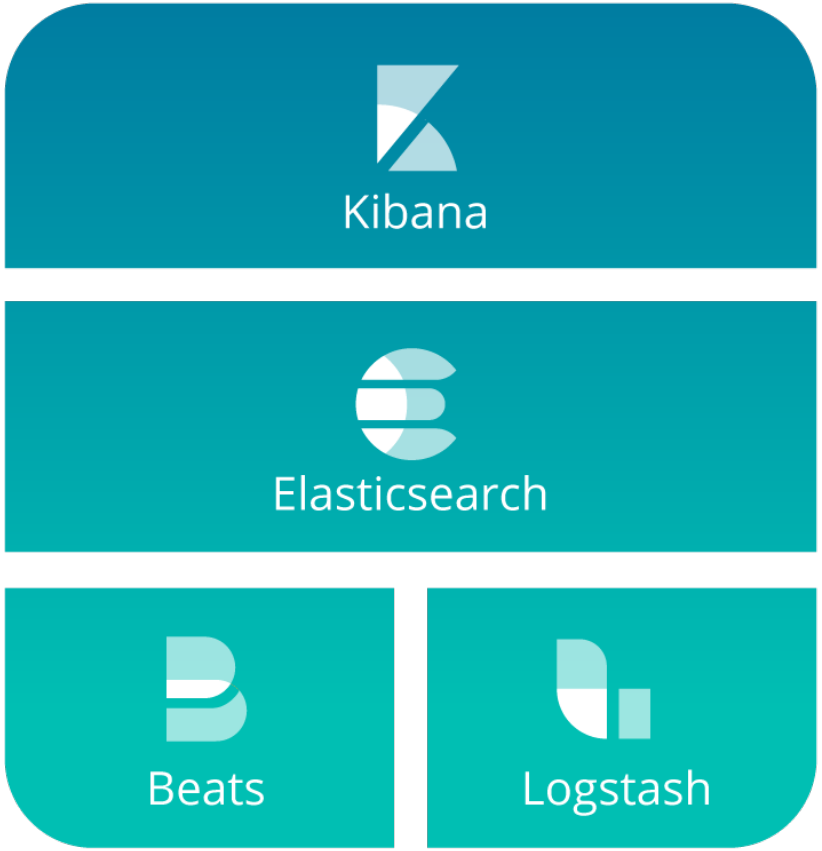
Workshop on Elastic Stack

Confidential information, for internal use only

Elastic Stack

100% open source
No enterprise edition
All new versions with 6.0



ASHNIK elastic

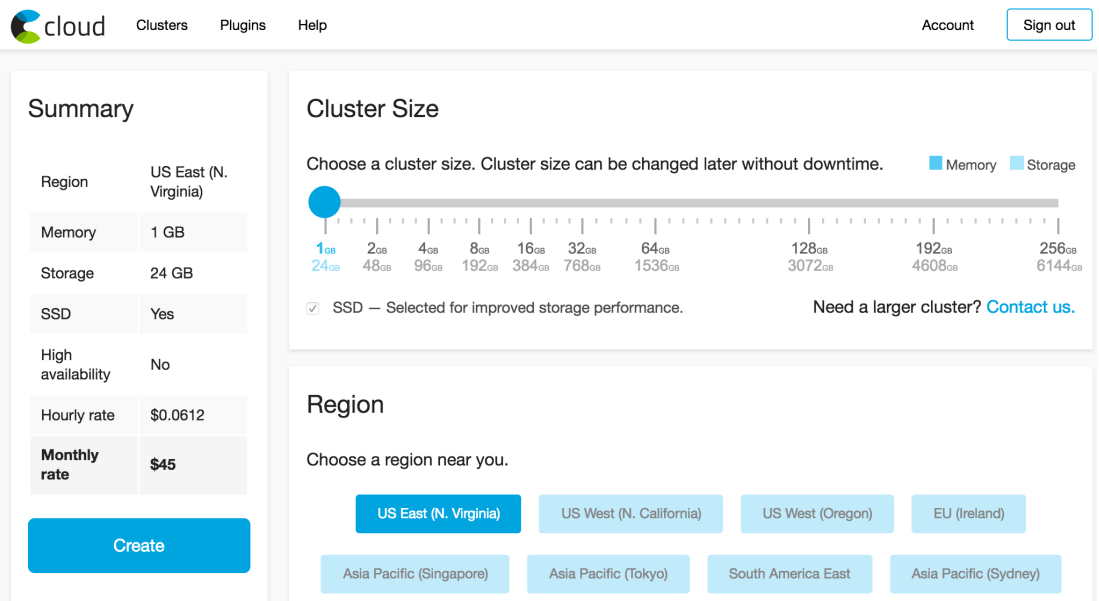
X-Pack

Single install
Extensions for the Elastic Stack
Subscription pricing

- **Security**
- **Alerting**
- **Monitoring**
- **Reporting**
- **Graph**
- **Machine Learning**

Elastic Cloud

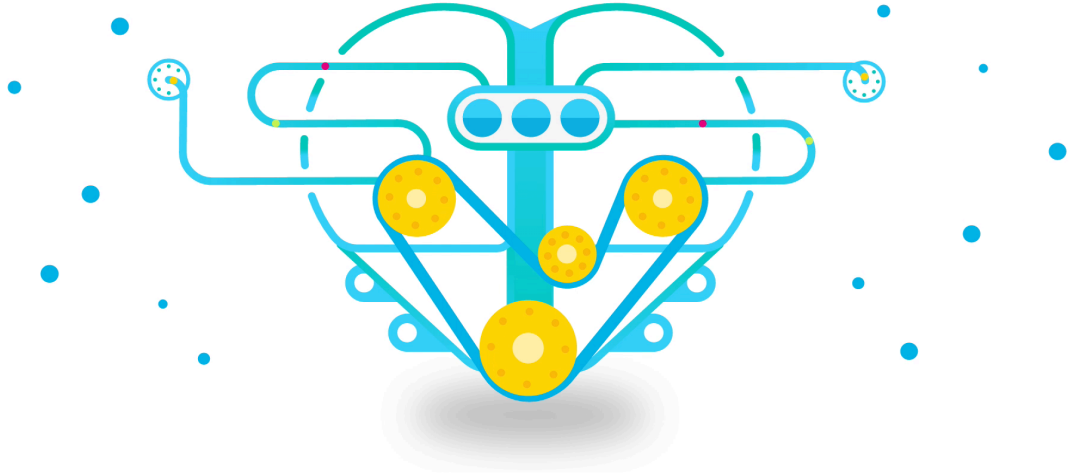
Hosted Elasticsearch & Kibana
Includes X-Pack features
Starts at $45/mo



Available in AWS today
Available in Google Cloud Platform (soon)
Available as a private cloud/on-premise solution
(Elastic Cloud Enterprise)

Elasticsearch

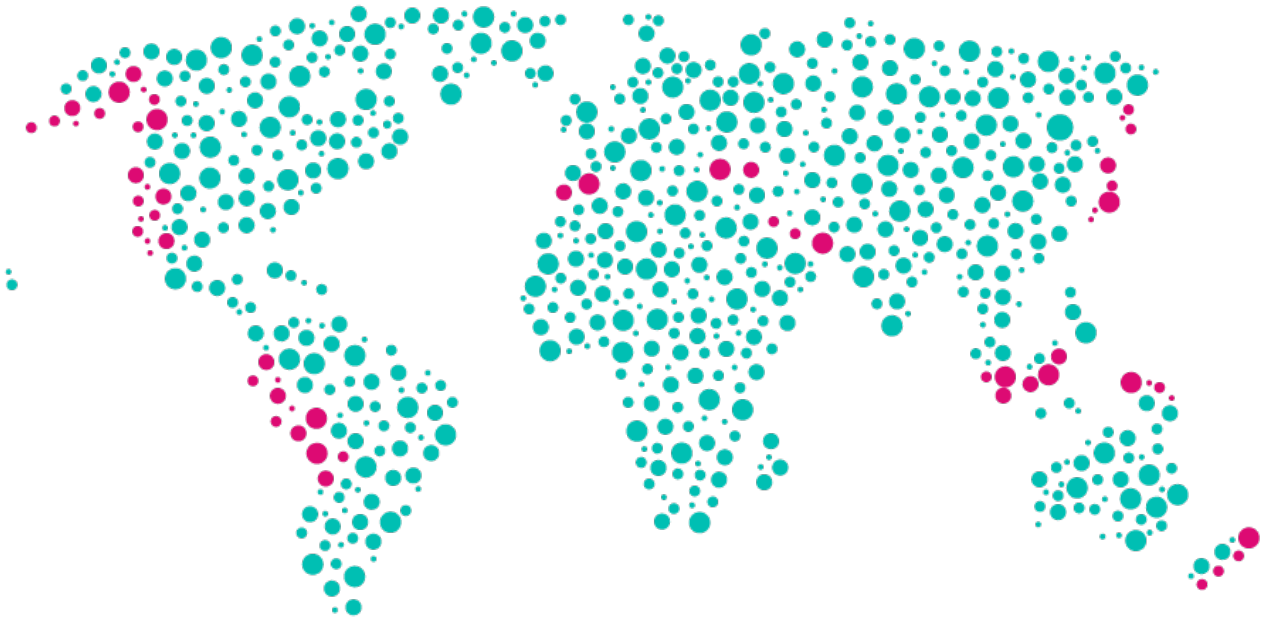
Heart of the Elastic Stack



Distributed, Scalable	**High-availability**	**Multi-tenancy**
Developer Friendly	**Real-time, Full-text Search**	**Aggregations**

Kibana

Window into the Elastic Stack



Visualize and analyze	**Geospatial**	**Customize and Share Reports**
Graph Exploration	**UX to secure and manage the Elastic Stack**	**Build Custom Apps**

kibana

Web Analytics | New | Add | Save | Open | Share | Options | Reporting | August 20th 2016, 09:00:00.000 to August 29th 2016, 08:59:03

*

- Discover
- Visualize
- Dashboard
- Graph
- Monitoring
- Timelion
- Management
- Dev Tools

Total Visitors

541,194

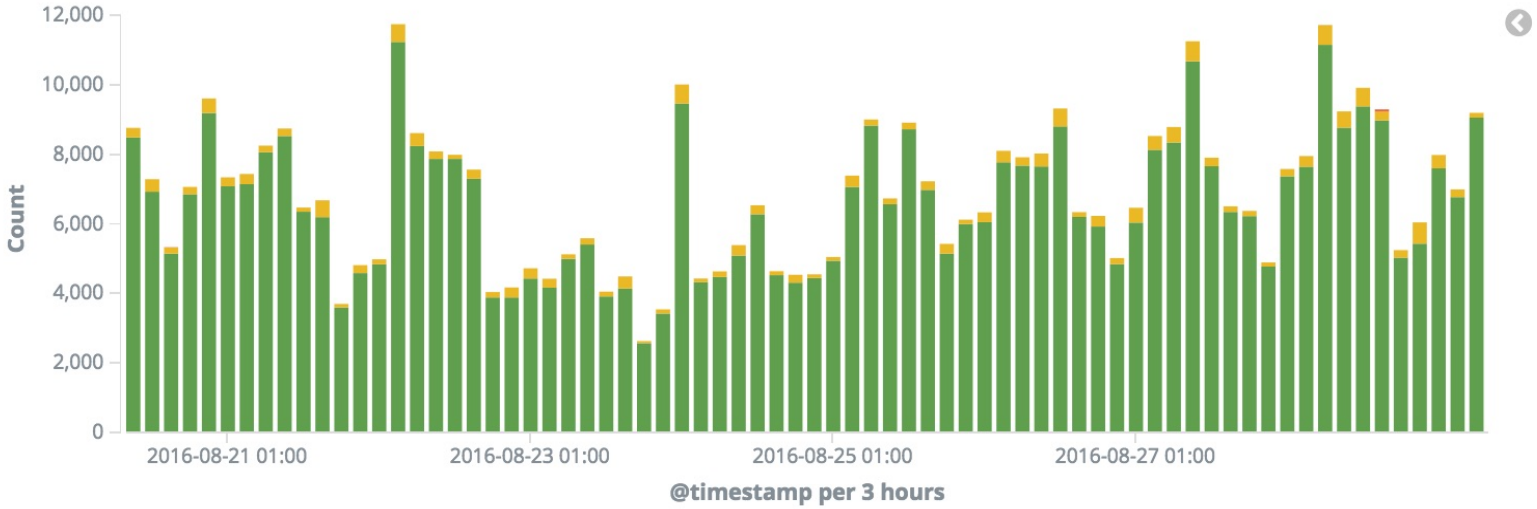
Total IPs

Unique Visitors

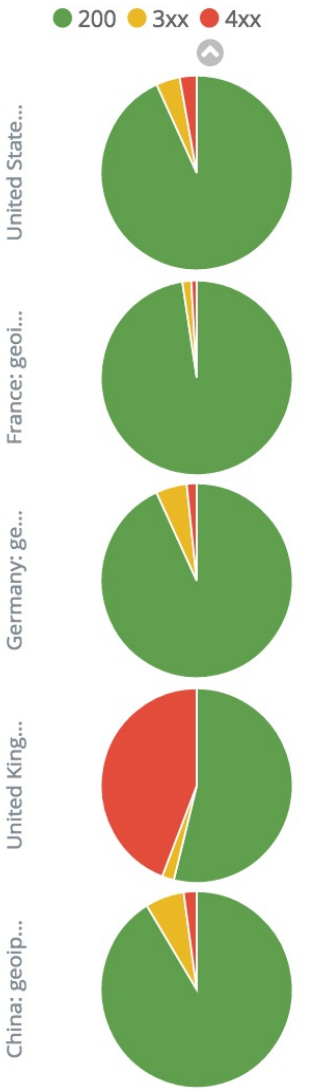
10,279

Unique IPs

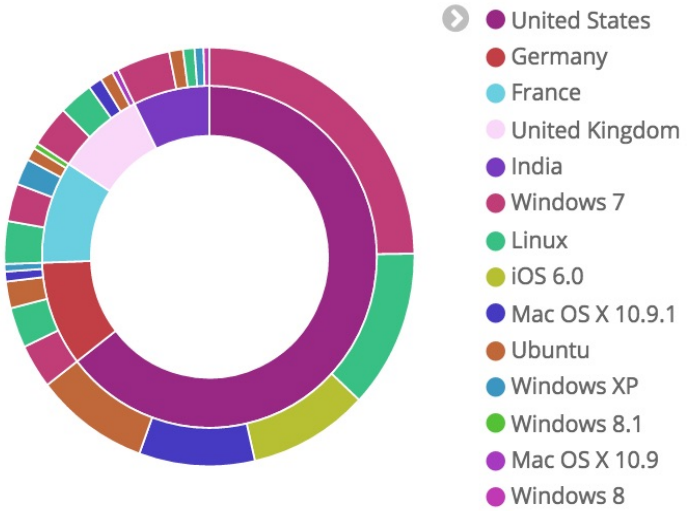
Response Code by Time



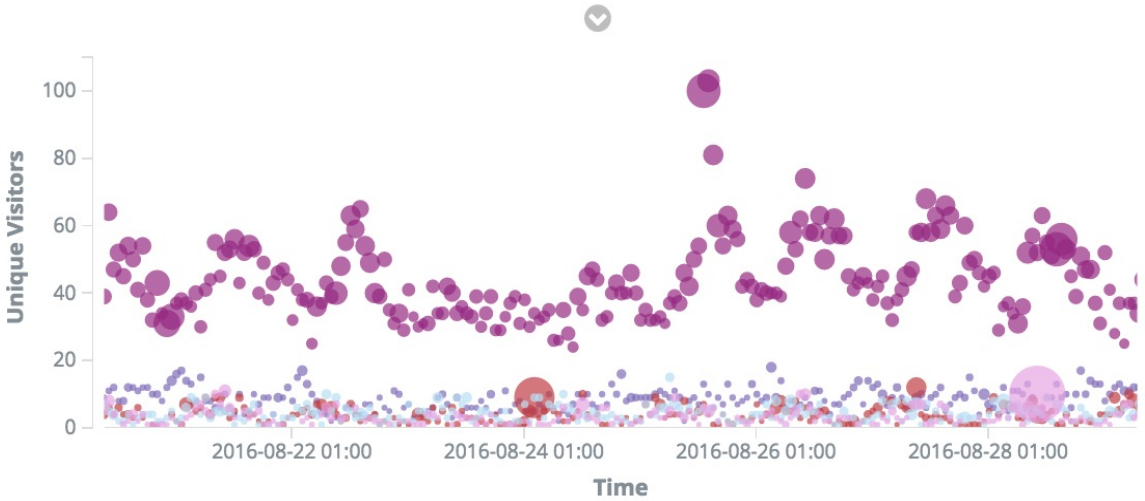
Response Code by Countr



Traffic by Country & OS



Bytes vs. Time



Unique Visits by City

City	Unique Visitors	Total Visitors
Beijing	346	8,232

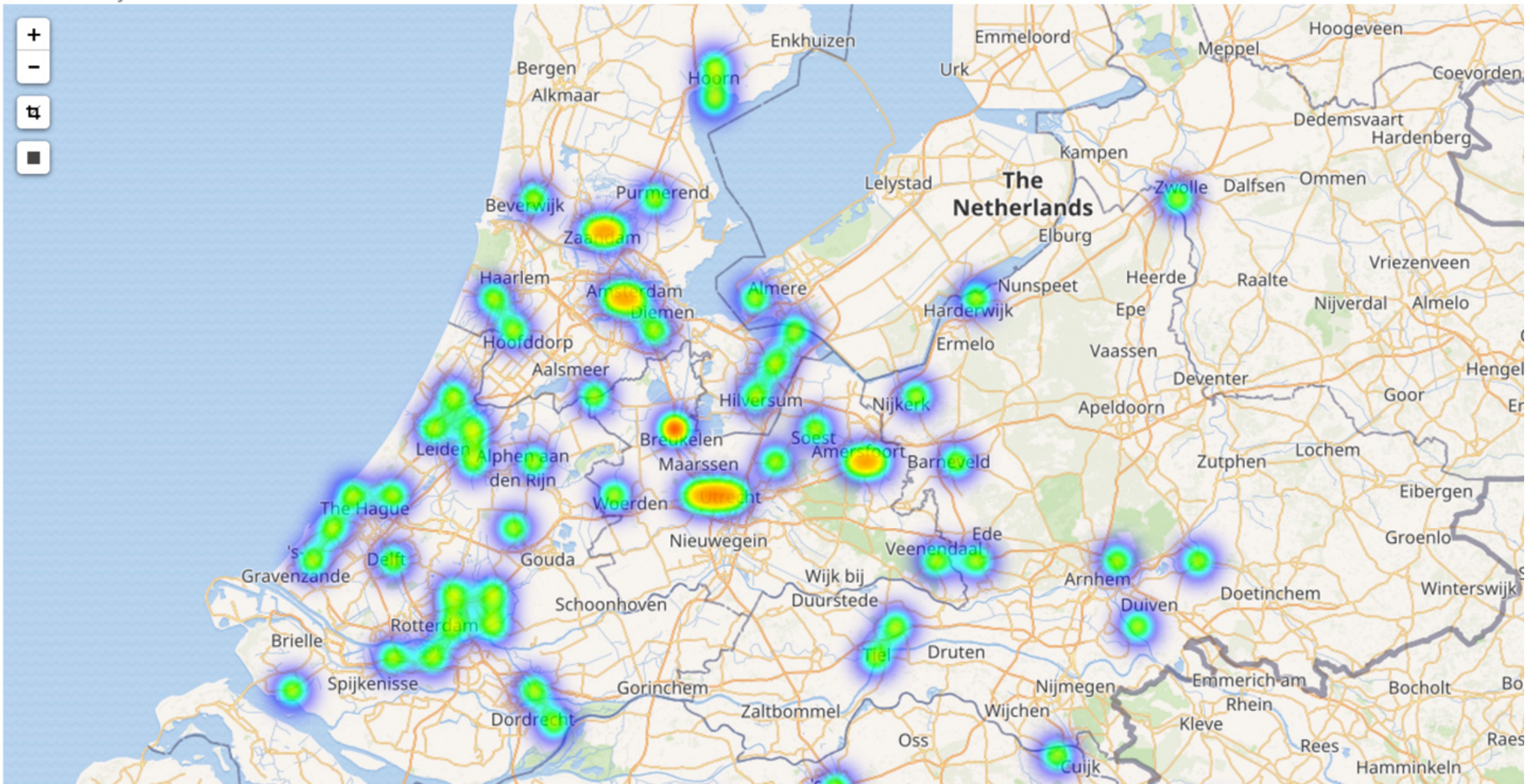
Traffic vs. Location

- elastic
- Logout

New Add Save Open Share Options Reporting Last 20y

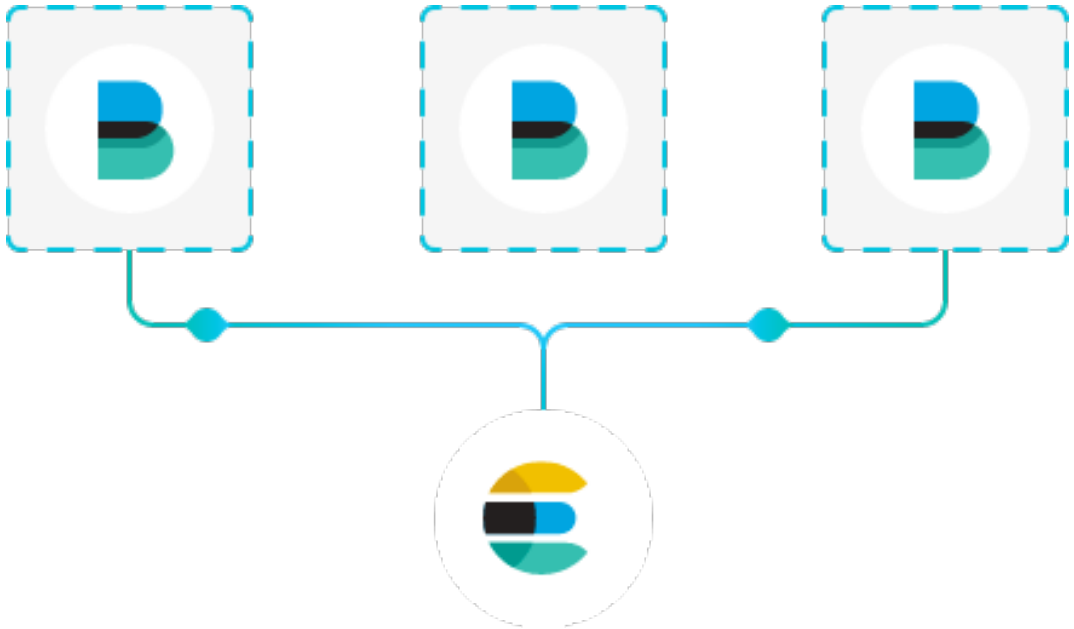
*

Web Traffic by Location



Beats

Window into the Elastic Stack



Ship data from the source	Ship and centralize in Elasticsearch	Ship to Logstash for transformation and parsing
Ship to Elastic Cloud	Libbeat: API framework to build custom beats	30+ community Beats

ASHNIK elastic

FILEBEAT
Log Files

METRICBEAT
Metrics

PACKETBEAT
Network Data

WINGLOGBEAT
Window Events

More than 30 community Beats and growing …

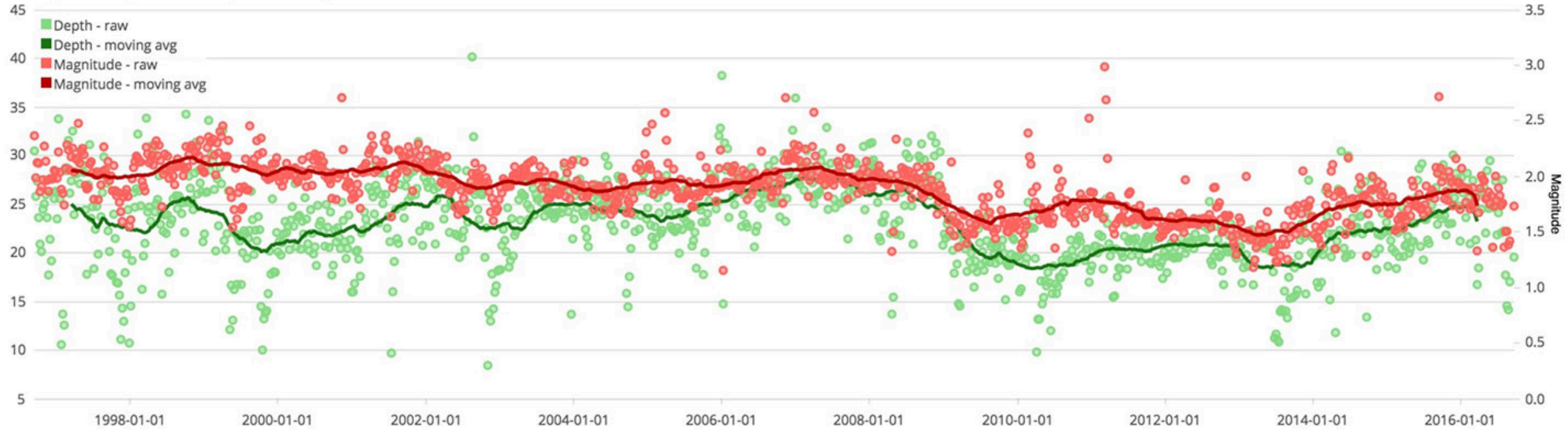
Apachebeat, dockbeat, httpbeat, mysqlbeat, nginxbeat, redis beats, twitterbeat, and more

Metrics Dashboard

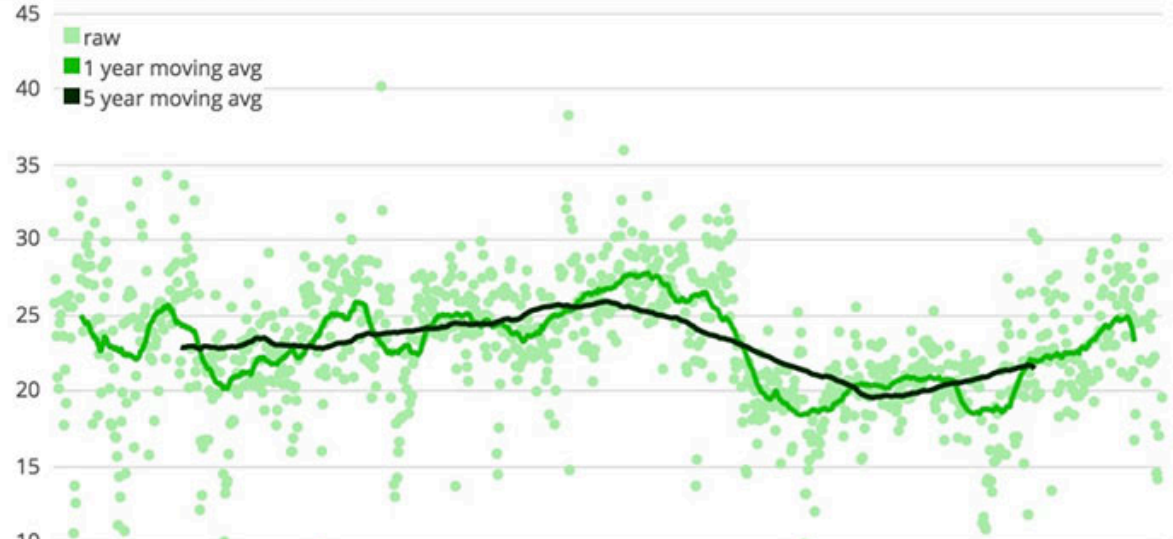
New Add Save Open Share Options Reporting Last 20y

*

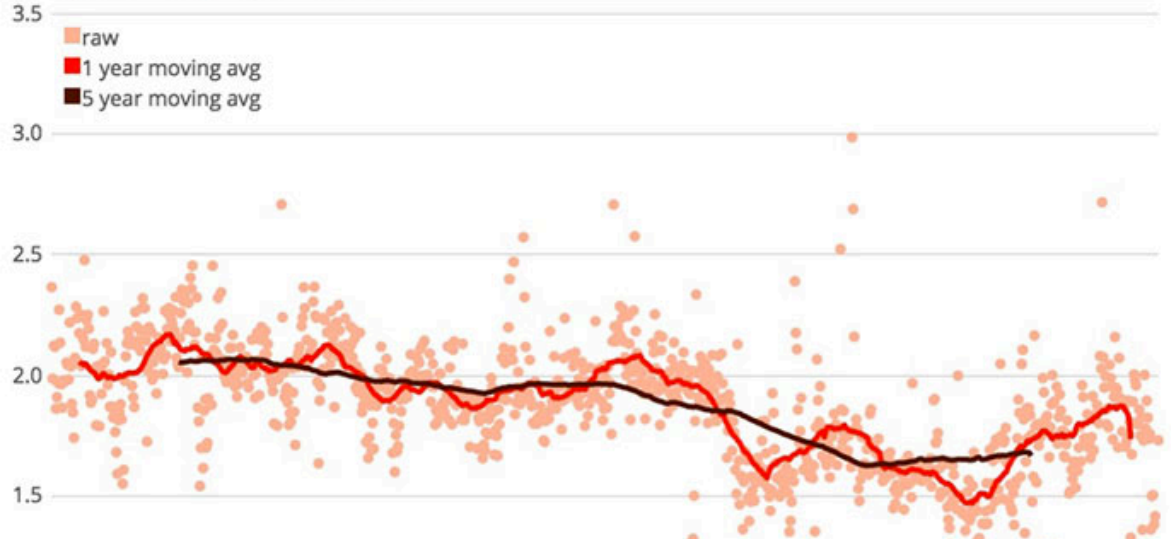
Earthquake - Magnitude vs Depth Mov. Avg



Earthquake - Depth Timeseries

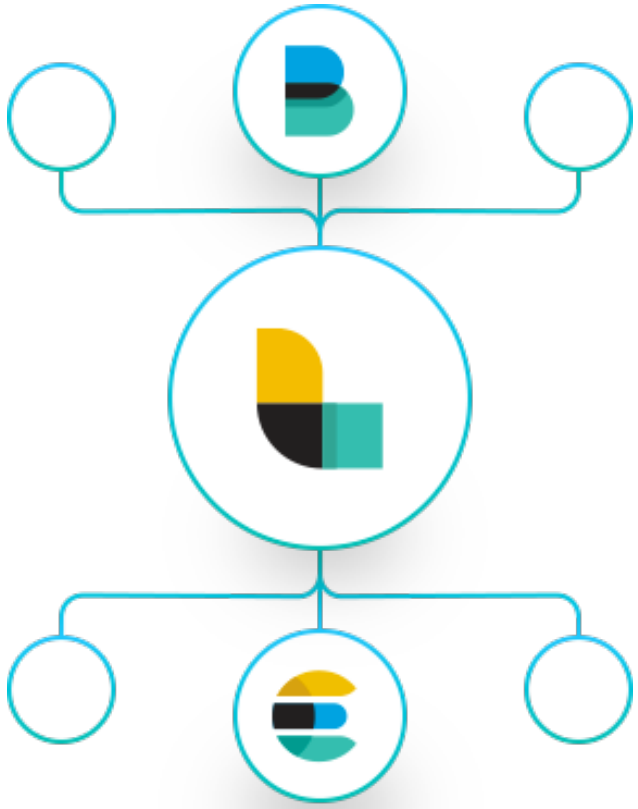


Earthquake - Magnitude Timeseries



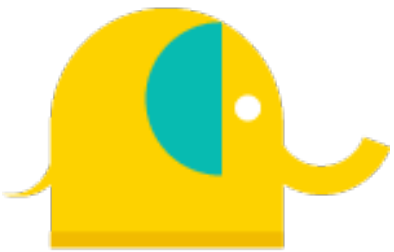
Logstash

Data processing pipeline



Ingest data of all shapes, sizes, and sources	**Parse and dynamically transform data**	**Transport data to any output**
Secure and encrypt data inputs	**Build your own pipeline**	**More than 200+ plugins**

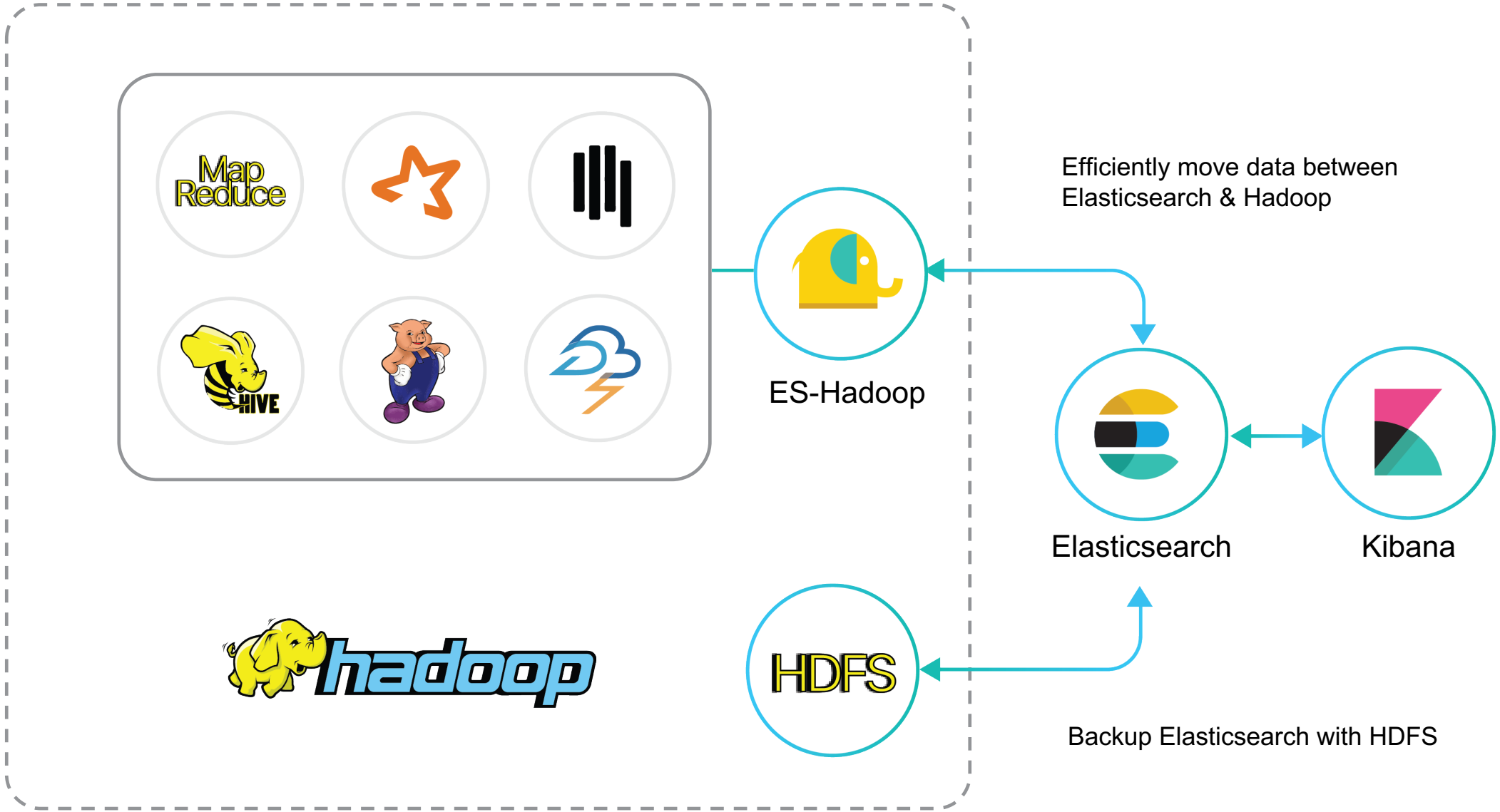
ASHNIK elastic



ES-Hadoop

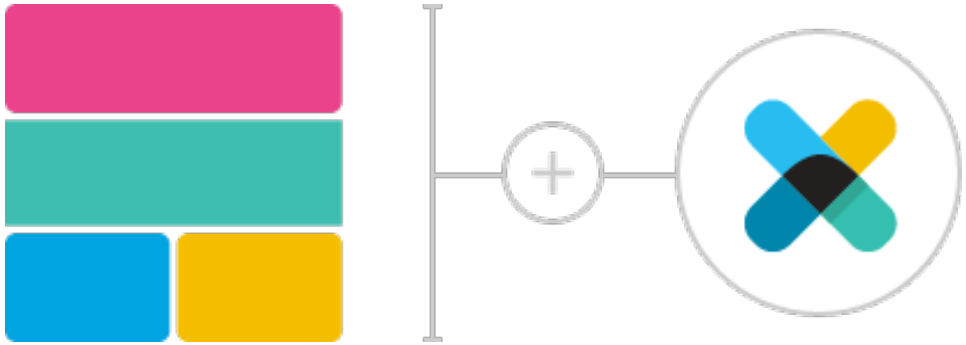
Elasticsearch for Hadoop

Two-way connector	**Index Hadoop data in Elasticsearch**	**Enable real-time search capabilities**
Visualize HDFS data in Kibana	**Snapshot and restore with HDFS**	**Support for Spark, Storm MapReduce, and more**


Map Reduce
HIVE
hadoop
ES-Hadoop
HDFS
Efficiently move data between Elasticsearch & Hadoop
Elasticsearch
Kibana
Backup Elasticsearch with HDFS

X-Pack

Extensions for the Elastic Stack



Security	**Alerting**	**Monitoring**
Reporting	**Graph Analytics**	**Single Install, included in Elastic Subscription**

X-Pack

Security

AUTHENTICATION

- Username and password
- Integrate with authentication systems
- Create a custom realm to authenticate users

AUTHORITIZATION

- Manage users and roles
- Assign permissions and privileges

ADDITIONAL CONTROLS

- SSL/TLS encryption
- IP filtering
- Field and document level security
- Audit logging

kibana

Management / Elasticsearch / Roles / Edit

Users Roles

Discover
Visualize
Dashboard
Graph
Monitoring
Timelion
Management
Dev Tools

Return to All Roles | Delete | Save

Edit Role

Name marketing-ops

Cluster Privileges

☐ **all** ☐ **monitor** ☐ **manage** ☐ **manage_security** ☐ **manage_index_templates**

Run As Privileges

Add a user...

Indices Privileges

twitter* × facebook* × downloads* ×

Add a query...

read ×

- all
- manage
- monitor
- index
- create
- delete
- write
- delete_index

elastic
Logout
Collapse

X-Pack

Alerting

SETUP ALERTS

- Create Watches to detect changes in your data
- Trigger automatic notifications
- Setup nested alerts
- Store and track alert history

NOTIFY AND INTEGRATE

- Email
- Slack
- Pagerduty
- Hipchat or JIRA
- Other monitoring systems

kibana

Discover
Visualize
Dashboard
Monitoring
Timelion
Management
Dev Tools

elastic
Logout
Collapse

Alert History New Add Save Open Share Options Reporting September 22nd 2016, 12:41:37.264 to September 22nd 2016, 13:06:12.313

*

Alert Conditions Evaluated

173

# of alerts evaluated

Alerts Triggered

22

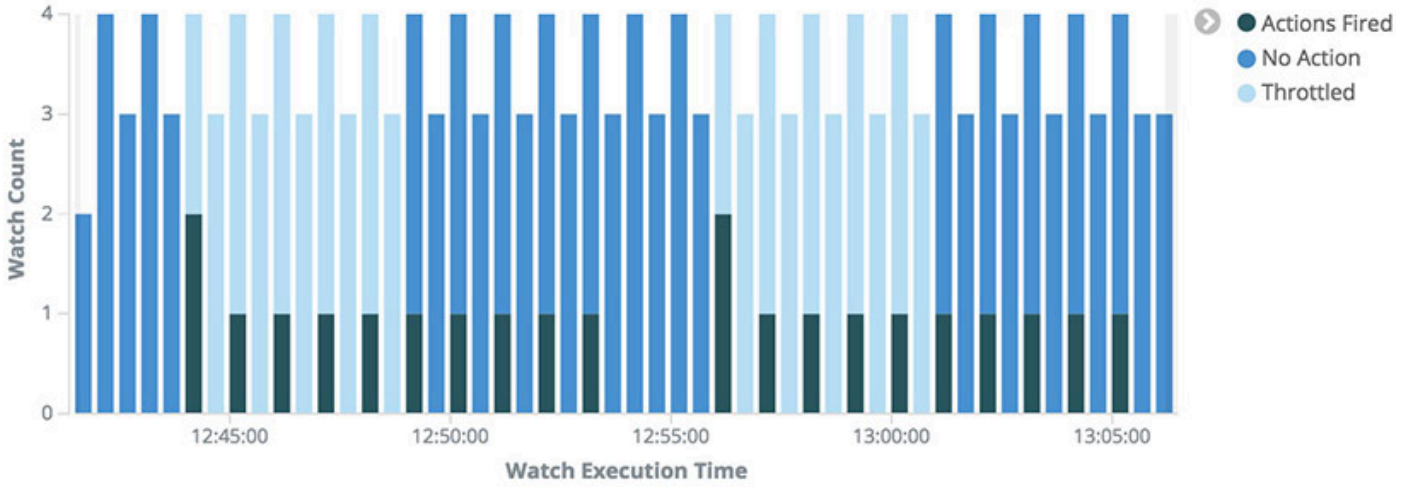
# of alerts fired

Alerts Throttled

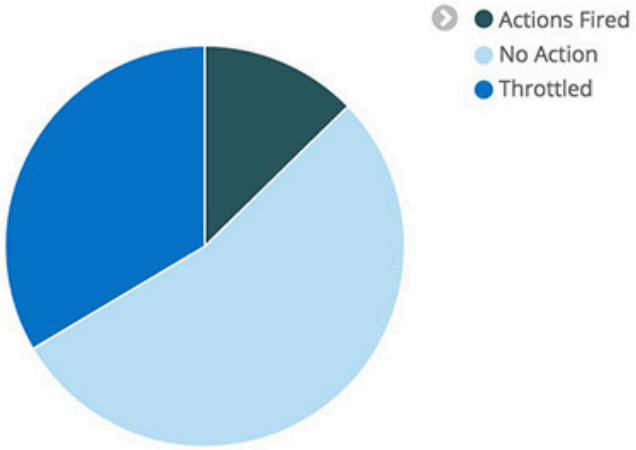
58

# of alerts fired

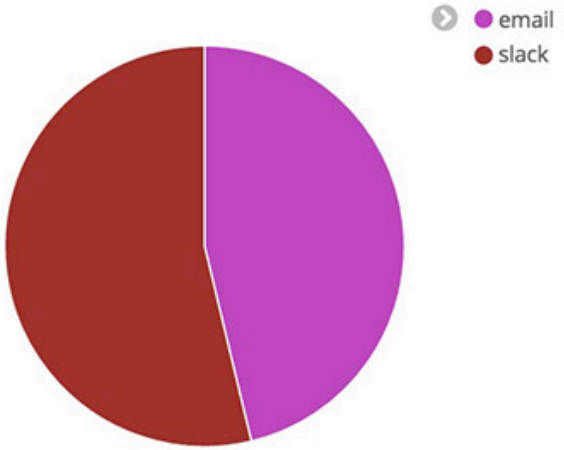
Alert Status over Time



Actions



Execution Status



Most Frequent Watches with Met Condition

watch_id: Descending	Count
dir-traversal	60
port-scan	20

X-Pack

Monitoring

MONITOR CLUSTER HEALTH

- Prebuilt Kibana dashboards to monitor the performance of the Elastic Stack
- Get vital statistics at various levels -- cluster, node, and indices

OPTIMIZE CLUSTER PERFORMANCE

- Multicluster support to compare health and performance of multiple clusters
- Analyze historical or real-time data for root cause analyses
- Utilize analyses to proactively optimize and improve cluster performance
- Configure data retention policy

kibana

Discover
Visualize
Dashboard
Graph
Monitoring
Timelion
Management
Dev Tools

elastic
Logout
Collapse

Clusters

10 seconds | Last 1 hour

Filter Clusters

3 of 3

Name	Status	Nodes	Indices	Data	Kibana	License
prod01	⚠	24	192	73.1 GB	1	Platinum Expires 1 Feb 17
Monitor *	✓	2	8	218.5 MB	1	Platinum Expires 1 Feb 17
dev-test	⚡	19	233	181.5 GB	1	Trial Expires 10 Nov 16

kibana

Clusters / production / Elasticsearch

5 seconds | Last 15 minutes

Overview | Indices | Nodes

- Discover
- Visualize
- Dashboard
- Monitoring
- Timelion
- Management
- Dev Tools

Nodes:	Indices:	Memory:	Total Shards:	Unassigned Shards:	Documents:	Data:	Uptime:	Version:
3	**3**	**2GB / 6GB**	**22**	**0**	**174,544,212**	**101GB**	**118 hours**	**5.0.0-beta1**

Nodes

Filter Nodes

3 of 3

Name	Status	CPU Usage	JVM Memory	Load Average	Disk Free Space	Shards
★ prod-1 127.0.0.1:9300	✓	0% ↓ 2% max 1% min	40% ↑ 43% max 16% min	2.82 ↓ 5.48 max 2.5 min	309.6 GB ↓ 309.7 GB max 309.5 GB min	7
prod-2 127.0.0.1:9301	✓	2% ↓ 2% max 1% min	16% ↓ 42% max 16% min	2.82 ↓ 5.48 max 2.5 min	309.6 GB ↓ 309.7 GB max 309.5 GB min	8
prod-3 127.0.0.1:9302	✓	1% ↓ 3% max 1% min	32% ↓ 43% max 17% min	2.82 ↓ 5.28 max 2.38 min	309.6 GB ↓ 309.7 GB max 309.5 GB min	7

- elastic
- Logout
- Collapse

kibana

Clusters / production / Kibana / Instances

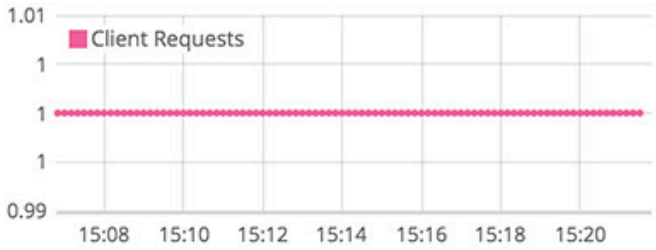
5 seconds Last 15 minutes

ensheneer.attlocal.net

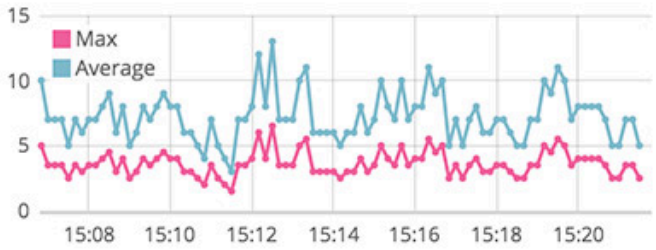
- Discover
- Visualize
- Dashboard
- Graph
- Monitoring
- Timelion
- Management
- Dev Tools

localhost:5552 OS Free Memory: **696.8 MB** Version: **5.0.0-beta1**

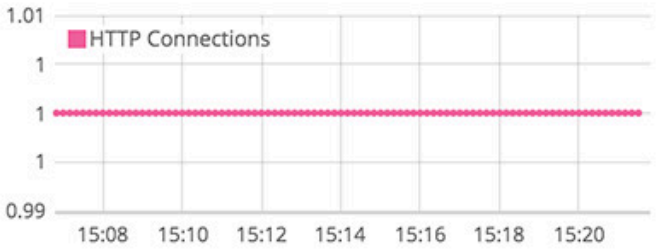
Client Requests



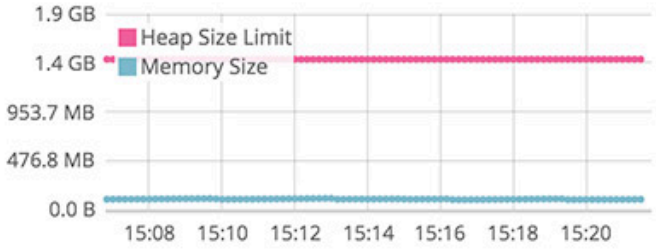
Client Response Time (ms)



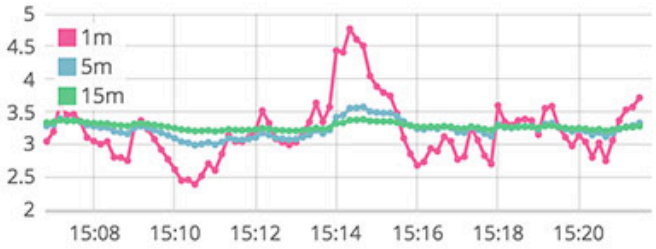
HTTP Connections



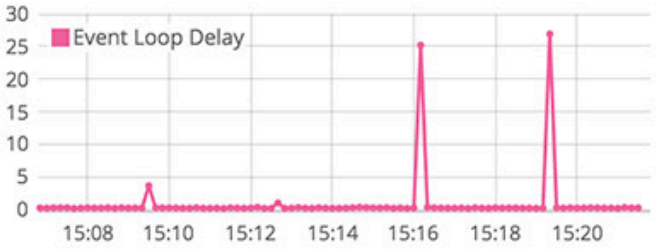
Memory Size (GB)



System Load



Event Loop Delay (ms)



- elastic
- Logout
- Collapse

X-Pack

Reporting

AUTOMATE SCHEDULING

- Email recurring status updates daily, weekly, monthly, etc.
- Combine reporting with X-Pack alerting capabilities to trigger conditional reports

SHARE AND COLLABORATE

- Export any Kibana visualization or dashboard
- Print-optimized and PDF formatted
- Download and share past reports

kibana

Management / Kibana

Index Patterns | Saved Objects | Reporting | Advanced Settings

- Discover
- Visualize
- Dashboard
- Graph
- Monitoring
- Timelion
- Management
- Dev Tools

Generated Reports

Filter Reports: ☑ **Only show my reports**

Document	Added	Status	Actions
Metrics Dashboard dashboard	2016-09-30 @ 10:02 AM elastic	processing 2016-09-30 @ 10:02 AM	
Web Traffic by Location visualization	2016-09-30 @ 10:01 AM elastic	completed 2016-09-30 @ 10:01 AM	
Traffic by Country & Device visualization	2016-09-30 @ 10:00 AM elastic	completed 2016-09-30 @ 10:01 AM	
Dashboard Overview visualization	2016-09-30 @ 10:00 AM elastic	completed 2016-09-30 @ 10:01 AM	
Web Analytics dashboard	2016-09-30 @ 9:59 AM elastic	completed 2016-09-30 @ 9:59 AM	
Alert History dashboard	2016-09-30 @ 9:57 AM elastic	completed 2016-09-30 @ 9:57 AM	
Metrics Dashboard dashboard	2016-09-23 @ 2:39 PM elastic	completed 2016-09-23 @ 2:39 PM	
Metrics Dashboard dashboard	2016-09-20 @ 7:44 PM elastic	completed 2016-09-20 @ 7:45 PM	

- elastic
- Logout
- Collapse

X-Pack

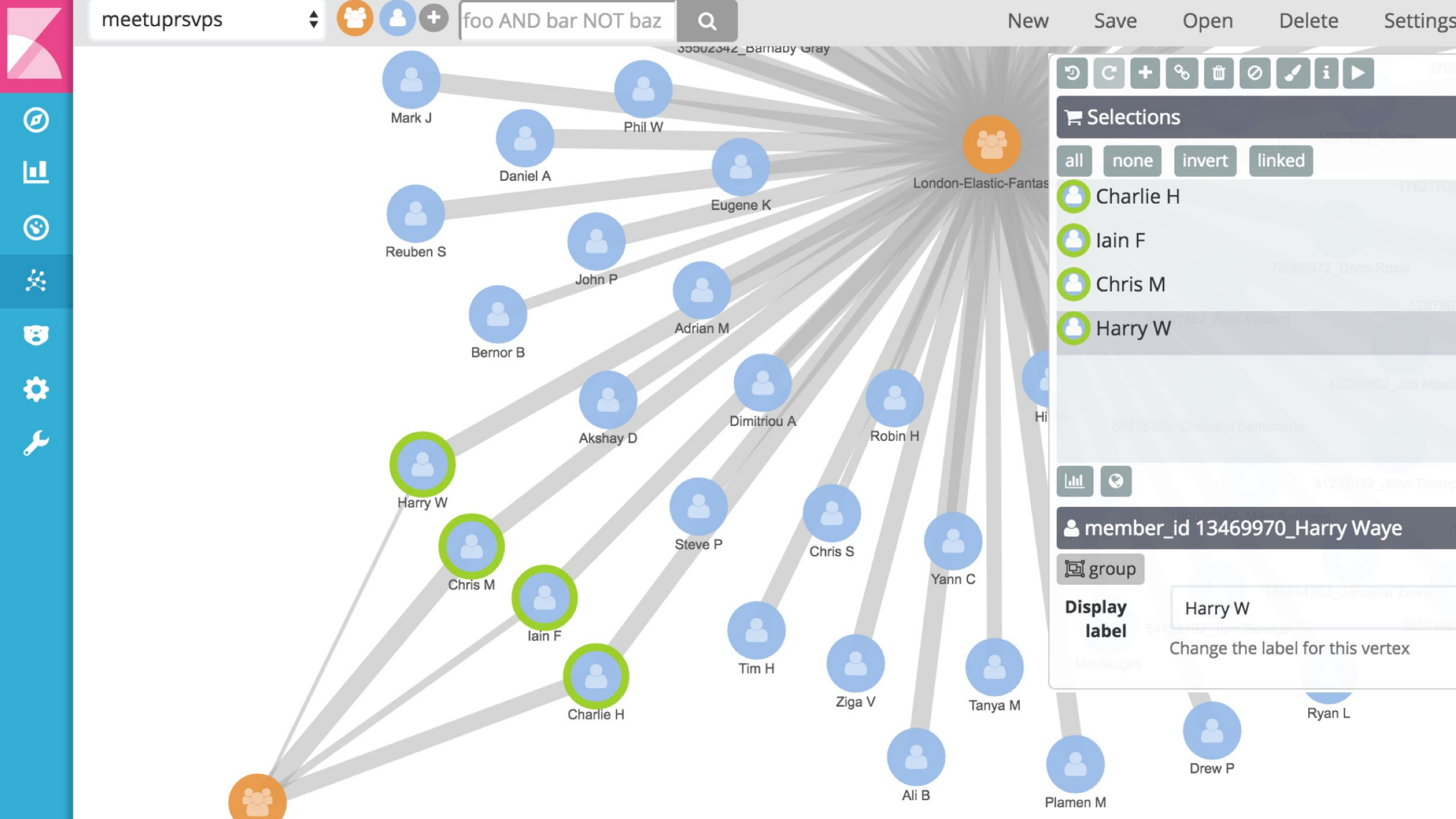
Graph

A NEW WAY TO EXPLORE DATA

- Uses relevance capabilities of Elasticsearch
- Discover linkages and connections
- Leverage API and UI-drive tool

EXTEND TO NEW USE CASES

- Fraud discovery
- Recommendations
- Cyber security
- Behavioral analyses


meetuprsvps
foo AND bar NOT baz
New
Save
Open
Delete
Settings
Mark J
Phil W
Daniel A
Eugene K
Reuben S
John P
Bernor B
Adrian M
Akshay D
Dimitriou A
Robin H
Harry W
Chris M
Steve P
Chris S
Yann C
Iain F
Tim H
Charlie H
Ziga V
Tanya M
Ali B
Plamen M
Drew P
Ryan L
London-Elastic-Fantas
Selections
all
none
invert
linked
Charlie H
Iain F
Chris M
Harry W
member_id 13469970_Harry Waye
group
Display label
Harry W
Change the label for this vertex

X-Pack

Machine Learning

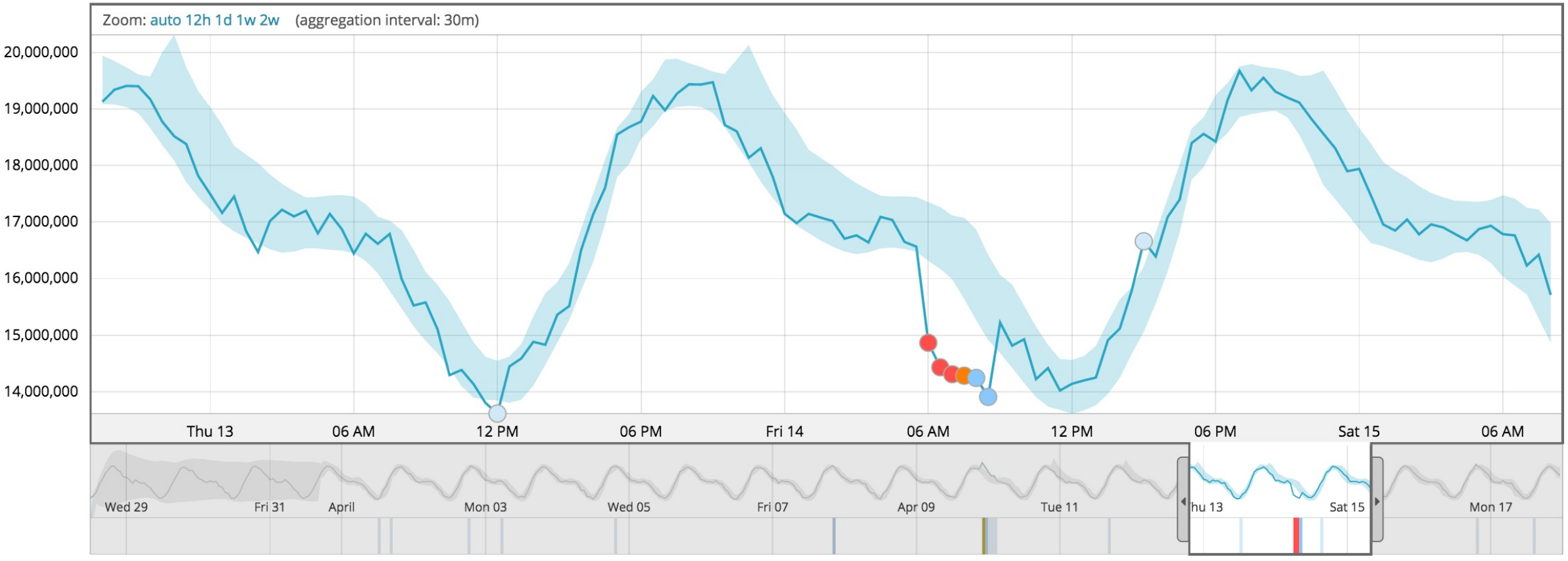
UNSUPERVISED MACHINE LEARNING

- Automatically detect anomalies
- Advanced correlation and categorization
- Identify root cause(s)
- Expose early warning signs

ENABLE NEW USE CASES

- Analyze time series data
- Expand security, IT Ops, fraud, finance, and many more use cases
- Available as beta in the 5.4 release

Time series analysis



Anomalies

Severity threshold: warning **Interval:** Auto

time	max severity	detector	actual	typical	description	job ID
April 14th 2017, 06:00	95	sum(total) (total-request)	14432600	16609200	↓ 1.2x lower	total-request
April 14th 2017, 07:00	83	sum(total) (total-request)	14310100	16421500	↓ 1.1x lower	total-request
April 14th 2017, 08:00	24	sum(total) (total-request)	13909700	15499400	↓ 1.1x lower	total-request
April 13th 2017, 12:00	1	sum(total) (total-request)	13615600	14113400	↓ Unusually low	total-request
April 14th 2017, 15:00	< 1	sum(total) (total-request)	16659900	15863700	↑ 1.1x higher	total-request

Job settings

Fields

event rate	Count
accept	Mean
deny	Mean
response	Mean
total	Sum

Bucket span

30m

Split Data

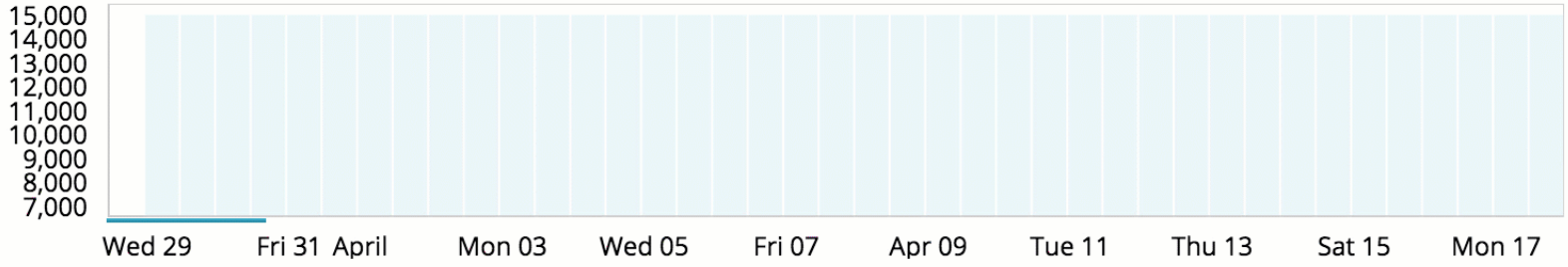
service.keyword

Key Fields

- host.keyword
- service.keyword

Results

Document count



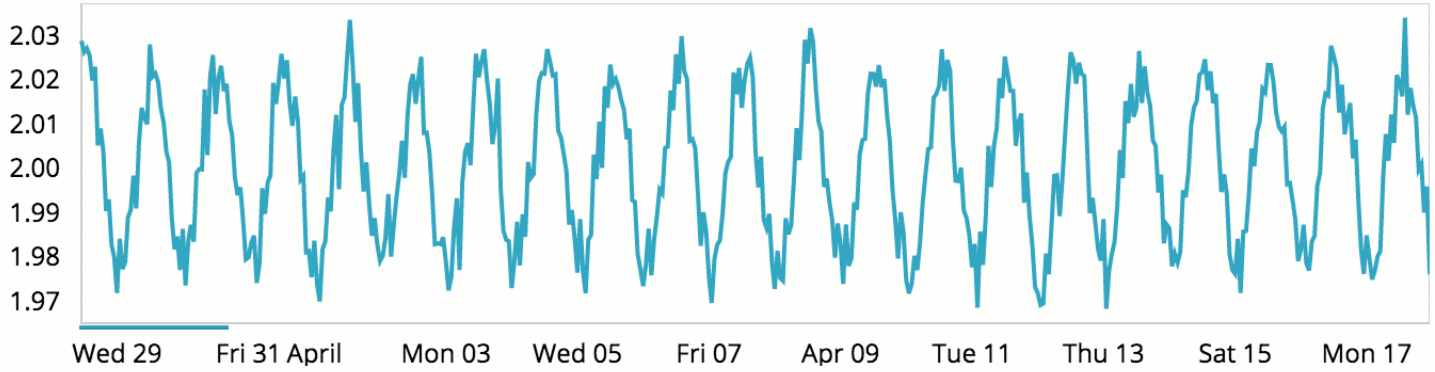
Data split by service.keyword

app_2

app_1

app_0

Mean response



Sum total

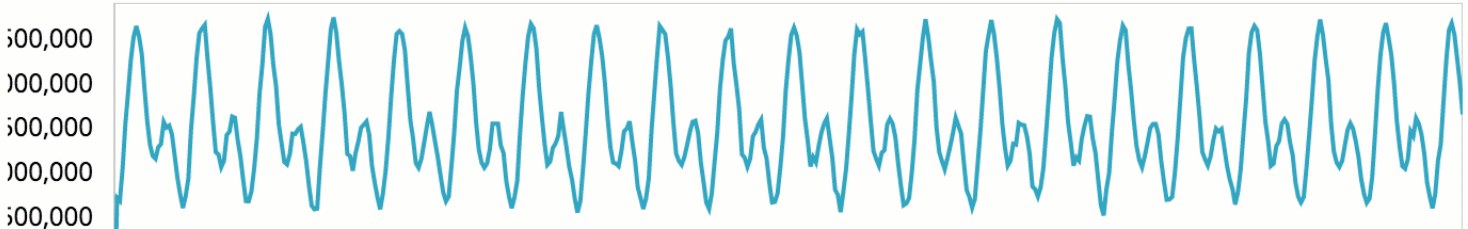


Chart interval: 1h Use full server* data

New job from index pattern server*

Aggregation

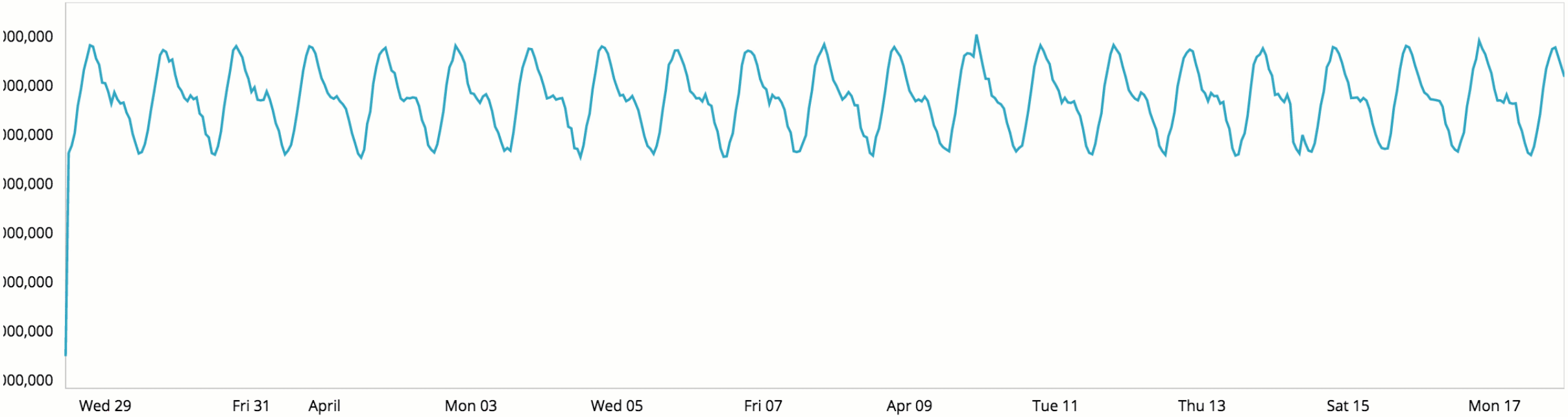
Sum

Field

total

Bucket span

60m



Name

total-requests

Description

Job description

Advanced

Create Job

host.keyword

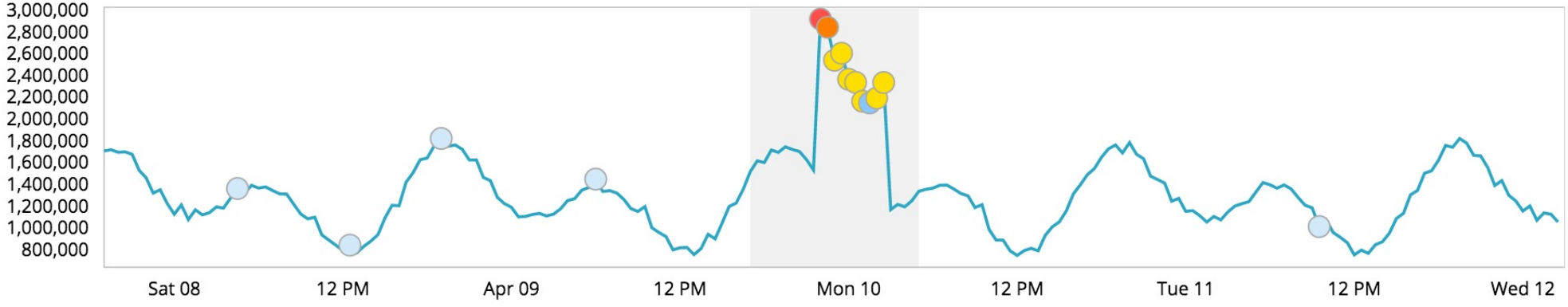
server_2 97 826

server_1 94 808

server_3 71 263

Anomalies

sum(total) (metric-by-appl) - service.keyword app_4



Severity threshold: major **Interval:** Auto

time	max severity	detector	found for	influenced by	actual	typical	description	job ID
April 9th 2017, 22:00	81	sum(total) (metric-by-appl)	app_4	host.keyword: server_2 host.keyword: server_1 service.keyword: app_4	2922690	1483330	2x higher	metric-by-appl

Description:
critical anomaly in sum(total) (metric-by-appl) found for service.keyword app_4

Details on highest severity anomaly:

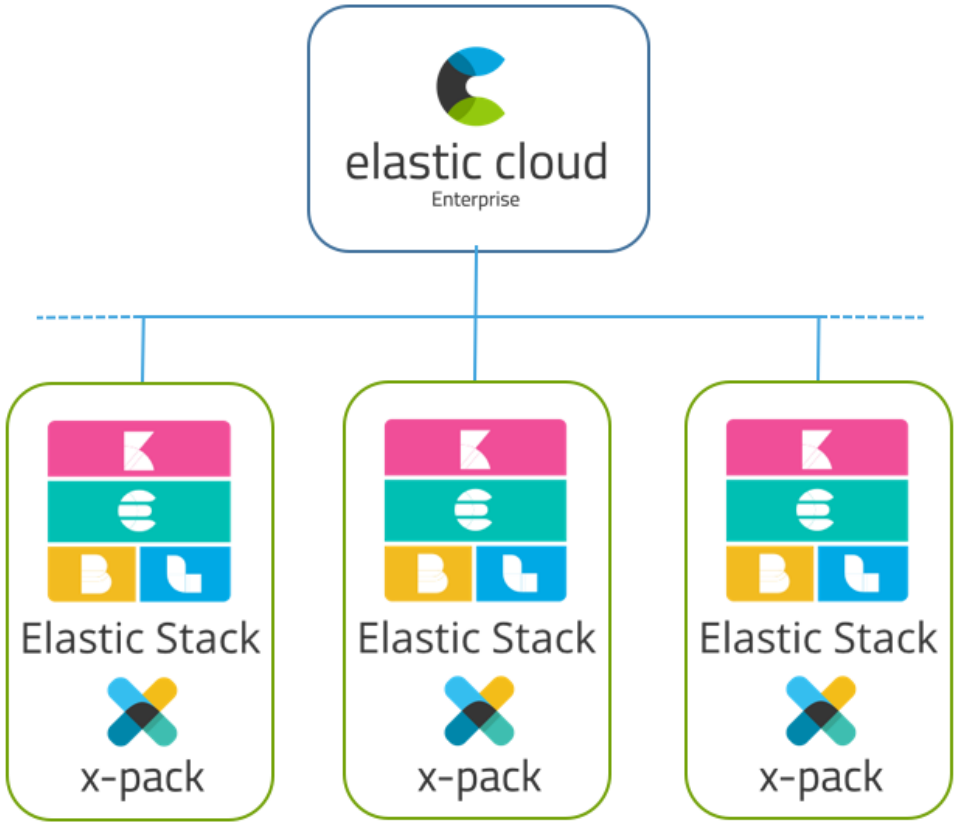
service.keyword:	app_4
time:	April 9th 2017, 22:00:00 to April 9th 2017, 22:30:00
function:	sum
fieldName:	total
actual:	2922690
typical:	1483330
job ID:	metric-by-appl
probability:	2.24588e-25

Influenced by:

host.keyword	server_2
host.keyword	server_1
service.keyword	app_4

Elastic Cloud Enterprise

Provision and manage multiple Elastic Stack environments; Expose logging as a service to your entire organization



Clusters / 5.1 test

Overview Manage Security Activity Snapshots Console

Search for anything...

Zones: **3** Nodes per Zone: **3** Memory per node: **1 GB** Total memory: **9 GB** Edit Cluster

Endpoints

Elasticsearch Kibana

Availability zones and nodes

eu-west-1b	eu-west-1a	eu-west-1c
Node #3 v5.1.1 Memory pressure: 30 % of 1 GB Disk used: 0 MB of 24 GB Stop routing / Pause	Node #7 v5.1.1 Memory pressure: 14 % of 1 GB Disk used: 0 MB of 24 GB Stop routing / Pause	Node #9 v5.1.1 Memory pressure: 47 % of 1 GB Disk used: 47 MB of 24 GB Stop routing / Pause
Node #2 v5.1.1 Memory pressure: 61 % of 1 GB Disk used: 46 MB of 24 GB Stop routing / Pause	Node #6 v5.1.1 Memory pressure: 54 % of 1 GB Disk used: 48 MB of 24 GB Stop routing / Pause	Node #10 v5.1.1 Memory pressure: 51 % of 1 GB Disk used: 1 MB of 24 GB Stop routing / Pause
Node #4 v5.1.1 Memory pressure: 38 % of 1 GB Disk used: 45 MB of 24 GB Stop routing / Pause	Node #5 v5.1.1 Memory pressure: 29 % of 1 GB Disk used: 0 MB of 24 GB Stop routing / Pause	Node #8 v5.1.1 Memory pressure: 36 % of 1 GB Disk used: 1 MB of 24 GB Stop routing / Pause

Elasticsearch Kibana

Search for anything...

Add filter | Elasticsearch version (5.1.1) ✖

Create cluster

Cluster (showing all 8 matching clusters)	Status	Version	Zones	Nodes per Zone	Memory per node	Total memory
5.1 fresh (2e9785)	✓	5.1.1	3	1	1 GB	3 GB
5.1 test (463668)	✓	5.1.1	3	3	1 GB	9 GB
5.1.1-QA (858edc)	✓	5.1.1	2	1	8 GB	16 GB
5.1.1-qa-small (9e9fbf)	✓	5.1.1	2	1	1 GB	2 GB
8263ecf4105b8b16a5c81045fd7b4f71 (8263ec)	✓	5.1.1	1	1	1 GB	1 GB
c58e6934c472a00e07de8e756e5f3673 (c58e69)	✓	5.1.1	1	1	1 GB	1 GB
fadbd71e6b01cafdb3b9a687369bc94b (fadbd7)	✓	5.1.1	1	1	1 GB	1 GB
ff9edb7647b2a760670dd0c3a21dc018 (ff9edb)	✓	5.1.1	1	1	1 GB	1 GB

Allocators • Runners • Proxies • Versions Admin Settings

Search for anything...

Available capacity

Zone	elasticsearch	elasticsearch_data	elasticsearch_master_only	kibana	non_ssd	
eu-west-1a	333.38 GB	293.81 GB	43.63 GB	43.63 GB	43.63 GB	289.75
eu-west-1b	541 GB	508.25 GB	33.5 GB	33.5 GB	33.5 GB	507.5
eu-west-1c	343.13 GB	290.13 GB	89.63 GB	89.63 GB	89.63 GB	253.5

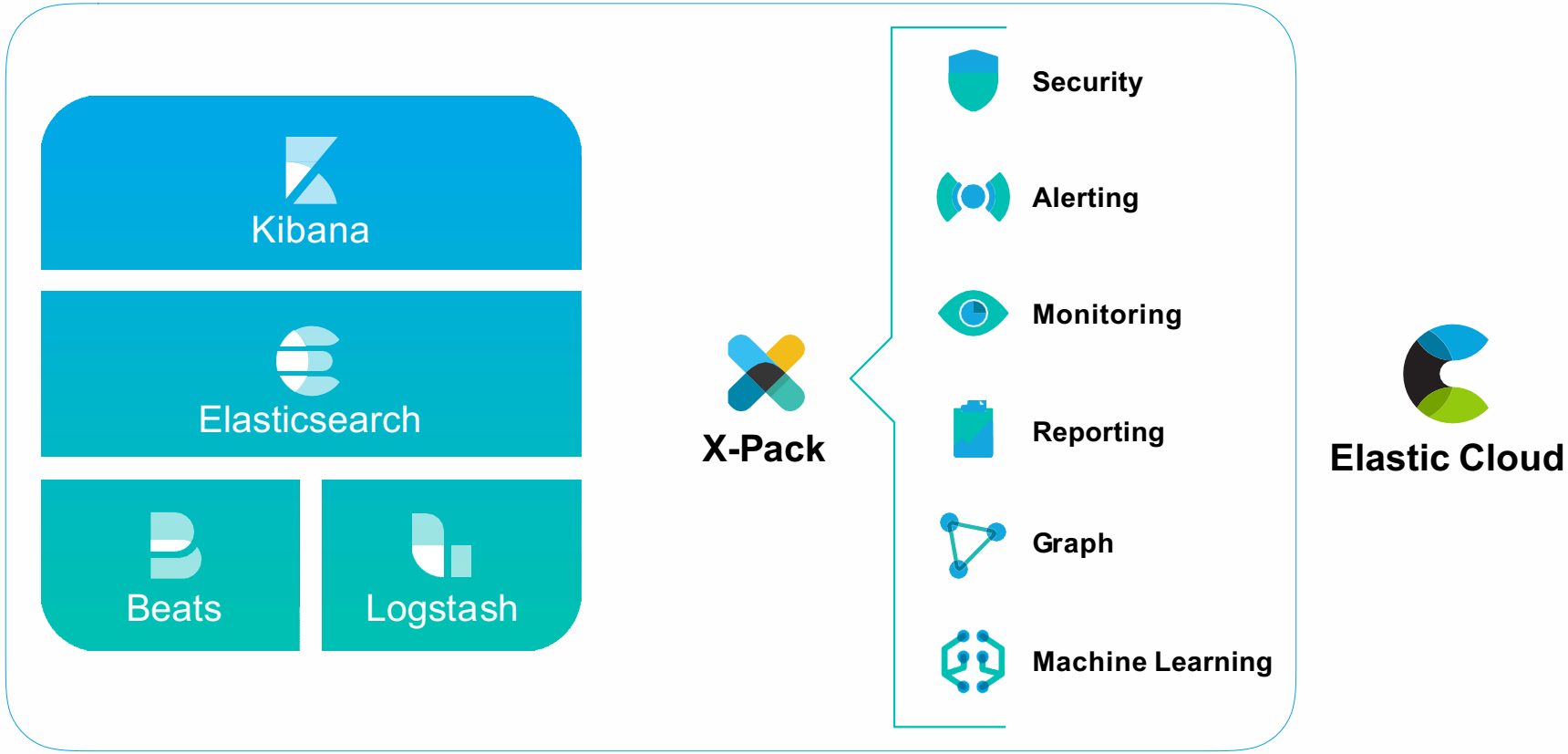
eu-west-1a

Allocator	Status	Available capacity	Node distribution
[illegible]	✓	234 GB	
[illegible]	✓	27 GB	
[illegible]	✓	23.25 GB	
[illegible]	✓	19 GB	
[illegible]	✓	9.25 GB	
[illegible]	✓	4.5 GB	
[illegible]	✓	3.19 GB	
[illegible]	✓	3.19 GB	
[illegible]	✓	2 GB	
[illegible]	✓	1.56 GB	
[illegible]	✓	1.25 GB	
[illegible]	✓	1.13 GB	
[illegible]	✓	1 GB	
[illegible]	✓	512 MB	

Architecture

Confidential information, for internal use only

The Elastic Stack



Confidential information, for internal use only

Elastic Stack

Beats

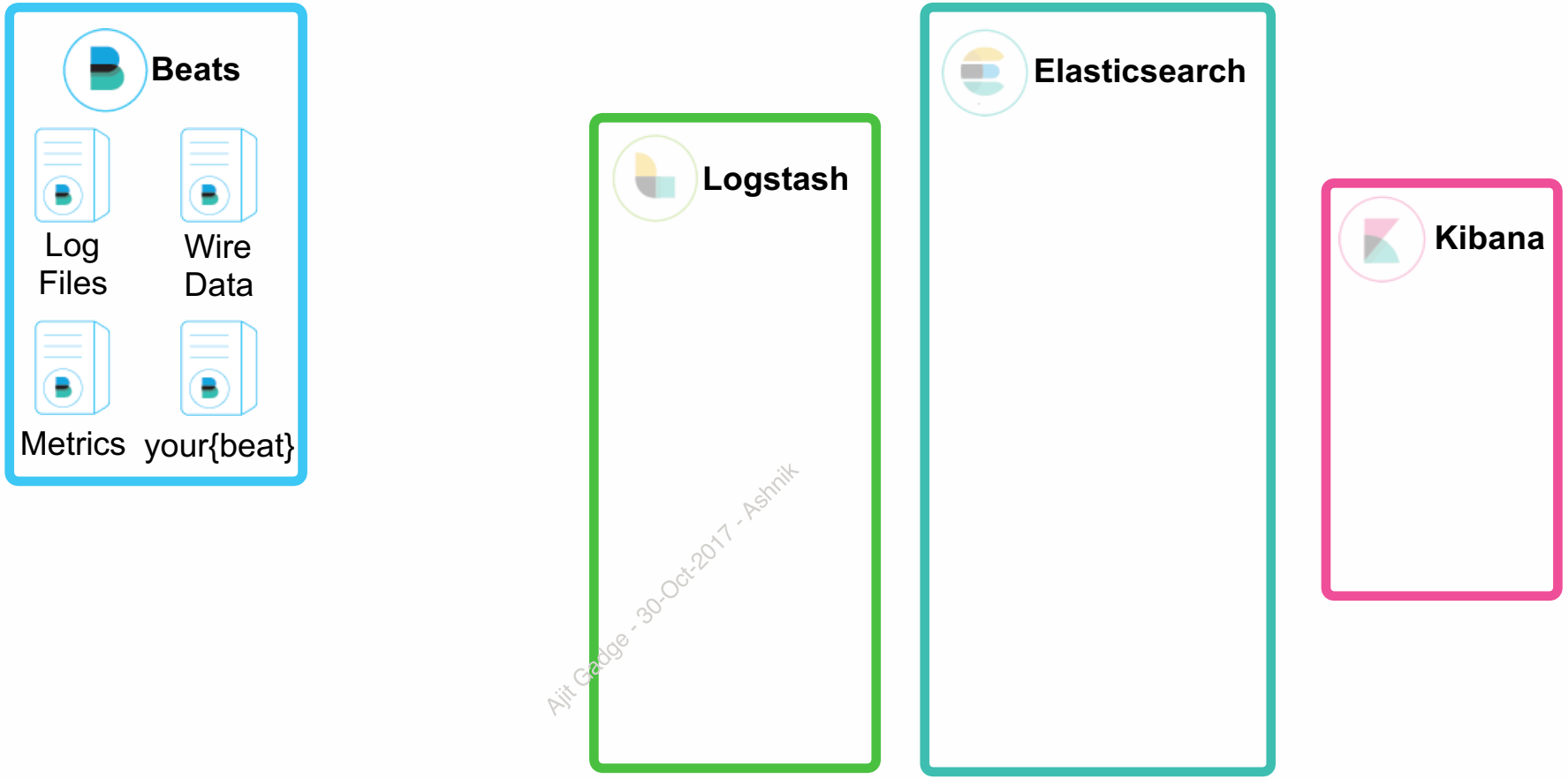
Logstash

Elasticsearch

Kibana

Confidential information, for internal use only

Beats

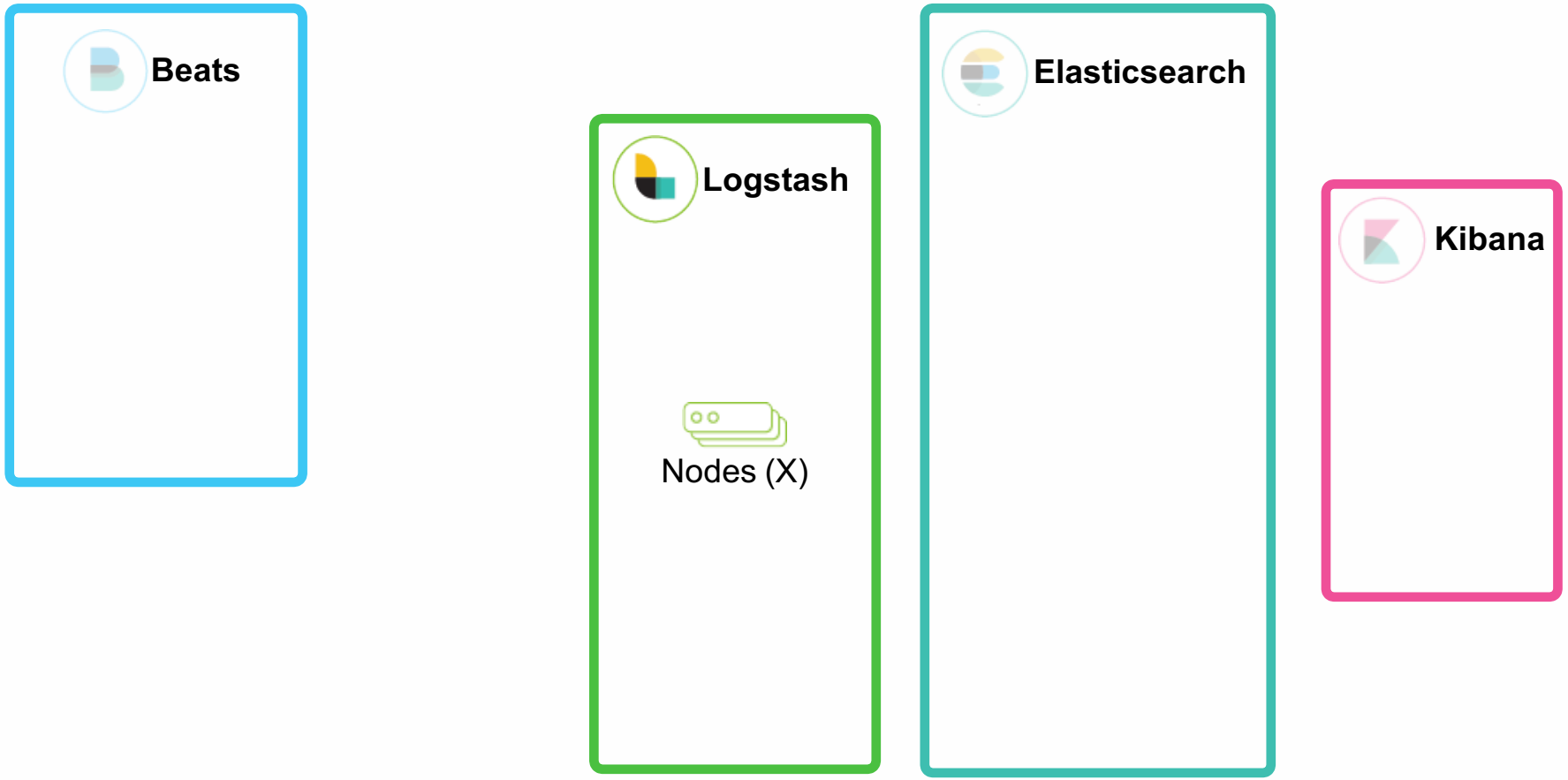


- Data Shipper
- Light Weight
- Application Side Component

Ajit Gadge - 30-Oct-2017 - Ashnik

Confidential information, for internal use only

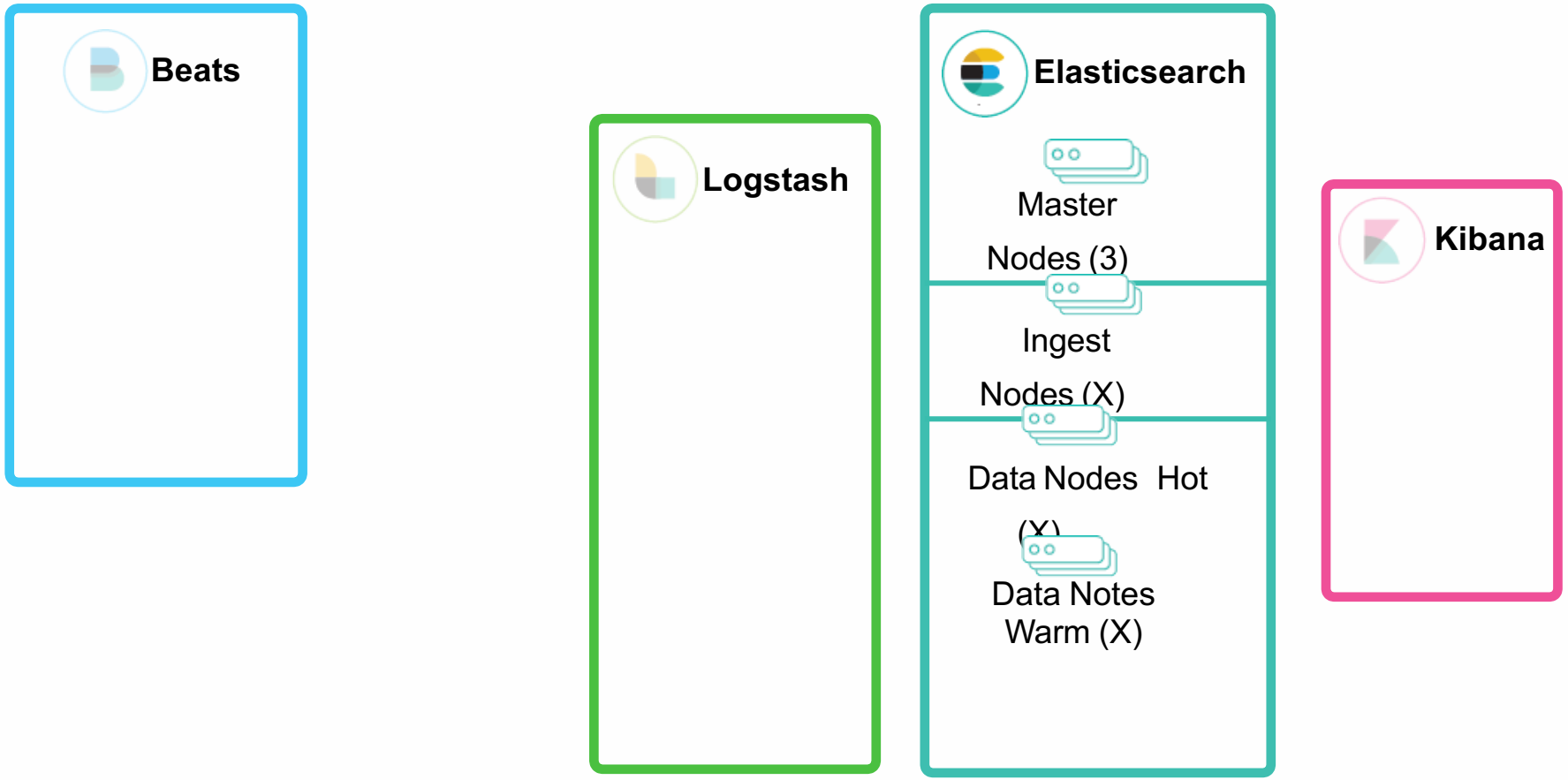
Logstash



- Data Collector/Processor
- Heavy Weight
- Server Side Component

Confidential information, for internal use only

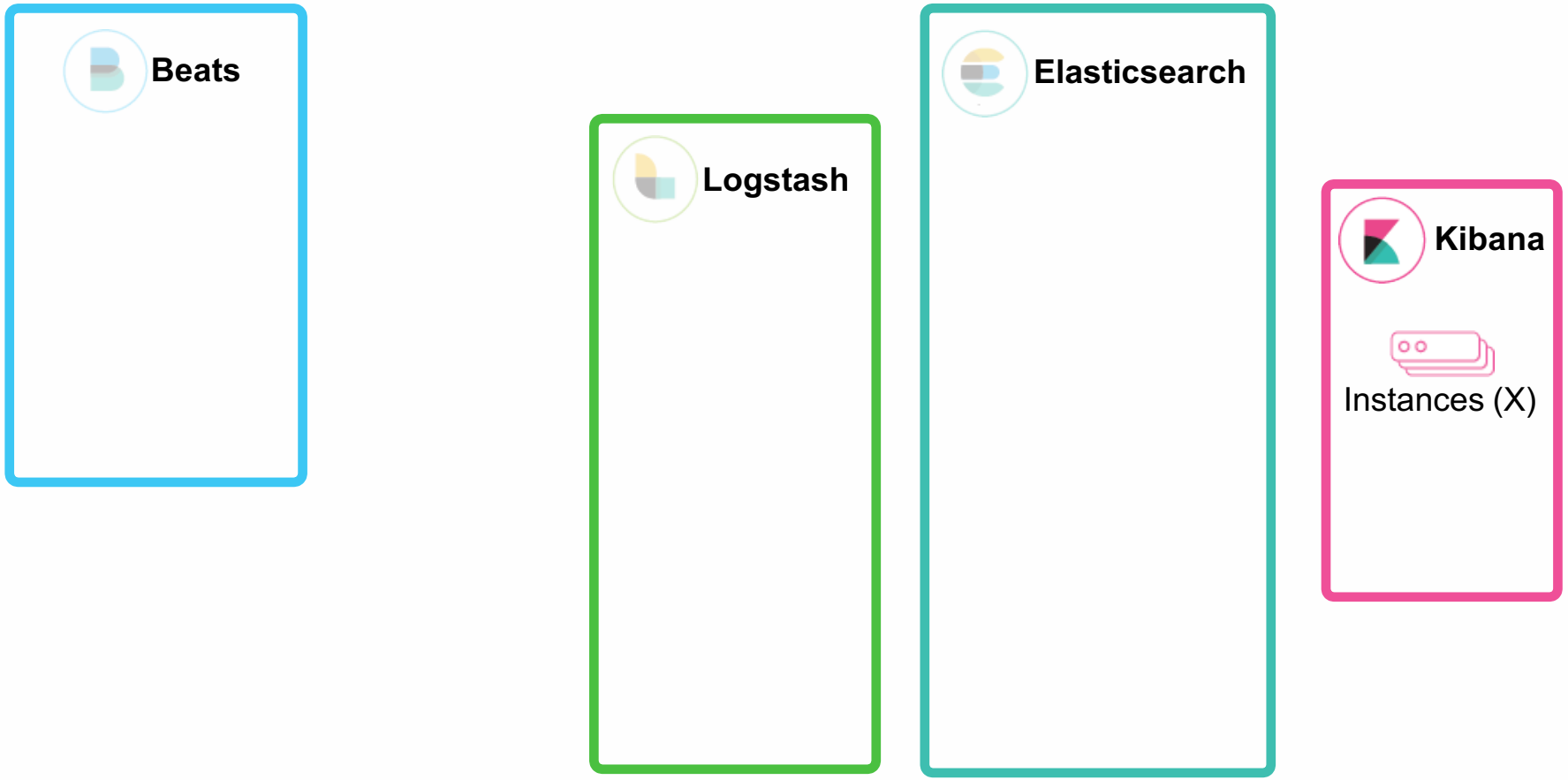
Elasticsearch



- Data Platform
- Really Fast
- HTTP + JSON

Confidential information, for internal use only

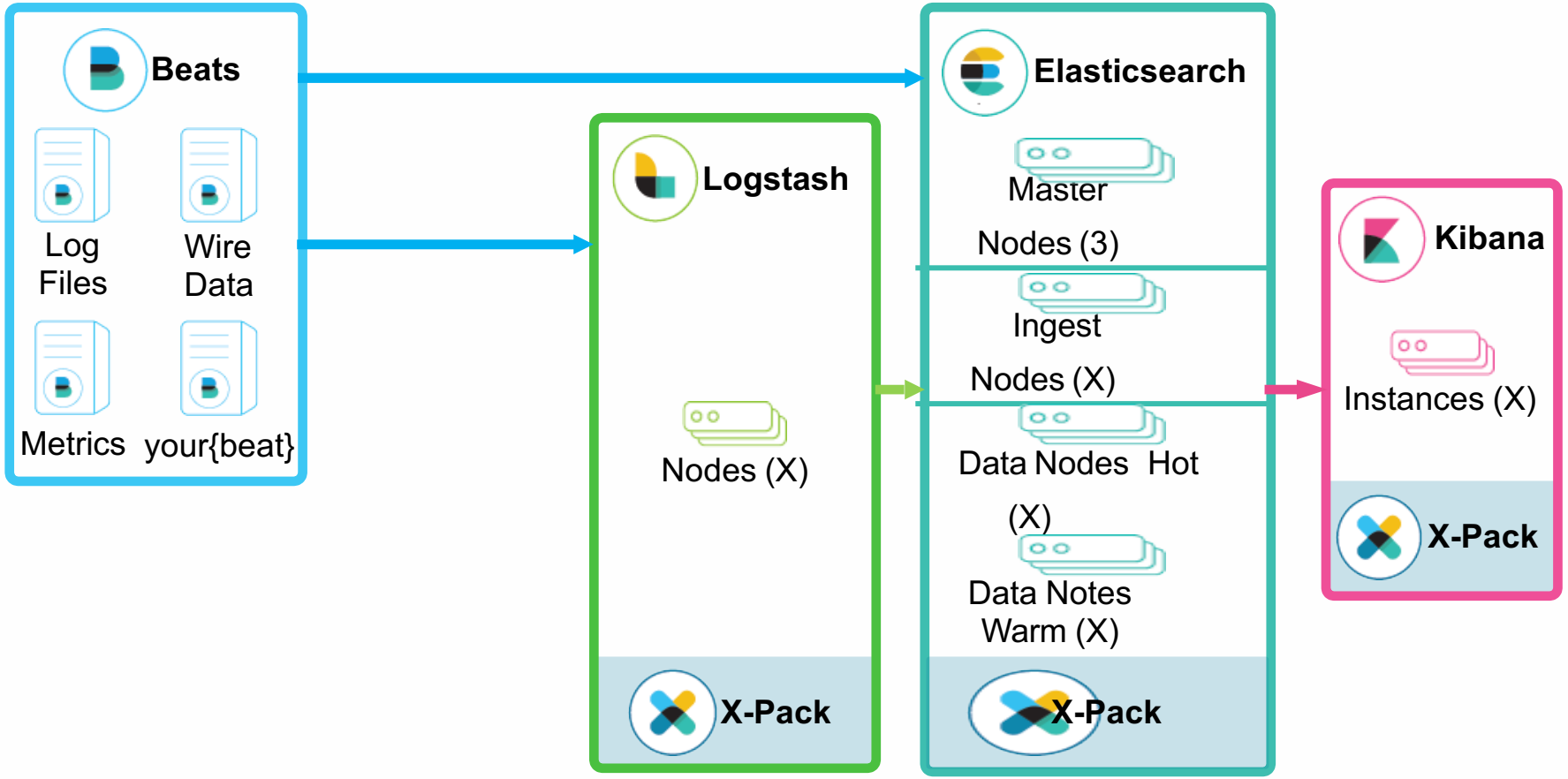
Kibana



- Data Visualization
- Stack Configuration
- Elastic Stack UI

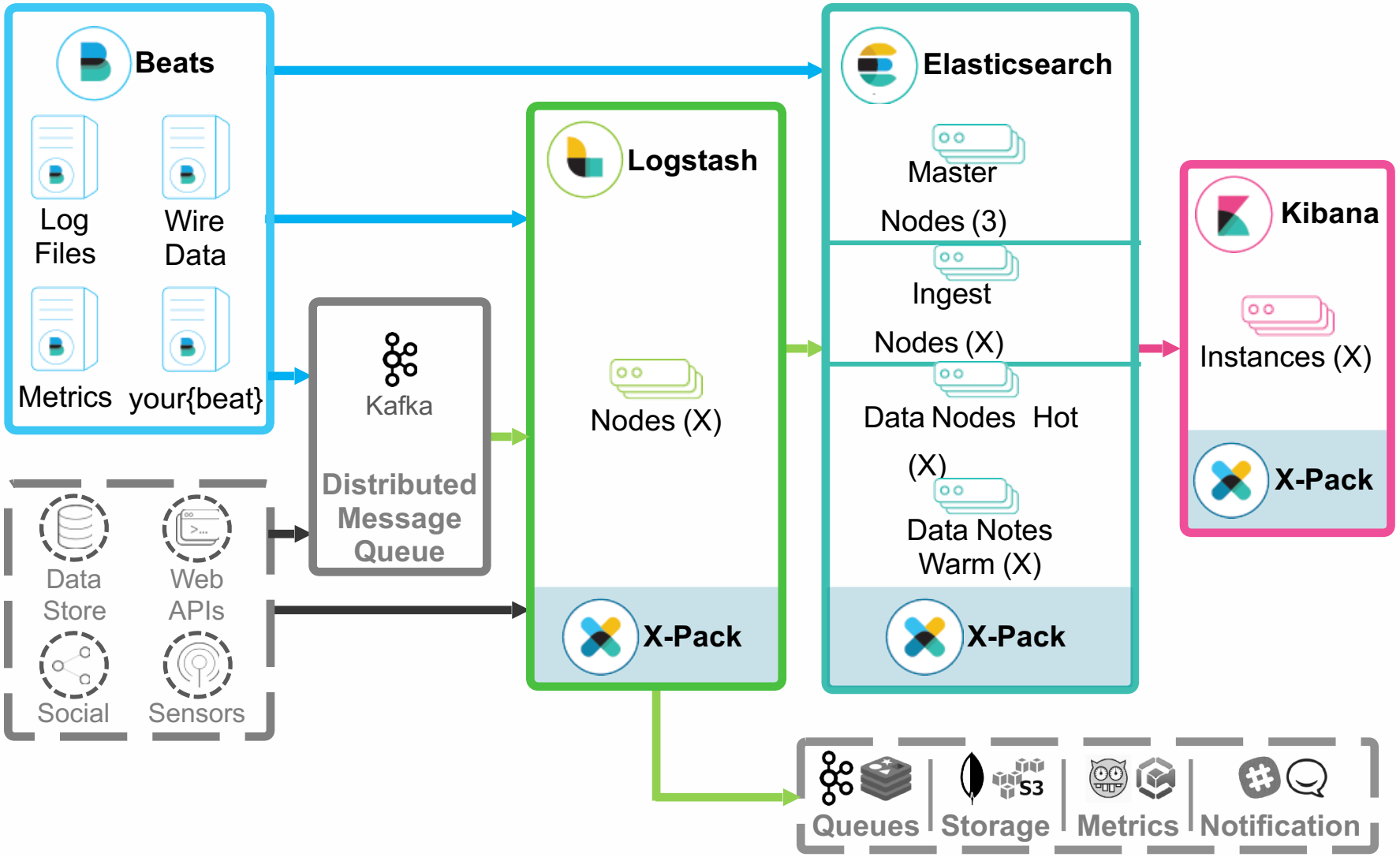
Confidential information, for internal use only

Elastic Stack



Confidential information, for internal use only

The Elastic Journey of an Event



Confidential information, for internal use only

Understanding of Event

Confidential information, for internal use only

Understanding Event Data

- Data can be divided into two main categories:
  - **fixed-size data**
    - predictable (small or large)
    - easier
  - **event-based data**
    - unpredictable
    - harder
- Even though you can use the Elastic Stack to analyze fixed-size data, this course focuses on event-based data

Confidential information, for internal use only

Event

- *"An action or occurrence recognized by software, often originating asynchronously from the external environment, that may be handled by the software."* (Wikipedia)
- Any piece of information that has a **timestamp** and **data:**
  - log line
  - tweet
  - CPU
  - temperature
  - ...

Confidential information, for internal use only

Event Lifecycle

1. **Genesis**: an event occurs
2. **Record**: device or application creates a record of that event
3. **Transport & Store**: move logs to alternate storage (archival, indexing, etc)
4. **Search & Analyze**: search for events, analyze data for trends and anomalies
5. **Archive**: long-term storage of events (often for regulatory reasons)
6. **Purge**: delete old data which is no longer required

Confidential information, for internal use only

Event-Based Data

- Events are represented as ***documents***
  - Typically events represent a real-world event that occurred
    - Cannot be repeated (at least not the exact same event)
    - Once occurred, it cannot be modified (it happened in the past and we don't support time travel…. yet)
- Each event (document) is associated with a ***timestamp***
- Documents keep on "flowing" in
  - Not only do we not support time travel, we cannot even stop time from ticking

Confidential information, for internal use only

Event-Based Data

- The nature of time-based data
  - Scales with time - always increasing
  - Impossible to predict the scale of future events
- Dictates the nature of how we use time-based data
  - We cannot keep all events for all time (not enough room in the world)
  - We mostly care about recent events compared to old events
  - Old needs to be defined as it is relative to the problem you're facing

Confidential information, for internal use only

ElasticSearch

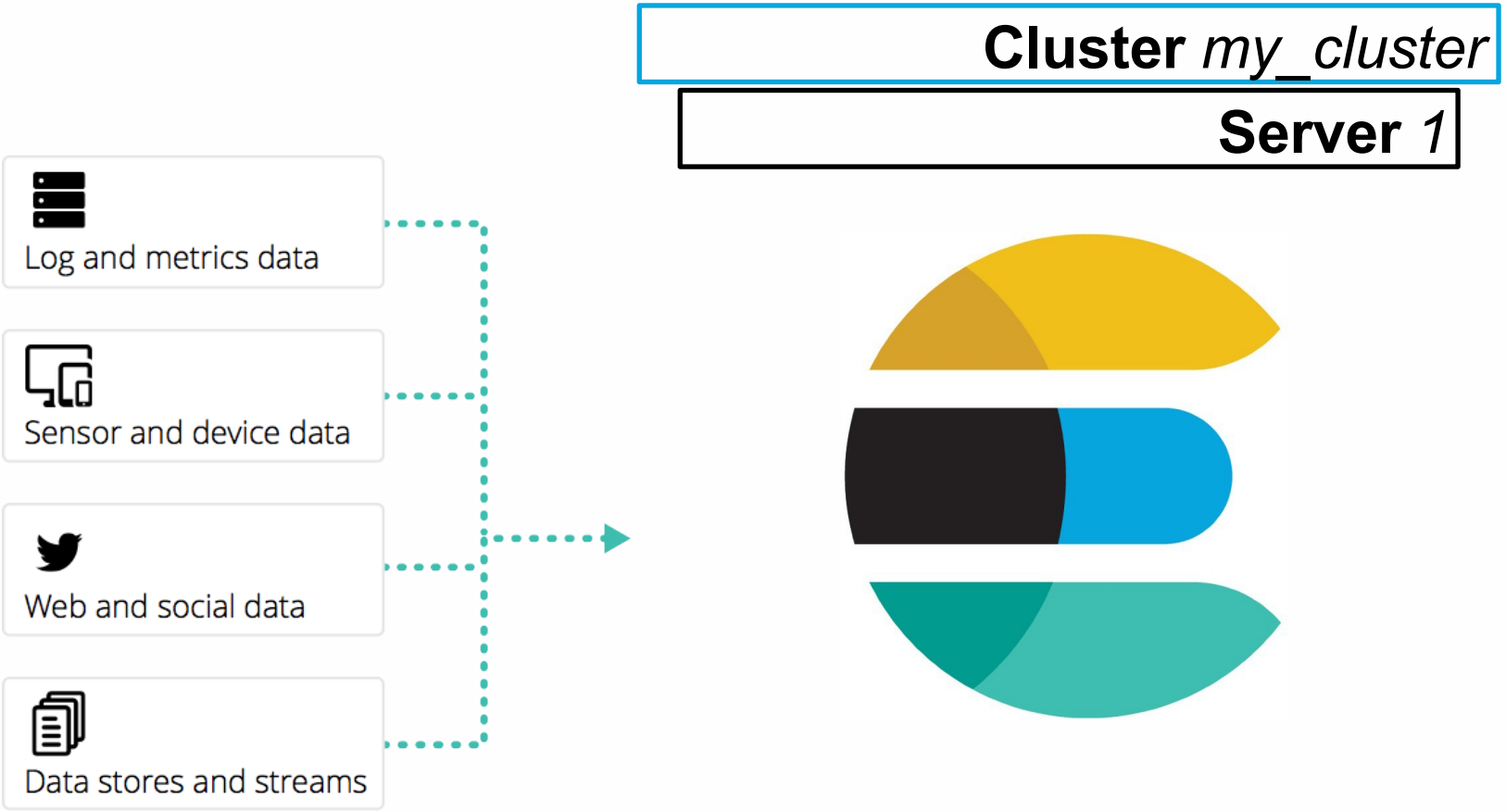
Confidential information, for internal use only

Elasticsearch

- What
  - Highly scalable open-source search and analytics engine
  - search: full-text, structured, geo, suggestions, highlighting, ...
  - analytics: trends, statistics, summarizations, anomalies, ...
  - real time & near-real time
- Why
  - Easy to use: simple RESTful API
  - log and event data analysis
  - full text search
  - alerting and classification
  - recommendations

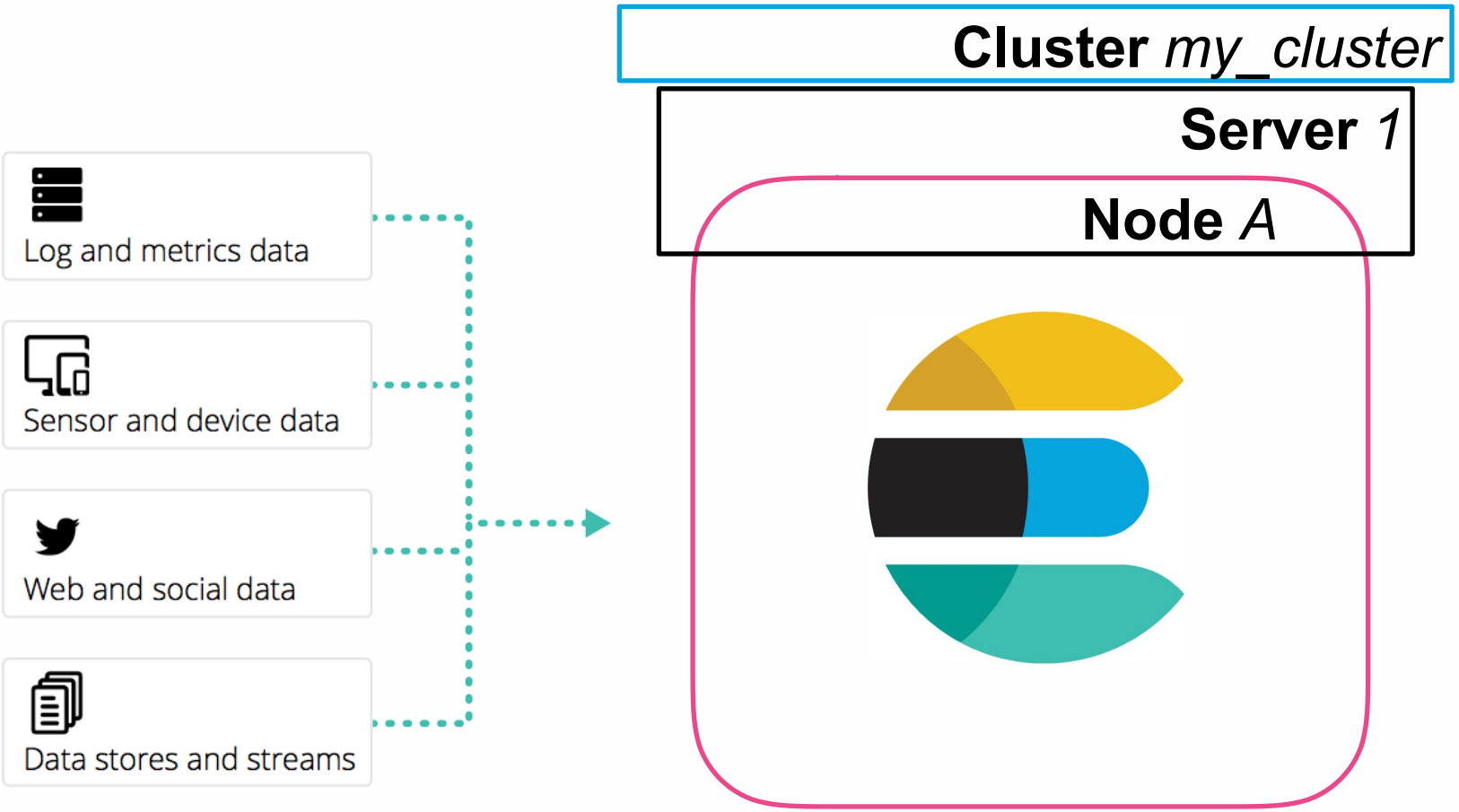
Confidential information, for internal use only

Terminology



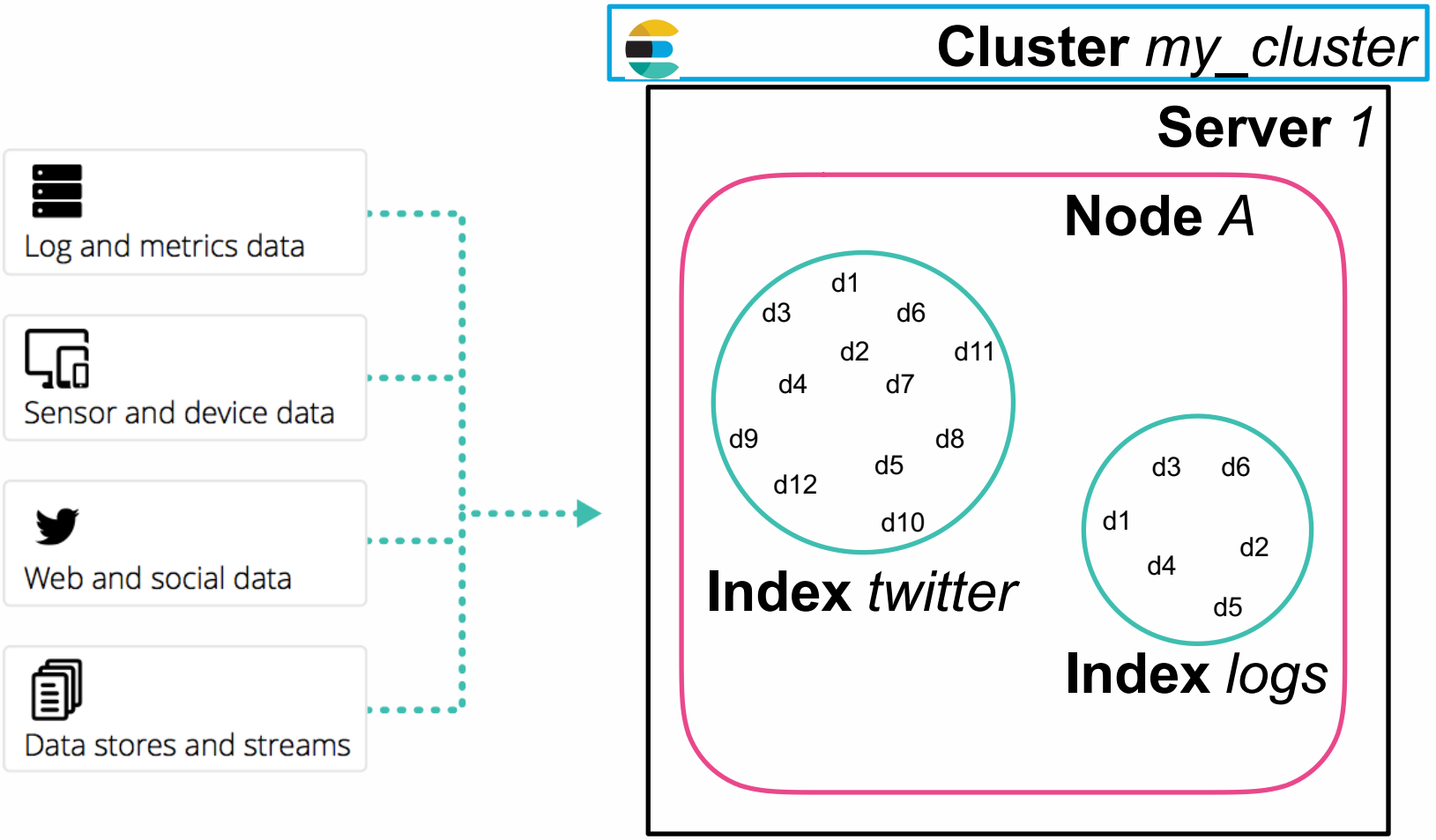
Confidential information, for internal use only

Terminology



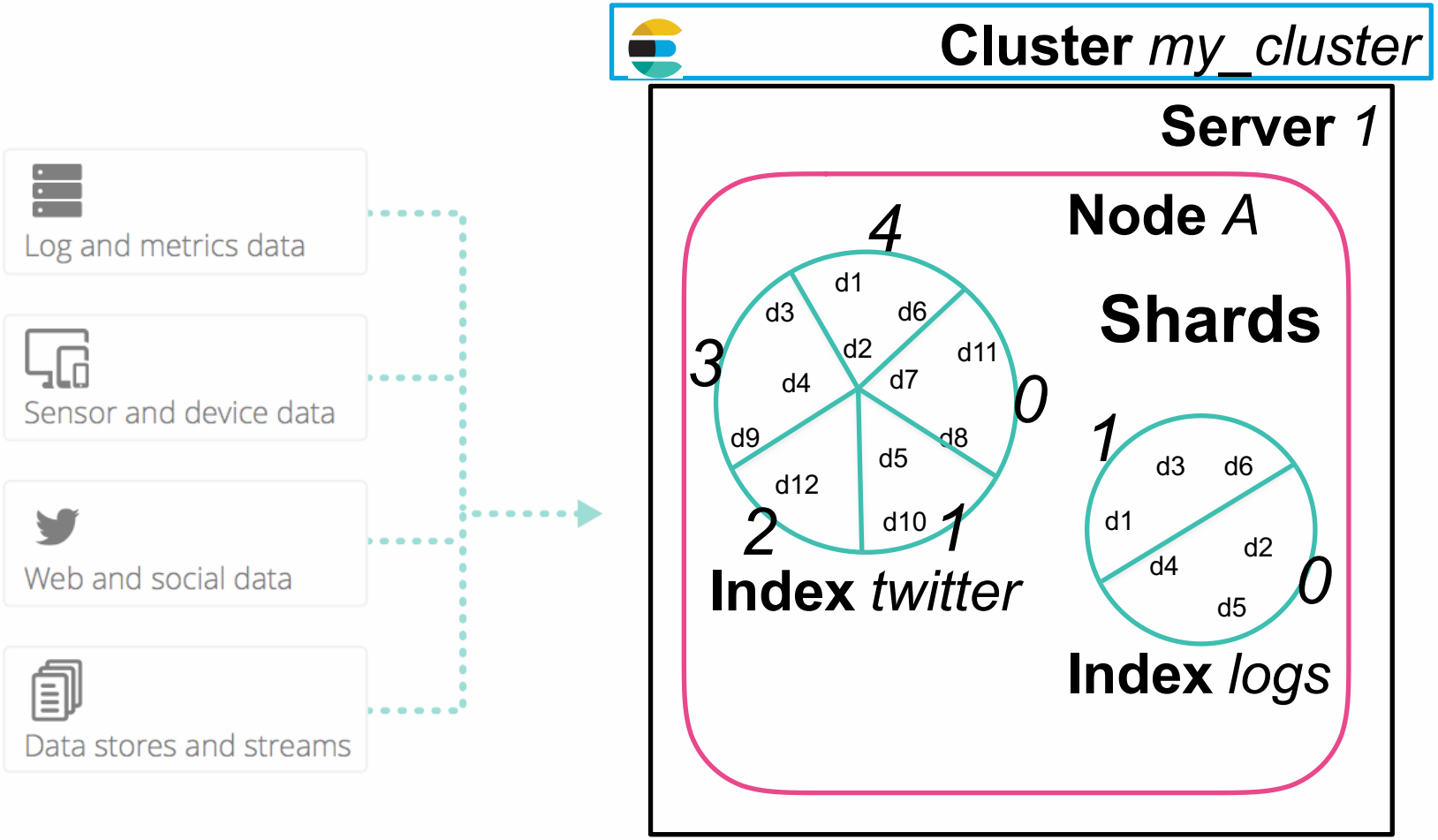
Confidential information, for internal use only

Terminology



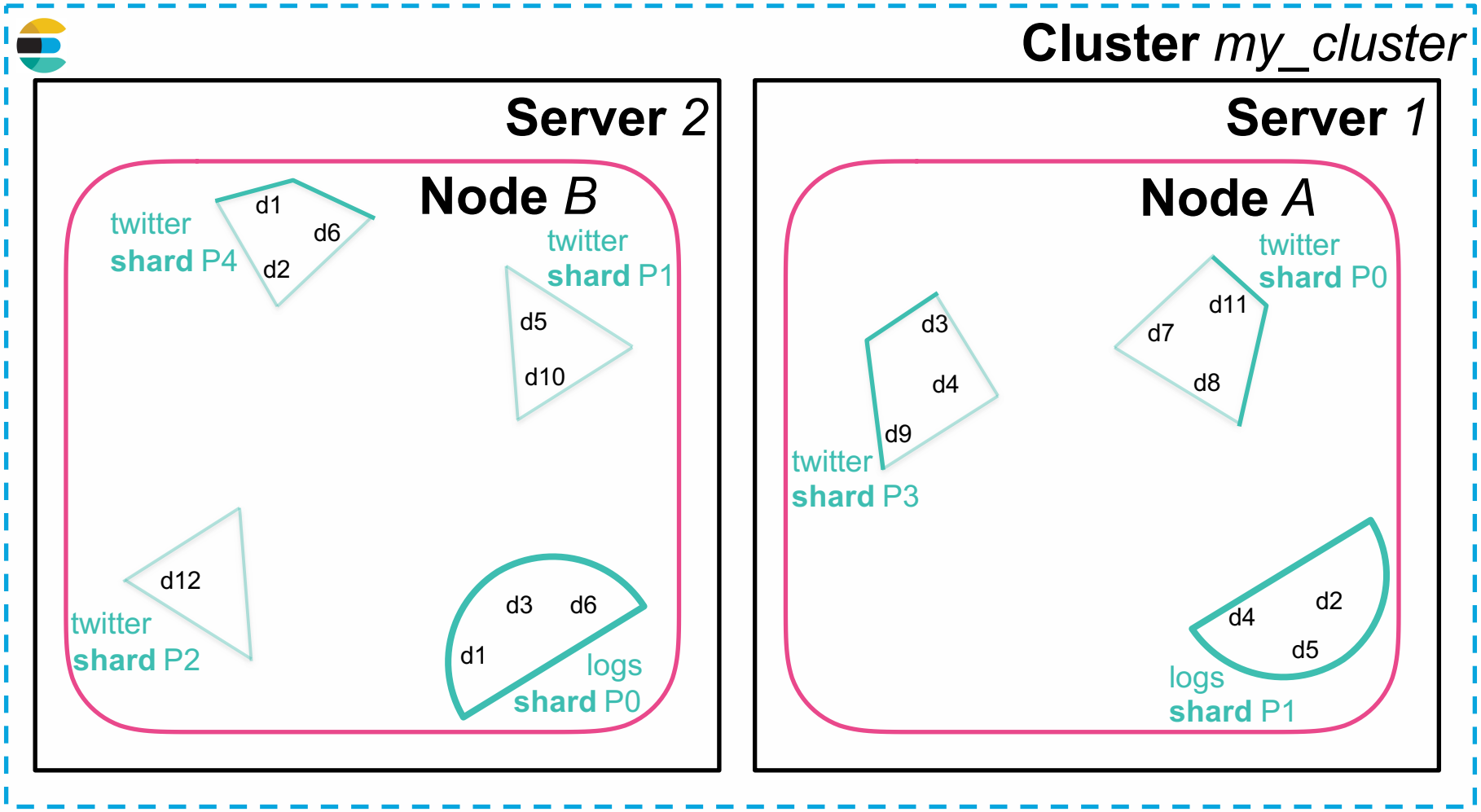
Confidential information, for internal use only

Partition



Confidential information, for internal use only

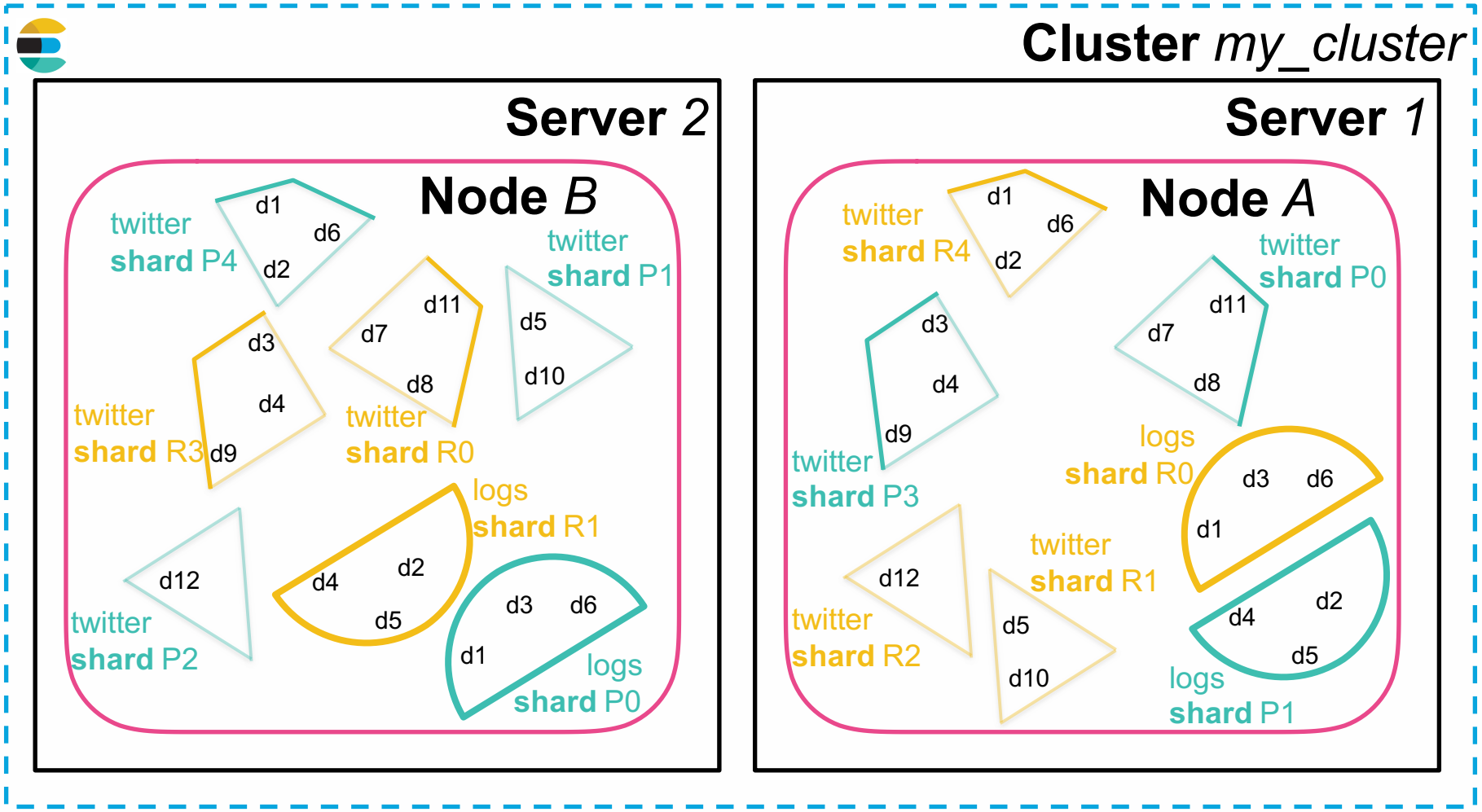
Distribution



Confidential information, for internal use only

Replication

- Primaries
- Replicas



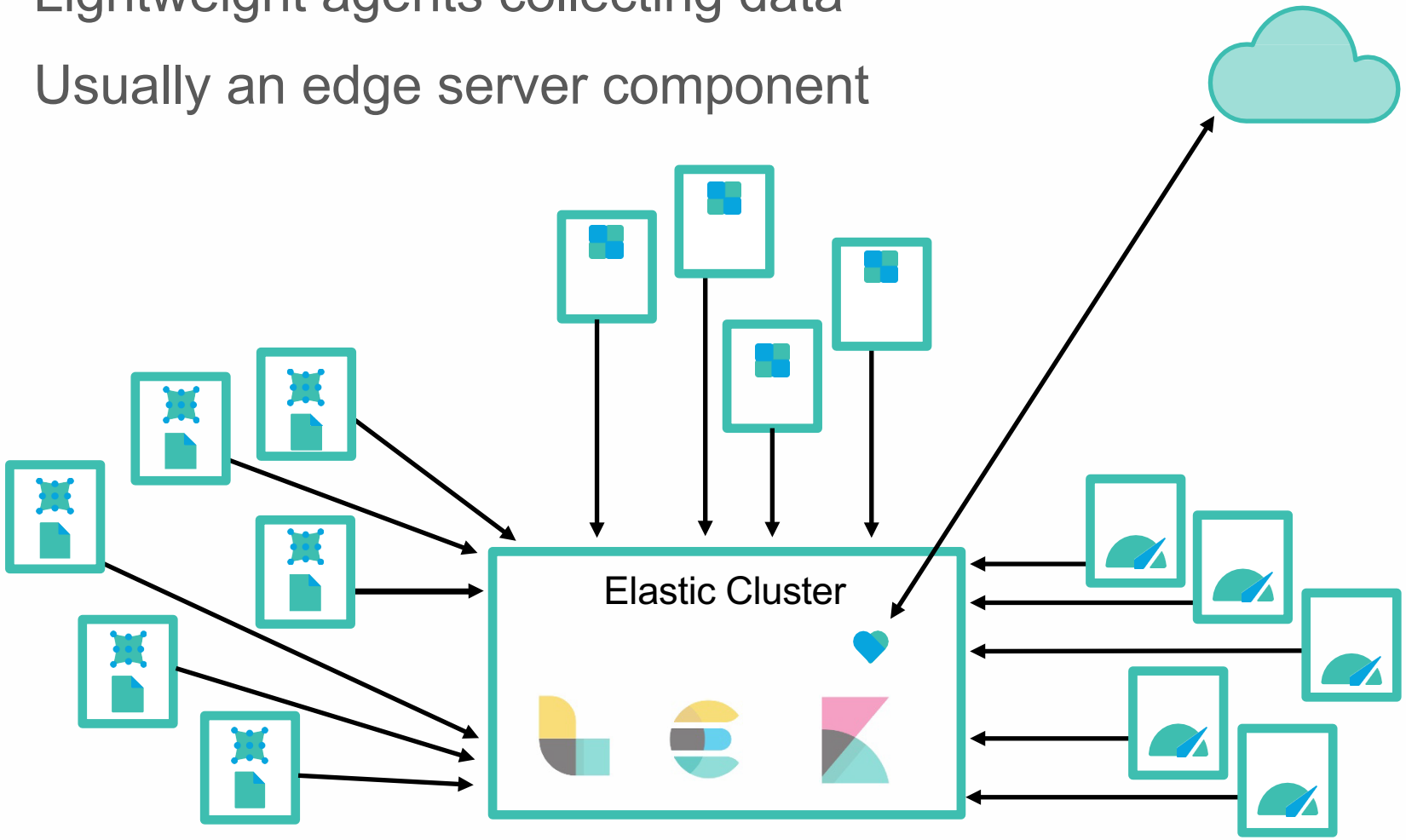
Confidential information, for internal use only

Beats

Confidential information, for internal use only

Beats

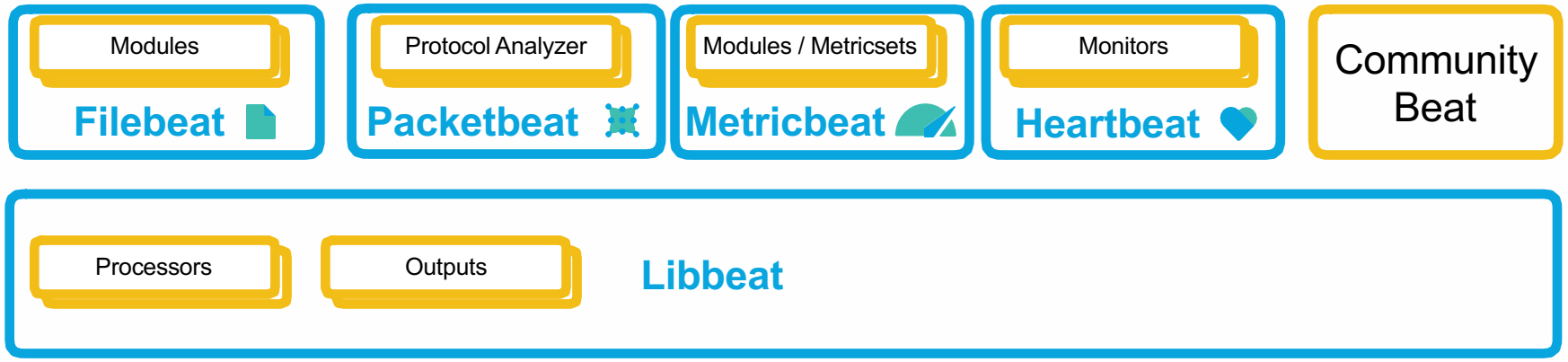
- Open source data shippers
- Lightweight agents collecting data
- Usually an edge server component



Confidential information, for internal use only

Beats

- Multiple flavors

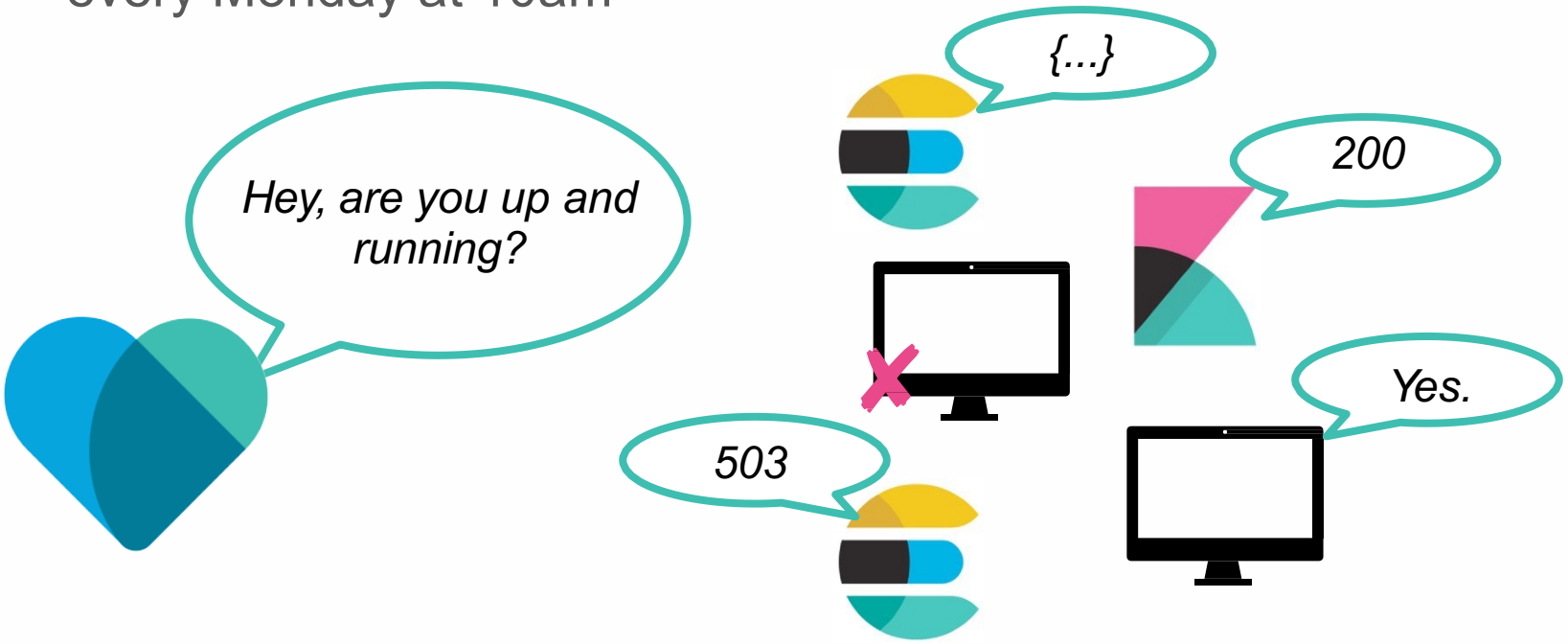


- Every Beat has two main components:
  - data collector (flavor: file, packet, metric, heart, ...)
  - data processor and publisher (**Libbeat**)

Confidential information, for internal use only

Heartbeat

- Tells you whether one or more services are reachable
- Monitors services based on a cronjob-fashion specified schedule, for example:
  - every 10 minutes
  - every Monday at 10am



Confidential information, for internal use only

Heartbeat Monitors

- ICMP (v4 and v6) Echo Requests
  - requires root access
- TCP
  - verify the endpoint by sending and/or receiving a custom payload
- HTTP
  - verify that the service returns the expected response, such as a specific status code, response header, or content
- Both TCP and HTTP monitors support SSL/TLS and some proxy settings.

Confidential information, for internal use only

Heartbeat Config

- The config file (**heartbeat.yml**) is in the heartbeat folder

```
heartbeat.monitors:

- type: icmp
  schedule: '*/5 * * * * * *'
  hosts: ["myhost"]

- type: tcp
  schedule: '@every 5s'
  hosts: ["otherhost:12345"]
  mode: any

output.elasticsearch
:
  hosts: ["monitorcluster:9200"]
```

Monitors myhost using ICMP exactly every 5s (10:00:00, 10:00:05, and so on)

Monitors otherhost:1234 using TCP every 5s from start time

Store data in Elasticsearch at monitorcluster:9200

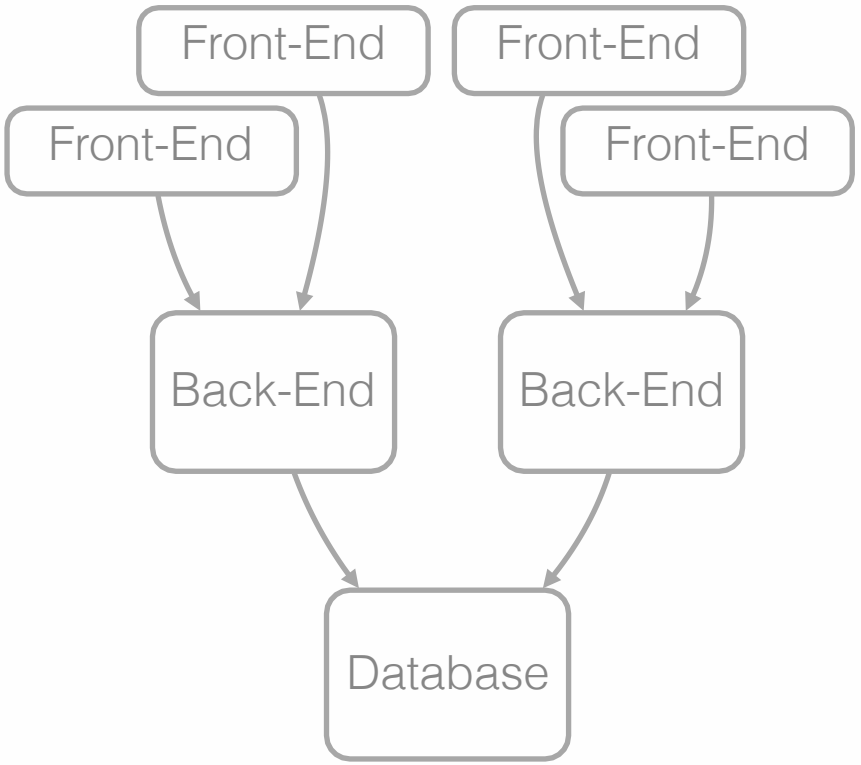
Confidential information, for internal use only

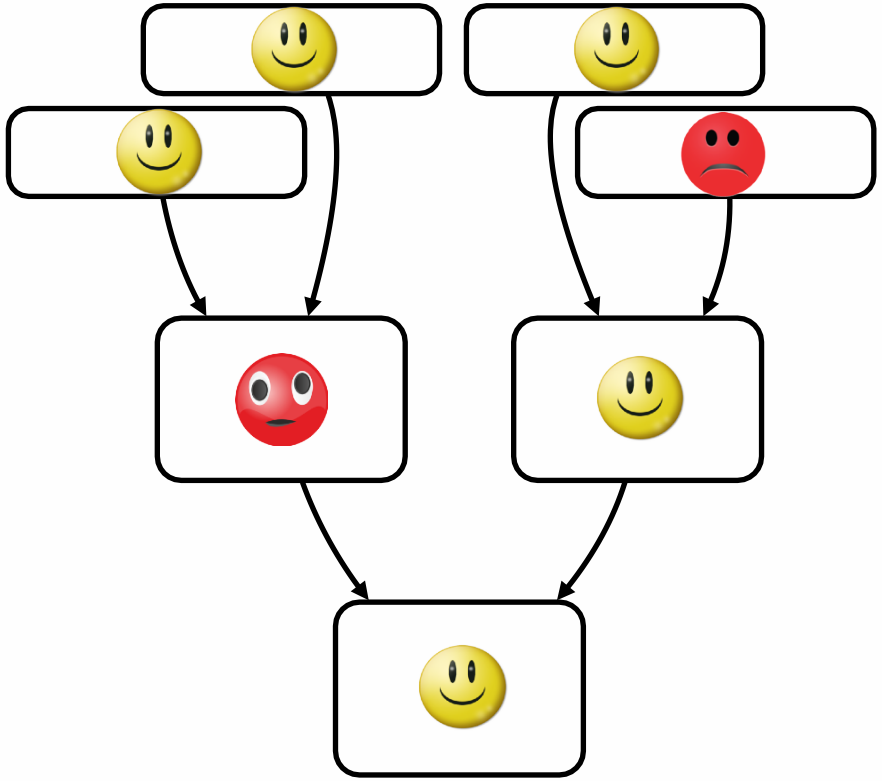
Heartbeat Config Options

- `type`
  - the type of monitor to run - one of: **icmp**, **tcp**, **http**
- `schedule`
  - cron-like expression that specifies the task schedule
- `mode`
  - **`any`**: pings only one IP address for a hostname (default)
  - **`all`**: pings all resolvable IPs for a hostname
- `timeout`
  - the total running time for each ping test (default 16s)

Confidential information, for internal use only

Why Metrics?





Confidential information, for internal use only

Metricbeat

- Collects metrics from the operating system and from services running on the server

System module

Nginx

MongoDB

MySQL

PostgreSQL

Redis

ZooKeeper

Apache

Confidential information, for internal use only

Why Metricbeat?

- Multiple metrics in one event
- Sends more than just numbers
- No aggregations when data is fetched
- Metricbeat error events

Confidential information, for internal use only

Metricbeat

- Metricbeat consists of **modules** and **metricsets**
- **Module**
  - defines the basic logic for collecting data from a specific service
  - specifies details about the service, including how to connect and how often to collect metrics
  - each module has one or more metricsets
- **Metricset**
  - periodically fetches and structures the data
  - retrieve a list of multiple related metrics in a single request

Confidential information, for internal use only

Metricbeat

- You can config multiple modules of the same type
- The period is defined per configured module
- Metricbeat reuses connections whenever possible
- If the service is not reachable, Metricbeat generates an error event
  - 10s is the default `timeout`
- Metricbeat sends the events asynchronously
  - event retrieval is not acknowledged
  - if the configured output is not available, events may be lost

Confidential information, for internal use only

Configuration

- metricbeat.yml

```
metricbeat.config.modules:
  # Glob pattern for configuration loading
  path: ${path.config}/modules.d/*.yml
```

- modules.d/

```
ubuntu@ip-172-31-0-79:~/metricbeat$ ls modules.d/
aerospike.yml.disabled      golang.yml.disabled      memcached.yml.disabled   rabbitmq.yml.disabled
apache.yml.disabled         haproxy.yml.disabled     mongodb.yml.disabled     redis.yml.disabled
ceph.yml.disabled           http.yml.disabled        mysql.yml.disabled       system.yml
couchbase.yml.disabled      jolokia.yml.disabled     nginx.yml.disabeld       vsphere.yml.disabled
docker.yml.disabled         kafka.yml.disabled       php_fpm.yml.disabled     windows.yml.disabled
dropwizard.yml.disabled     kibana.yml.disabled      postgresql.yml.disabled  zookeeper.yml.disabled
elasticsearch.yml.disabled  kubernetes.yml.disabled  prometheus.yml.disabled
ubuntu@ip-172-31-0-79:~/metricbeat$
```

system is the only one enabled by default

Confidential information, for internal use only

Configuration Example

- apache.yml.disabled

```
- module: apache
  metricsets: ["status"]
  period: 10s

  # Apache hosts
  hosts: ["http://127.0.0.1"]
```

Confidential information, for internal use only

Metricbeat System Module

- How do we monitor system resources?

CPU

Mem

diskIO

filesystem

load

network

cores

processes

Confidential information, for internal use only

modules.d/system.yml

```
- module: system
  period: 10s
  metricsets:
  - cpu
  - load
  - memory
  - network
  - process
  - process_summary
  processes: ['.*']
  process.include_top_n
  : by_cpu:    5   # include top 5 processes by CPU
    by_memory:     # include top 5 processes by memory

- module: system
  period: 1m
  metricsets:
  - filesystem
  - fsstat
  processors:
  - drop_event.when.regexp:
      system.filesystem.mount_point:
                                        '^/(sys|cgroup|proc|dev|etc|host|lib)($|/)
      '
```


Confidential information, for internal use only

Metricbeat System Module (load)

metadata

metricbeat module information

```
{
  ...
  "metricset":{
    "module": "system",
    "name": "load",
    "rtt": 77
  },
  "system": {
    "load": {
      "1": 0.58,
      "5": 0.39,
      "15": 0.14,
      "cores": 4,
      "norm": {
        "1": 0.145,
        "5": 0.0975,
        "15": 0.035
      }
    }
  }
}
```

This is an event from the **system** module that contains the **load** information and the **Round Trip Time** to collect the event was 77us

metricset data

Confidential information, for internal use only

Metricbeat System Module (cpu)

```
{
  ...
  "metricset":{
    "module": "system",
    "name": "cpu",
    "rtt": 162
  },
  "system": {
    "cpu": {
      "cores": 4
      "iowait": {
        "pct": 0
      },
      "total": {
        "pct": 0
      },
     "user": {
        "pct": 0
      },
      "system": {
        "pct": 0
      },
      ...
} } }
```

This is an event from the **system** module that contains a **cpu** information and the **Round Trip Time** to collect the event was 162us

cpu data

Confidential information, for internal use only

Metricbeat System Module (process)

```
{
  ...
  "metricset":{
    "module": "system",
    "name": "process",
    "rtt": 30105
  }
  "system": {
    "process": {
      "cmdline":
             "/sbin/init",
      "cpu": {
        ...
      },
      "ppid": 0,
      "username": "root",
      "pid": 1,
      "pgid": 1,
      "memory": {
        ...
      },
      "name": "systemd",
      "state": "sleeping",
      "cgroup": {
        ...
} } } }
```

This is an event from the **system** module that contains a **process** information and the **Round Trip Time** to collect the event was ~30ms

process data

Confidential information, for internal use only

Metricbeat System Module (filesystem)


```
{
  ...
  "metricset":{
    "module": "system",
    "name": "filesystem",
    "rtt": 597
  },
  "system": {
    "filesystem": {
      "available": 1534222336,
      "free_files": 1924385,
      "used": {
        "pct": 0.0271,
        "bytes": 42668032
      },
      "files": 1924910,
      "type": "tmpfs",
      "mount_point": "/run",
      "free": 1534222336,
      "device_name": "tmpfs",
      "total": 1576890368
    }
  }
}
```


This is an event from the **system** module that contains a **filesystem** information and the **Round Trip Time** to collect the event was 597us

filesystem data

Confidential information, for internal use only

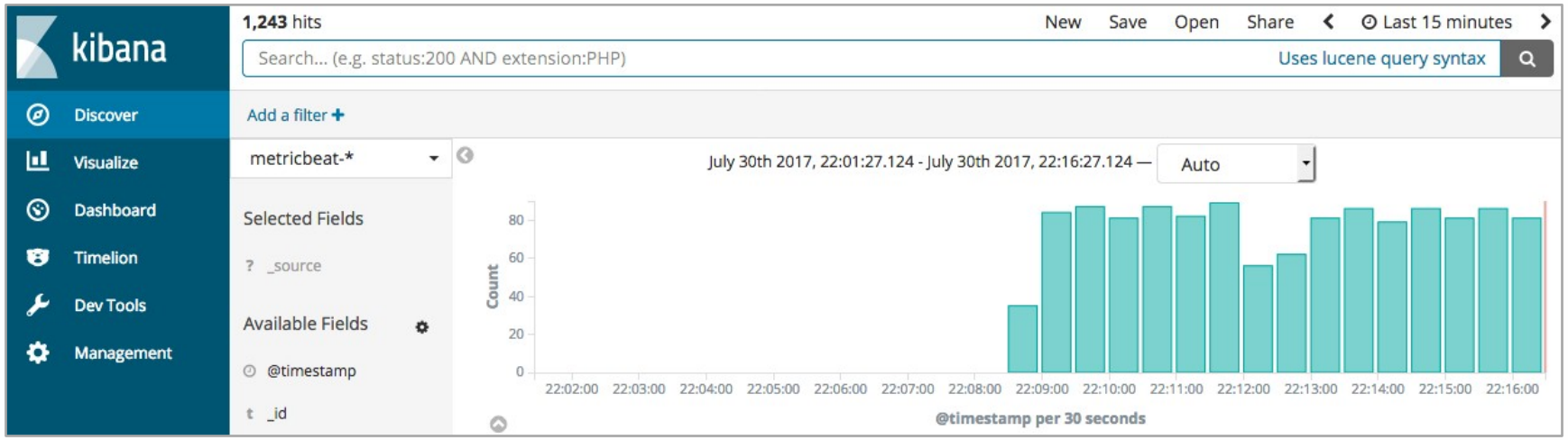
Metricbeat Dashboards

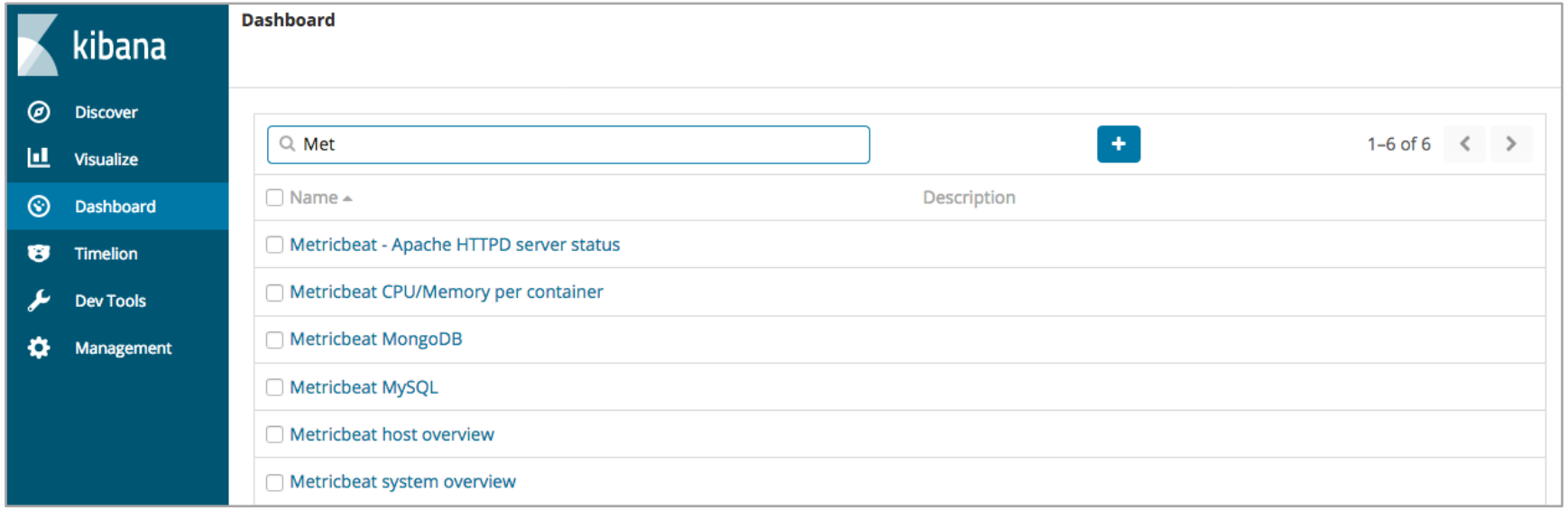
- Easy to start monitoring your servers in Kibana
- Pre-created Metricbeat dashboards
- Customize them to meet your needs
- Load dashboards:
  - metricbeat.yml `setup.dashboards.enabled: true`

    OR
  - command `./metricbeat setup --dashboards`
- It also creates a `metricbeat-*` index pattern.

Confidential information, for internal use only

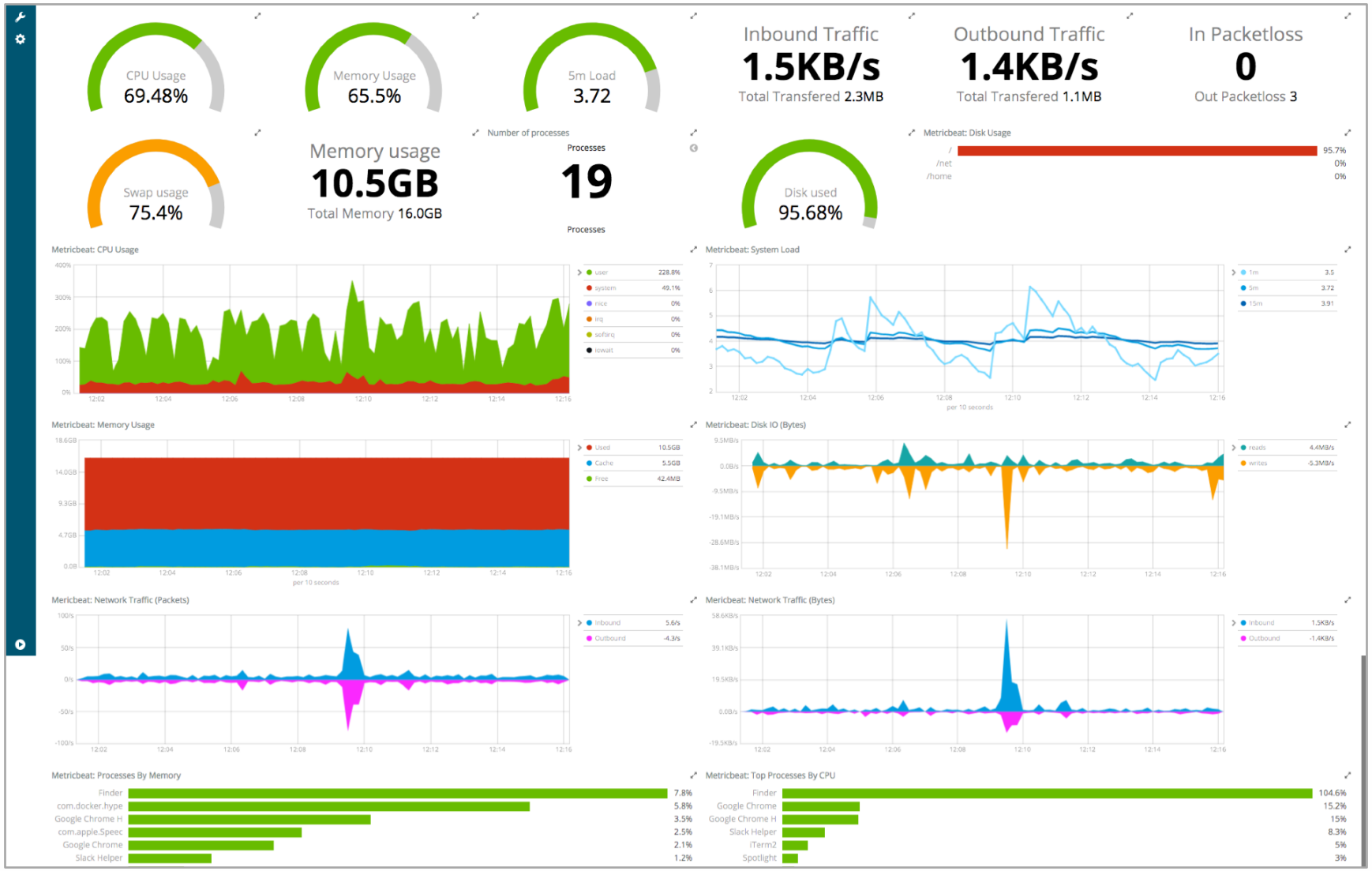
Metricbeat Dashboards





Confidential information, for internal use only

Metricbeat Dashboards



Confidential information, for internal use only

Filebeat

- Installed as an agent on all servers from which you want to gather logs
- Monitors log directories or specific log files
- Tails these files
- Ship log entries to:
  - Elasticsearch
  - Logstash
  - Kafka
  - Redis

Confidential information, for internal use only

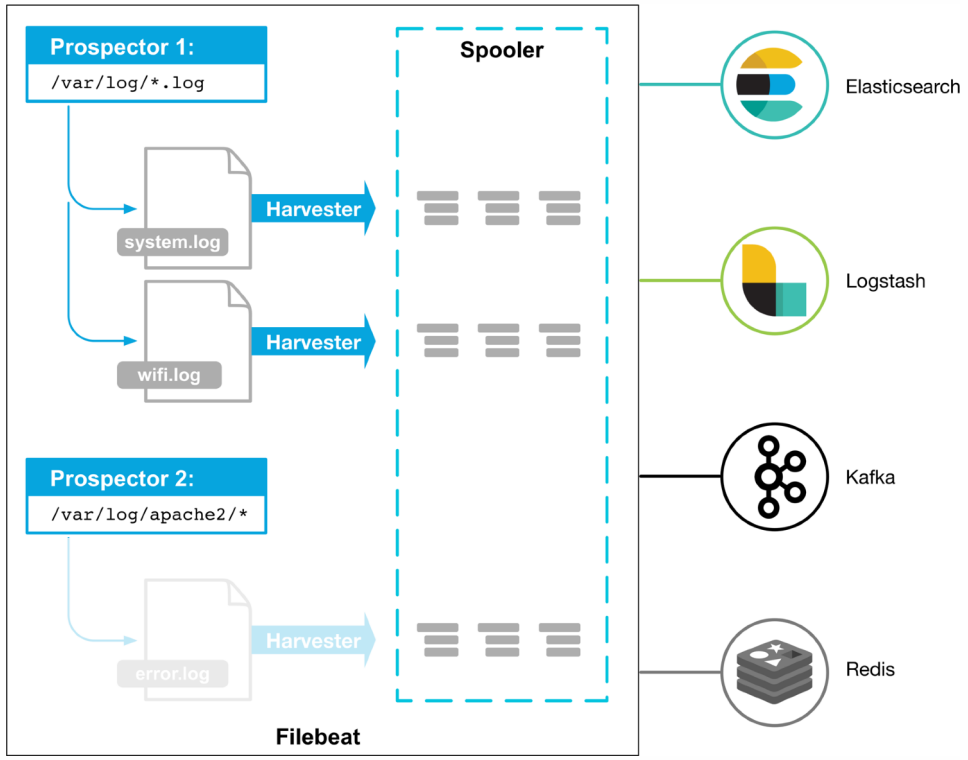
Why Filebeat?

- Lightweight executable
  - no JVM
    - Filebeat runs as a binary so no runtime is needed
  - easy to deploy across many architectures
  - scale your data shippers independently of your processors

Confidential information, for internal use only

Prospectors

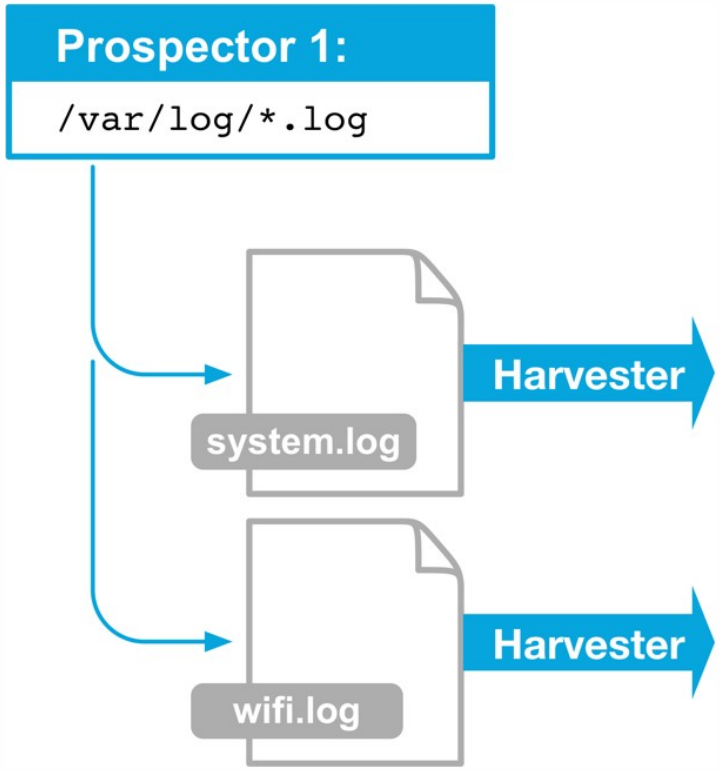
- Each instance of Filebeat can be configured with one or more ***prospectors***
- Each prospector can be configured to monitor one or more file paths



Confidential information, for internal use only

Harvesters

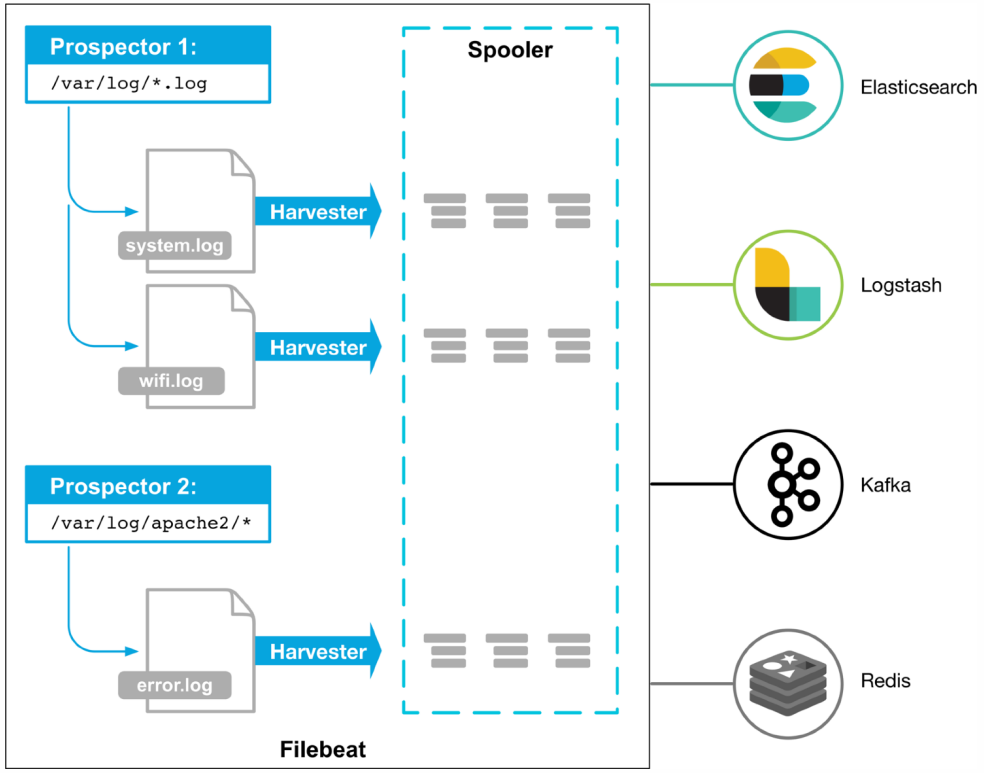
- For each file that a prospector locates, Filebeat starts a ***harvester***
- Each harvester reads a single file for new content



Confidential information, for internal use only

Spooler

- Each harvester sends the events to the ***spooler***
- The spooler aggregates the events and sends the aggregated data to the output that has been configured



Confidential information, for internal use only

Filebeat Configuration

- Default configuration file is **filebeat.yml**
- Sample configuration:

```
filebeat.prospectors:
- input_type: log
  paths:
    - /var/log/*.log

output.elasticsearch:
  hosts: ["localhost:9200"]
```

Confidential information, for internal use only

Prospector Include/Exclude

- Besides processors, Filebeat allows you to use regular expression to filter exported data:

```
filebeat.prospectors
- paths:
    - /var/log/myapp/*.log
  include_lines: ['^ERR', '^WARN']
```

export any lin es that start with "ERR" or "WARN"

```
filebeat.prospectors
- paths:
    - /var/log/myapp/*.log
  exclude_lines: ['^DBG']
```

drop any lines that start with "DBG"

```
filebeat.prospectors
- paths:
    - /var/log/myapp/*
  exclude_files: ['\.gz$']
```

ignore all the f iles that the name ends in .gz

Confidential information, for internal use only

What is Packetbeat?

- Lightweight data shipper that captures data from application layer protocols
- Capture network traffic with minimal to no overhead
- Not in the application request/response path so no risk to breaking running applications
- Can record network traffic at OS or hardware level

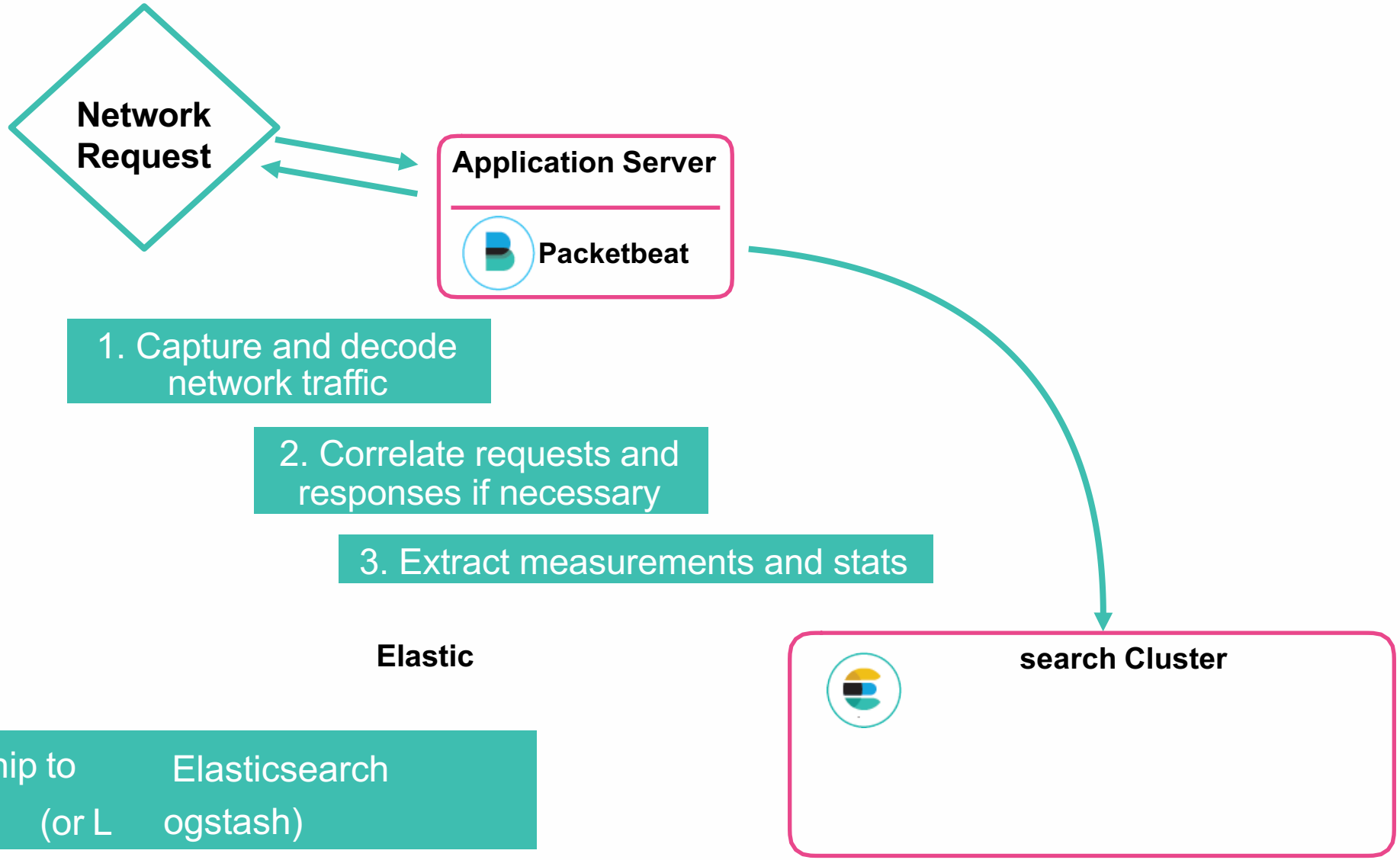
Confidential information, for internal use only

Packetbeat Use Cases

- Security
  - Intrusion detection
  - Protocol violations
- Troubleshooting
  - Application network communications
  - Network issues
- Performance Analysis
  - Record network performance and detect bottlenecks
  - Understand how much time is spent communicating between networked processes and applications

Confidential information, for internal use only

How Packetbeat Works



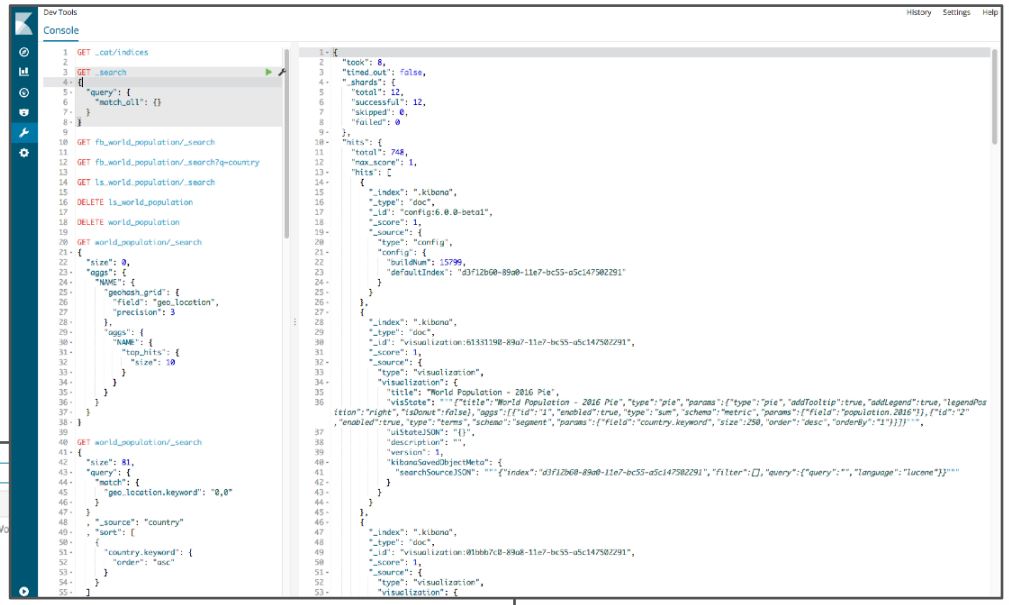
Confidential information, for internal use only

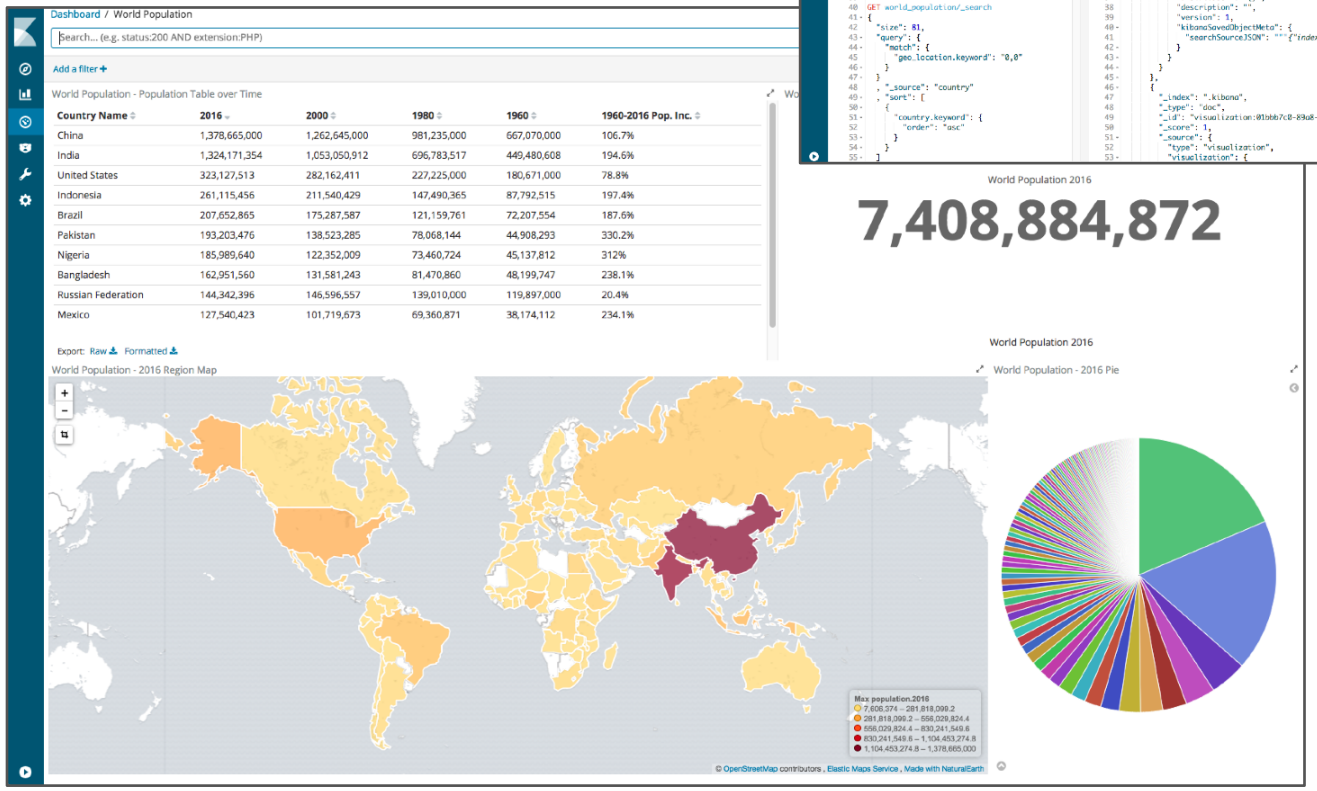
Kibana

Confidential information, for internal use only

Kibana

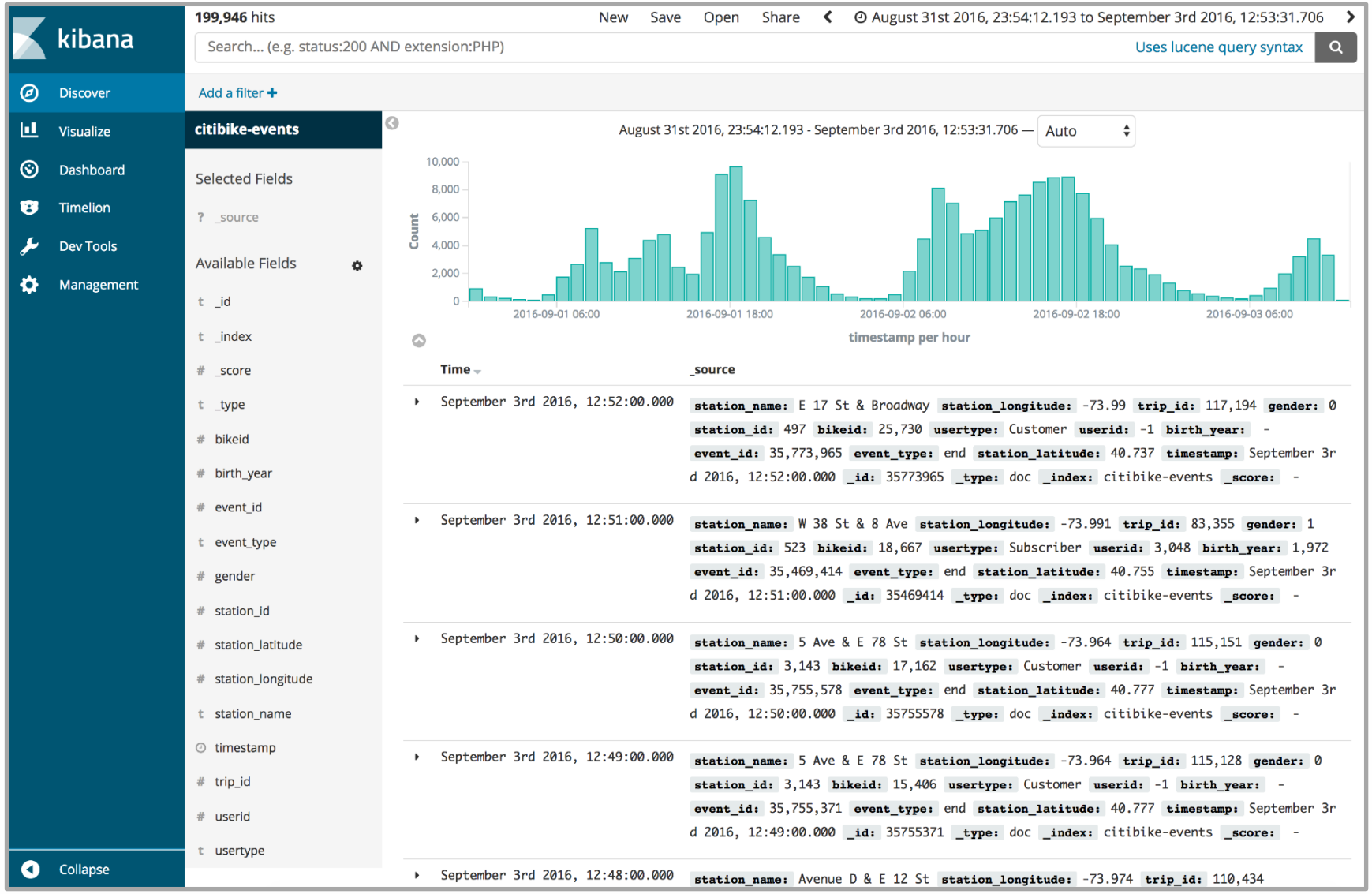
- Data Visualization
- Stack Configuration
- Elastic Stack UI





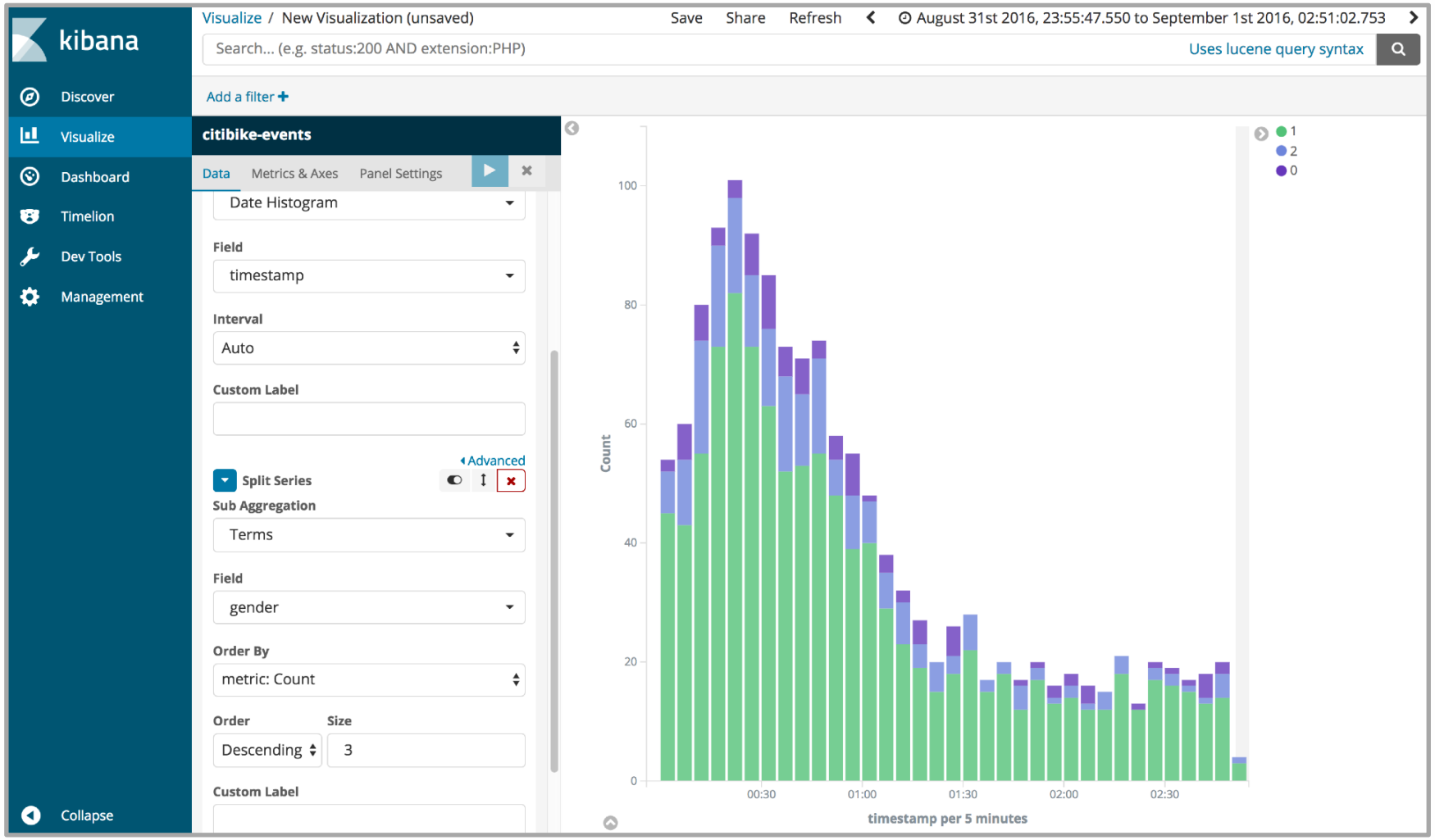
Confidential information, for internal use only

Discover



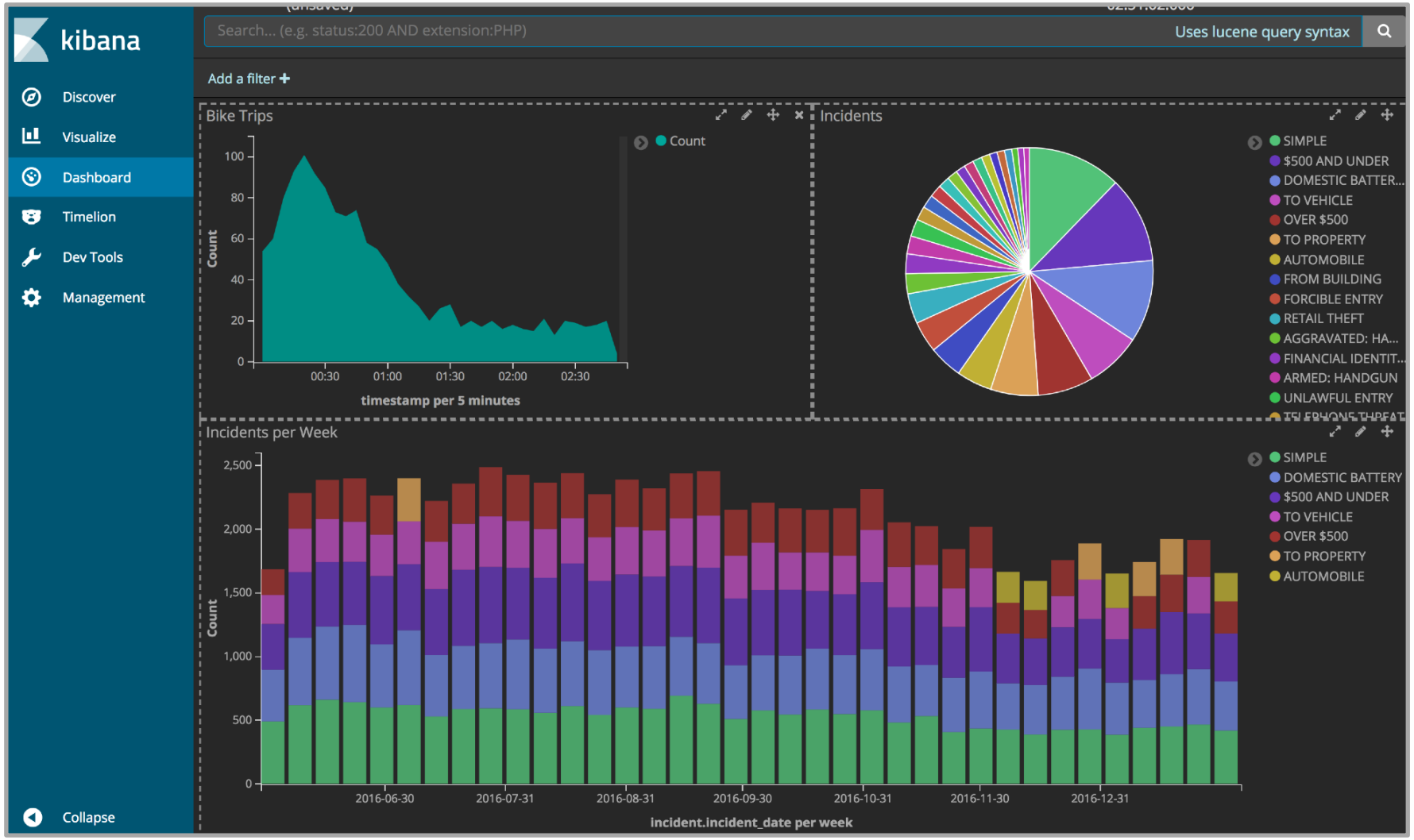
Confidential information, for internal use only

Visualization



Confidential information, for internal use only

Dashboard



Confidential information, for internal use only

Index Pattern

kibana

Discover
Visualize
Dashboard
Timelion
Dev Tools
Management

Collapse

Management / Kibana

Index Patterns Saved Objects Advanced Settings

+ Create Index Pattern

citibike-events
crimes*

citibike-events

Time Filter field name: timestamp

This page lists every field in the **citibike-events** index and the field's associated core type as recorded by Elasticsearch. While this list allows you to view the core type of each field, changing field types must be done using Elasticsearch's Mapping API

fields (19) | scripted fields (0) | source filters (0)

Filter | All field types

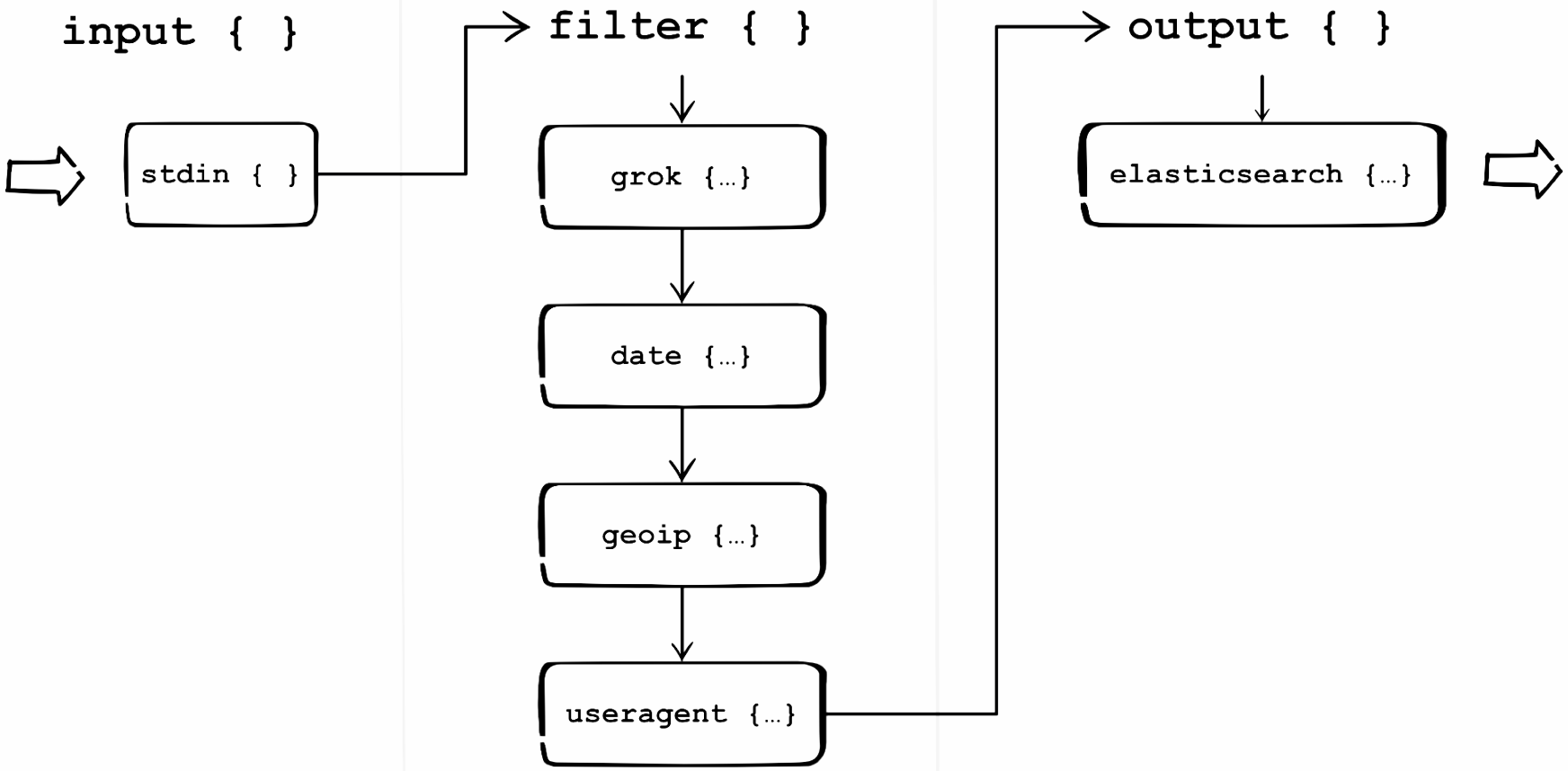
name	type	format	searchable	aggregatable	excluded	controls
_id	string		✔	✔		
_index	string		✔	✔		
_score	number					
_source	_source					
_type	string		✔	✔		
bikeid	number		✔	✔		
birth_year	number		✔	✔		
event_id	number		✔	✔		
event_type	string		✔	✔		
gender	number		✔	✔		
station_id	number		✔	✔		
station_latitude	number		✔	✔		
station_longitude	number		✔	✔		
station_name	string		✔			

Confidential information, for internal use only

Logstash

Confidential information, for internal use only

Logstash Pipeline



Confidential information, for internal use only

Logstash Terms and Definitions

- Event
  - any piece of data that has a timestamp and some data
- Plugins
  - working units that handle events

Confidential information, for internal use only

Plugins

- Inputs
  - capture a data stream and turn it into events
- Filters
  - process events
- Outputs
  - send events to destinations
- Codec (operate as part of an Input or Output filter)
  - changes the data representation of an event

Confidential information, for internal use only

Input Configuration

```
input {
  stdin {}
}
```

plugin name

```
input {
  beats {
    port => 5044
  }
}
```

value

setting

```
input {
  twitter {
    consumer_key => "consumer_key"
    consumer_secret => "consumer_secret"
    oauth_token => "access_token"
    oauth_token_secret =>
    "access_token_secret"  keywords => [
    "elastic", "elastic stack" ]
  }
}
```


Confidential information, for internal use only

Filter Configuration

```
filter {}
```

plugin name

```
filter {
  mutate {
    remove_field => "credit_card"
  }
}
```

value

setting

```
filter {
  csv {
    source => "message"
    columns =>
    ["country","country_code","population"]
    separator => ";"
  }
}
```


Confidential information, for internal use only

Output Configuration

```
output  {
  stdout {}
}
```

plugin name

```
output {
  elasticsearch {
    hosts => [ "my_host1", "my_host2" ]
  }
}
```

value

setting

```
output {
  pagerduty {
      service_key => "my_pagerduty_api_key"
      details => {
        "message" => "Error in production!"
      }
    }
}
```

Confidential information, for internal use only

Simple Configuration

```
input {
  stdin { }
}

filter {}

output {
  stdout { codec => rubydebug }
}
```

Good to see if logstash executes without any problems in the machine

```
input {
  stdin { }
}

filter { }

output {
  elasticsearch {}
}
```

Good to see if logstash can write to Elasticsearch.

```
hosts => [ 'localhost:9200' ]
index => 'logstash-yyyy.MM.dd'
```

Confidential information, for internal use only

More Advanced Configuration

```
input {
  beats { port => 5044 }
}

filter {
  csv {
    source => "message"
    columns => ["country","country_code",  "dummy1", "population"]
  }

  if [country] == "Country" { drop {} }

  mutate { remove_field => "dummy1" }
}

output {
  stdout { codec => dots }

  elasticsearch {
    hosts => ["my_host1", "my_host2", "my_host3"]
    index => "ls_world_population"
  }
}
```

receives events from Beats

extracts 4 fields from the csv line

drops the header line

removes useless field

prints a dot for every processed event

outputs to a specific Elasticsearch index using round-robin

Confidential information, for internal use only