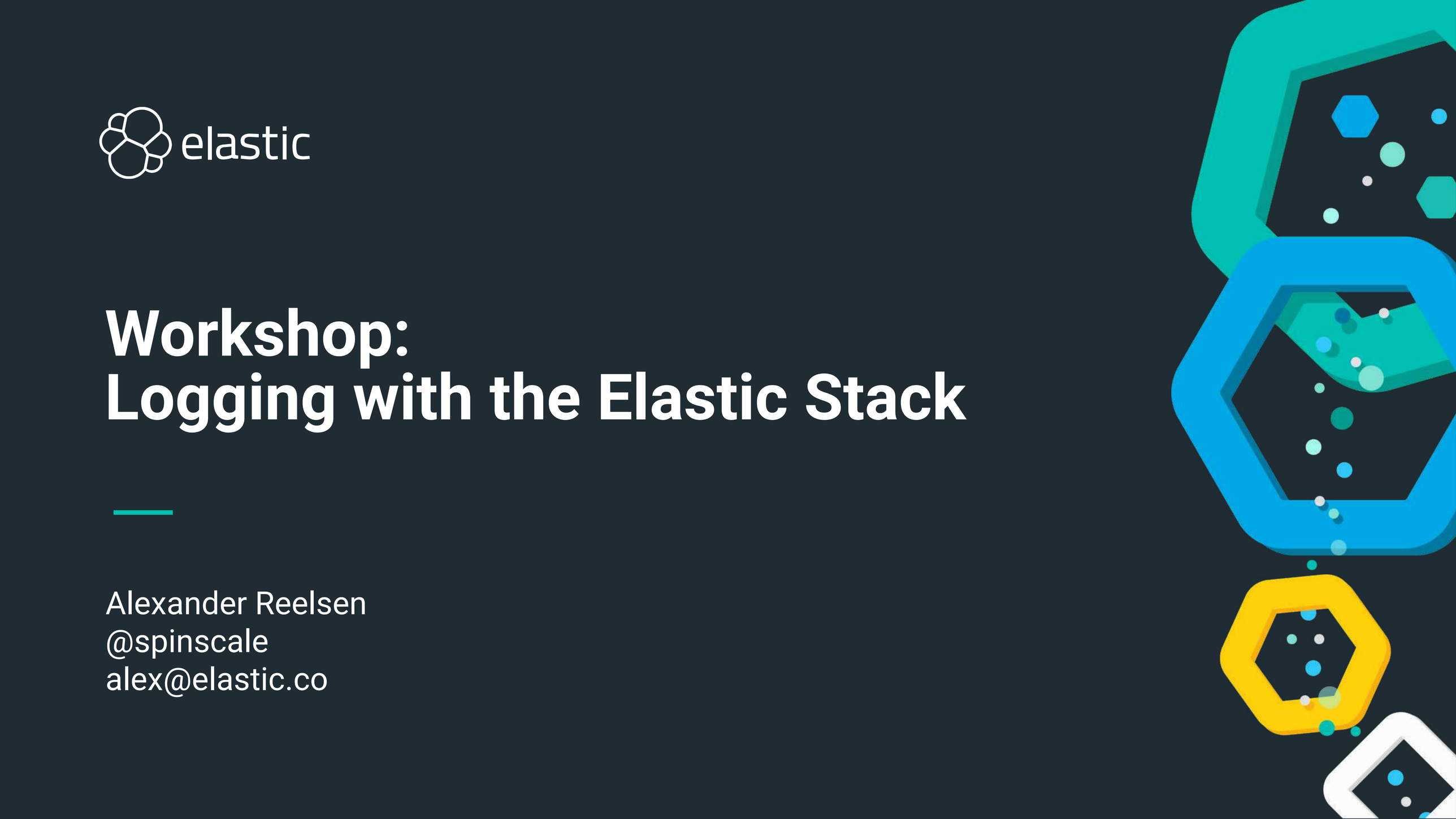
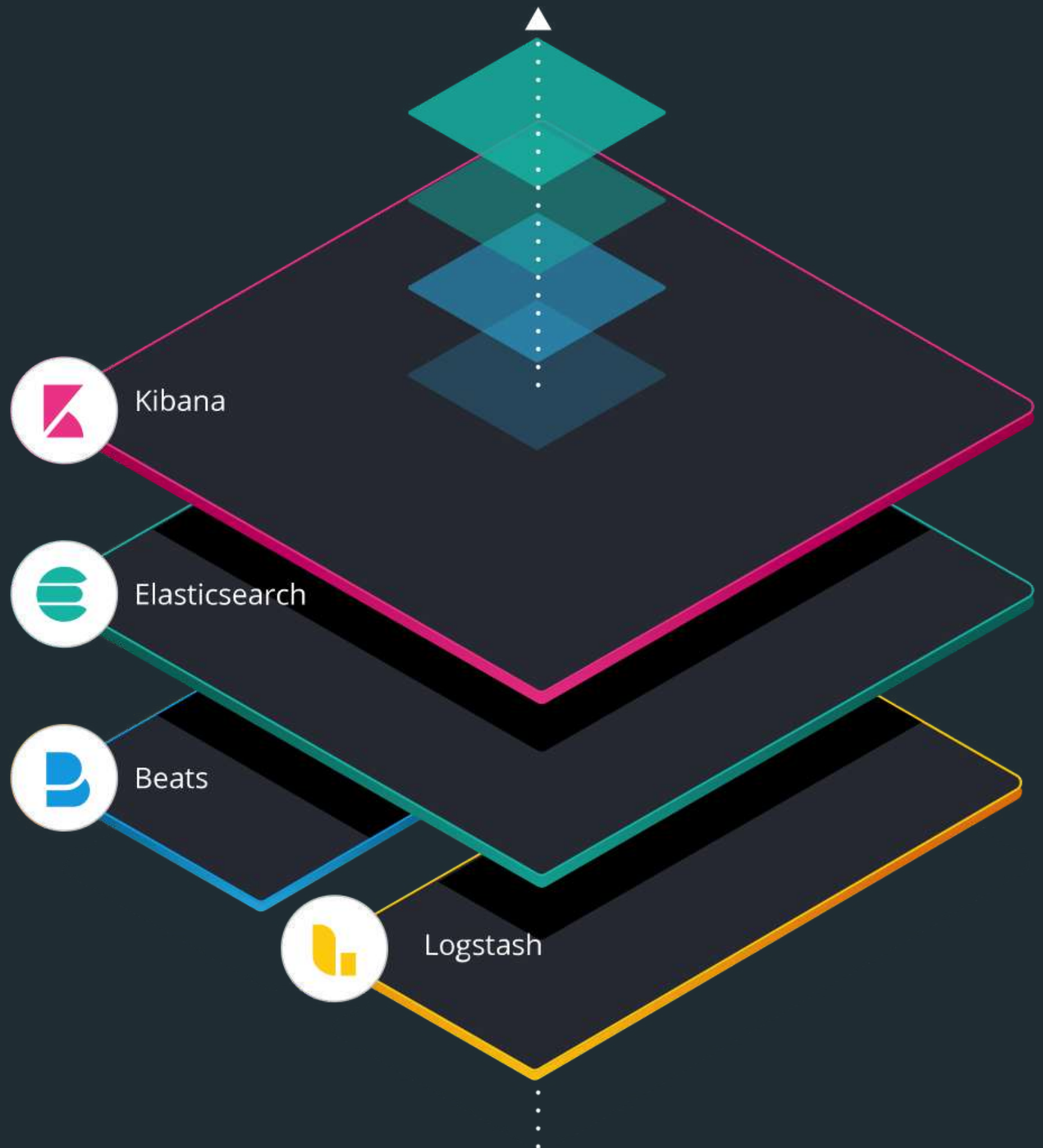


Workshop:
Logging with the Elastic Stack

Alexander Reelsen
@spinscale
alex@elastic.co



elastic

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

elastic

Prerequisites

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

elastic

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}{q3RjloGxRdm176yLo9d9UA}{Vq4FpFklRbCyVFAVKU7ukQ}
{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

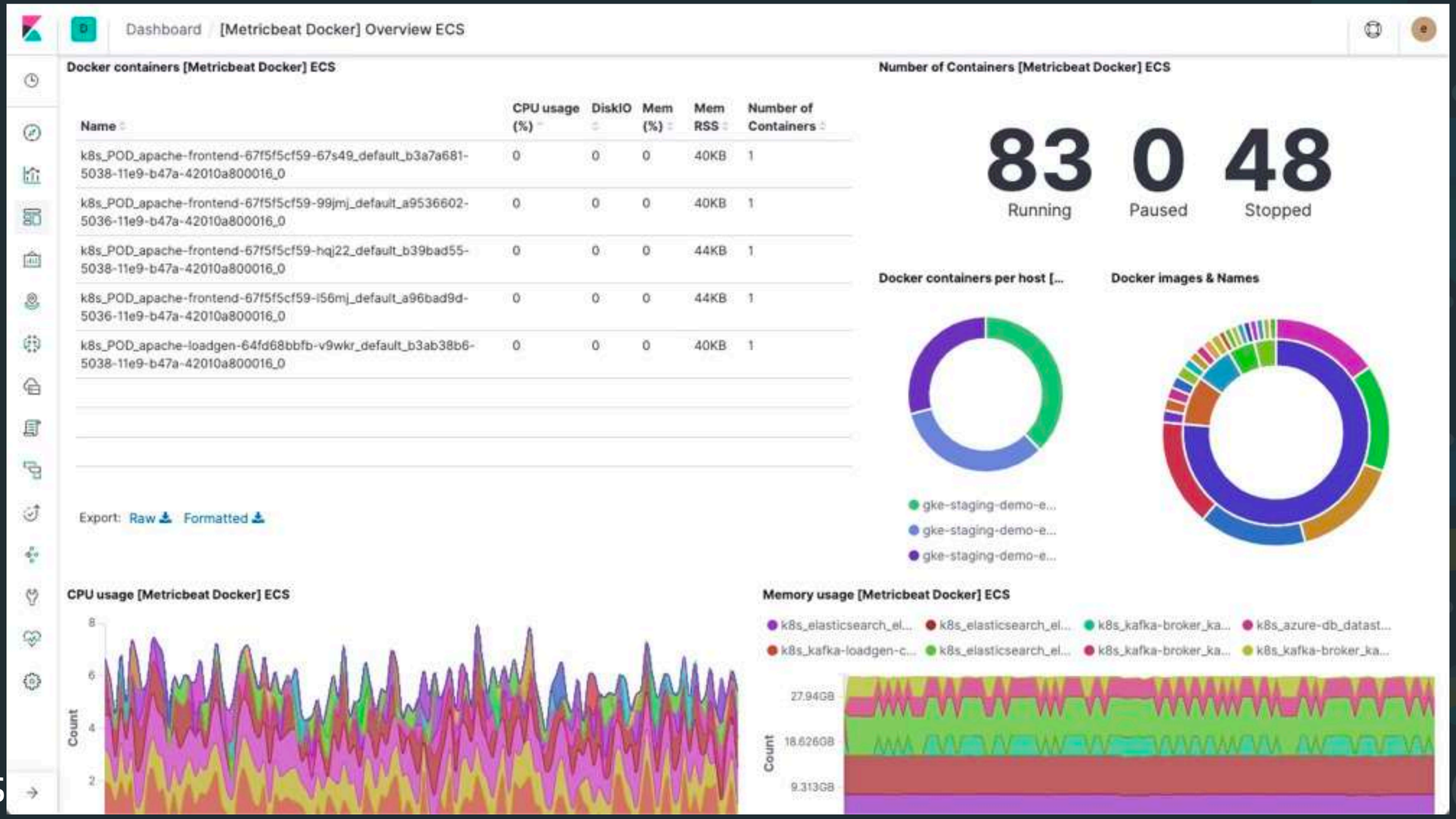
Time series have a lifecycle

- 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

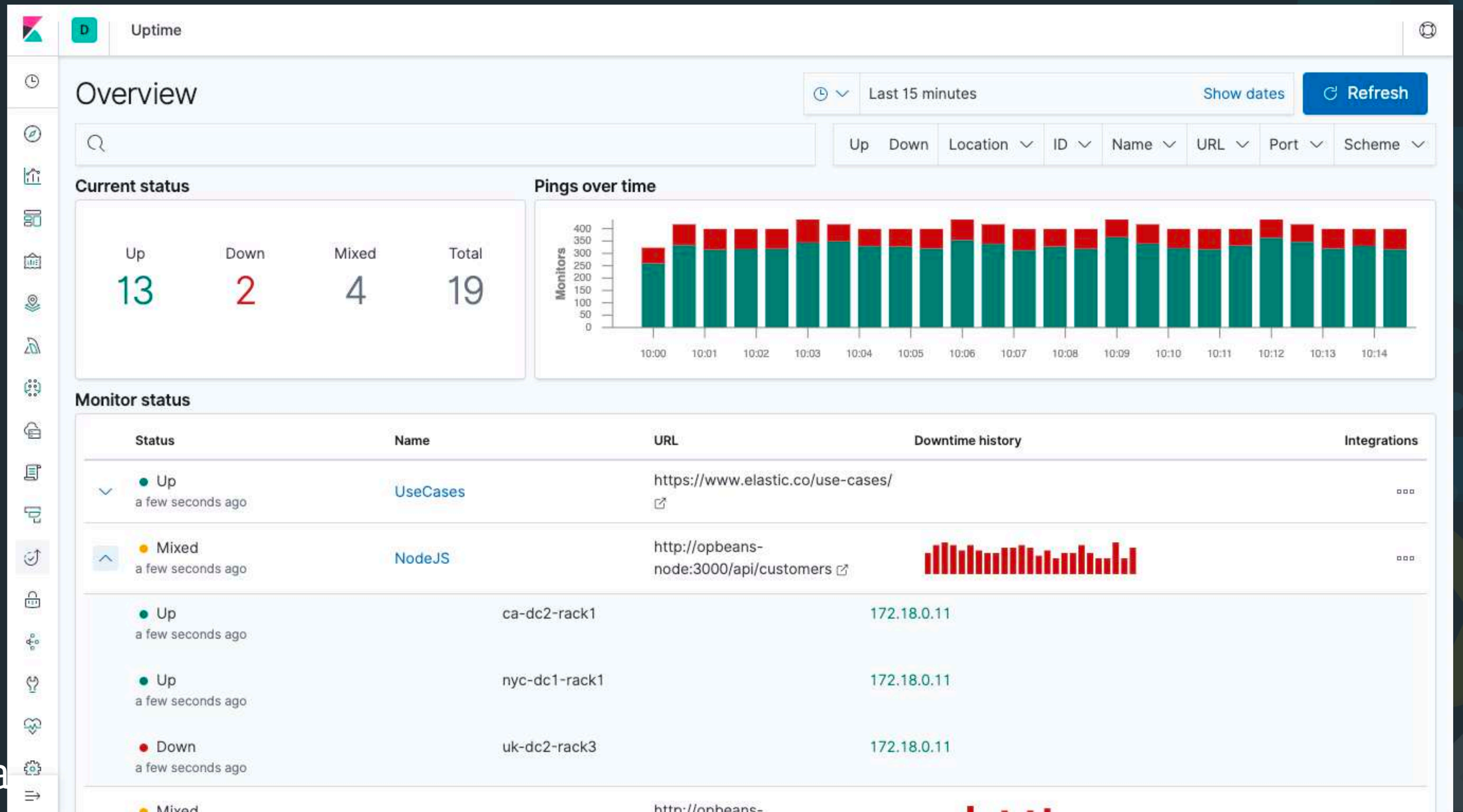
- Max response time per 10 minute window since yesterday
- Documents: All documents from yesterday till now
- Aggregate in 10 minute buckets (6*24)
- For each bucket, extract max value

Dashboards & Time Series

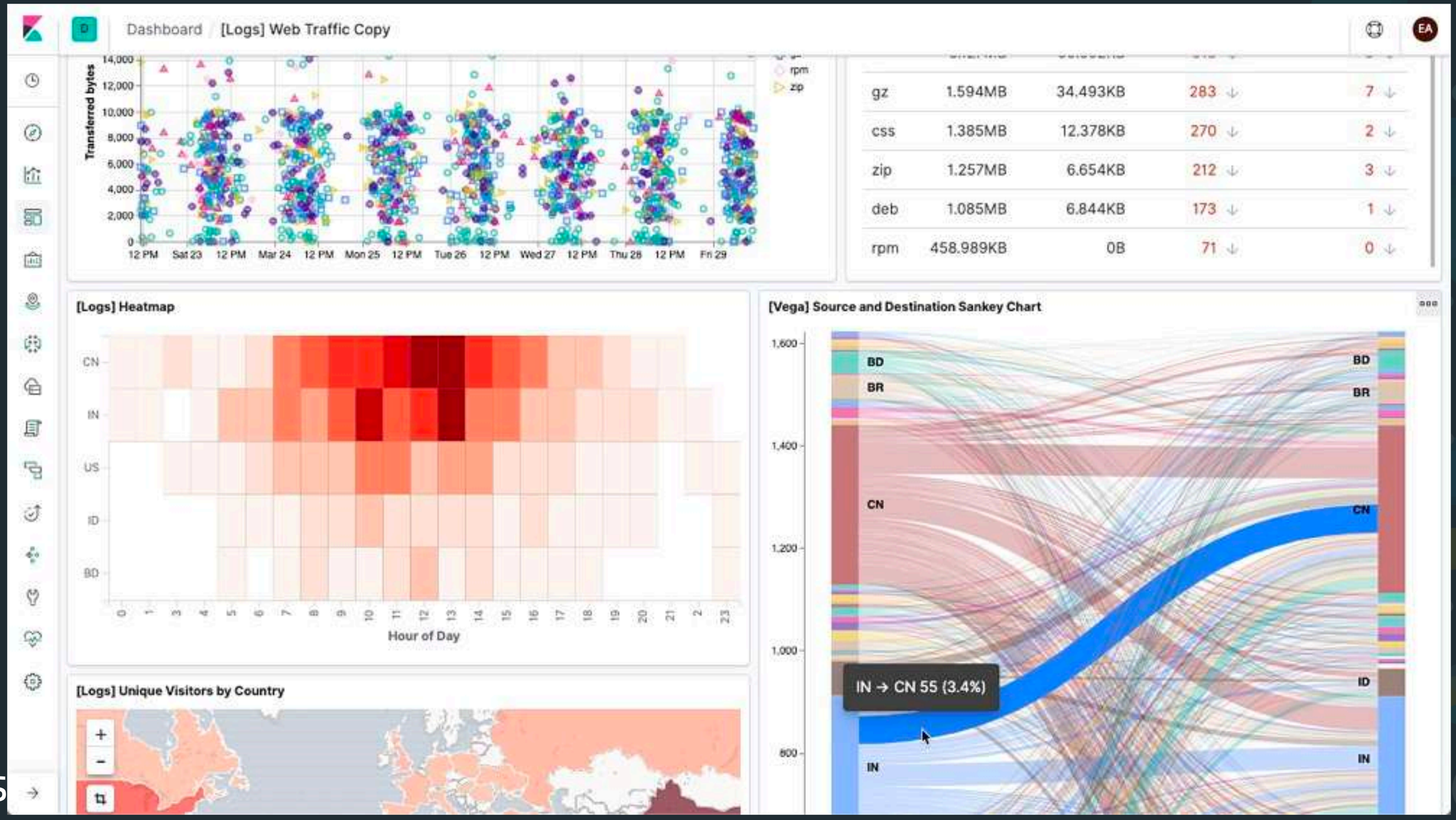


Dashboards & Time Series

Dashboards & Time Series



Dashboards & Time Series



Dashboards & Time Series

Standardizing data

What is ECS?

edit

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.

elastic

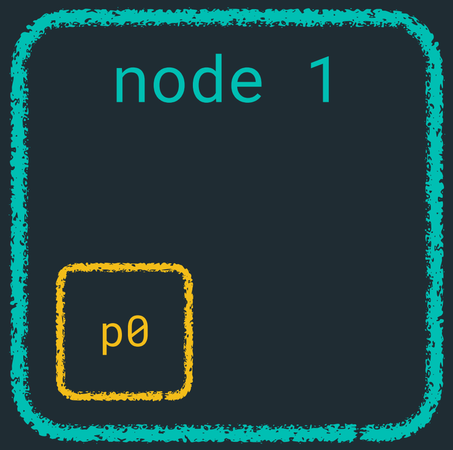
elastic

Elasticsearch overview

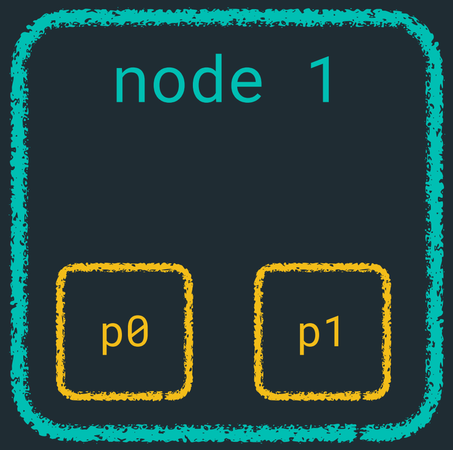
Elasticsearch in 10 seconds

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

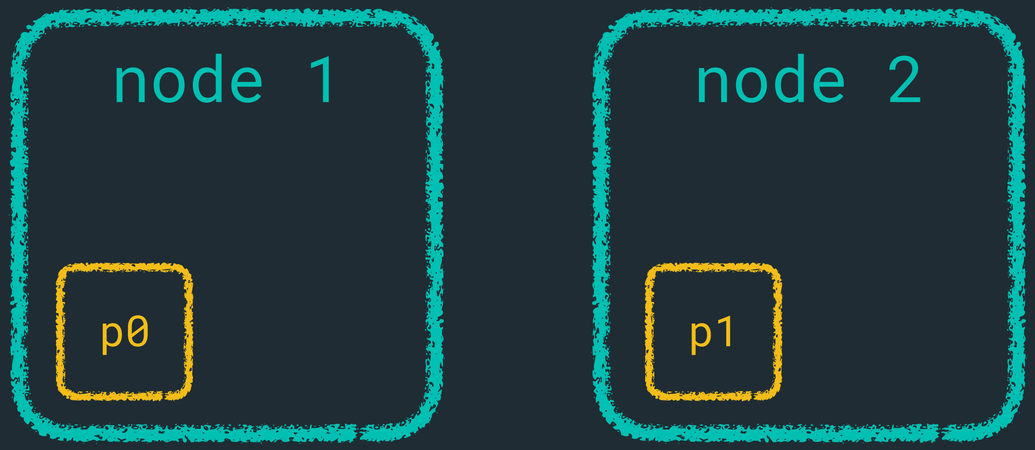
Elasticsearch - a distributed system



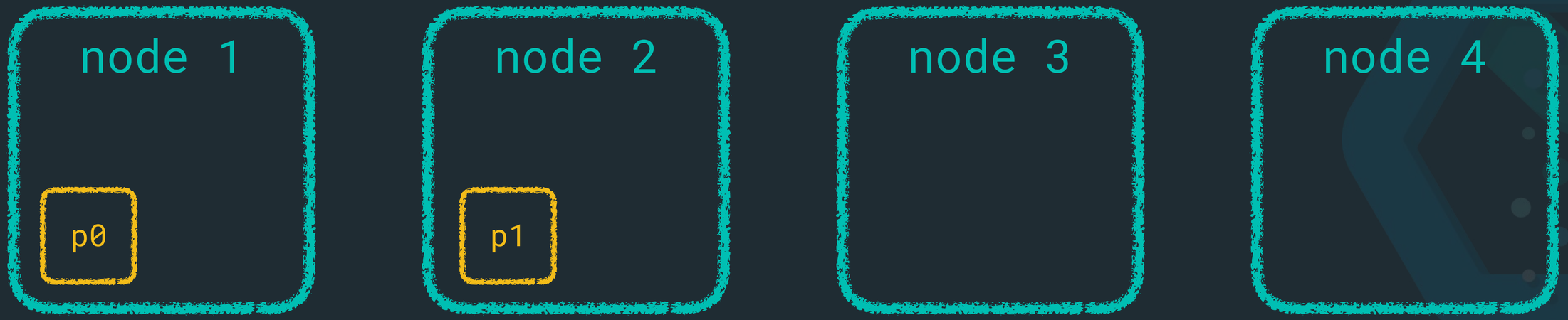
Elasticsearch - a distributed system



Elasticsearch - a distributed system

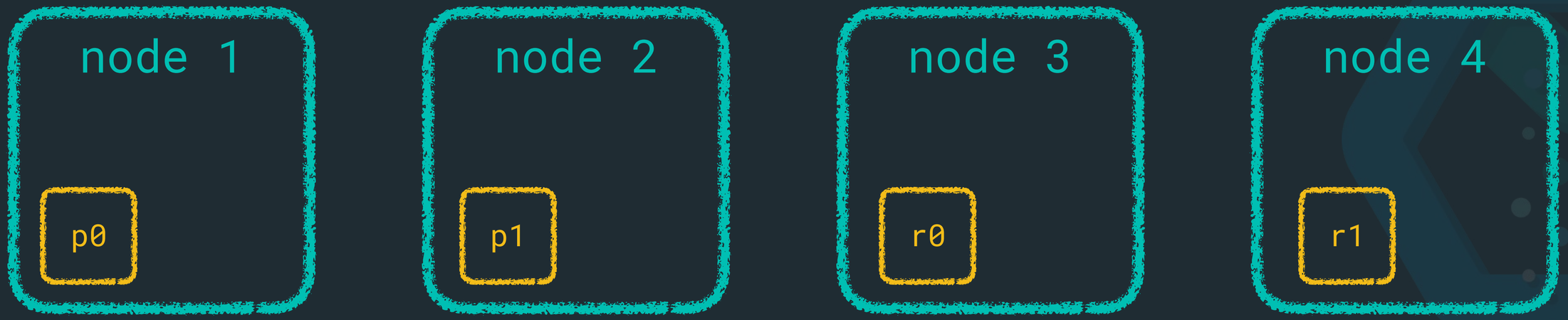


Elasticsearch - a distributed system



elastic

Elasticsearch - a distributed system



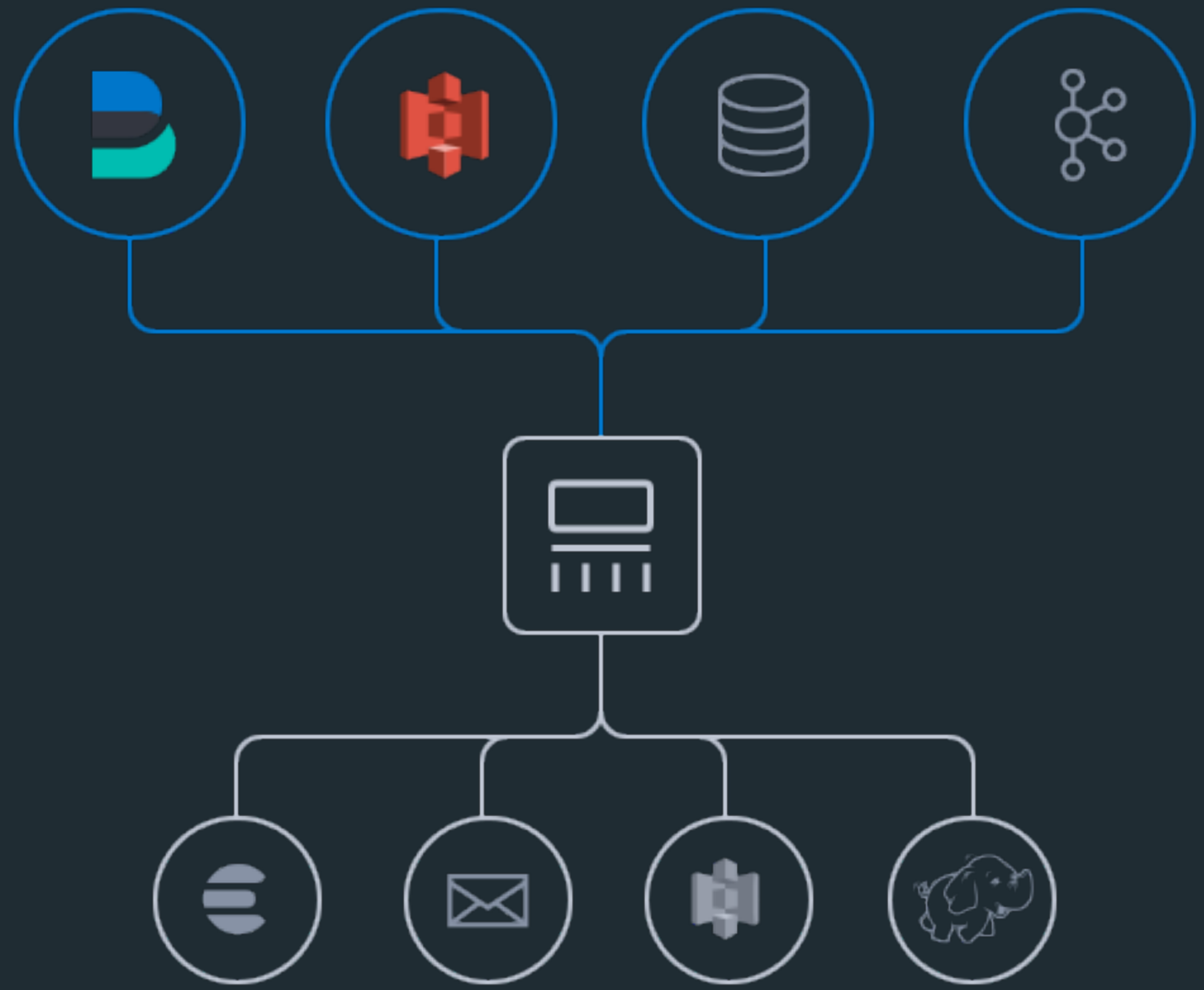
elastic

Ingest overview

Ingestion

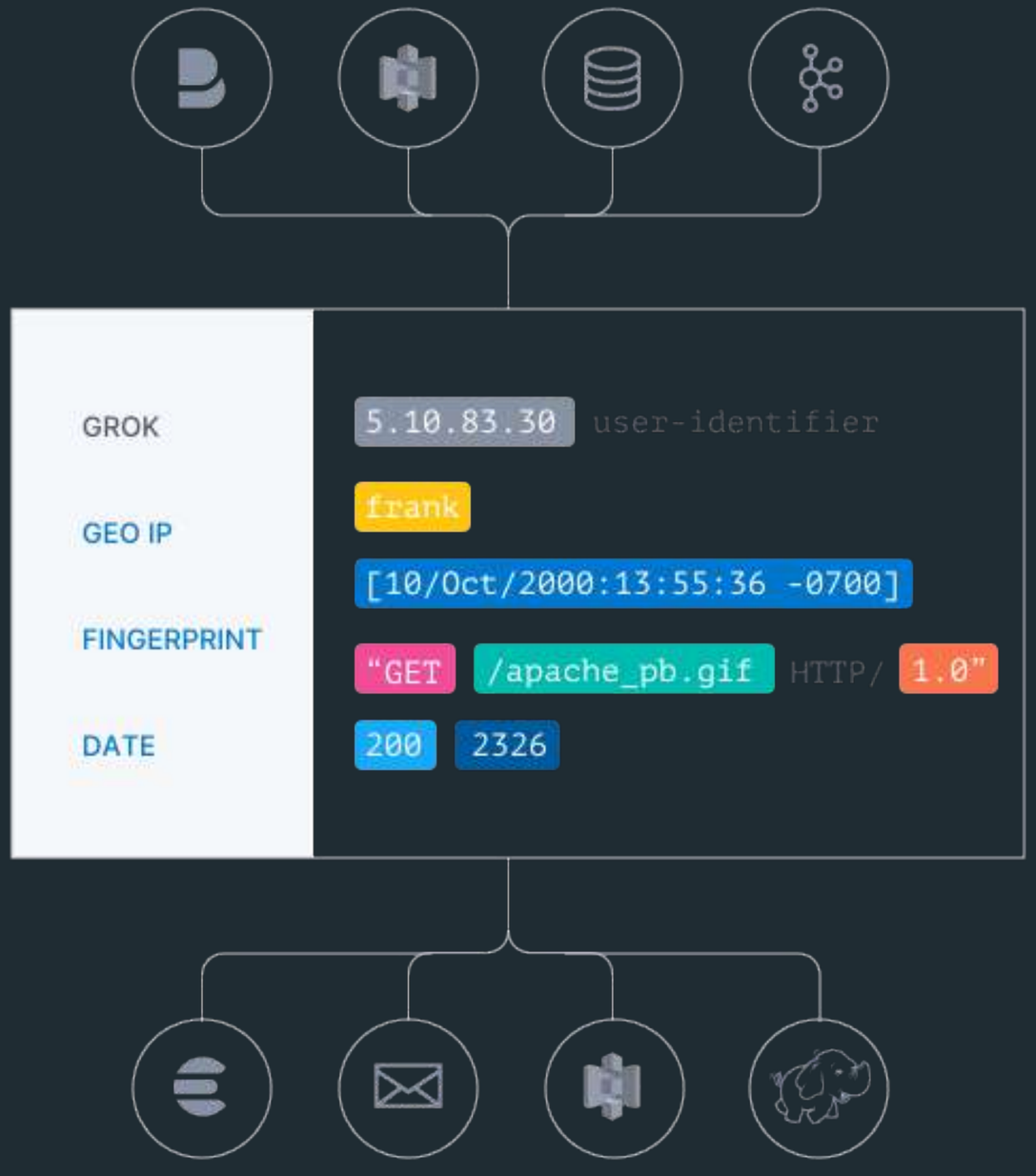
- Logstash: extensible dynamic data collection
- Beats: specialized single purpose data shipper
- your own rolled integration, it's all HTTP!

Logstash



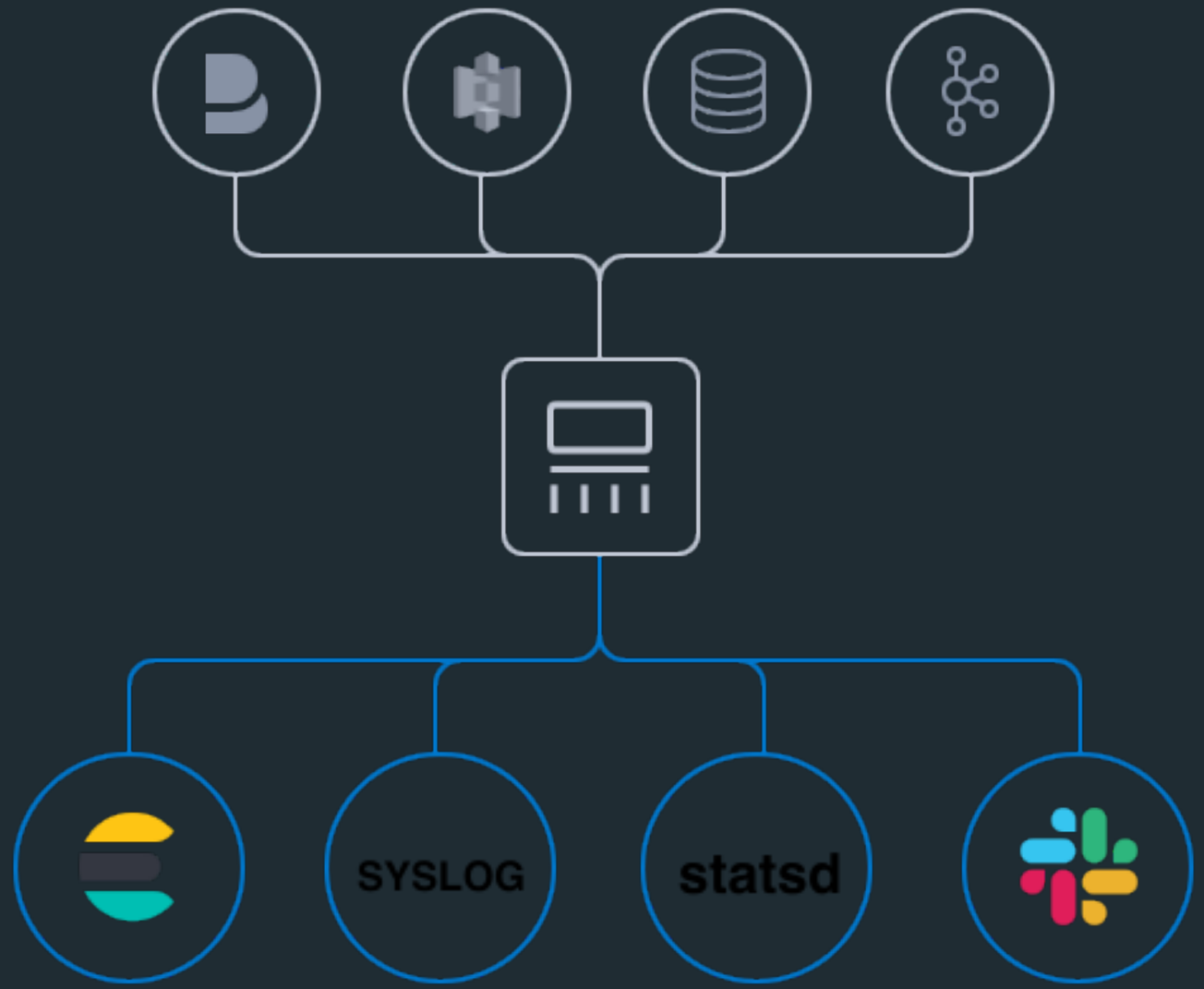
elastic

Logstash



elastic

Logstash



elastic

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)

elastic

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

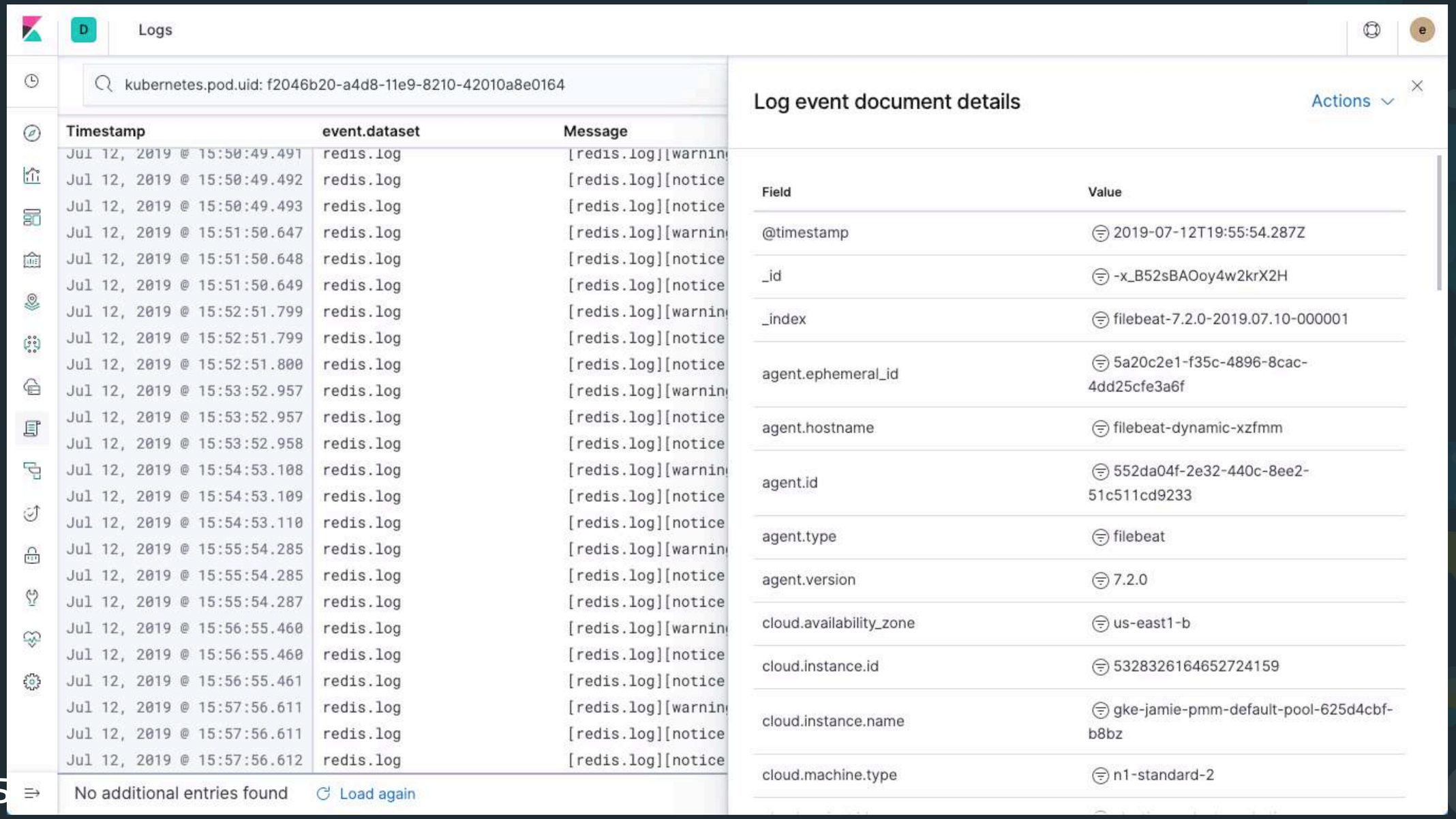
elastic

Solutions

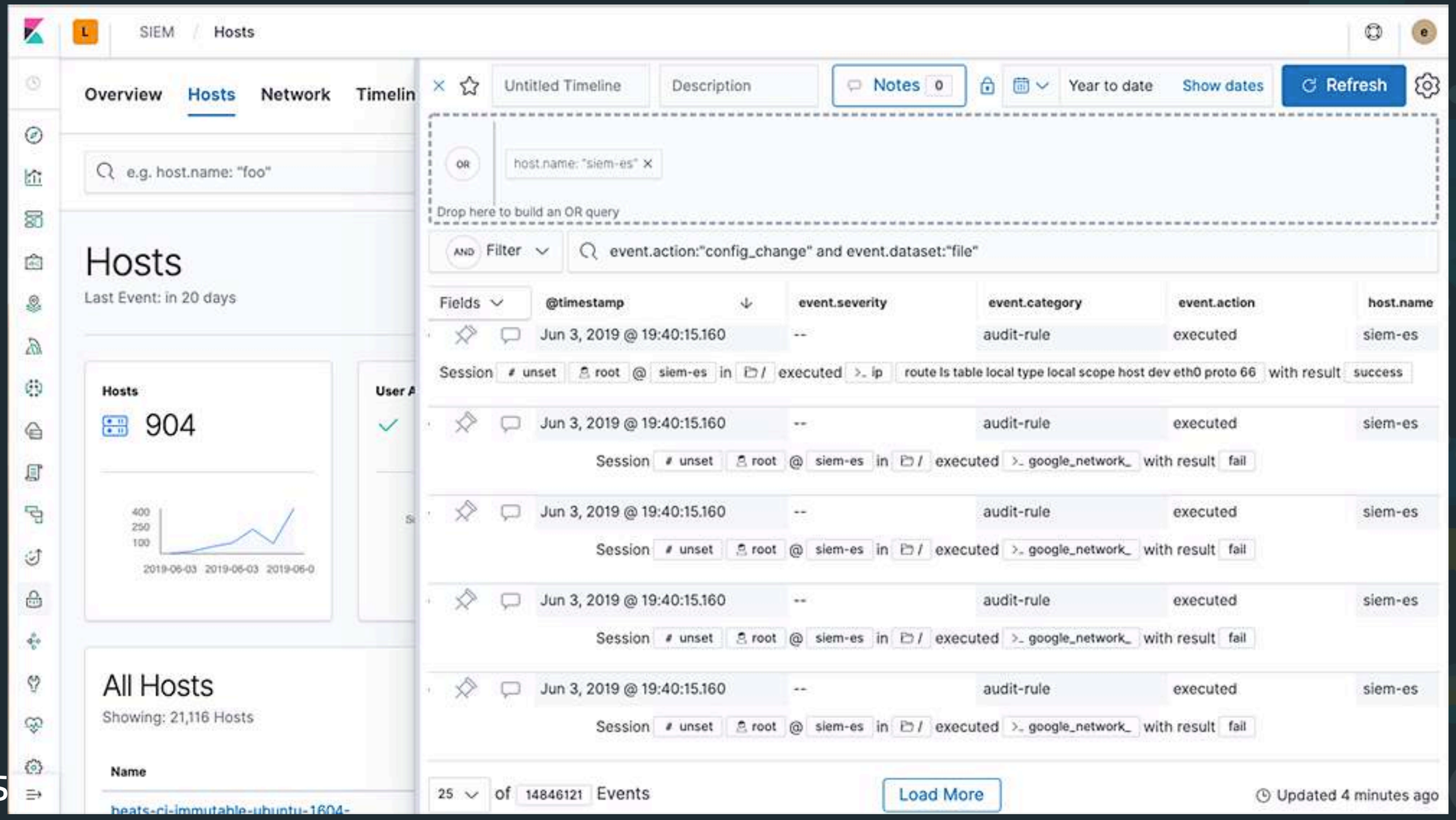
Elastic APM

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

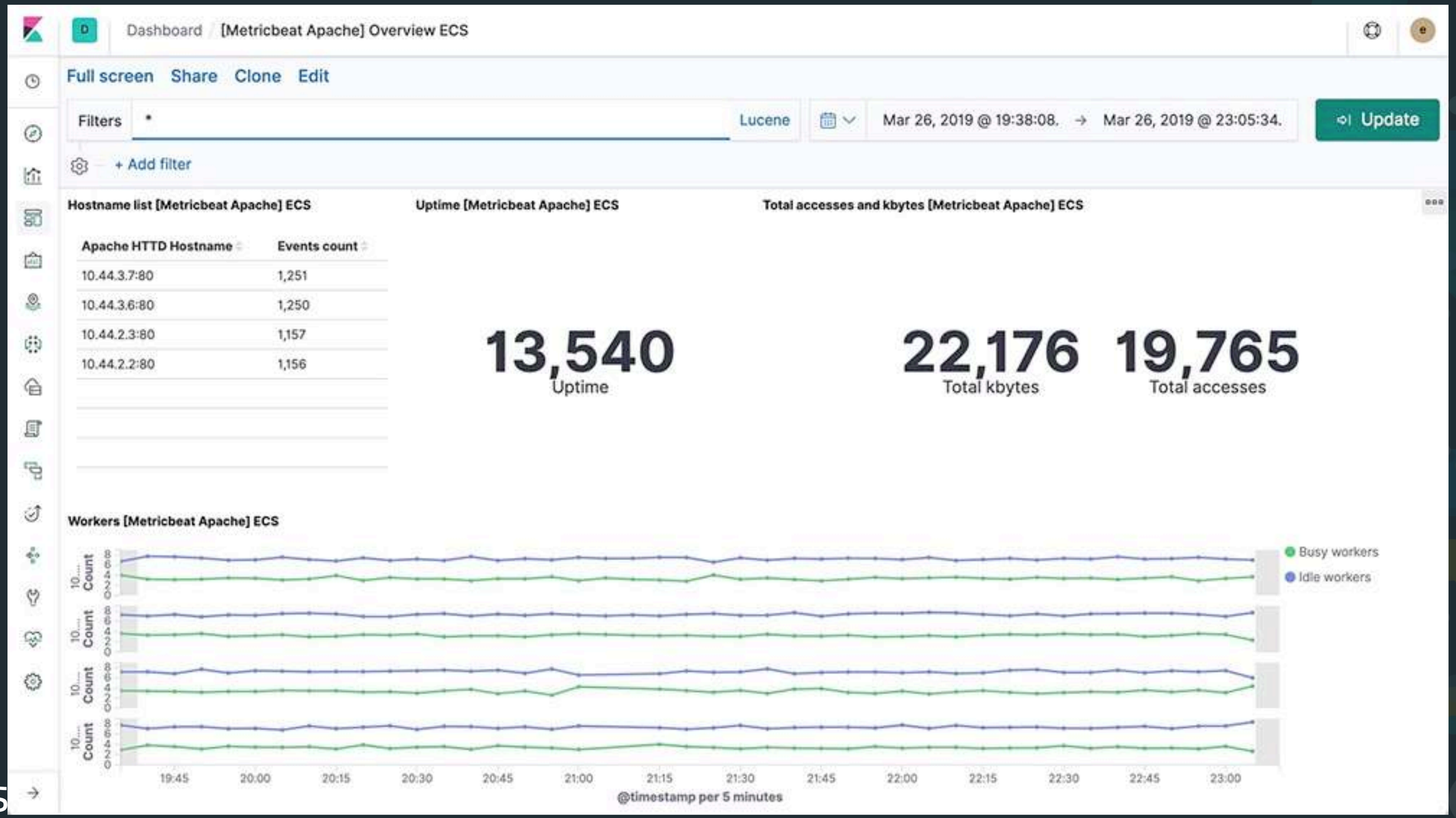
Elastic Logs



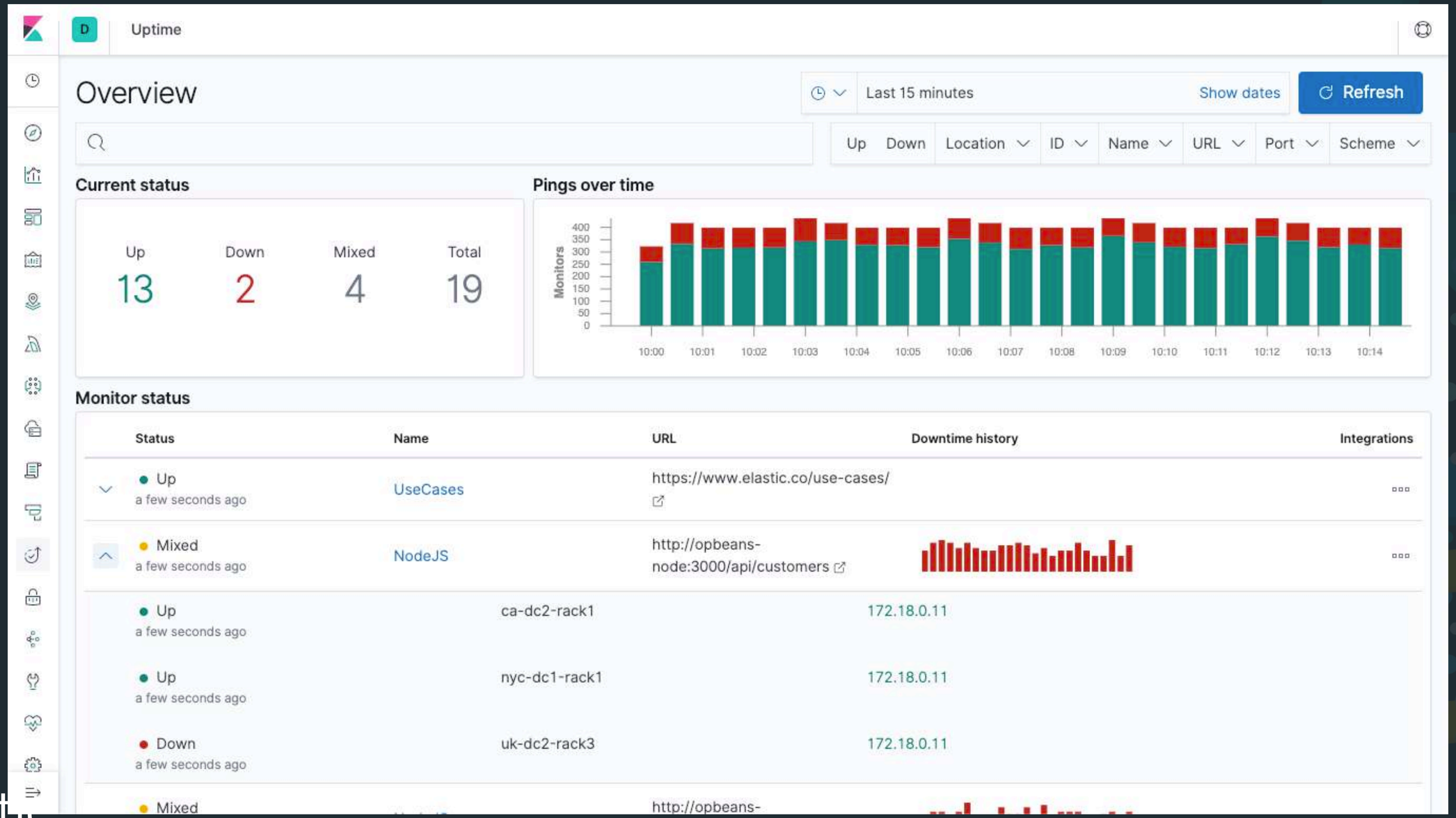
Elastic SIEM



Elastic Metrics



Elastic Uptime



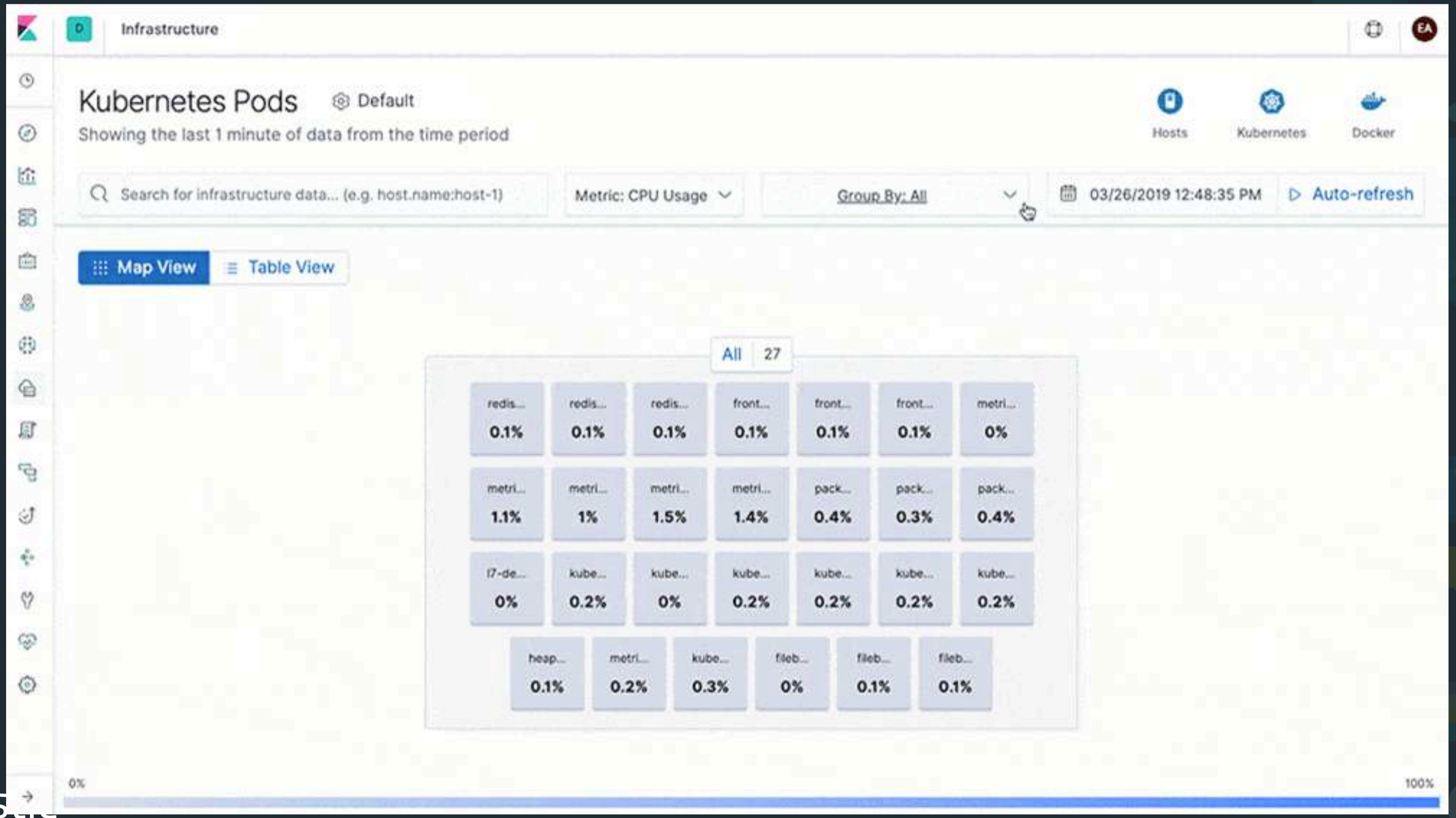
elastic

Elastic Uptime

```
heartbeat.monitors:
- type: http
  schedule: '@every 5s'
  urls:
    - "https://discuss.elastic.co/"
    - "https://www.elastic.co"
    - "https://demo.elastic.co"
  check.response.status: 200
- type: http
  schedule: '@every 5s'
  urls:
    - "http://www.elastic.co"
  check.response.status: 301
- type: http
  schedule: '@every 5s'
  urls:
    - "https://www.elastic.co/solutions/apm"
  check.response:
    status: 200
    body: "Open Source Application Performance Monitoring"
```

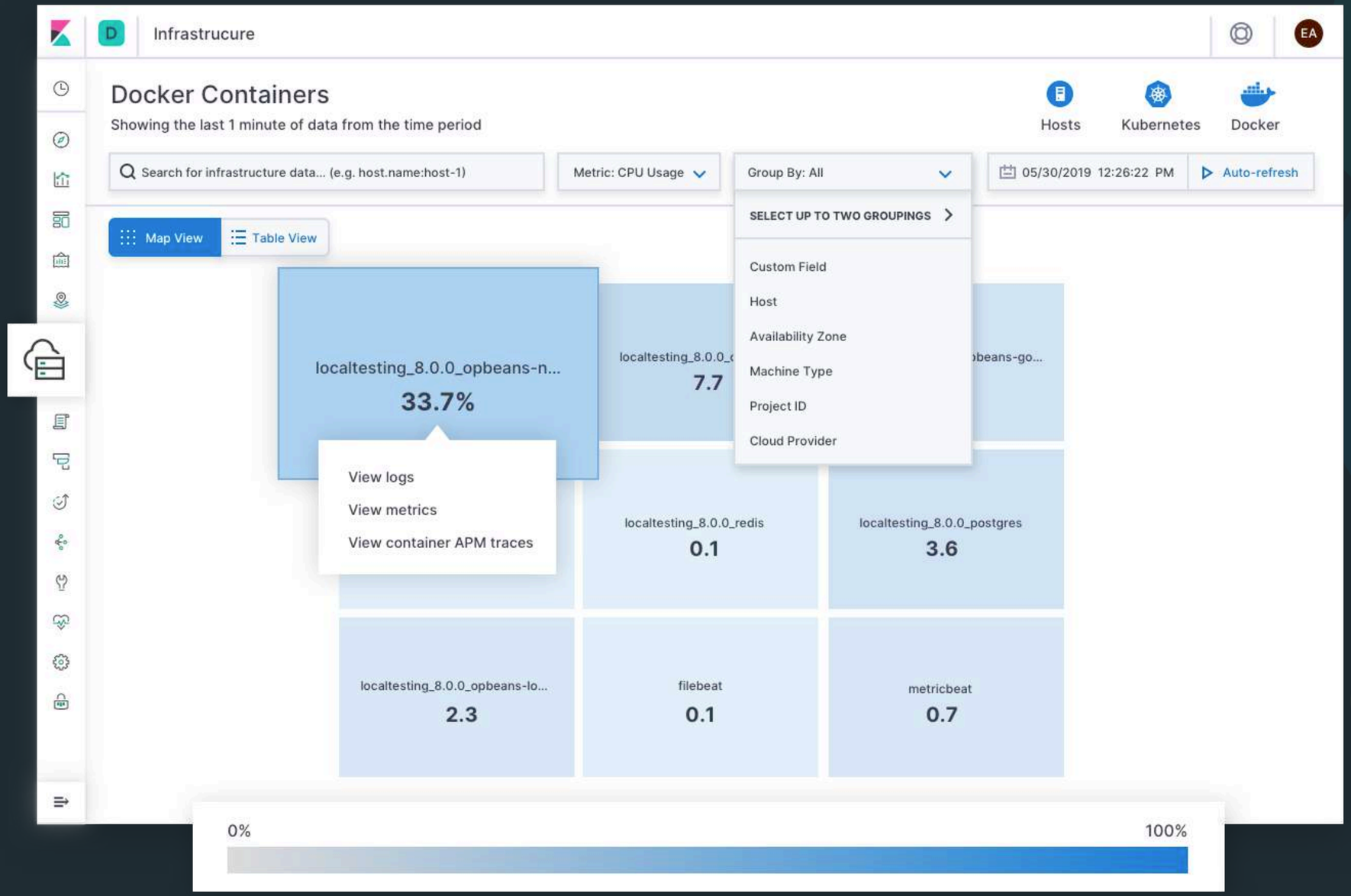
elastic

Elastic Infrastructure



elastic

Elastic Infrastructure



elastic

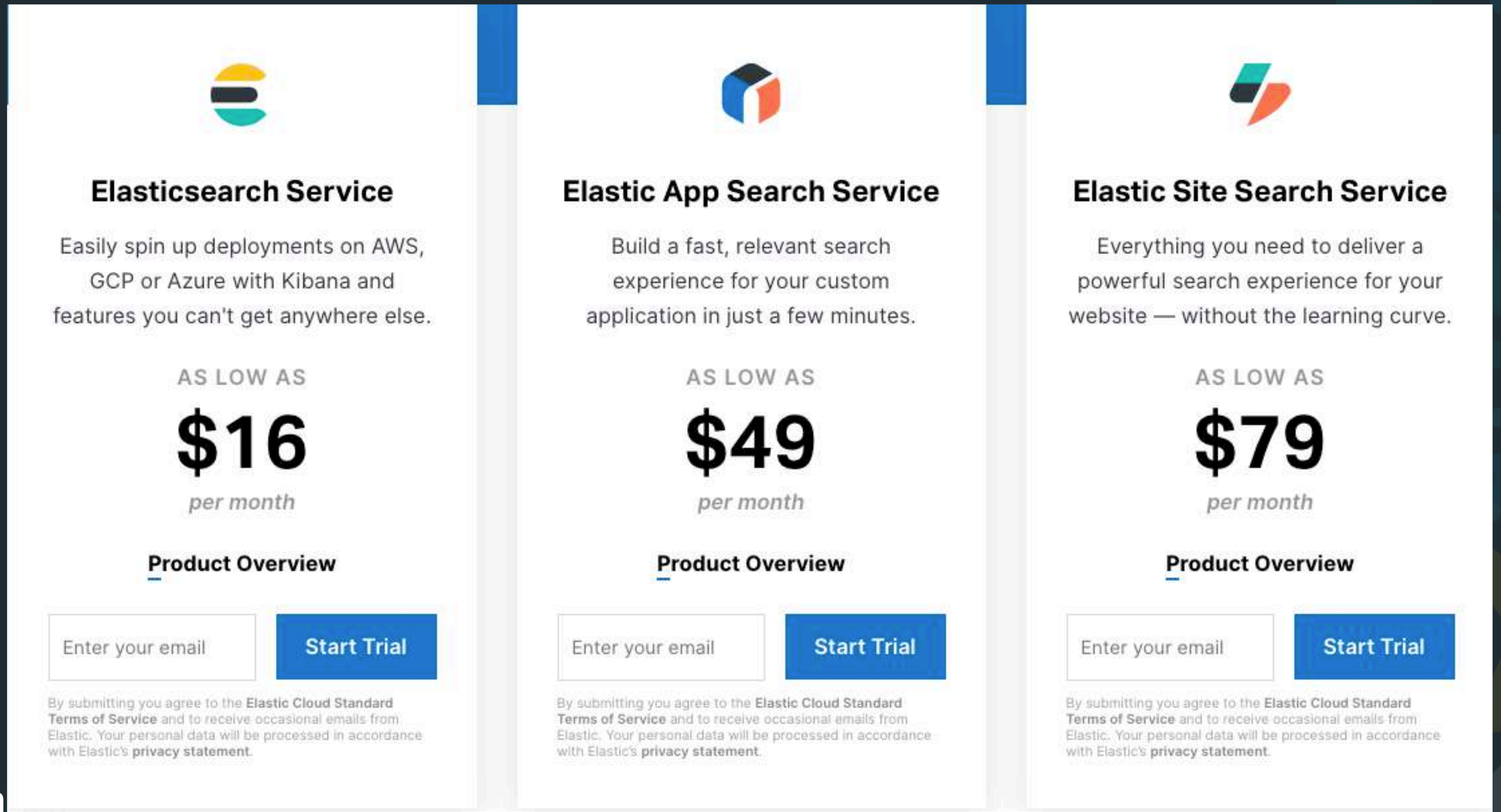
elastic

Deployment options

Distributions

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

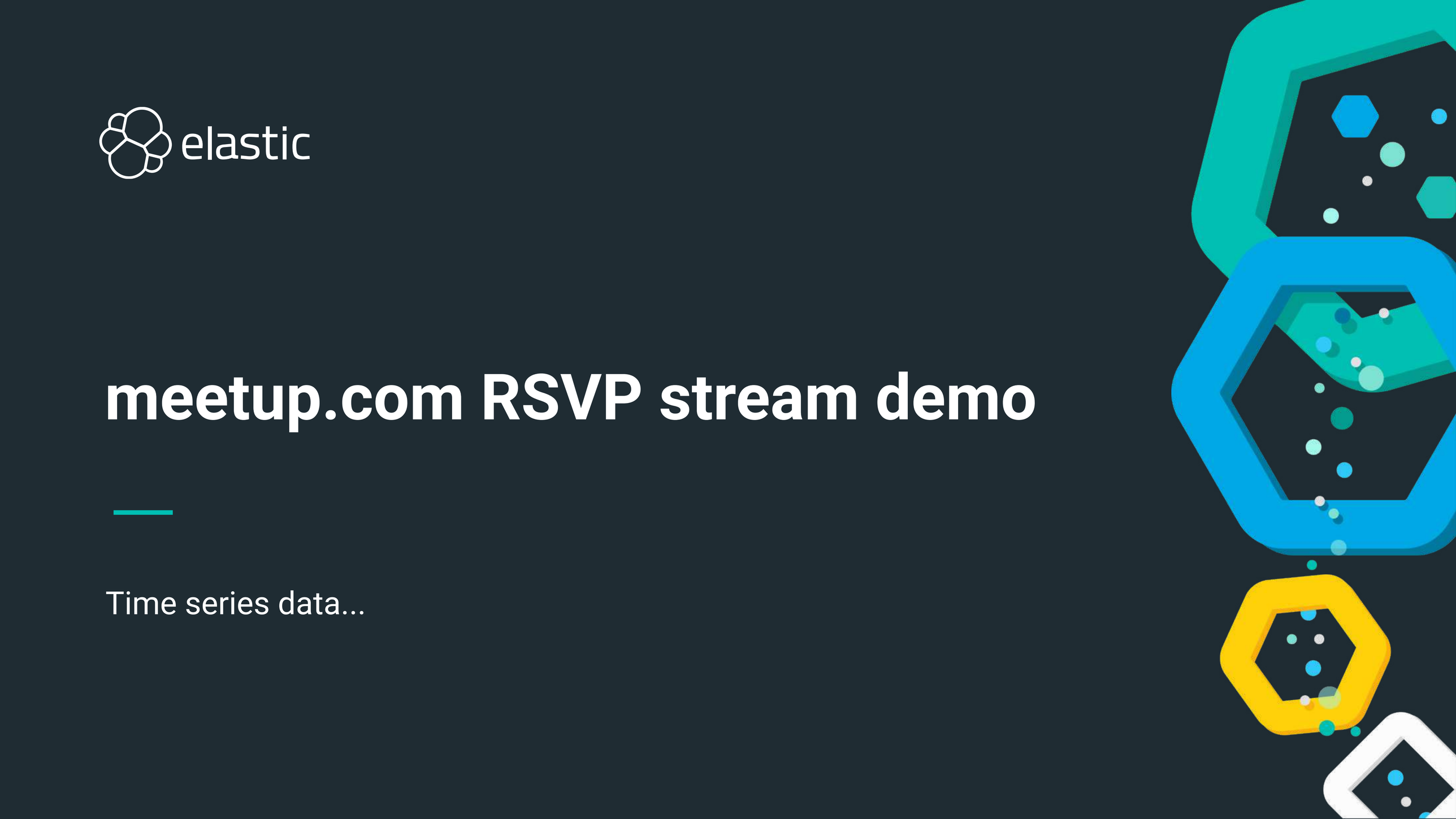
Elastic Cloud



elastic

Elastic Cloud Enterprise

elastic



meetup.com RSVP stream demo

Time series data...

elastic

logging workshop demo

start your engines...

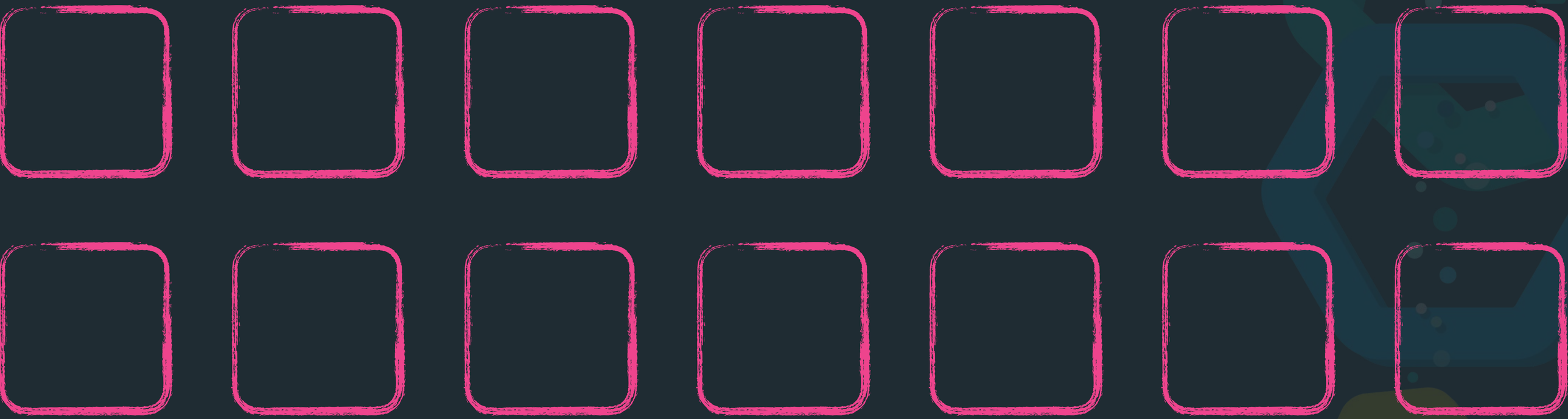
elastic

Logging patterns

Time based data

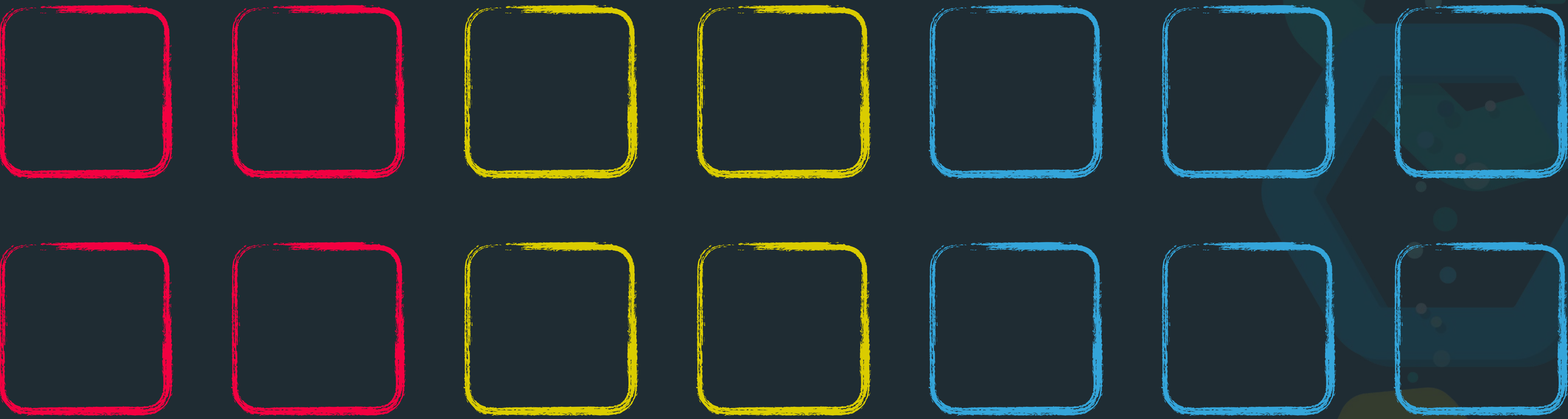
- time based data has properties
- current data gets indexed
- more recent data gets searched more
- old data is still required 'just in case'

Homogeneous architecture



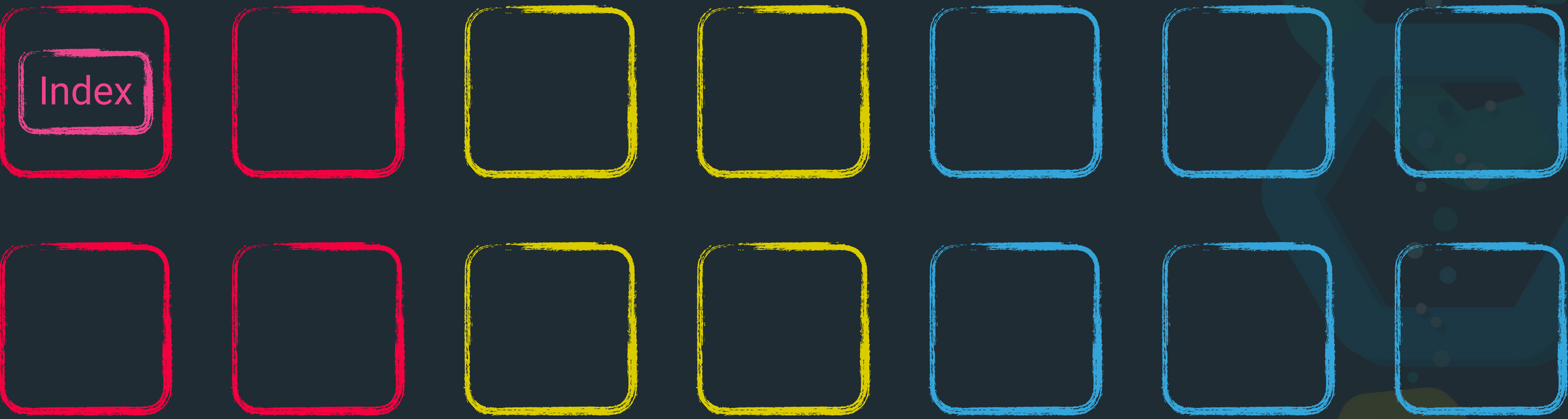
elastic

Hot warm architecture



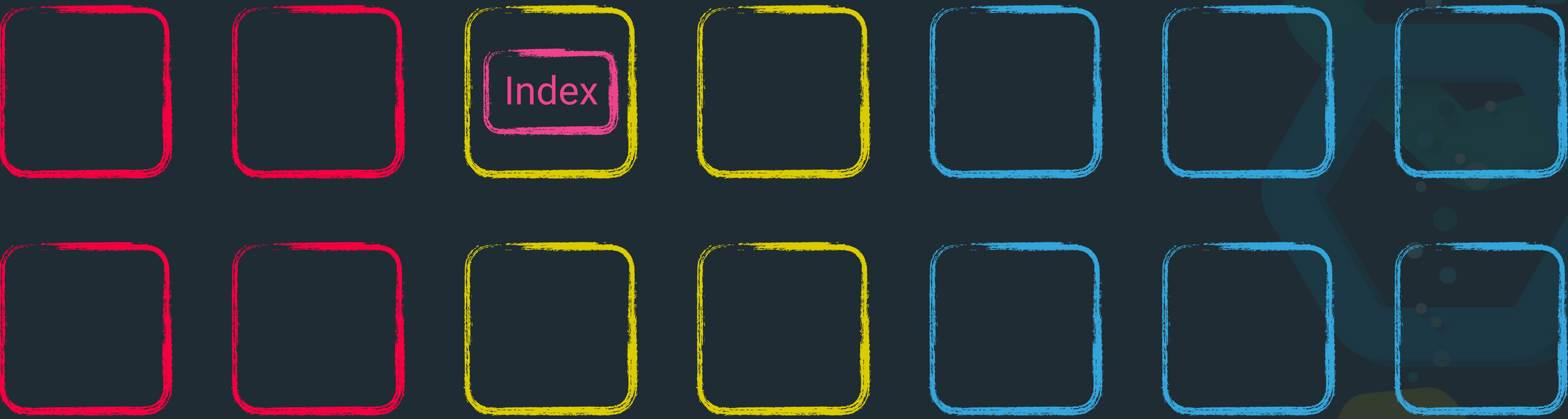
elastic

Hot warm architecture



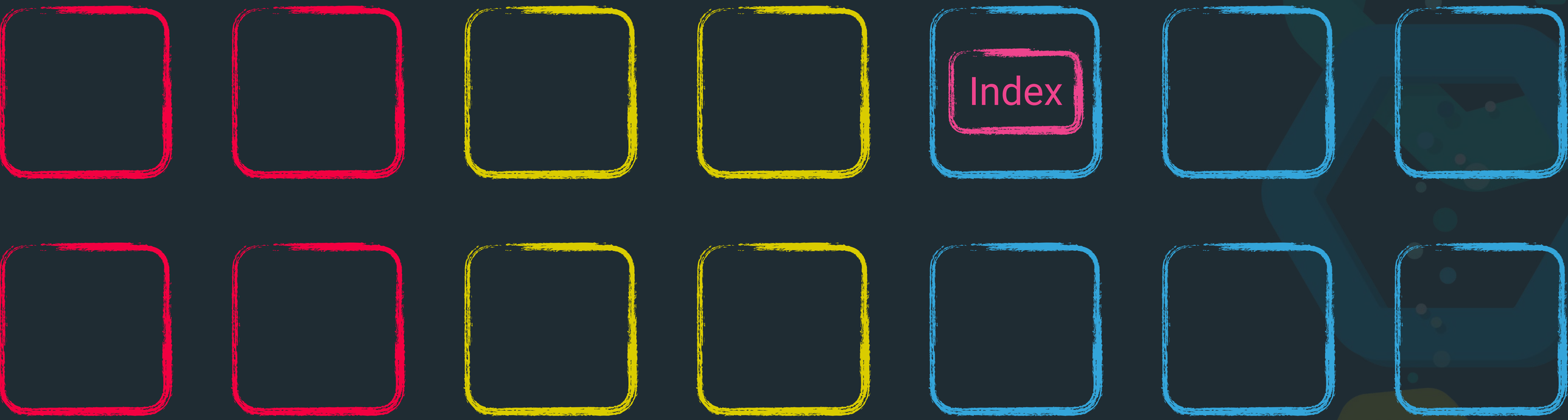
elastic

Hot warm architecture



elastic

Hot warm architecture



elastic

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

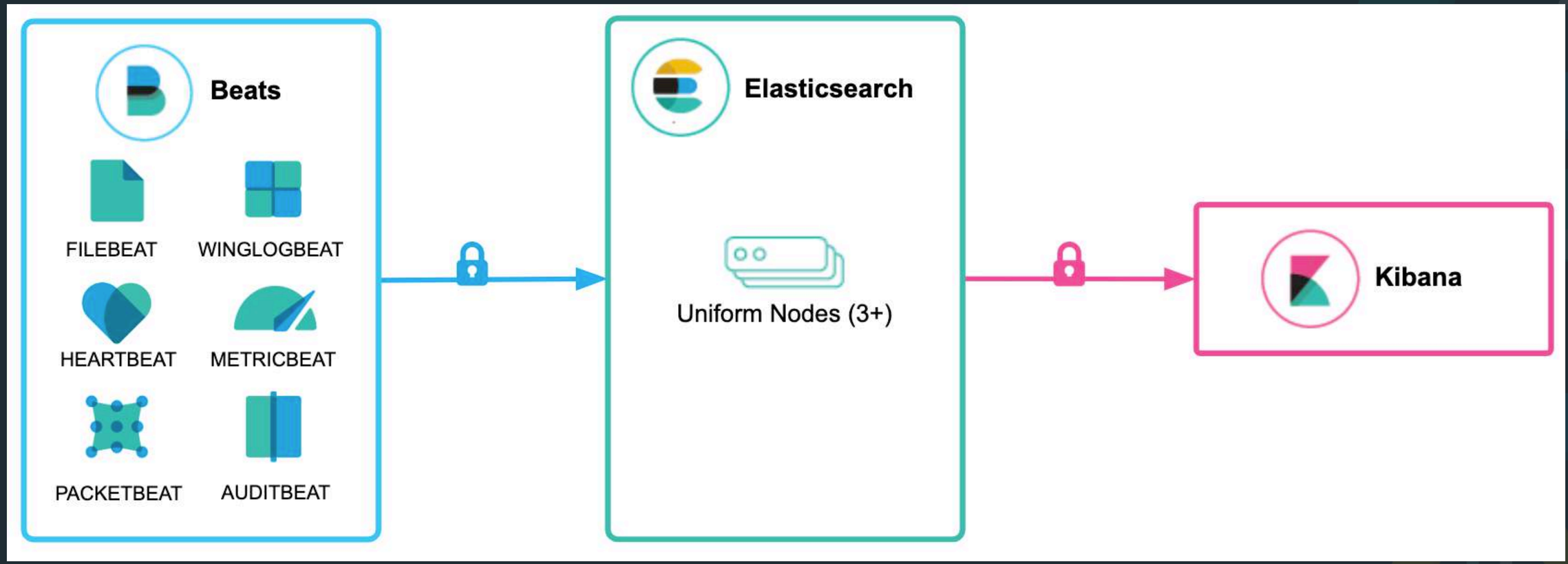
More lifecycle topics

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

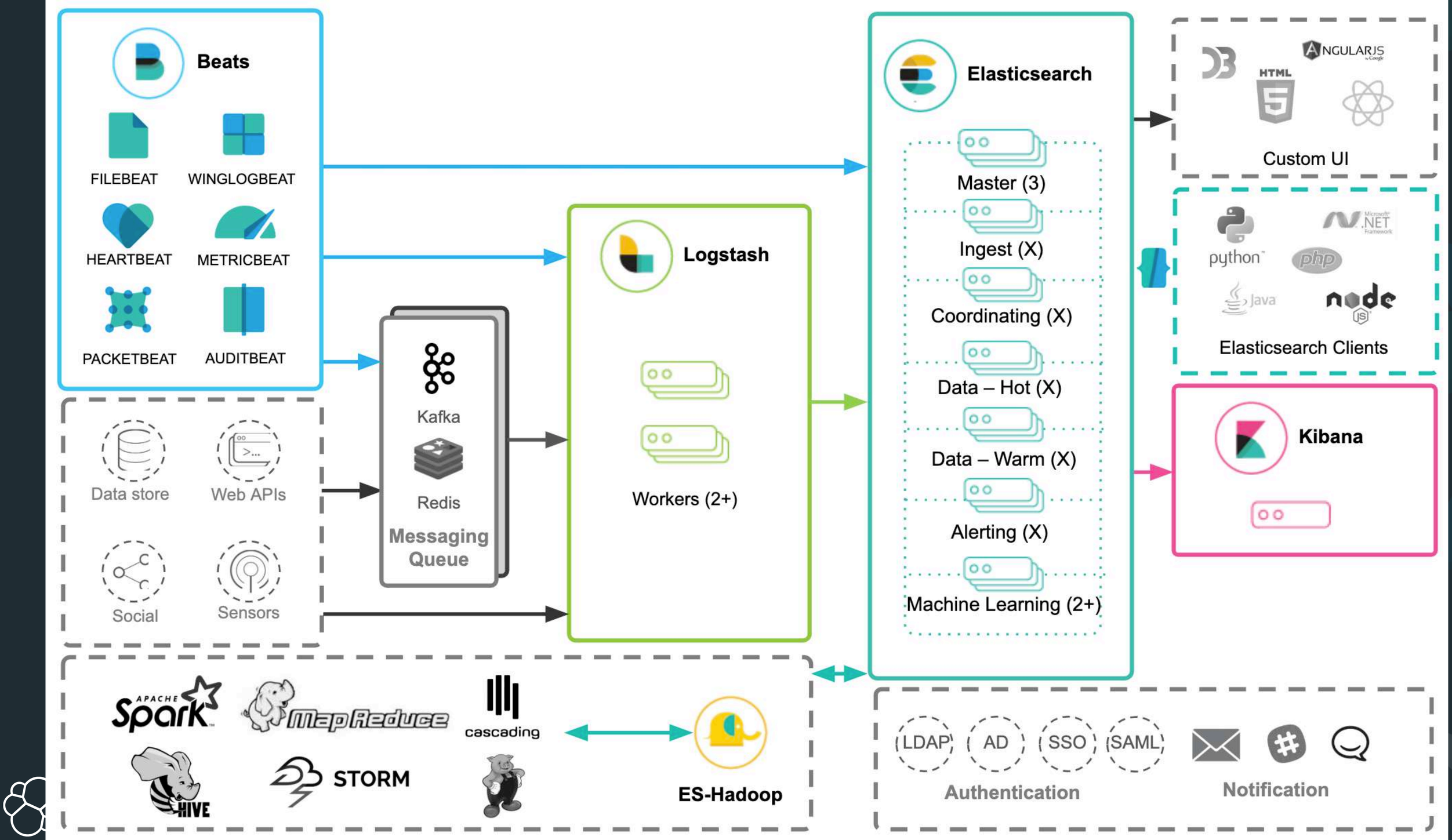
elastic

Architecture patterns

Start small



elastic


Beats
FILEBEAT
WINGLOGBEAT
HEARTBEAT
METRICBEAT
PACKETBEAT
AUDITBEAT
Data store
Web APIs
Social
Sensors
Kafka
Redis
Messaging Queue
Logstash
Workers (2+)
Elasticsearch
Master (3)
Ingest (X)
Coordinating (X)
Data – Hot (X)
Data – Warm (X)
Alerting (X)
Machine Learning (2+)
Custom UI
Elasticsearch Clients
Kibana
APACHE Spark
MapReduce
cascading
STORM
HIVE
ES-Hadoop
LDAP
AD
SSO
SAML
Authentication
Notification

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

elastic

Deploy Elasticsearch and Kibana in 3 Minutes or Less

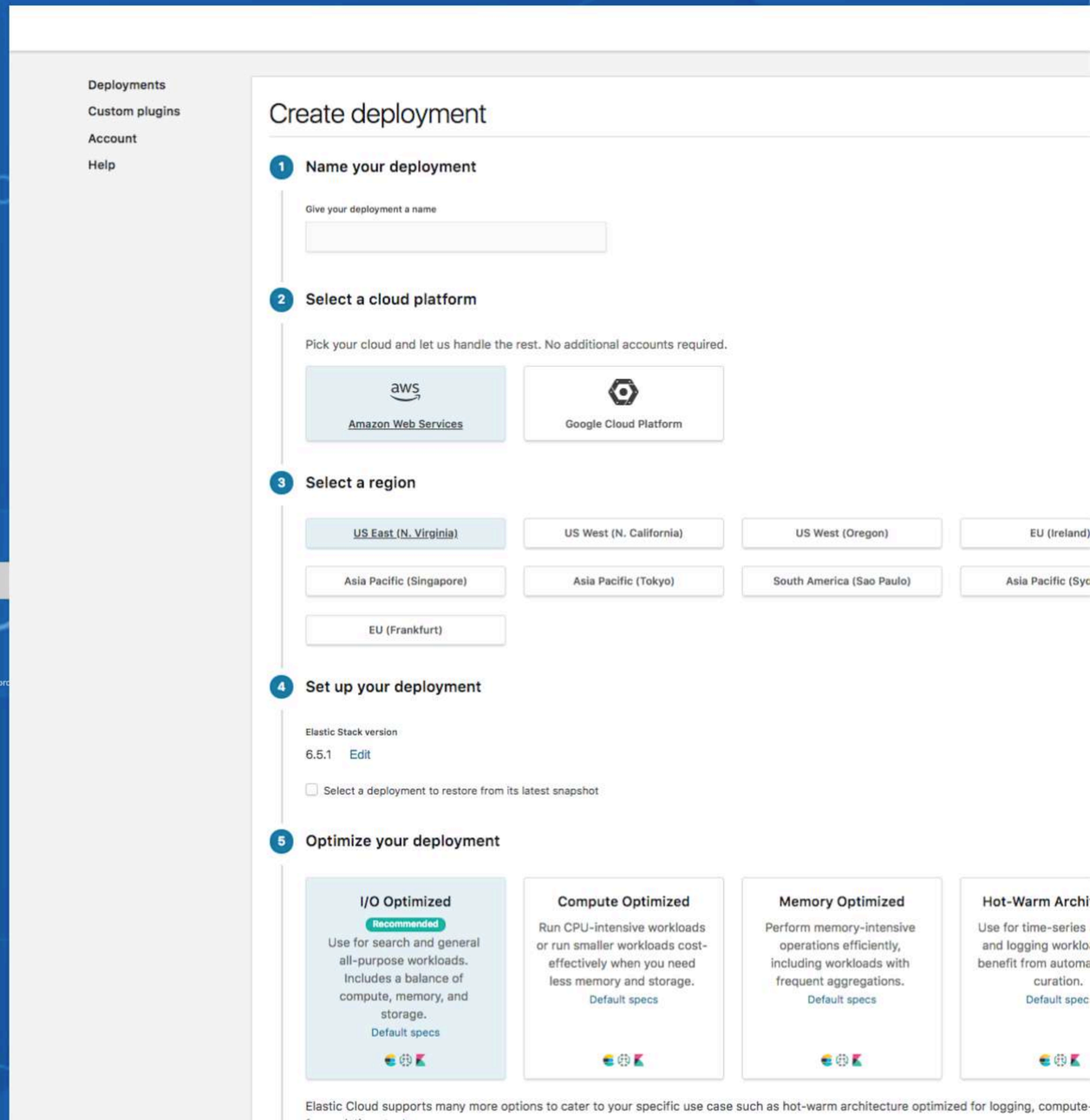
We hope you learned something new at our "Centralize you logs with Elastic Stack" Workshop — put your knowledge to the test and try it out on Elasticsearch Service.

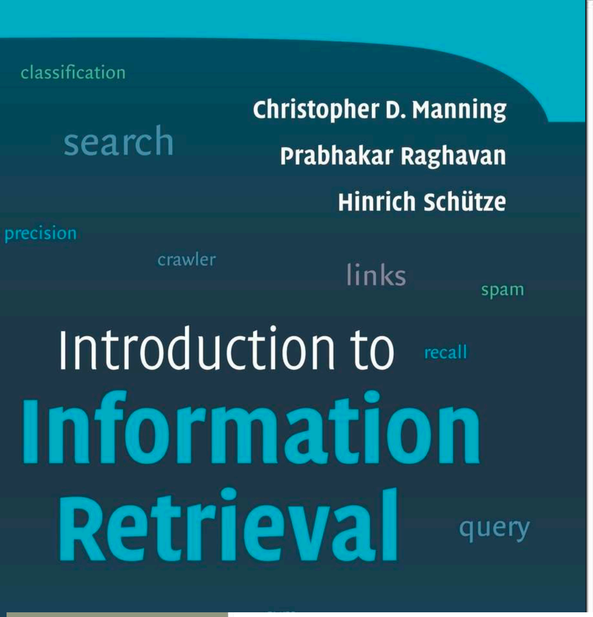
- ✓ Get a 30-day free trial (vs. standard 14-day)
- ✓ No credit card required
- ✓ Get the latest versions, powerful features, and optimized deployment templates for your use case.

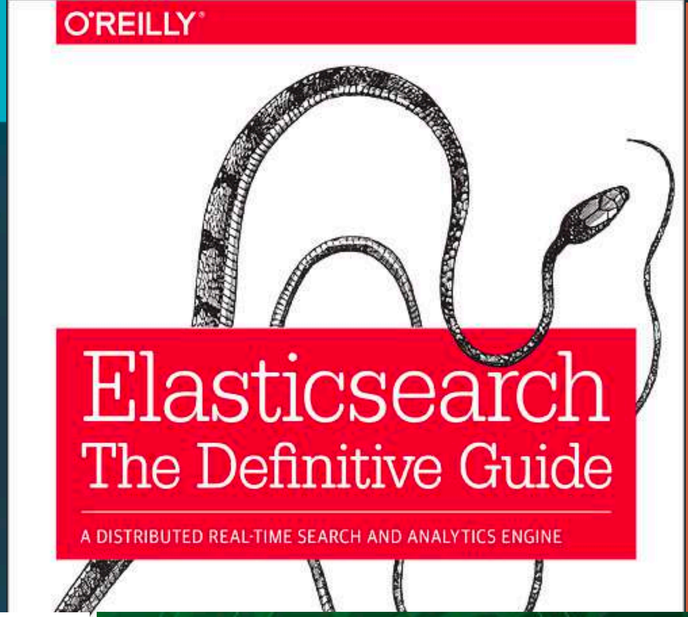
Enter your email

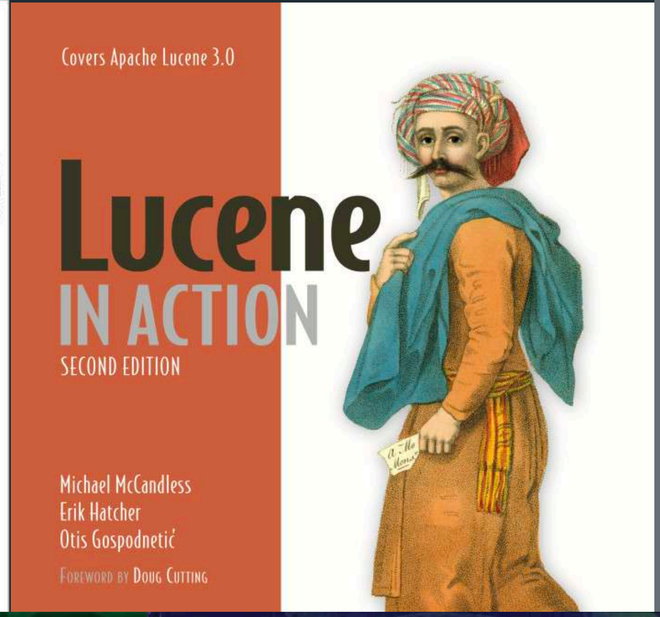
Start Free Trial

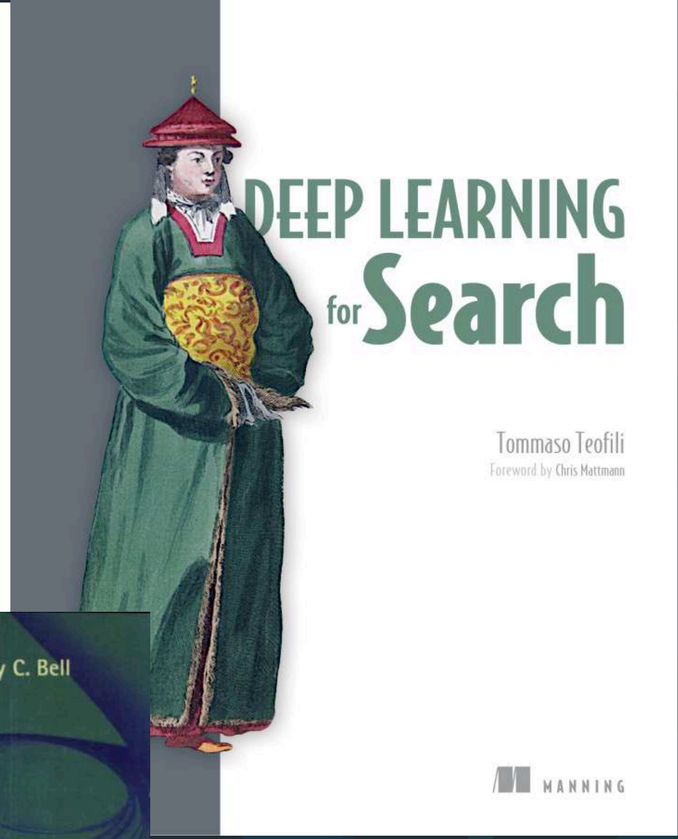
By submitting you agree to the Elastic Cloud Standard Terms of Service and to receive occasional emails from Elastic. Your personal data will be processed in accord

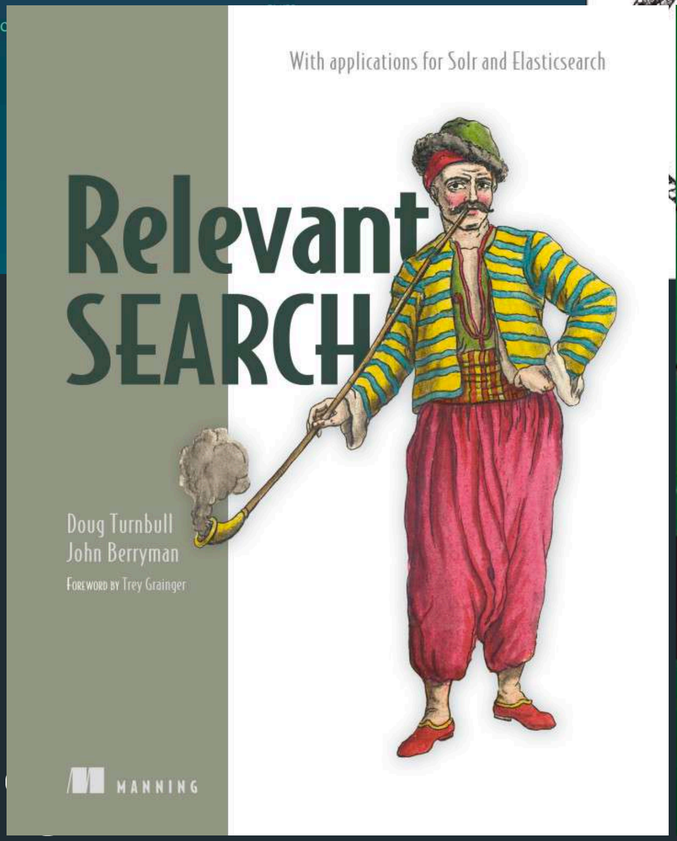


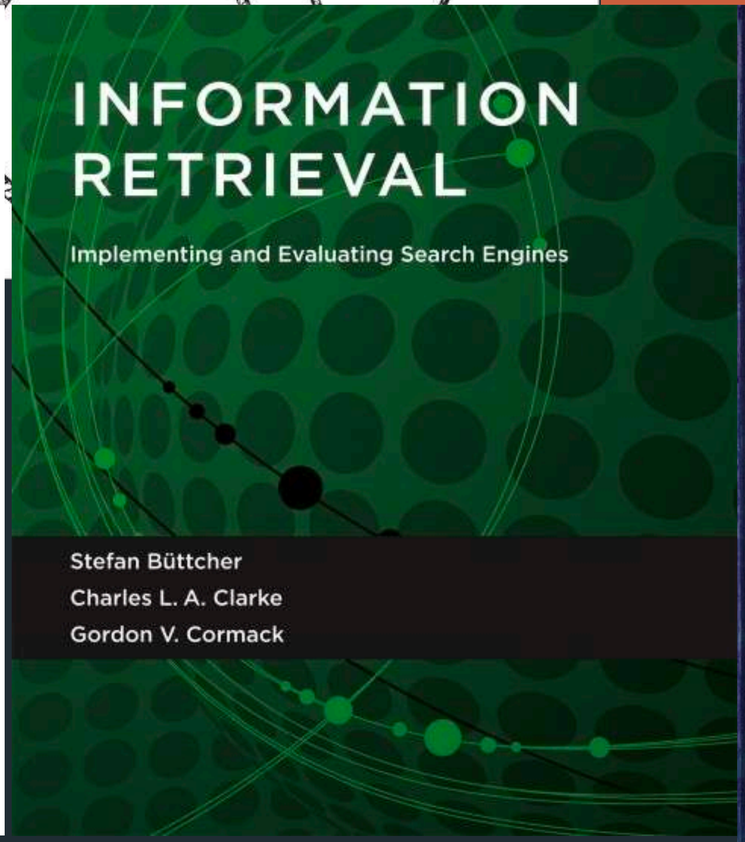

classification
search
Christopher D. Manning
Prabhakar Raghavan
Hinrich Schütze
precision
crawler
links
spam
Introduction to
recall
Information
Retrieval
query

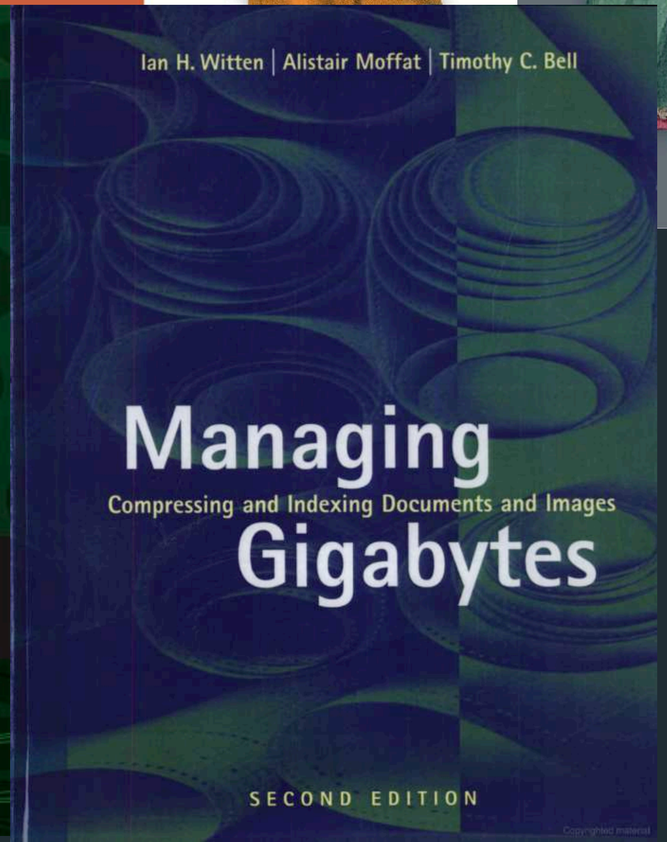

O'REILLY
Elasticsearch
The Definitive Guide
A DISTRIBUTED REAL-TIME SEARCH AND ANALYTICS ENGINE


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
Compressing and Indexing Documents and Images
Gigabytes
SECOND EDITION

elastic

Q & A