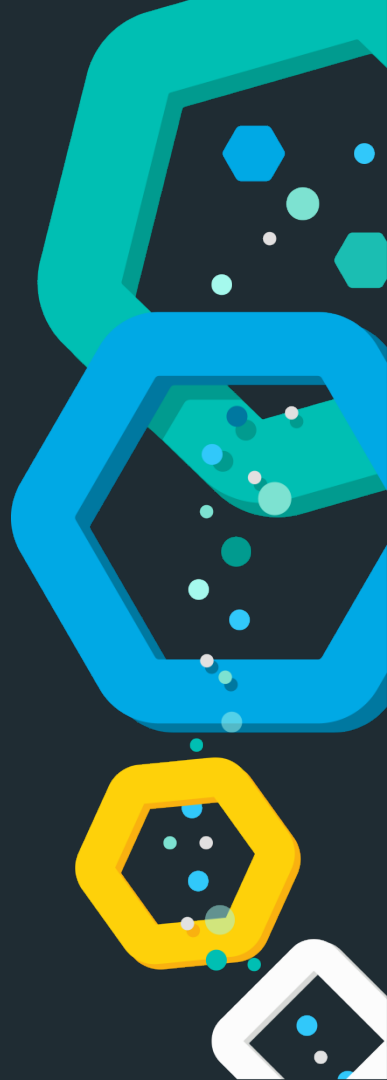




Ingest

David Pilato, Developer | Evangelist
Paris, 31 Janvier 2017



Data Ingestion

The process of collecting and importing data for immediate use in a datastore



simple things should be simple



Ingest Technologies

Elasticsearch

APIs
Ingest Node



Beats

Lightweight Data
Shippers



Logstash

Centralized Data
Collection Engine



ES-Hadoop

Hadoop
Ecosystem
Connector



Elastic Ingestion Technologies

Dev Tools

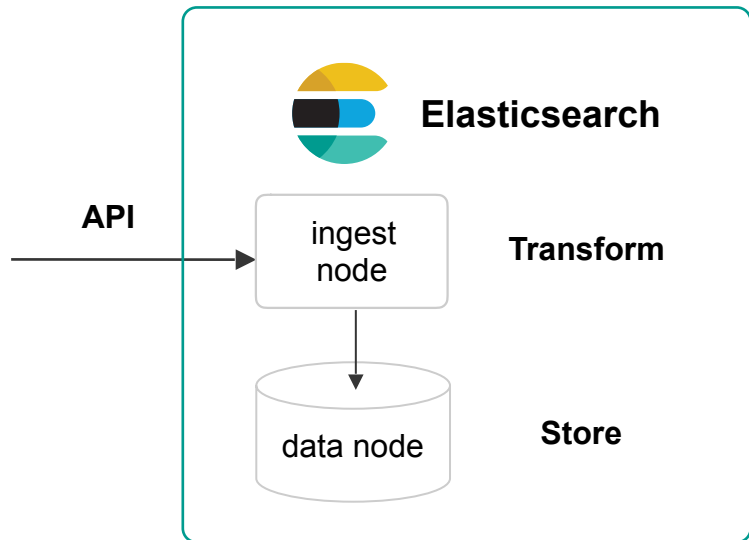
Console Search Profiler Grok Debugger

```
1 PUT _ingest/pipeline/my-pipeline-id
2 {
3   "description": "describe pipeline",
4   "processors": [
5     {
6       "set": {
7         "field": "foo",
8         "value": "bar"
9       }
10    }
11  ]
12 }
```

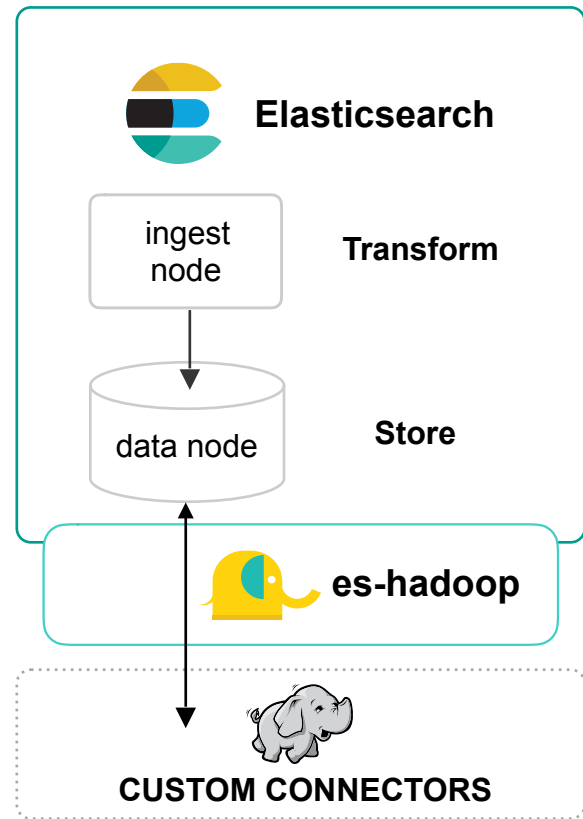
1 {
2 "acknowledged": true
3 }

Ingest Node

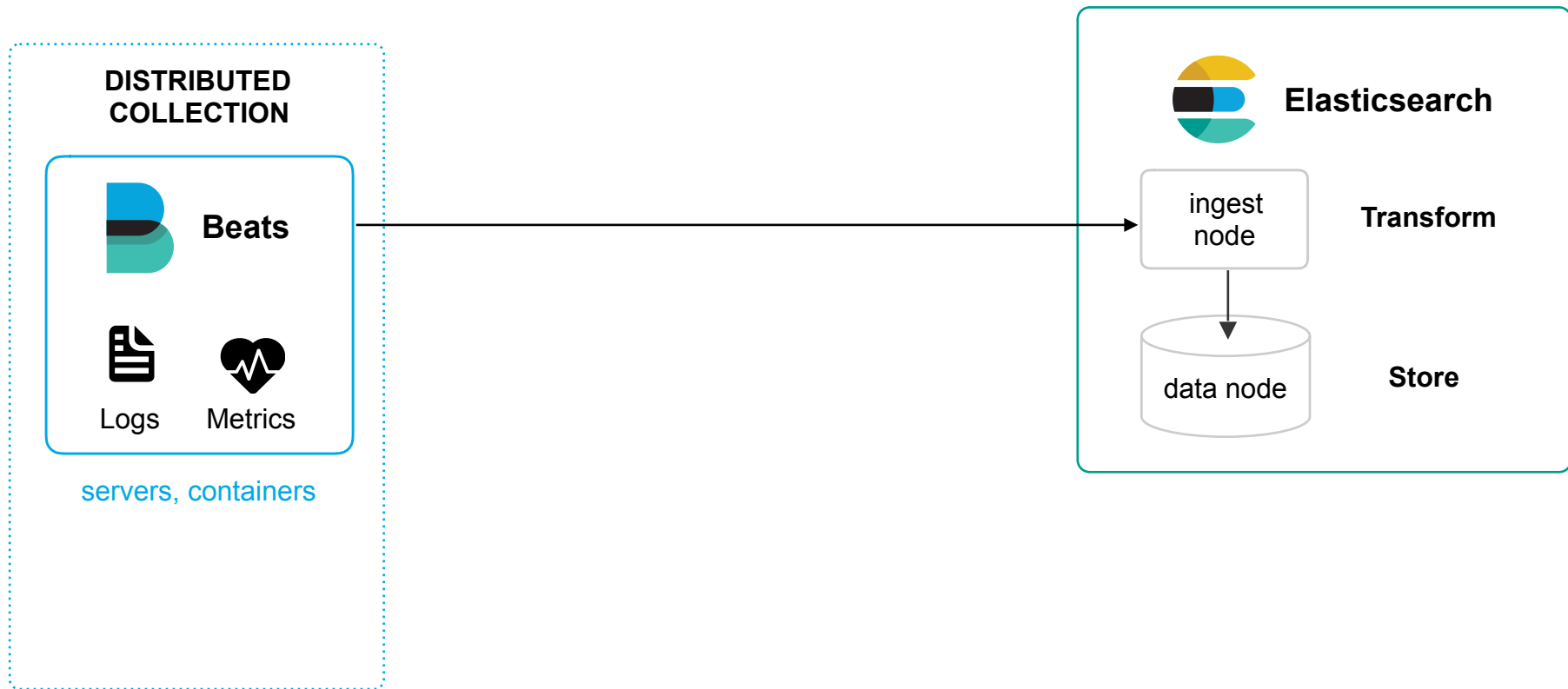
- Pipeline Definition
- + Ingest APIs
 - Accessing Data in Pipelines
 - Handling Failures in Pipelines
- Processors
 - Append Processor
 - Convert Processor
 - Date Processor
 - Date Index Name Processor
 - Fail Processor
 - Foreach Processor
 - Grok Processor
 - Gsub Processor
 - Join Processor
 - JSON Processor
 - KV Processor
 - Lowercase Processor
 - Remove Processor
 - Rename Processor
 - Script Processor
 - Set Processor
 - Split Processor
 - Sort Processor
 - Trim Processor
 - Uppercase Processor
 - Dot Expander Processor



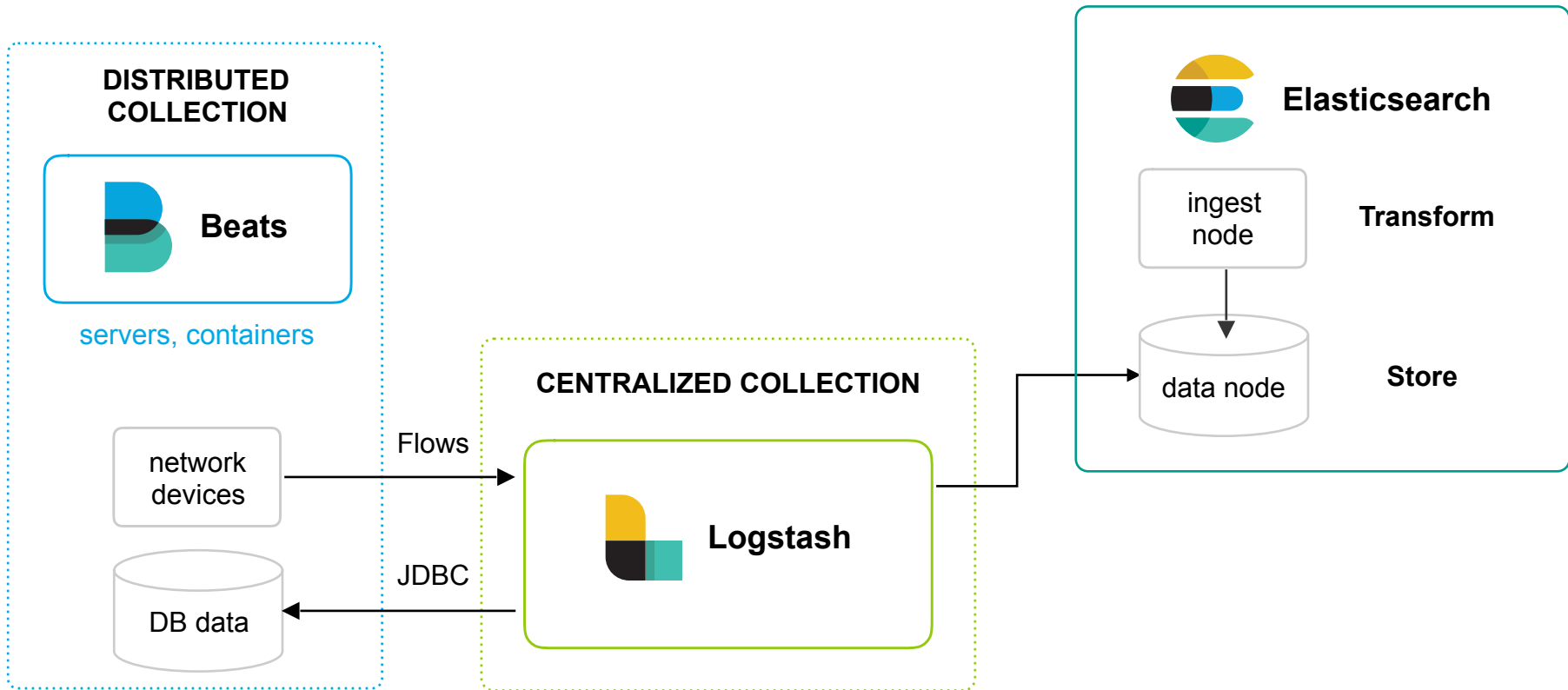
Elastic Ingestion Technologies



Elastic Ingestion Technologies



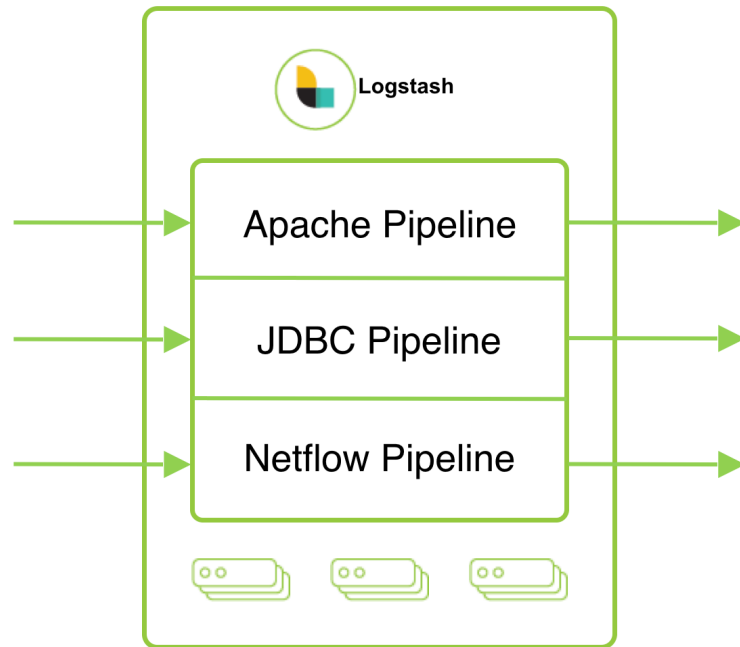
Elastic Ingestion Technologies



Ingestion Architecture

Scalable and robust centralized ETL

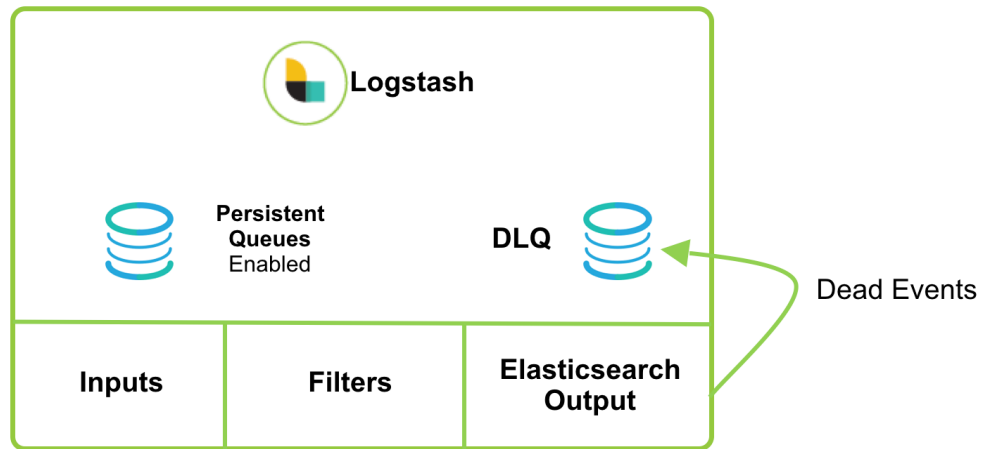
- Java event rewrite
- Multiple pipelines



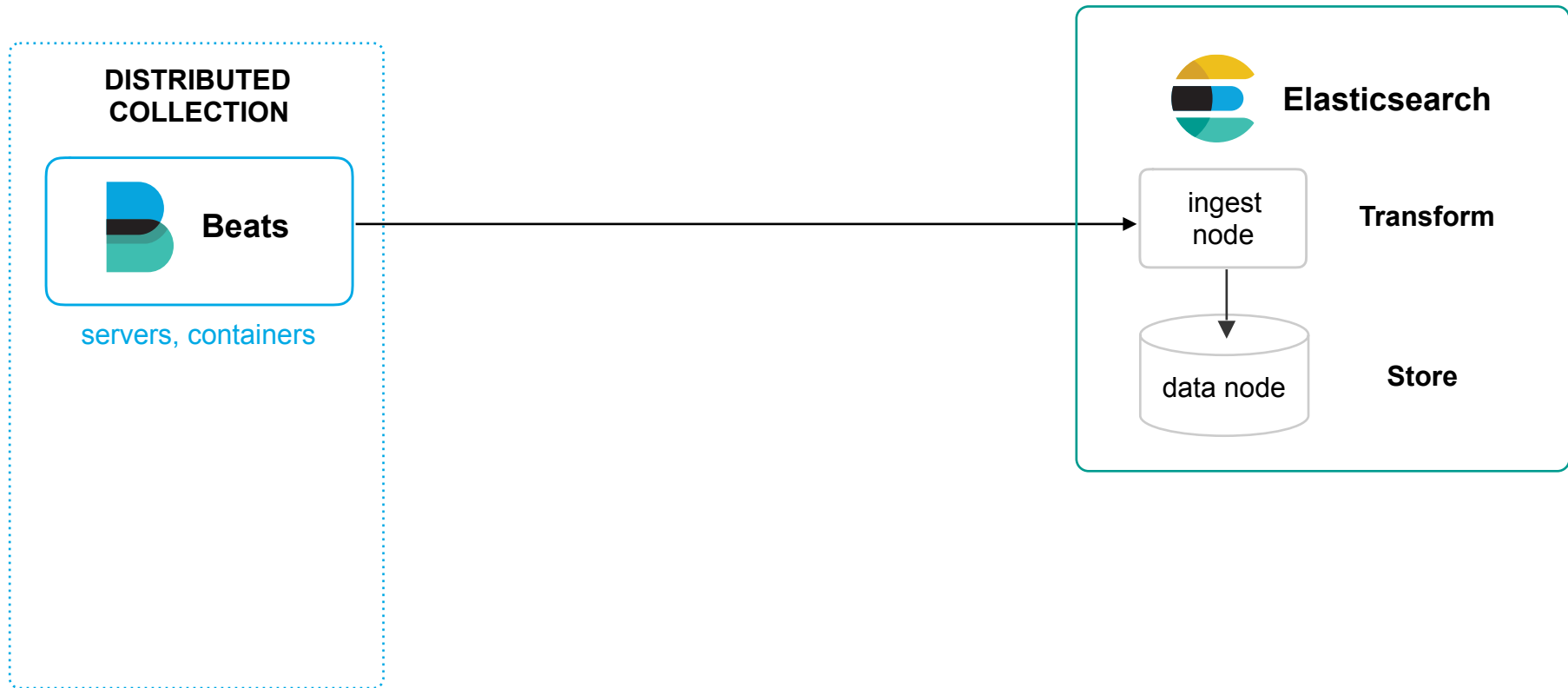
Ingestion Architecture

Scalable and robust centralized ETL

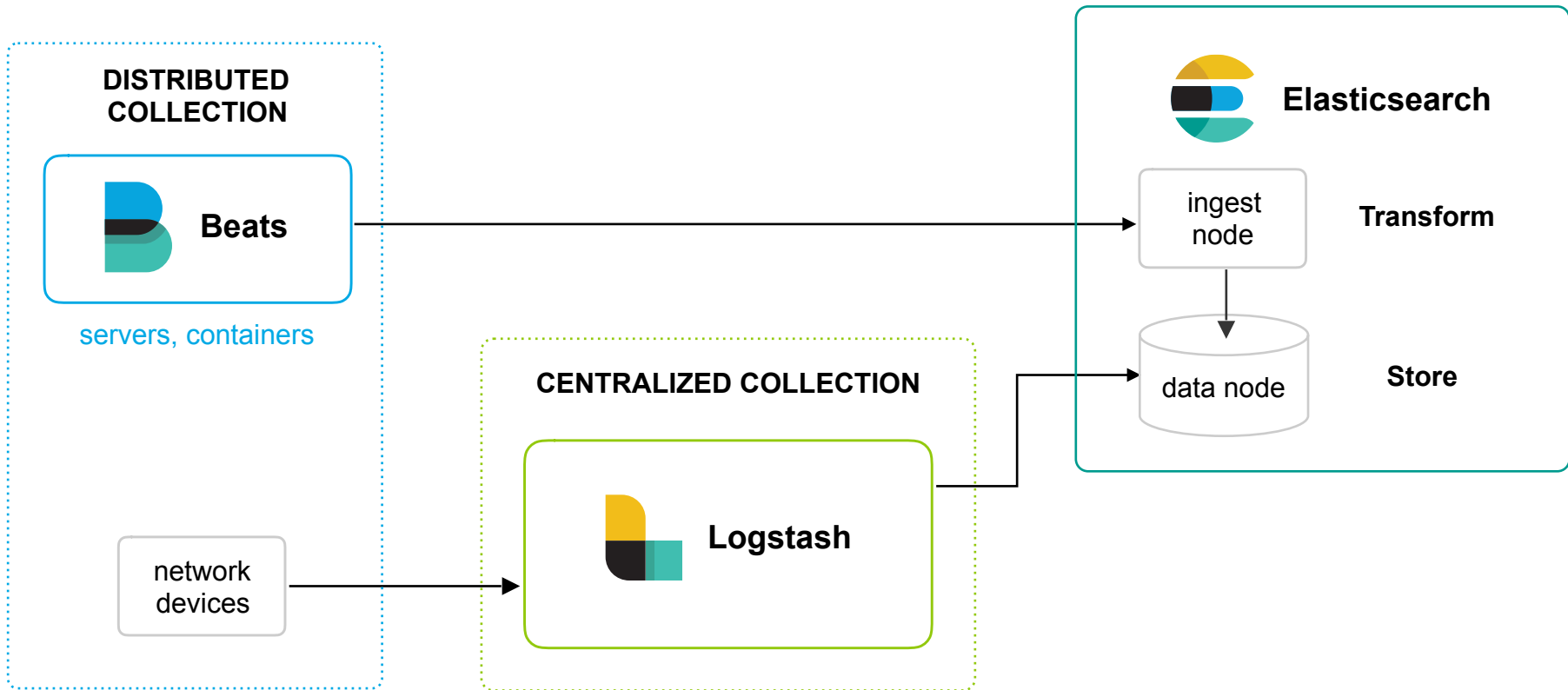
- Persistent queues
- Dead letter queues



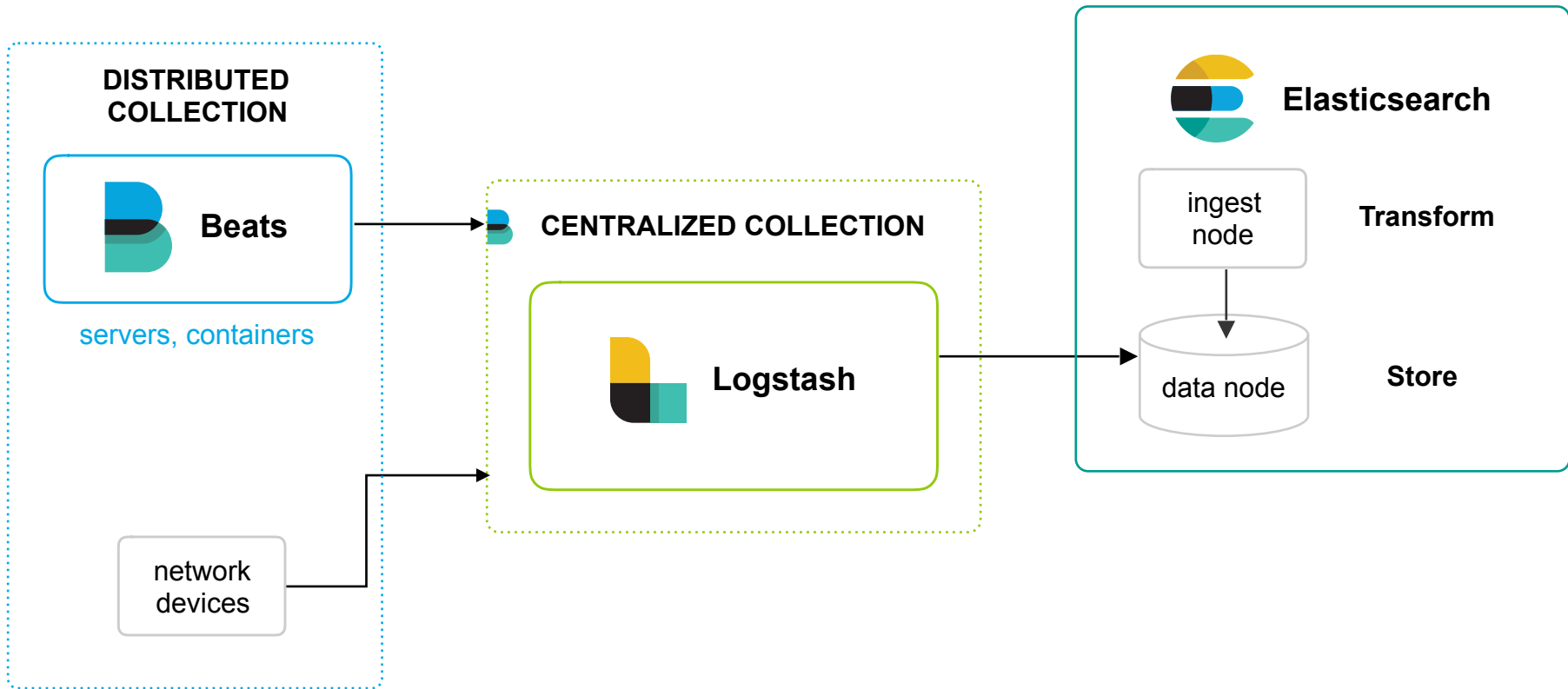
Cooperative Ingestion



Elastic Ingestion Technologies

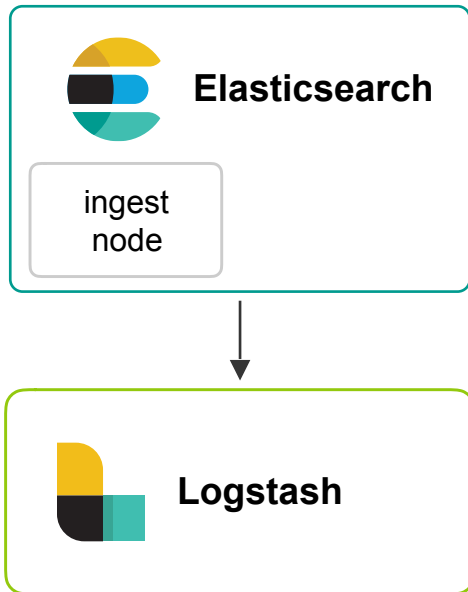


Elastic Ingestion Technologies



Easy migration between ingest technologies

Ingest Node to
Logstash
conversion tool



```
$ bin/ingest-convert.sh --input=file:///tmp/ingest-nginx.json --output=file:///tmp/logstash-nginx.conf
$ cat /tmp/logstash-nginx.conf
filter {
  grok {
    match => {
      "request" => "%{WORD:request_action} %{DATA:request1} HTTP/%{NUMBER:http_version}"
    }
  }

  geoip {
    source => "remote_ip"
    target => "geoip"
  }

  date {
    match => [
      "time",
      "dd/MMM/YYYY:HH:mm:ss Z"
    ]
    target => "@timestamp"
    locale => "ENGLISH"
  }
}
output {
  elasticsearch {
    hosts => "localhost"
  }
}
$
```



Use Cases & Data Sources



Use Cases & Data Sources

Logging

Common Log Formats

System
Web Servers
Queues

Metrics

Turnkey Monitoring

Infrastructure
Containers
Databases

Security

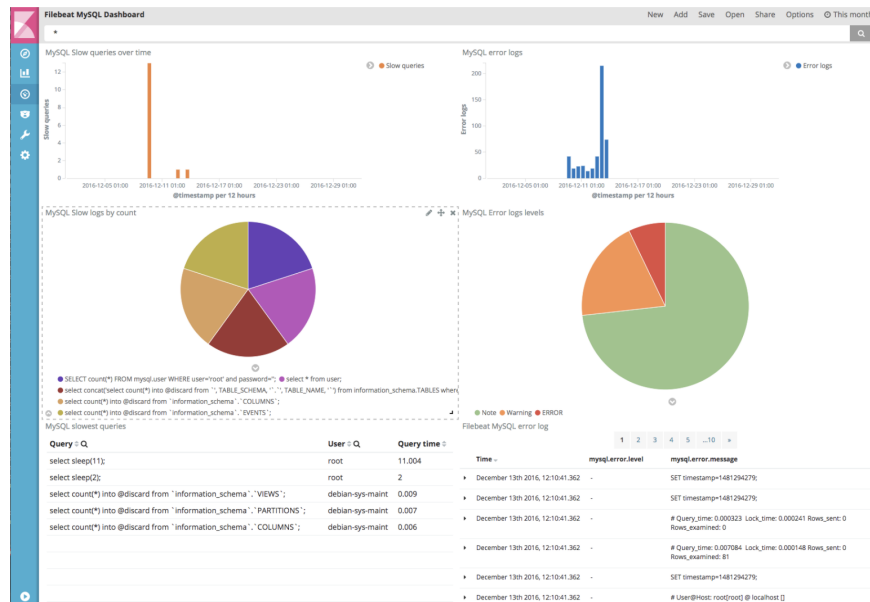
SecOps Dashboards

Audit
Firewalls, IDS/IPS
SIEM Augmentation

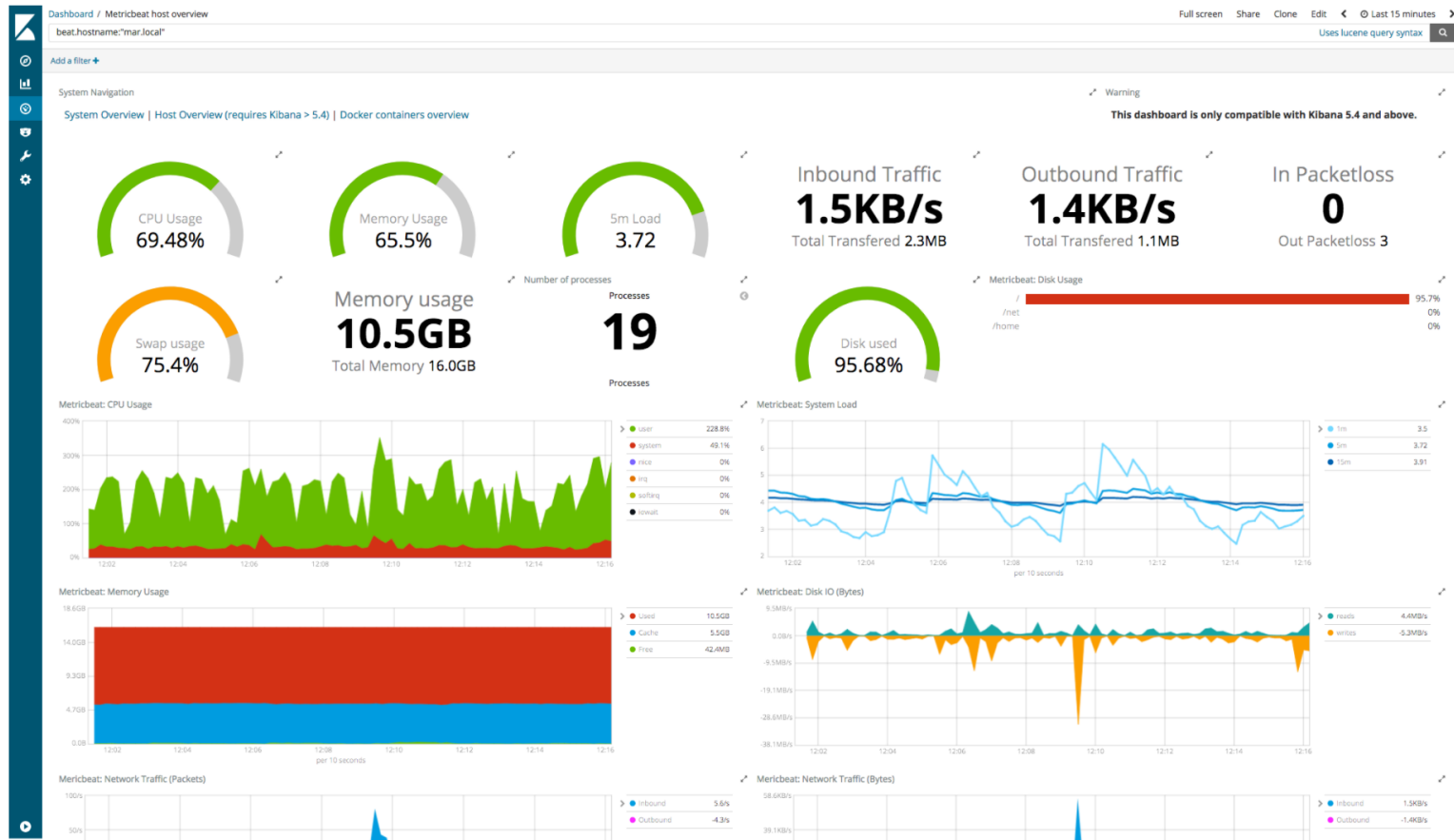
Modules: Data sources made easy

- Collect specific type of data
- Parse and enrich it
- Default dashboards, alerts, ML jobs

`./filebeat -e -modules=system -setup`



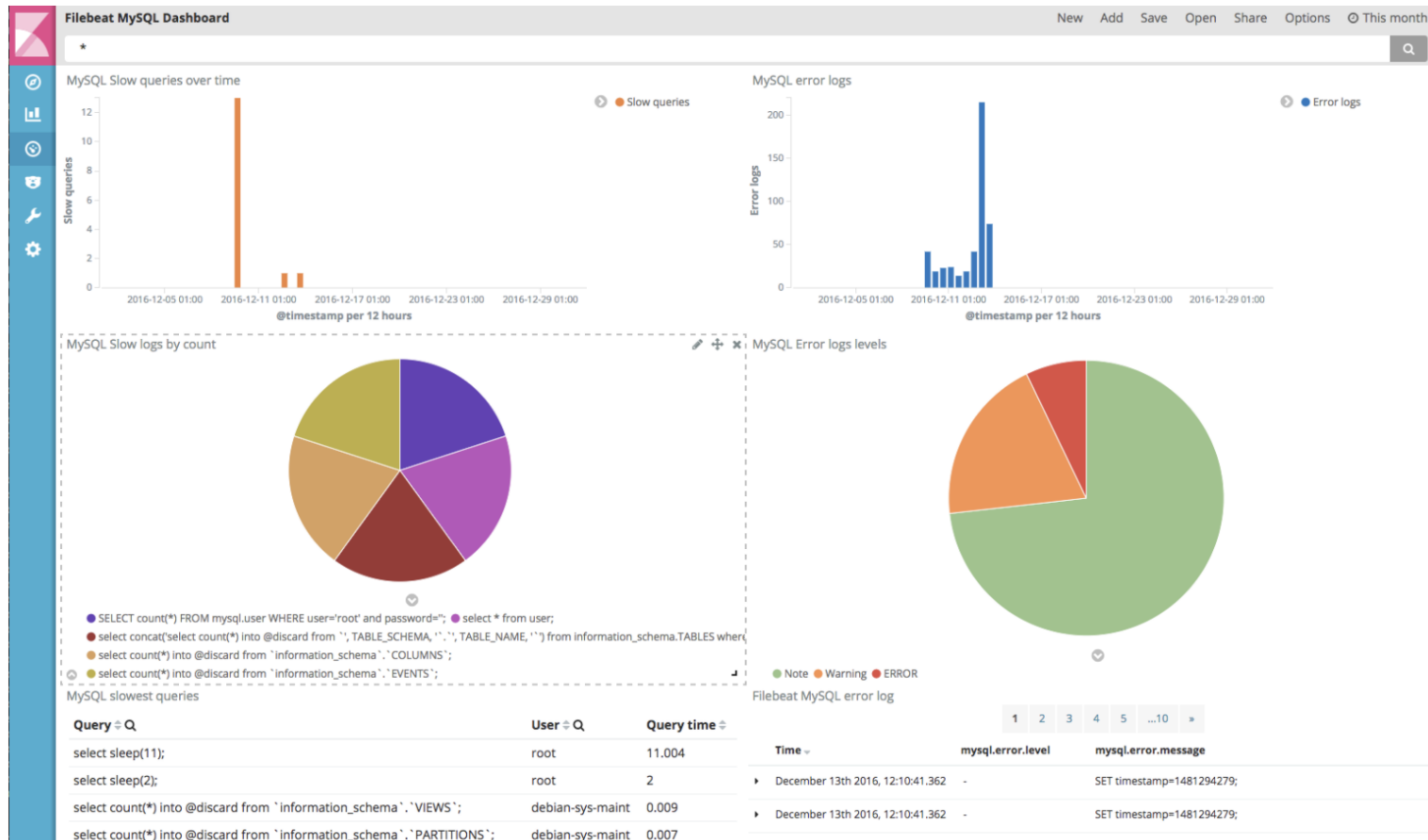
Metricbeat modules (introduced in 5.0)



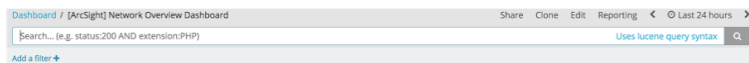
Aerospike
Apache
Ceph
Couchbase
Docker
Dropwizard
Elasticsearch
Golang
Graphite
HAProxy
HTTP
Jolokia
Kafka
Kibana
Kubernetes
Memcached
MongoDB
MySQL
Nginx
PHP_FPM
PostgreSQL
Prometheus
RabbitMQ
Redis
System
vSphere
Windows
ZooKeeper

Filebeat modules (introduced in 5.3)

Apache2
Auditd
Icinga
Kafka
MySQL
Nginx
PostgreSQL
Redis
System



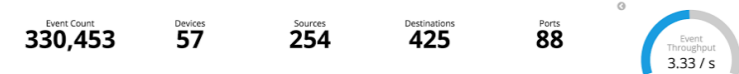
Logstash modules (introduced in 5.6)



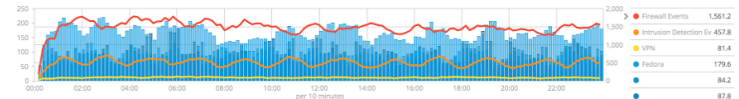
ArcSight: Dashboard Navigation

Network Overview | Network Suspicious Activity | Endpoint Overview | Endpoint OS Activity | DNS Overview

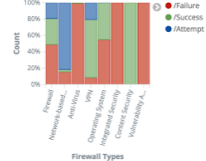
Device Metrics Overview [ArcSight]



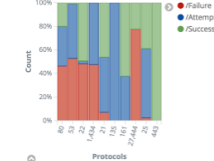
Events by Source [ArcSight]



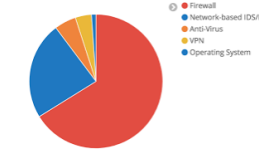
Outcome by Device Type [ArcSight]



Destination Ports by Outcome [ArcSight]



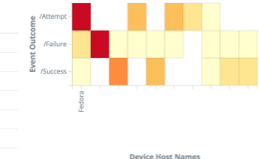
Device Type Breakdown [ArcSight]



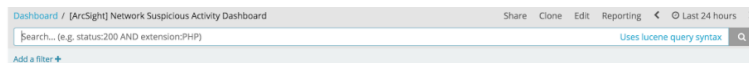
Top 10 Devices by Bandwidth [ArcSight]

Device	Source(s)	Destination(s)	Destination Ports	Bandwidth (Incoming)	Bandwidth (Outgoing)
Fedora	54	50	17	48,543,771	775,105,603
	1	96	1	1,827,840	1,763,328
	3	5	3	275,800	264,950
	1	0	1	775	456
	1	1	1	128	128
	1	2	1	0	0
	1	0	0	0	0

Top 10 Devices by Outcome [ArcSight]



Top 10 Application Protocols [ArcSight]



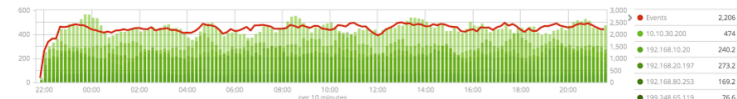
ArcSight: Dashboard Navigation

Network Overview | Network Suspicious Activity | Endpoint Overview | Endpoint OS Activity | DNS Overview

Device Metrics Overview [ArcSight]



Events by Source Addresses [ArcSight]



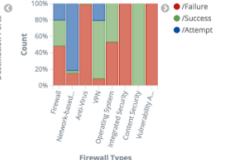
Events by Severity [ArcSight]



Unique Destinations and Ports by Source [ArcSight]



Outcome by Device Type [ArcSight]



Top 5 Sources by Destination Addresses [ArcSight]

Source Address	Destination Addresses	Event Count
10.0.111.46	96	2,688
209.128.98.73	96	2,591
193.115.143.134	21	1,870
10.0.112.6	9	1,176
10.0.112.24	9	3,386

Top 5 Sources by Destination Ports [ArcSight]

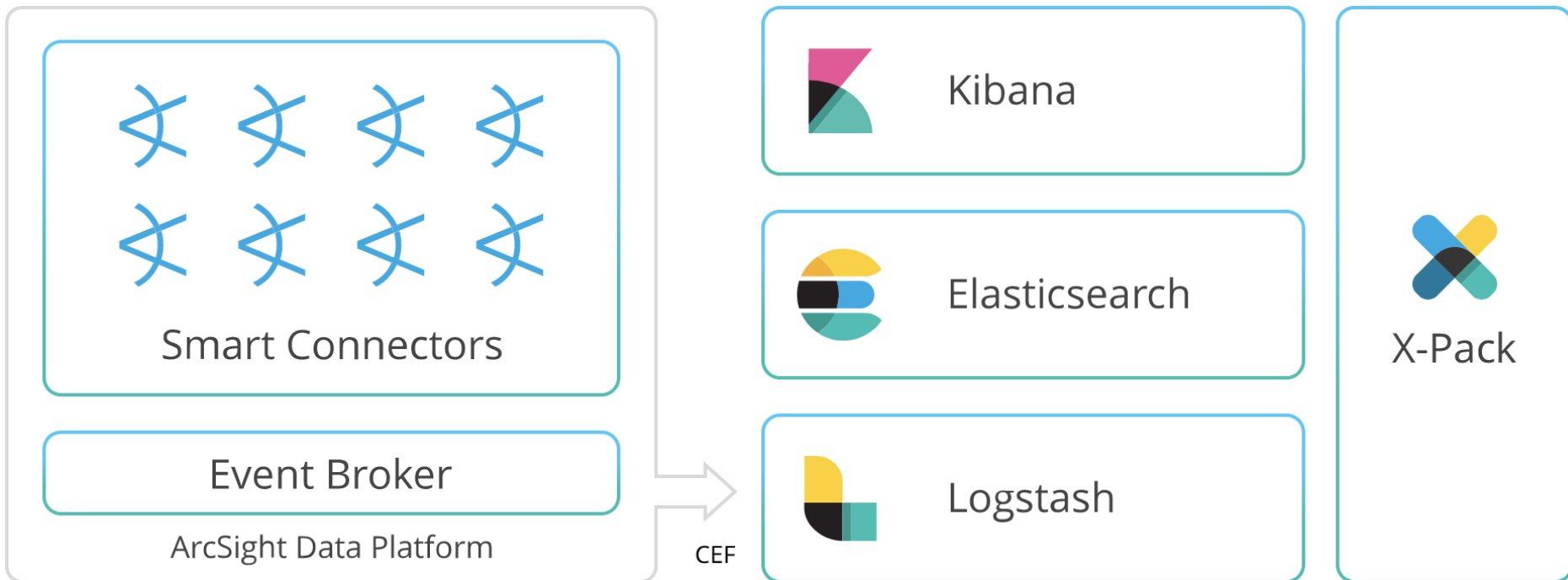
Source Address	Destination Ports	Event Count
10.0.112.251	10	360
199.248.65.119	9	13,009
10.0.112.208	7	249
10.0.112.6	6	1,176
111.111.250.143	5	9,097

Top 10 Destination Ports [ArcSight]



ArcSight
Netflow

ArcSight Module (Introduced in 5.6)



DEMO

Logging Data Sources

FILEBEAT



WINLOGBEAT



Infrastructure

System

- Linux / MacOS
- Windows Events

Containers

- Docker (6.0)
- Kubernetes (6.0)

Applications

Databases

- MySQL
- PostgreSQL (6.1)

Queues

- Kafka (6.1)
- Redis (6.0)

Web servers

- Apache
- Nginx

Other

- HAProxy*
- Zookeeper*

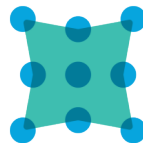
** Near-term roadmap*

Metrics & Event Data

METRICBEAT



PACKETBEAT



LOGSTASH



Infrastructure

System

- Linux
- MacOS
- Windows
- Perfmon (6.0)
- WMI*

Containers

- Docker
- Kubernetes (6.0)

Virtualization

- vSphere (6.0)

Cloud

- AWS
- GCP
- Azure*
- DigitalOcean

....

Network

- Netflow (5.6)
- Packets

Storage

- Ceph

** Near-term roadmap*

Metrics & Event Data

METRICBEAT HEARTBEAT LOGSTASH



Applications

Datastores

- MySQL
- PostgreSQL
- MongoDB
- Couchbase
- Aerospike (6.0)
- Graphite (6.1)

Queues

- Kafka
- Redis
- RabbitMQ (6.0)

Caches

- Memcached (6.0)

Uptime

- Heartbeat

Custom apps

- JMX/Jolokia
- PHP-FPM
- Golang (6.0)
- Dropwizard (6.0)

Web servers

- Apache
- Nginx

Other

- HAProxy
- Zookeeper
- Prometheus

** Near-term roadmap*

Security Data Sources

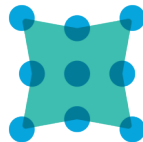
FILEBEAT



METRICBEAT



PACKETBEAT



LOGSTASH



Security

SIEM Augmentation

- ArcSight (5.6)
- more*

Audit

- Auditd
- Auditbeat (6.0)

Activity

Systems

- Access
- SSH

Applications

- Connections
- Users

Network

- IPs / GeoIP
- DNS Packets
- Netflow (5.6)
- Firewalls*
- IDS/IPS*

** Near-term roadmap*

Business Analytics

LOGSTASH



Structured

Databases

- JDBC input
- JDBC filter

Activity

SaaS services

- Salesforce
- Heroku
- Github
- Azure*

Social media

- Twitter

** Near-term roadmap*



Administration

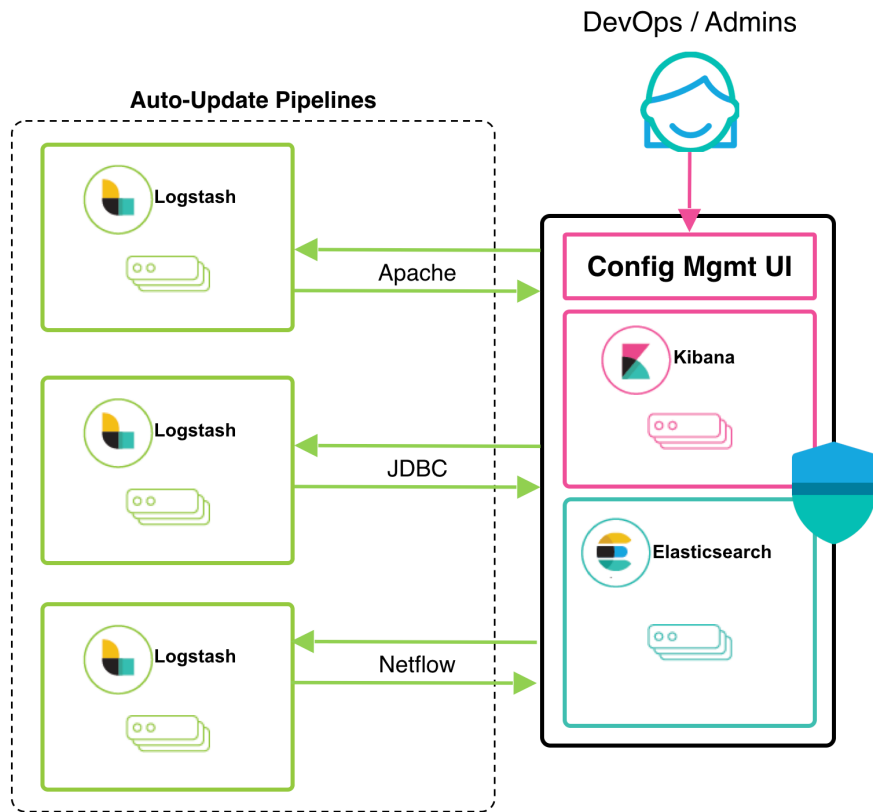


Monitoring & Management

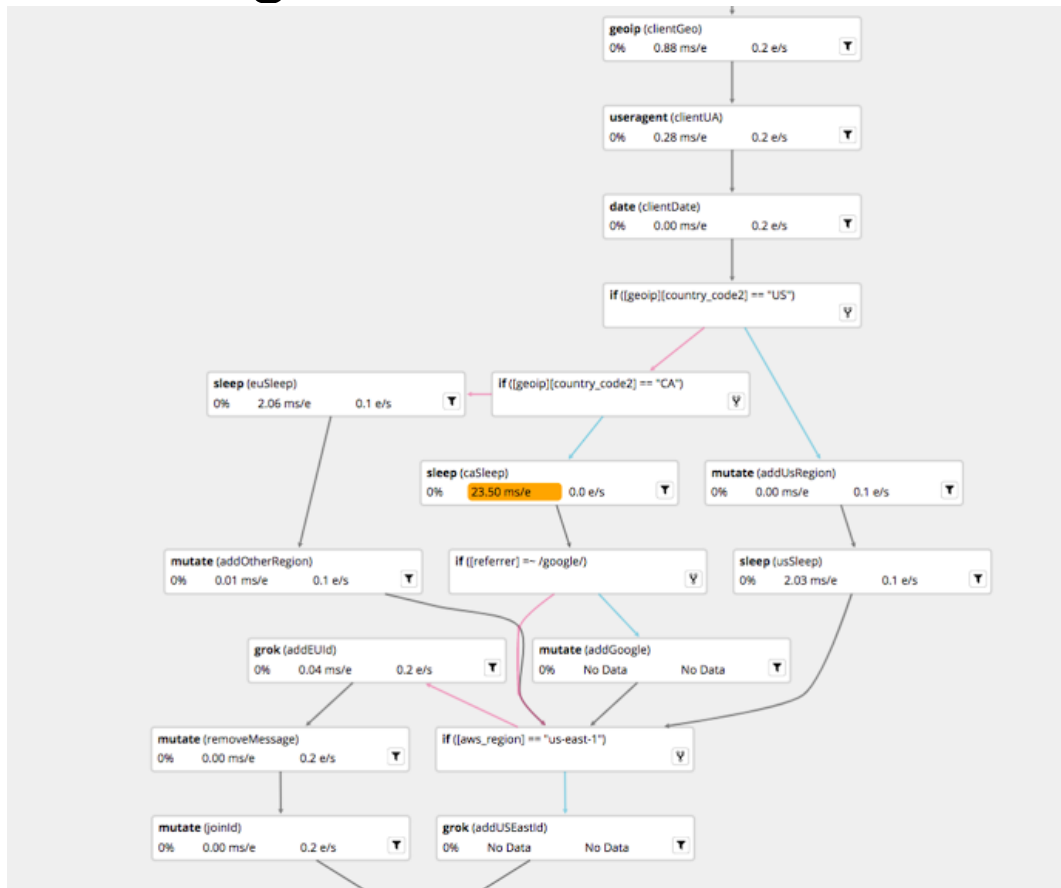
Logstash

- Centralized monitoring (5.3)
- Centralized management (6.0)

** Near-term roadmap*



Logstash Monitoring

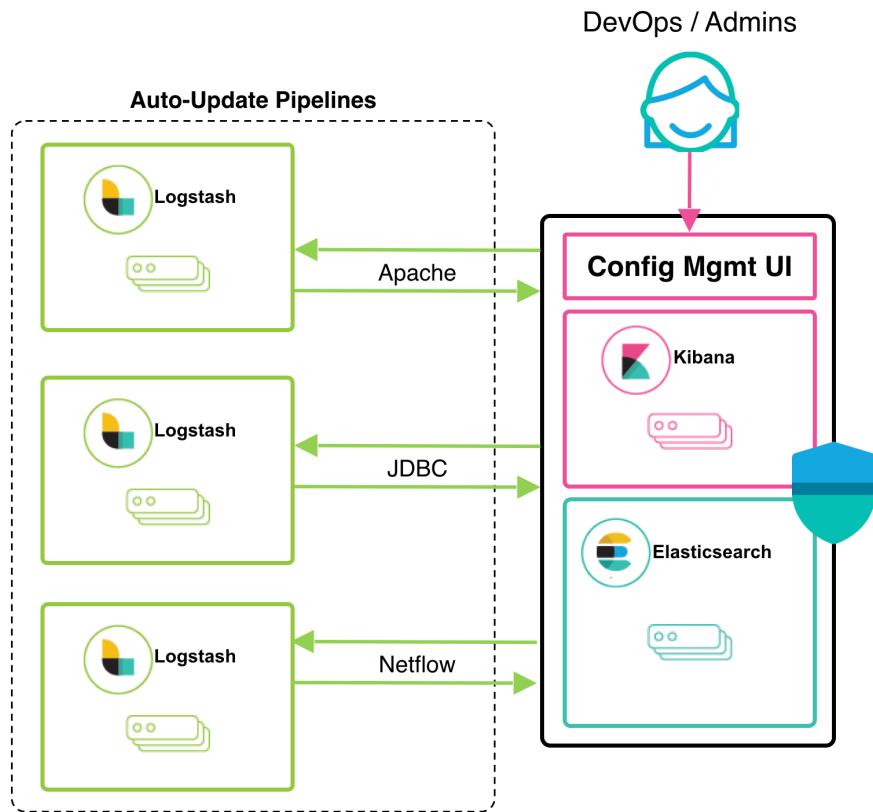


Monitoring & Management

Logstash

- Centralized monitoring (5.3)
- Centralized management (6.0)

** Near-term roadmap*



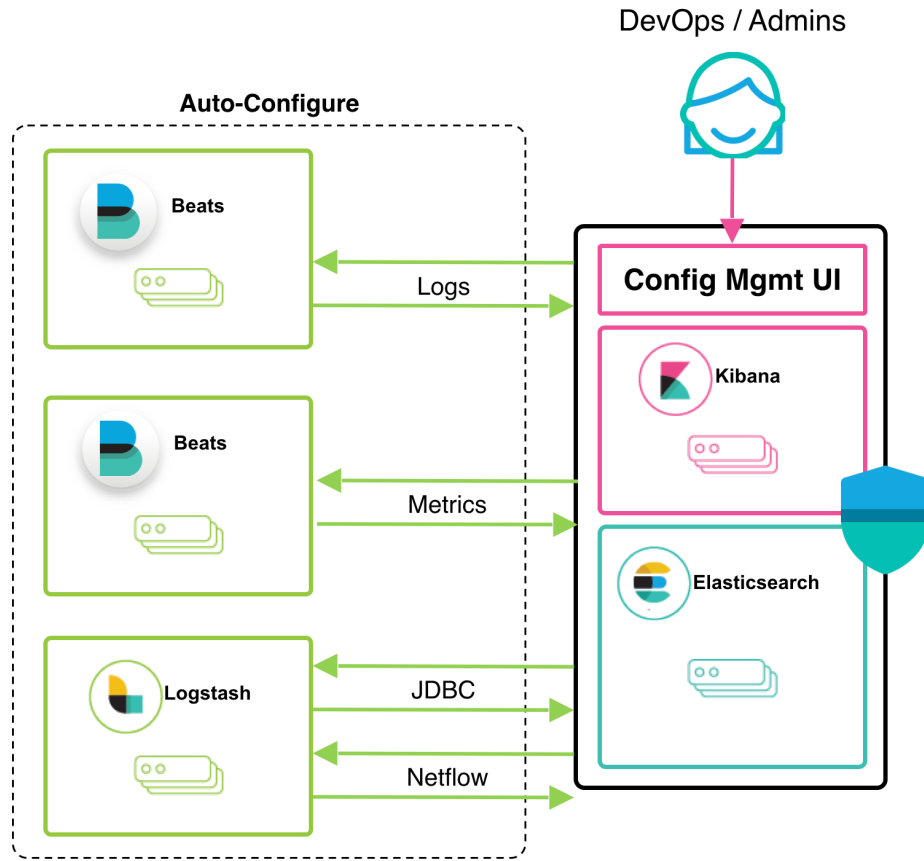
Monitoring & Management

Logstash

- Centralized monitoring (5.3)
- Centralized management (6.0)

Beats (Roadmap)

- Centralized monitoring
- Centralized management



Next steps

- Familiarize yourself with latest integrations (including in X-Pack)
- Watch UI roadmap for additional add-data workflows
- Come talk to us at the AMA booth



Thank You

Find me at AMA booth or email david@elastic.co

