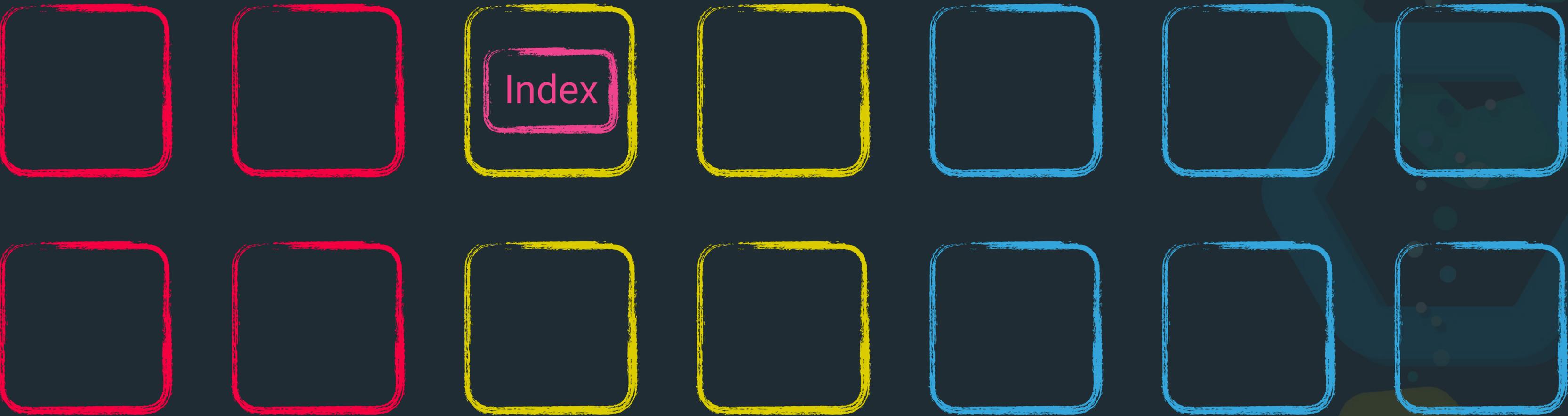
Hot warm architecture



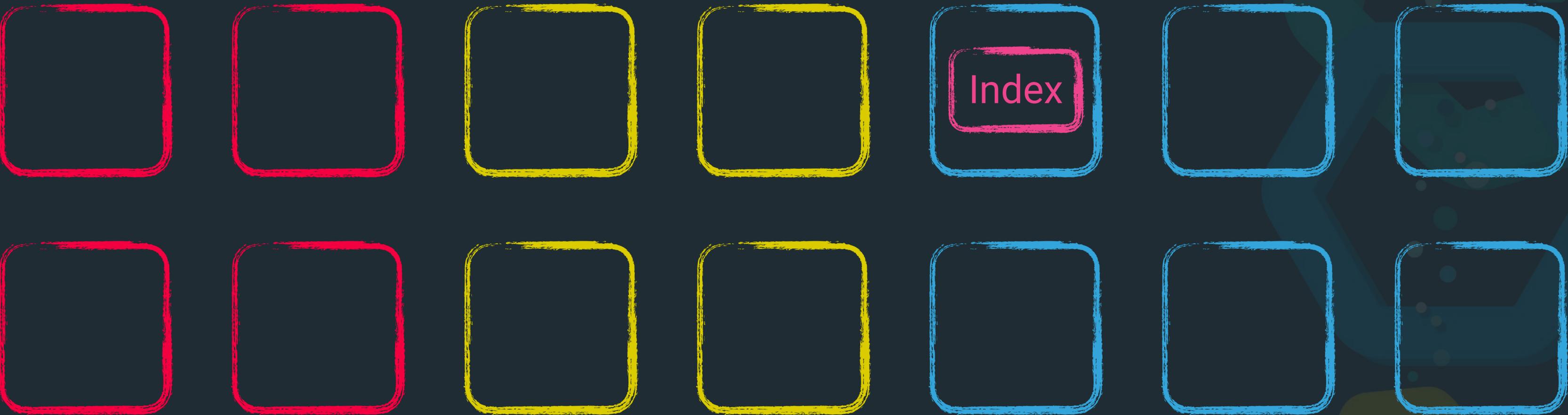
Hot warm architecture



Hot warm architecture



Hot warm architecture



Index Lifecycle Management

- Hot: read & write
- Warm: frequently read
- Cold: seldom read
- Delete: no longer needed

Index Lifecycle Management: Hot

- rollover
- set priority
- unfollow

Index Lifecycle Management: Warm

- set priority
- unfollow
- read-only
- allocate
- shrink
- forge merge

Index Lifecycle Management: cold

- set priority
- unfollow
- allocate
- freeze

More lifecycle topics

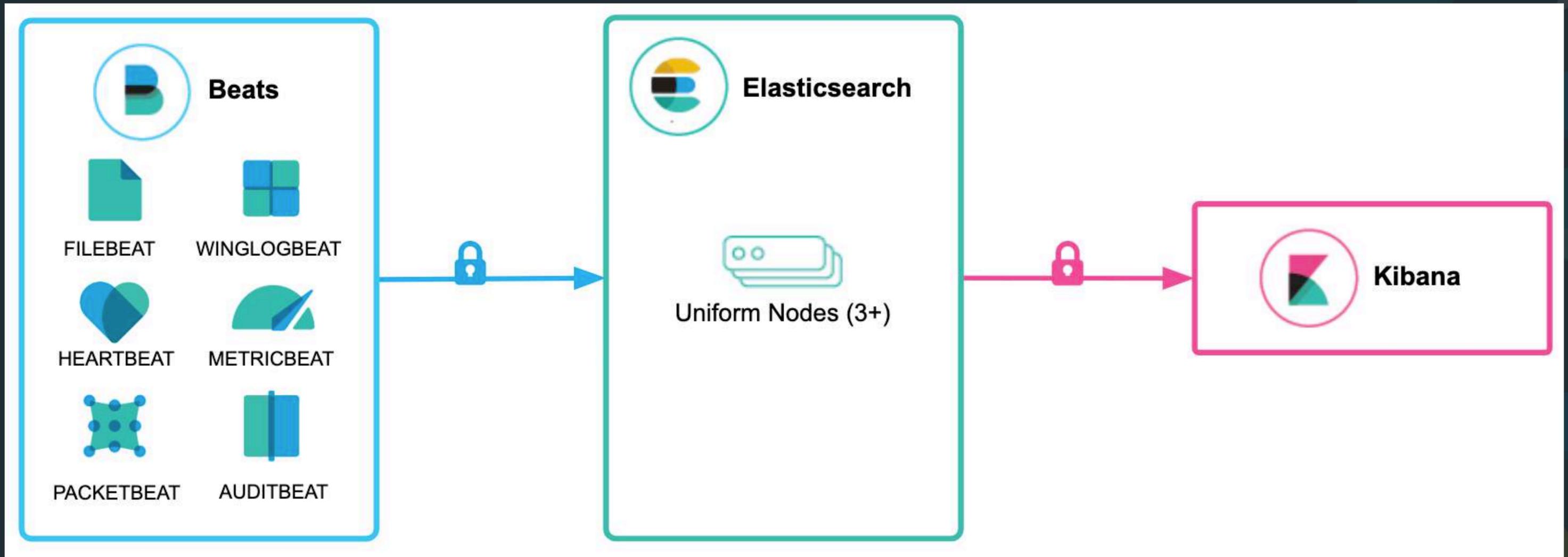
- SLM: create snapshots based on cron
- Rollup: Summarize and store historical data
- Transform: Pivot data to entity centric indices

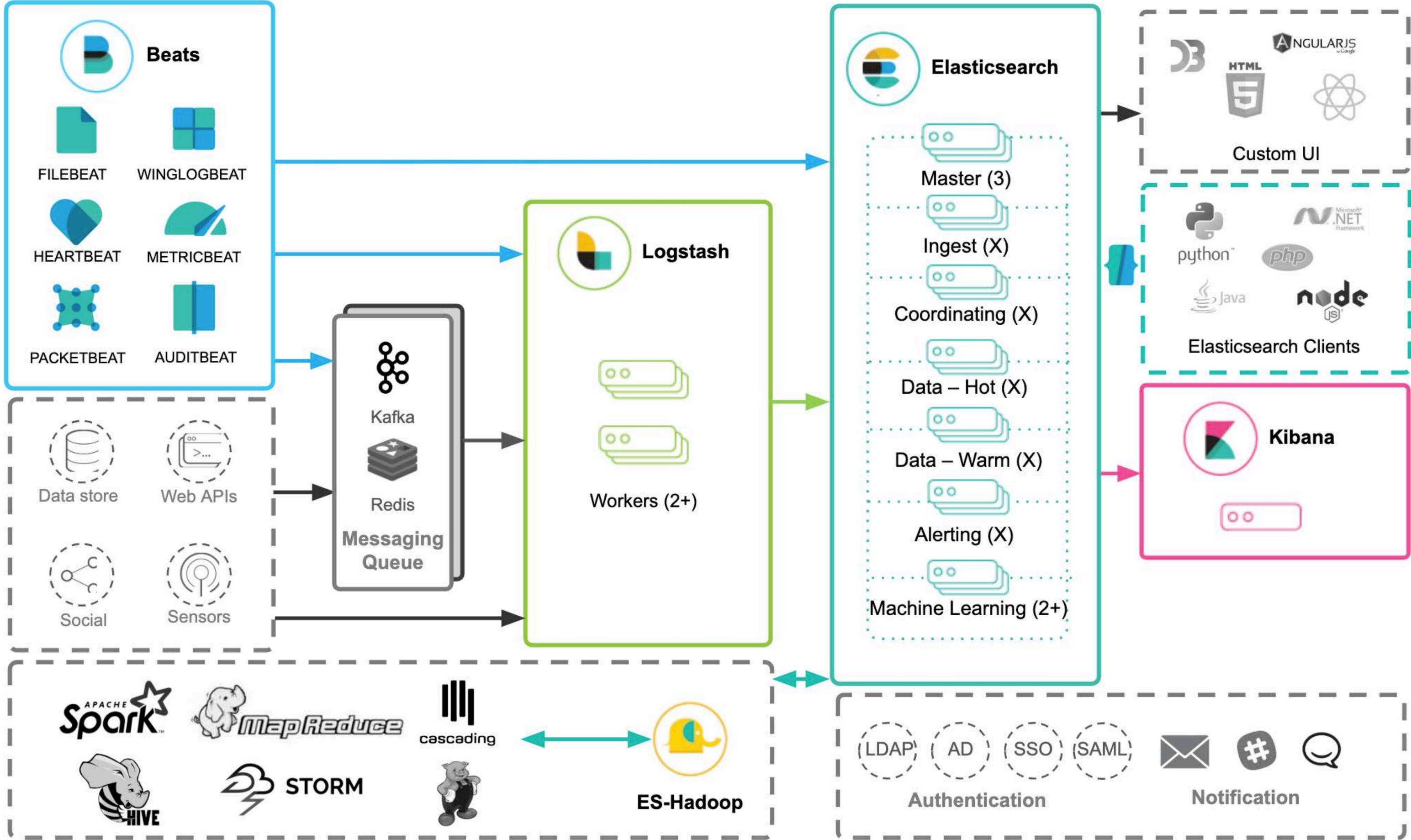


Architecture patterns



Start small





https://ela.st/cfcamp-workshop-munich



Deploy Elasticsearch and Kibana in 3 Minutes or Less

We hope you learned something new at our "Centralize your logs with Elastic Stack" Workshop — put your knowledge to the test and try it out on Elasticsearch Service.

- ✓ Get a 30-day free trial (vs. standard 14-day)
- ✓ No credit card required
- ✓ Get the latest versions, powerful features, and optimized deployment templates for your use case.

Enter your email

Start Free Trial

By submitting you agree to the [Elastic Cloud Standard Terms of Service](#) and to receive occasional emails from Elastic. Your personal data will be processed in accordance with our [Privacy Policy](#).

- Deployments
- Custom plugins
- Account
- Help

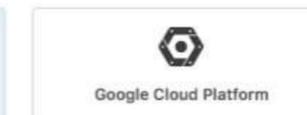
Create deployment

1 Name your deployment

Give your deployment a name

2 Select a cloud platform

Pick your cloud and let us handle the rest. No additional accounts required.



3 Select a region

US East (N. Virginia)

US West (N. California)

US West (Oregon)

EU (Ireland)

Asia Pacific (Singapore)

Asia Pacific (Tokyo)

South America (Sao Paulo)

Asia Pacific (Sydney)

EU (Frankfurt)

4 Set up your deployment

Elastic Stack version

6.5.1 [Edit](#)

Select a deployment to restore from its latest snapshot

5 Optimize your deployment

I/O Optimized

Recommended

Use for search and general all-purpose workloads. Includes a balance of compute, memory, and storage.

Default specs



Compute Optimized

Run CPU-intensive workloads or run smaller workloads cost-effectively when you need less memory and storage.

Default specs



Memory Optimized

Perform memory-intensive operations efficiently, including workloads with frequent aggregations.

Default specs



Hot-Warm Archi

Use for time-series and logging workloads that benefit from automatic data retention.

Default specs



Elastic Cloud supports many more options to cater to your specific use case such as hot-warm architecture optimized for logging, compute-optimized for analytics, and memory-optimized for machine learning.

classification

search

precision

crawler

links

spam

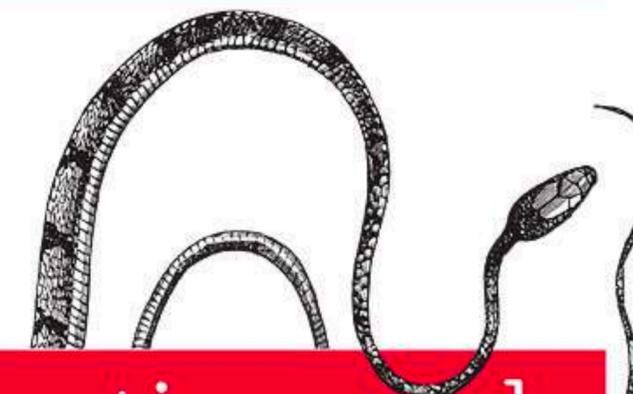
recall

query

Christopher D. Manning
Prabhakar Raghavan
Hinrich Schütze

Introduction to Information Retrieval

O'REILLY



Elasticsearch

The Definitive Guide

A DISTRIBUTED REAL-TIME SEARCH AND ANALYTICS ENGINE

Covers Apache Lucene 3.0

Lucene

IN ACTION

SECOND EDITION



Michael McCandless
Erik Hatcher
Otis Gospodnetić

FOREWORD BY DOUG CUTTING

DEEP LEARNING

for Search

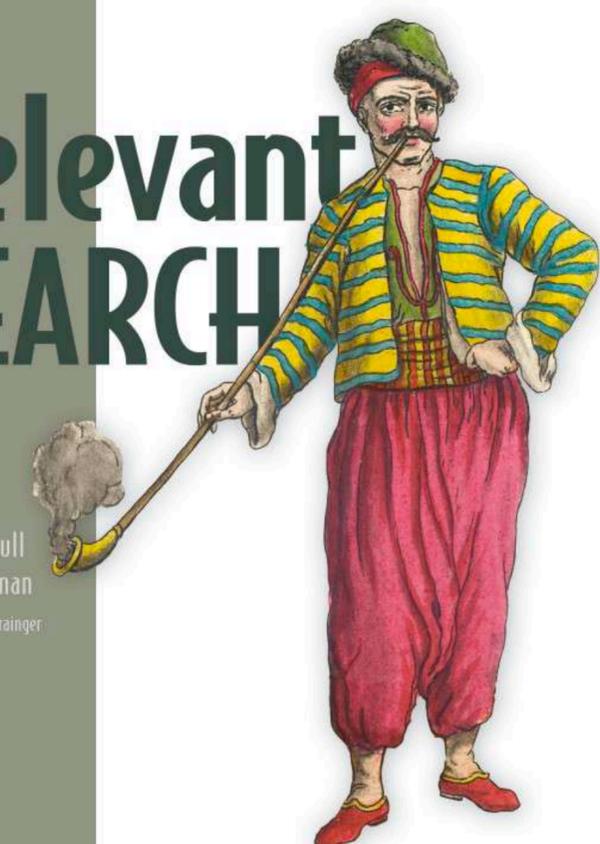


Tommaso Teofili
Foreword by Chris Mattmann

MANNING

With applications for Solr and Elasticsearch

Relevant SEARCH

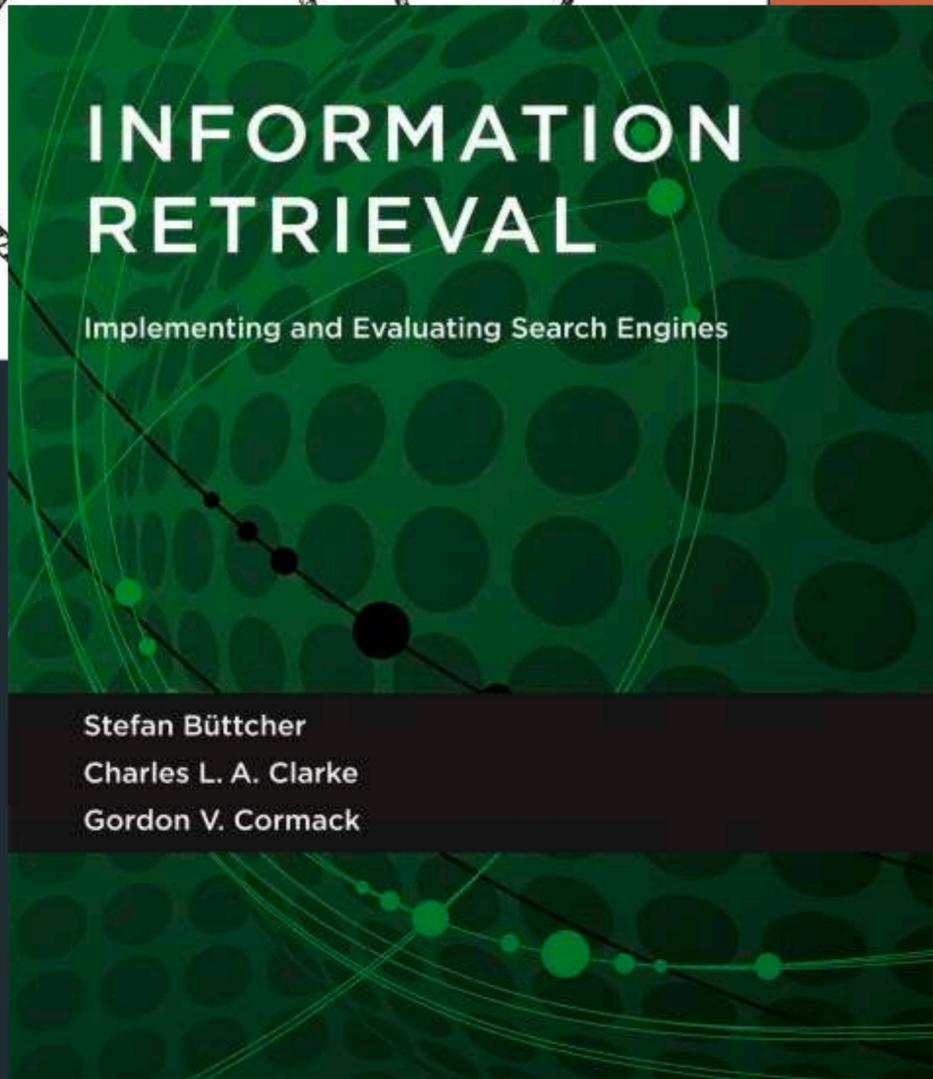


Doug Turnbull
John Berryman
FOREWORD BY TREY GRAINGER

MANNING

INFORMATION RETRIEVAL

Implementing and Evaluating Search Engines

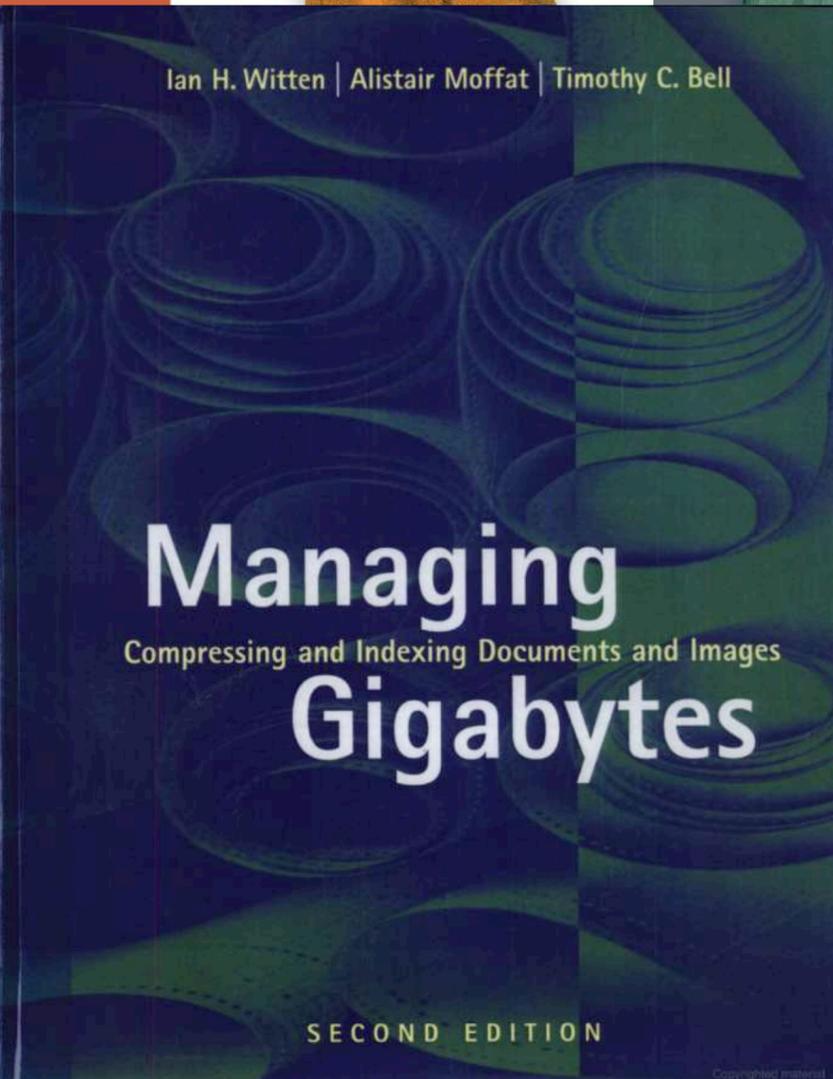


Stefan Büttcher
Charles L. A. Clarke
Gordon V. Cormack

Ian H. Witten | Alistair Moffat | Timothy C. Bell

Managing Gigabytes

Compressing and Indexing Documents and Images



SECOND EDITION



Q & A

