



Elastic(search)

Neues aus dem Maschinenraum

Alexander Reelsen
@spinscale
alex@elastic.co



What's the problem?

Elastic Makes Building Scalable, Real-Time Systems Simple

Handles Complex
& Diverse Data



Social



Location



User-
Activity



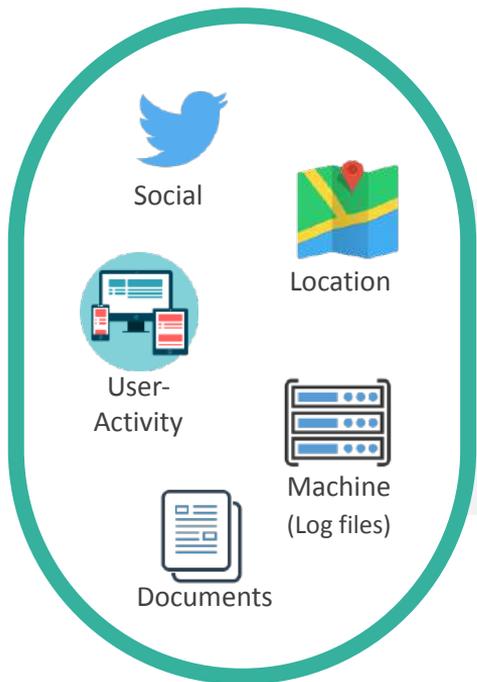
Machine
(Log files)



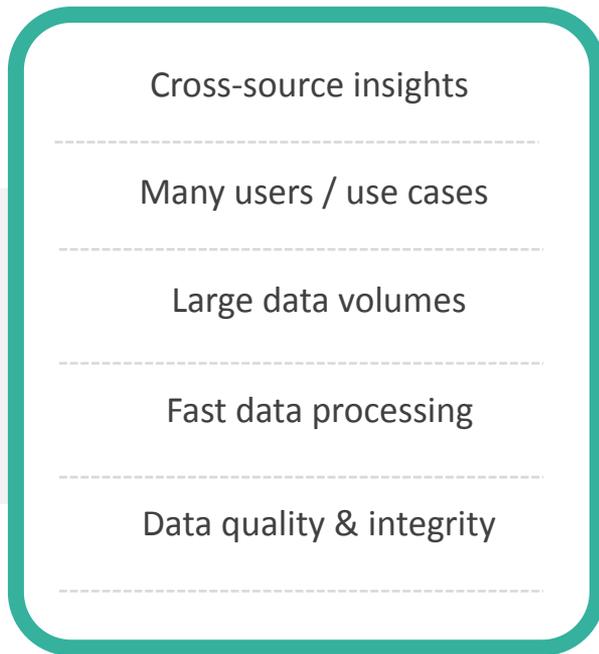
Documents

Elastic Makes Building Scalable, Real-Time Systems Simple

Handles Complex
& Diverse Data

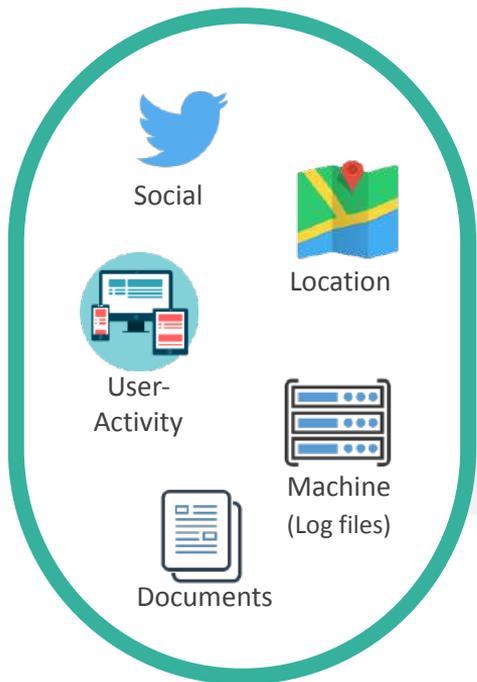


Meets Core
Developer Requirements

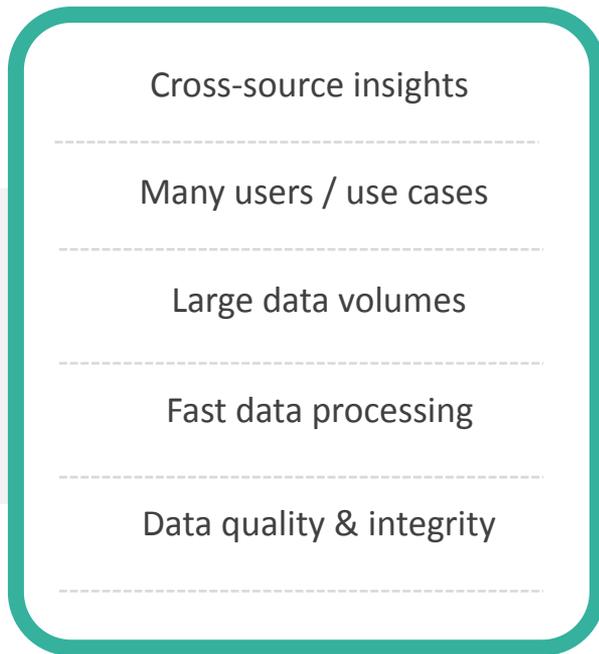


Elastic Makes Building Scalable, Real-Time Systems Simple

Handles Complex
& Diverse Data



Meets Core
Developer Requirements



Solves Critical
Use Cases



The Elastic Stack

 Elastic Stack

Plugins

Monitoring

Security

Alerting

User Interface

Kibana

Store, Index,
& Analyze

Elasticsearch

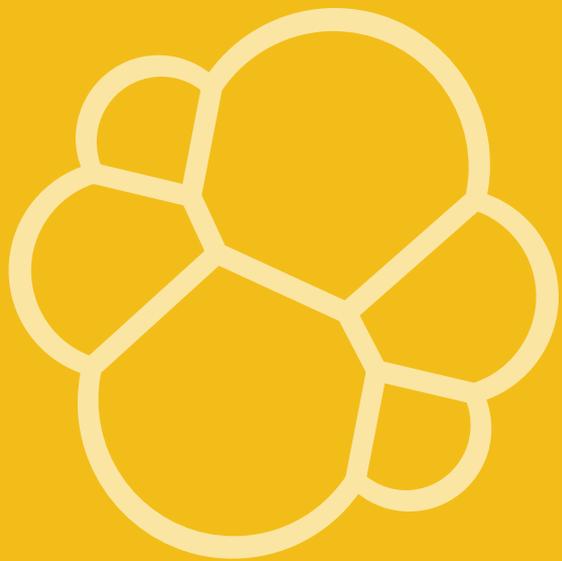
Ingest

Logstash

Beats

Hosted Service

Found: Elasticsearch as a Service



Elasticsearch

Elasticsearch: Store, Index, and Analyze

Distributed, scalable, and resilient

Designed for scale-out; high availability

Developer friendly

API-first; schemaless, native JSON & HTTP, client libraries

Real-time Search & Analytics

Real-time aggregations, geospatial, full-text search; query structured and unstructured data

Elasticsearch 2.x



Pipeline Aggregations

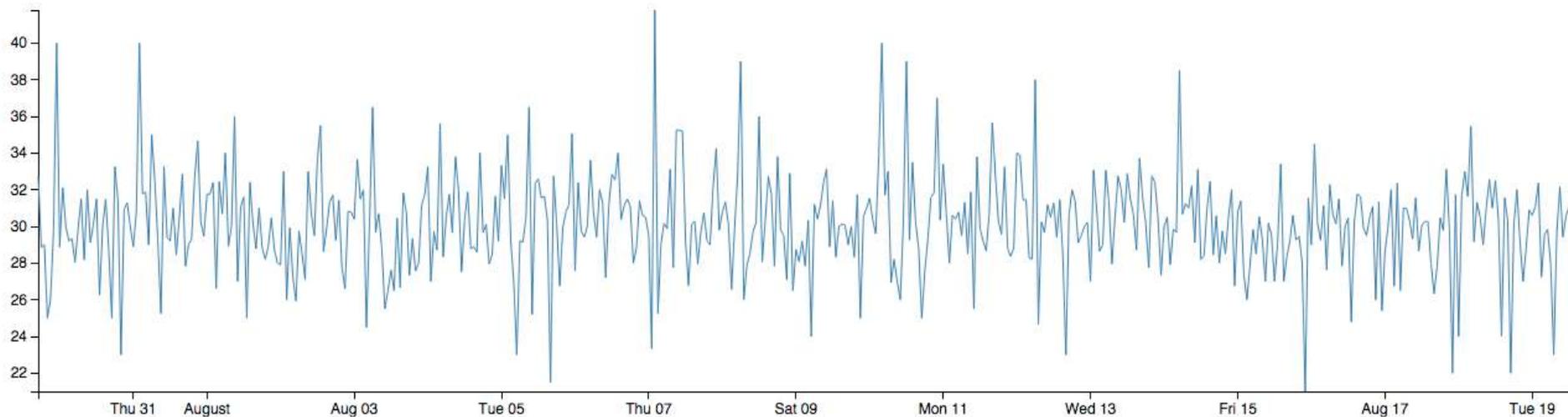


Query profiler

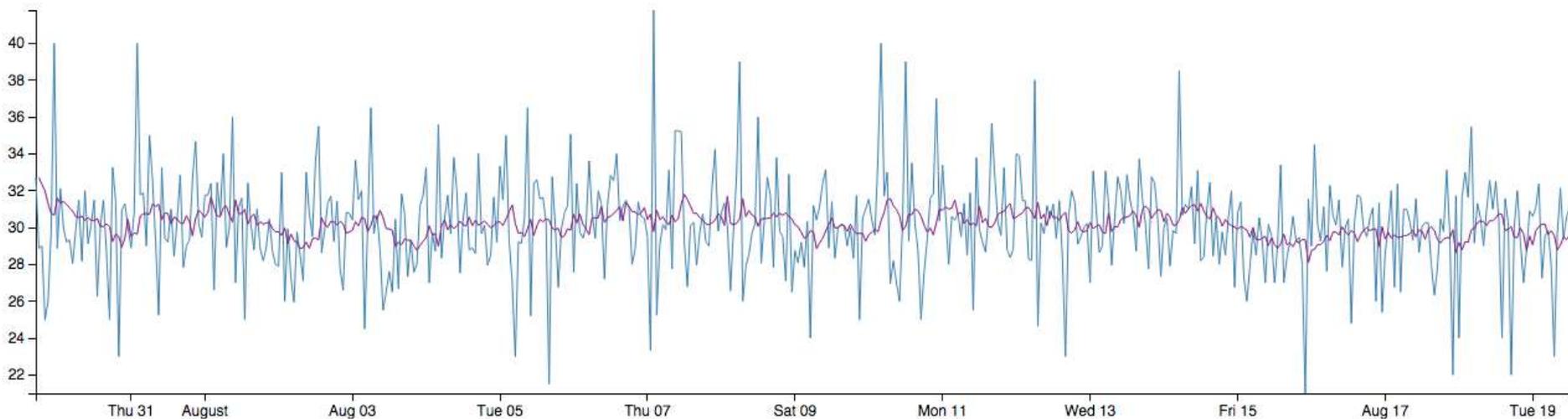


Plugins as first class citizen

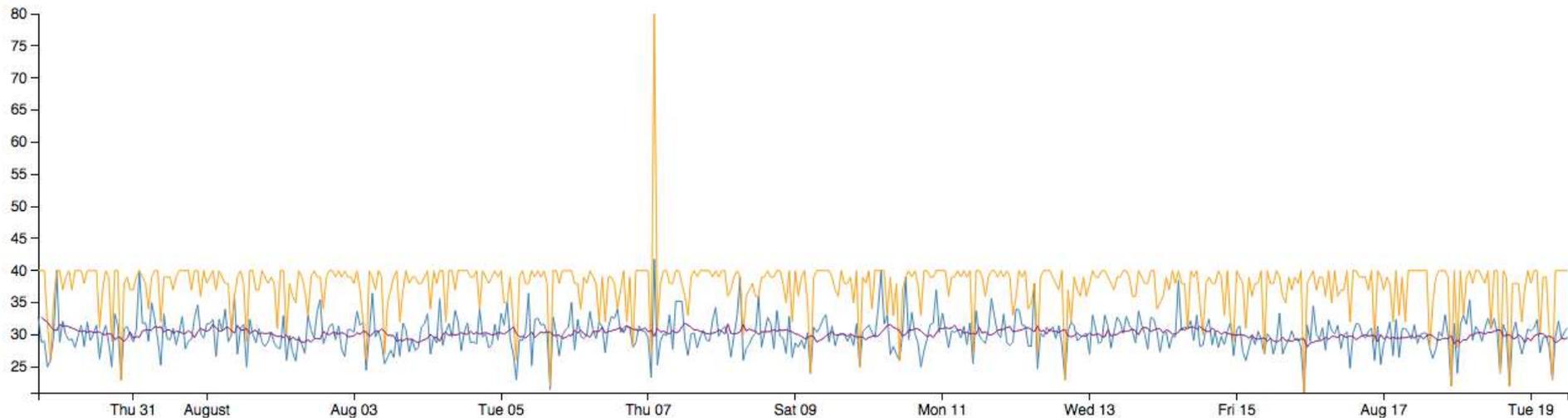
Pipeline Aggregations



Pipeline Aggregations



Pipeline Aggregations



Query profiler

```
GET /profile/data/_search
```

```
{  
  "profile": true,  
  "query": {  
    "match": {  
      "foo": "bar baz"  
    }  
  }  
}
```

Query profiler

```
{  
  "took": 2, "timed_out": false,  
  "_shards": { ... },  
  "hits": { ... },  
  "profile": {  
    "shards": [  
      {  
        "id": "[vj4imdlqQOK0Xj_n70xD_A][profile][0]",  
        "searches": [  

```

Query profiler

...

```
"searches": [ {  
  "query": [ {  
    "query_type": "BooleanQuery",  
    "lucene": "+(foo:bar foo:baz) #ConstantScore(_type:data)",  
    "time": "1.056684000ms",  
    "breakdown": {  
      ...
```

Elasticsearch 2.x



Allocation and recovery



Security Manager



Resilience

Elasticsearch 2.x



Mapping updates



Two phase query execution



Query/Filter caching

Elasticsearch 3.x



Task Management



Reindex



New scripting language

Elasticsearch 3.x



Strict settings



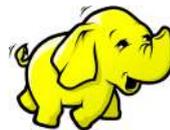
Improved suggester



Percolator

ES-Hadoop: Integrate with Hadoop, Spark & More

Real-time search on Hadoop data



Standalone, self-contained library on Hadoop



Access ES data bi-directionally



Support for MapReduce, Hive, Pig, Cascading, Spark, and Storm



Leverage HDFS to backup and archive ES data



Security for the Elastic Stack (Shield)

Simply Secure Elasticsearch

Username/password protection

Advanced Security When Needed

LDAP/AD integration

Role-based access control

Field and document level security

Encrypted communications

Audit logging



Alerting for the Elastic Stack (Watcher)

Alerts based on your data

Flexible Notifications

Wide range of use-cases

Integrations

Slack

Hipchat

Pagerduty

Email



Monitoring for the Elastic Stack (Marvel)

Monitor Metrics

Track real-time stats and metrics for all clusters and nodes

Diagnose Issues

Analyze historical or real-time data for root cause analyses

Optimize Performance

Utilize in-depth analyses to improve cluster performance



Elasticsearch as a Service (Found)

The only fully managed and hosted Elasticsearch product supported by the creators of Elasticsearch, Logstash and Kibana

Set up clusters in seconds

Dedicated memory and storage

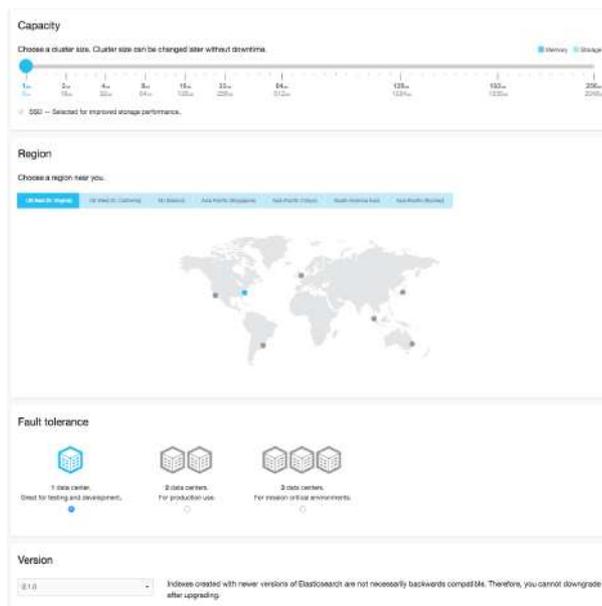
Native, unmodified Elasticsearch endpoint

High availability with replication

Simply scale up or down

Pre-integrated Elastic plugins

Enterprise, SLA-support



The screenshot displays the configuration interface for an Elasticsearch cluster. It is divided into several sections:

- Capacity:** A slider allows selecting a cluster size from 1m to 25m nodes. A note states, "Cluster size can be changed later without downtime." A radio button for "SSD" is selected, with a note: "Selected for improved storage performance."
- Region:** A prompt "Choose a region near you." is followed by a world map with several location markers.
- Fault tolerance:** Three options are shown with cube icons: "1 data center" (Direct for testing and development), "2 data centers" (For production use), and "3 data centers" (For mission-critical environments).
- Version:** A dropdown menu is set to "8.10". A warning note reads: "Indices created with newer versions of Elasticsearch are not necessarily backwards compatible. Therefore, you cannot downgrade after upgrading."

Getting up and running - easy!

```
unzip elasticsearch-2.x.y.zip ; cd elasticsearch 2.x.y  
bin/elasticsearch
```

```
bin/plugin install analysis-icu
```

```
bin/plugin install shield
```

```
bin/plugin install watcher
```

```
bin/plugin install marvel
```

Elasticsearch: Further info

<https://www.elastic.co/guide/en/elasticsearch/reference/2.2/search-profile.html>

<https://www.elastic.co/guide/en/elasticsearch/reference/2.2/search-aggregations-pipeline.html>

<https://www.elastic.co/blog/out-of-this-world-aggregations>

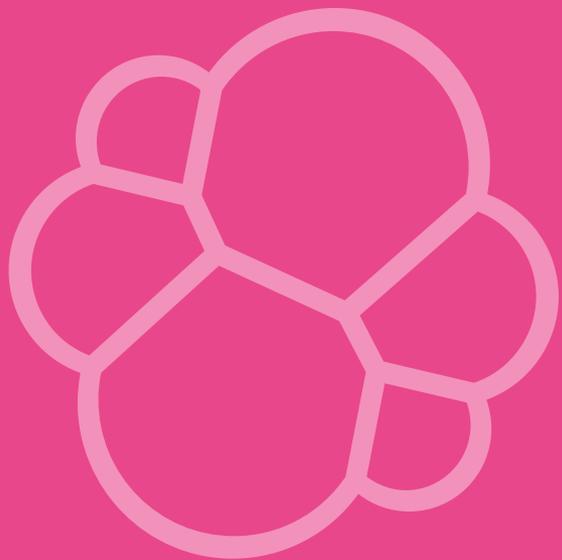
<https://www.elastic.co/blog/staying-in-control-with-moving-averages-part-1>

<https://www.elastic.co/blog/staying-in-control-with-moving-averages-part-2>

<https://www.elastic.co/blog/implementing-a-statistical-anomaly-detector-part-1>

<https://www.elastic.co/blog/implementing-a-statistical-anomaly-detector-part-2>

<https://www.elastic.co/blog/implementing-a-statistical-anomaly-detector-part-3>



Kibana

Kibana: A User Interface for All of Your Data



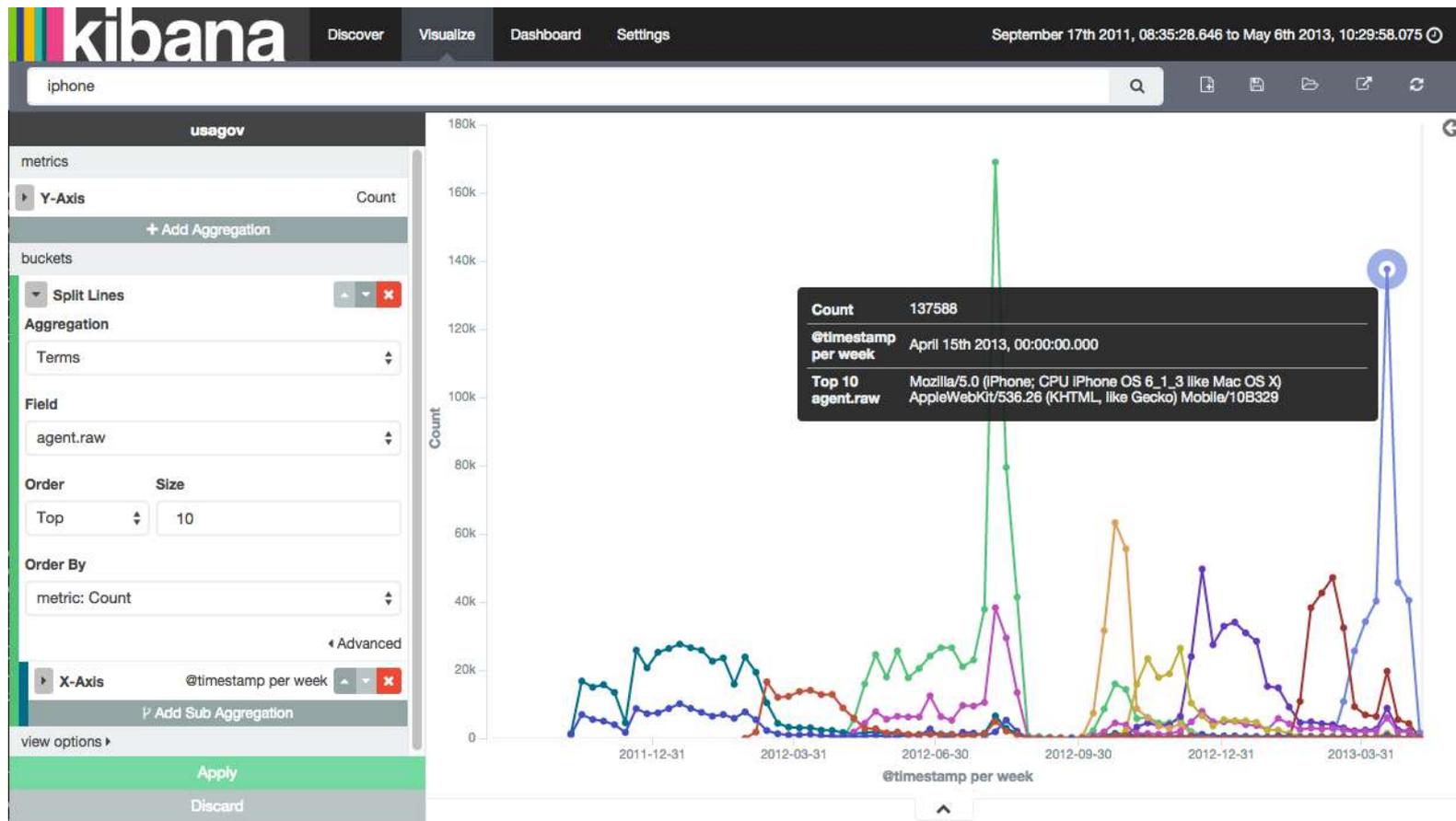
Explore and discover insights

Instant response at any scale

Build interactive dashboards

Share, embed, and integrate

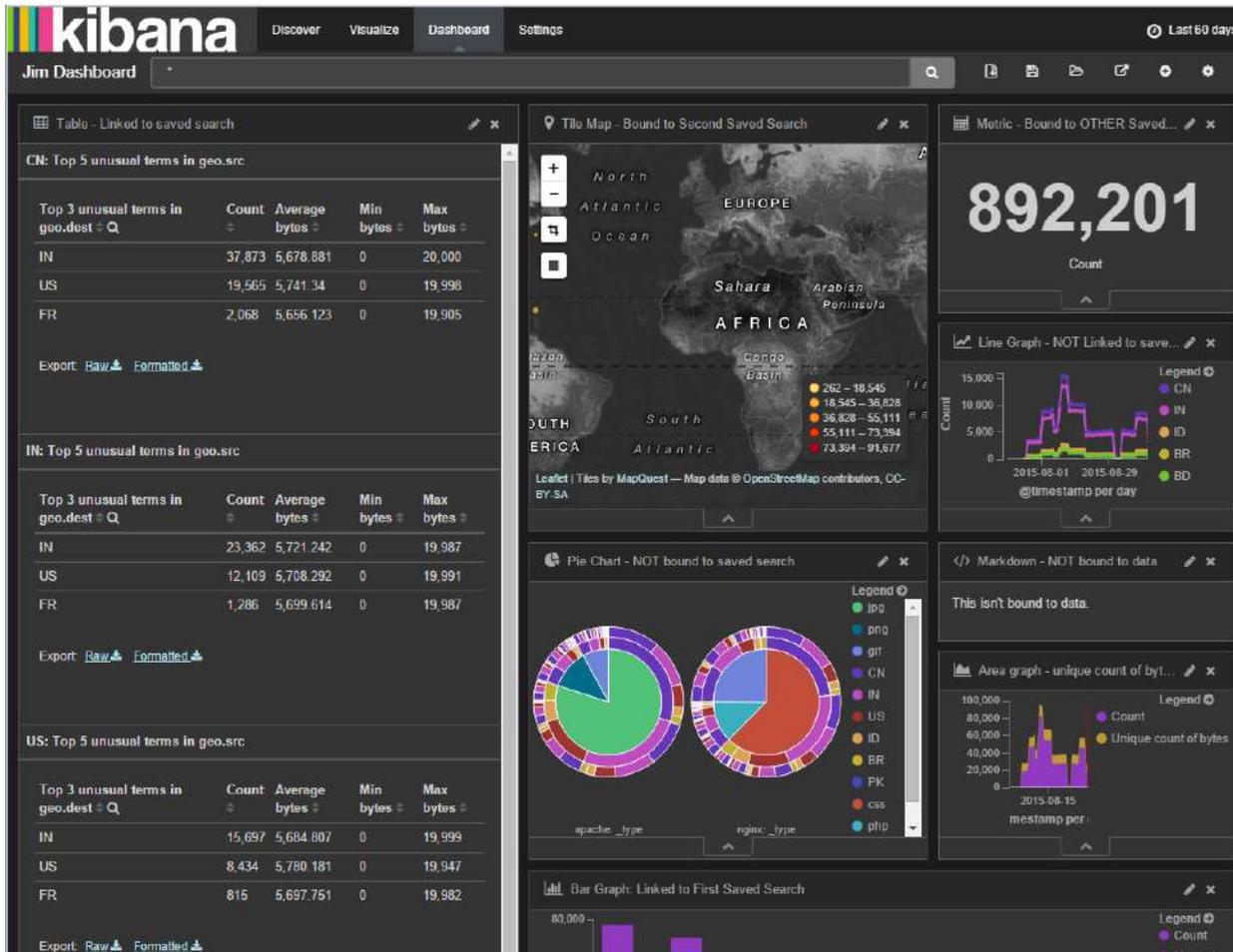
Single visualizations...



... form a dashboard



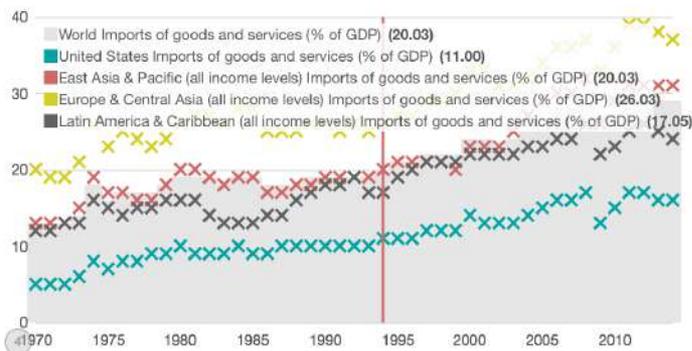
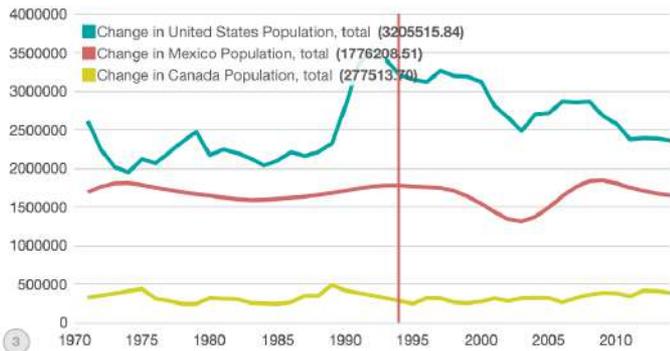
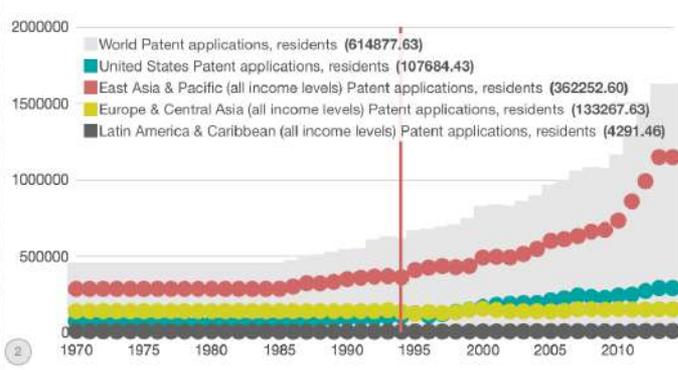
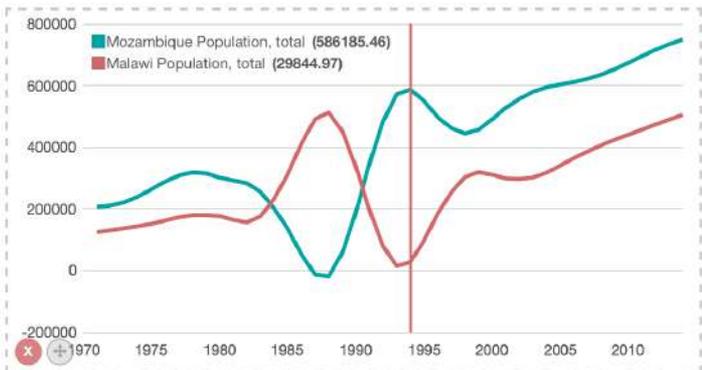
... the dark side



Timelion - Time series composer

 ~ 46 years ago to ~ 2 years ago

1y ▶ ☰



Timelion - Time series composer

Composable functions

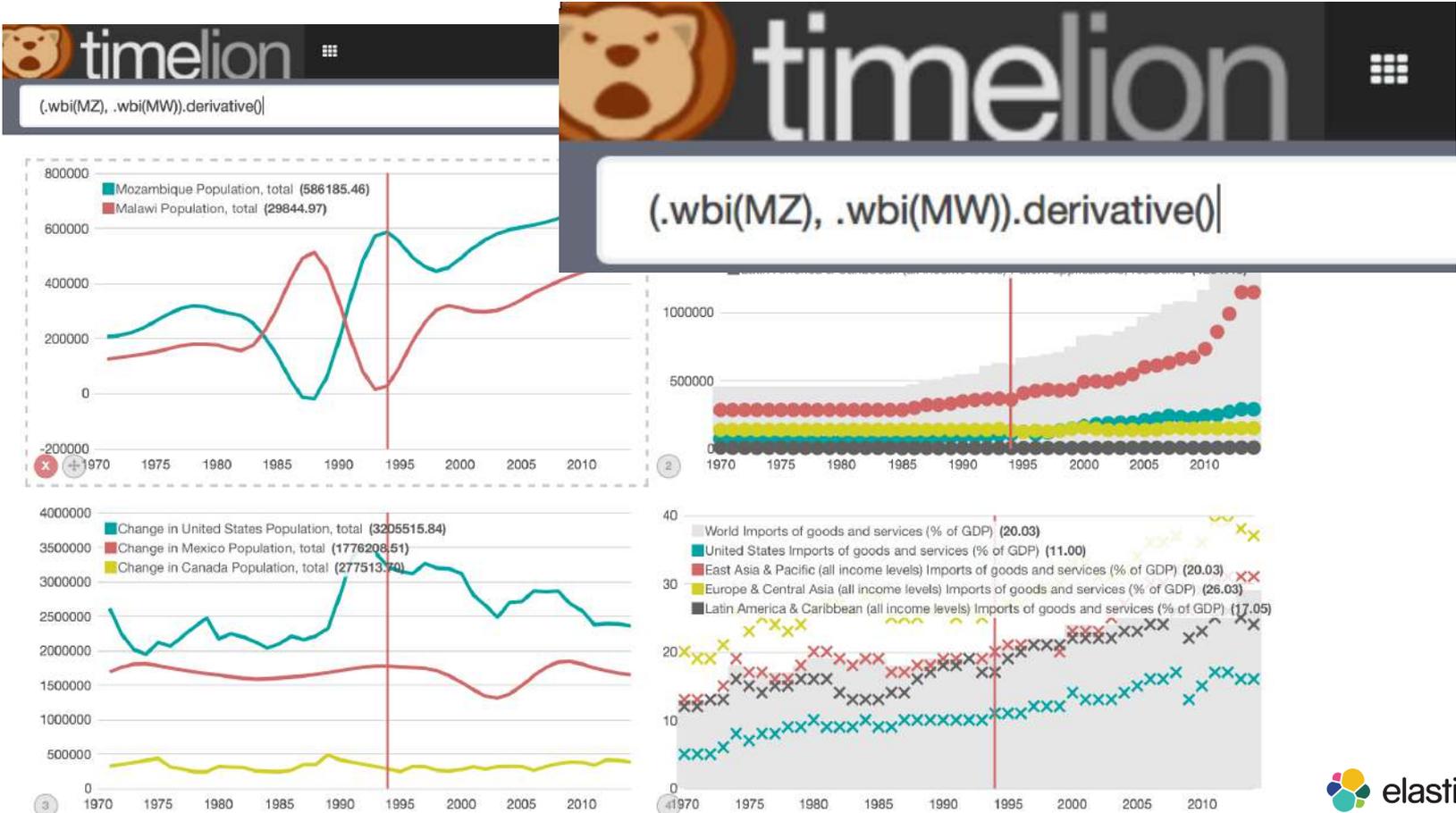
abs, derivative, cusum, divide, first, max, min,
movingaverage, movingstd, multiply, subtract, sum
bars, color, hide, label, legend, lines, points,
precision, yaxis
es, graphite, Quandl, worldbank, wbi

More info

<https://www.elastic.co/blog/timelion-timeline>

<https://www.youtube.com/watch?v=-sgZdW5k7eQ>

Timelion - Time series composer



Sense - The missing UI

```
1 # Delete all data in the `website` index
2 DELETE /website
3
4 # Create a document with ID 123
5 PUT /website/blog/123
6 {
7   "title": "My first blog entry",
8   "text": "Just trying this out...",
9   "date": "2014/01/01"
10 }
11
12 # Search!
13 GET website/_search
14 {
15   "query": {
16     "match": {
17       "title": "blog"
18     }
19   }
20 }
21
```

```
1 # DELETE /website
2 { }
5
6 # PUT /website/blog/123
7 { }
19
20 # GET website/_search
21 {
22   "took": 3,
23   "timed_out": false,
24   "_shards": {
25     "total": 5,
26     "successful": 5,
27     "failed": 0
28   },
29   "hits": {
30     "total": 0,
31     "max_score": null,
32     "hits": []
33   }
34 }
```

Sense - Features

Suggestions for all requests

Multiple Requests

Auto-Indent

Copy as cURL

Keyboard shortcuts

History

More info

<https://www.elastic.co/blog/sense-2-0-0-beta1>

<https://www.elastic.co/guide/en/sense/current/index.html>

Getting up and running - easy!

```
tar xvf kibana-...tar.gz ; cd kibana
```

```
bin/kibana
```

```
./bin/kibana plugin --install elastic/sense
```

```
./bin/kibana plugin --install elastic/timelion
```

Writing own plugins - easy!

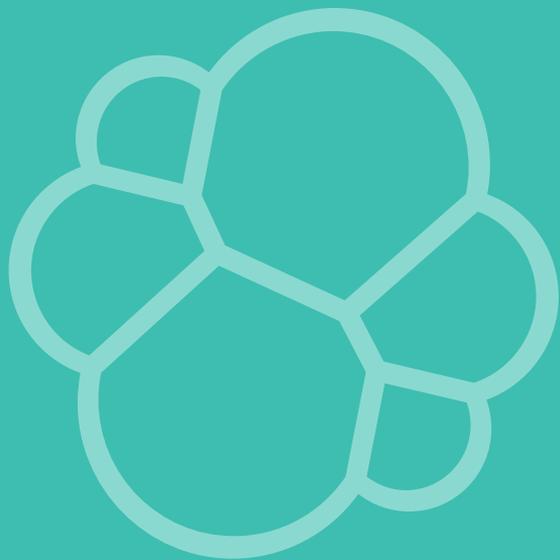
```
npm install -g yo
```

```
npm install -g generator-kibana-plugin
```

```
mkdir my-new-plugin
```

```
cd my-new-plugin
```

```
yo kibana-plugin
```



Logstash

Logstash: Collect, Enrich, and Transport

Collect data from many sources

Application

Social data

Infra/web/audit logs

Sensor data

Message queues

Documents

Transaction/wire



Open-source ETL engine with more than 200+ community extensible plugins

Enrich

Parse, transform, clean

Transport

Output to Elasticsearch and other systems

Logstash 2.x - Changes

Next generation Pipeline in 2.2

Better performance, works in micro batches, automatic worker scaling

Plugins

kafka input/output, JDBC input, HTTP input, WebHDFS output, Salesforce input, HTTP poller

Logstash 3.x



Persistent Queues



Performance



Clustering



Monitoring

Getting up and running - easy!

```
unzip logstash-2.X.Y.zip ; cd logstash-2.X.Y
```

```
bin/logstash -f logstash.conf
```

```
bin/plugin install logstash-output-jms
```

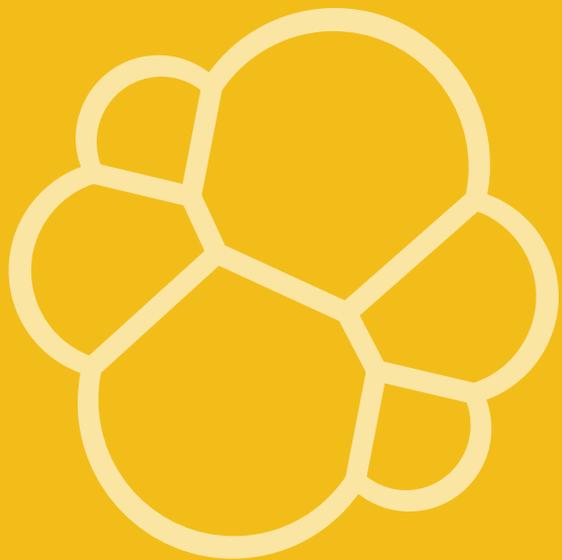
Writing own plugins - easy!

```
git clone https://github.com/logstash-plugins/logstash-input-example
```

```
git clone https://github.com/logstash-plugins/logstash-output-example
```

```
git clone https://github.com/logstash-plugins/logstash-filter-example
```

```
git clone https://github.com/logstash-plugins/logstash-codec-example
```



Beats

Beats: Lightweight Data Shippers

Libbeat

Library for forwarding host-based metrics to Elasticsearch

Packetbeat

Real-time network packet analytics for web, database, and any network protocols

Topbeat

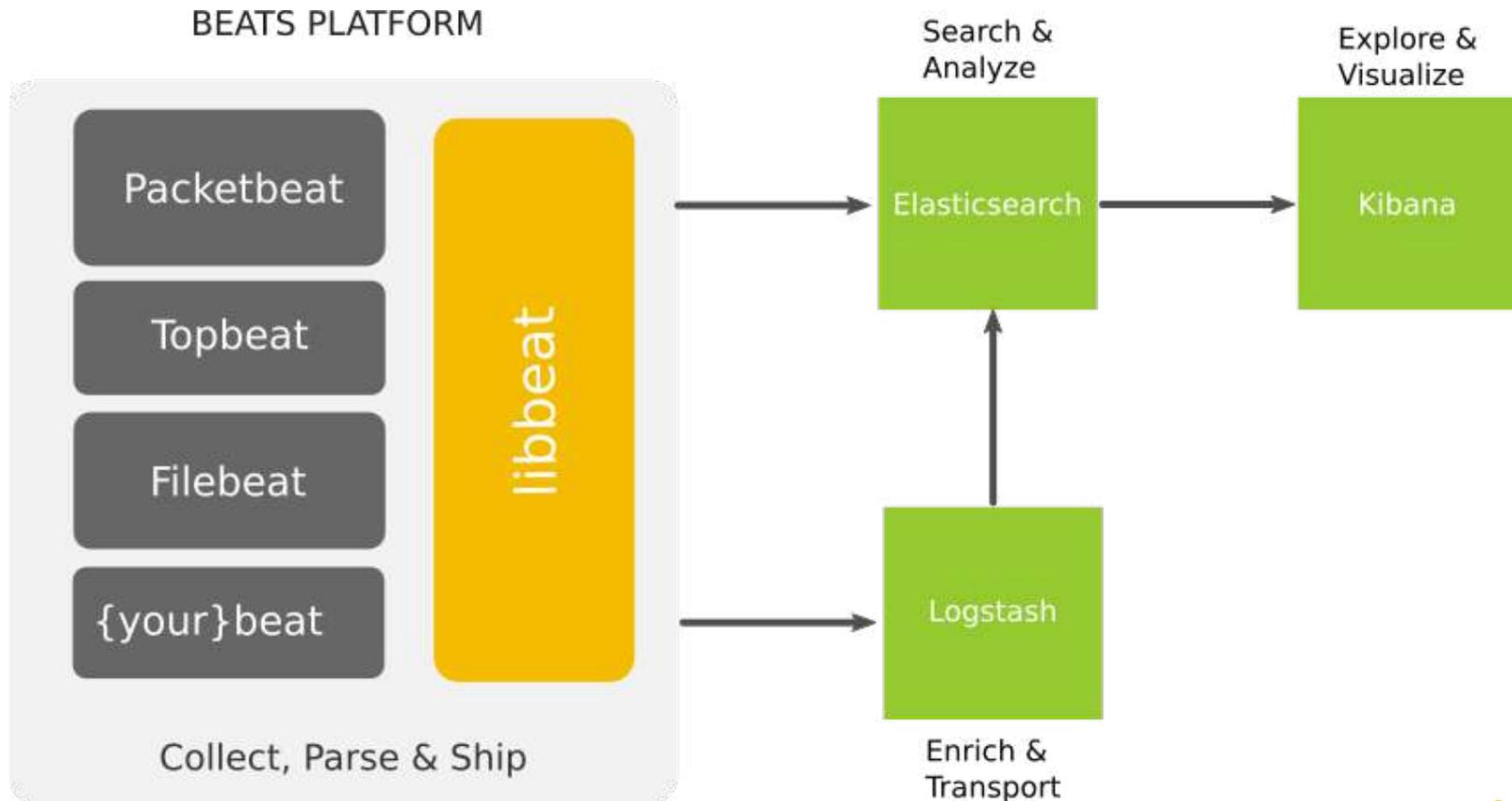
Gather resource utilization data such as CPU, memory, etc and ship it to Elasticsearch to analyze .

Filebeat

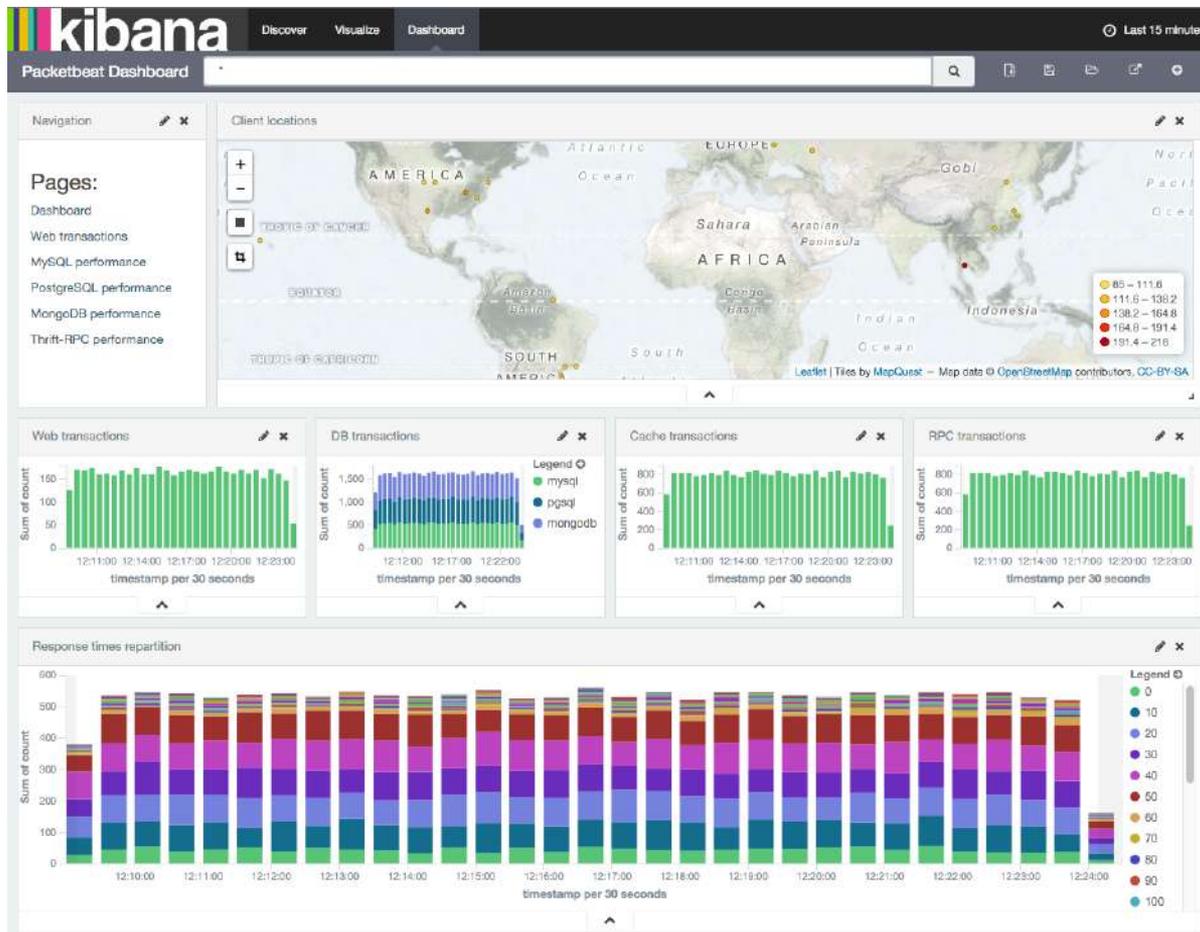
Next-generation Logstash forwarder to collect, pre-process, and forward log files.



Beats: Lightweight Data Shippers



Packetbeat



Packetbeat

Protocols

ICMP (v4 and v6), DNS, HTTP, Mysql, PostgreSQL, Redis, Thrift-RPC, MongoDB, Memcache

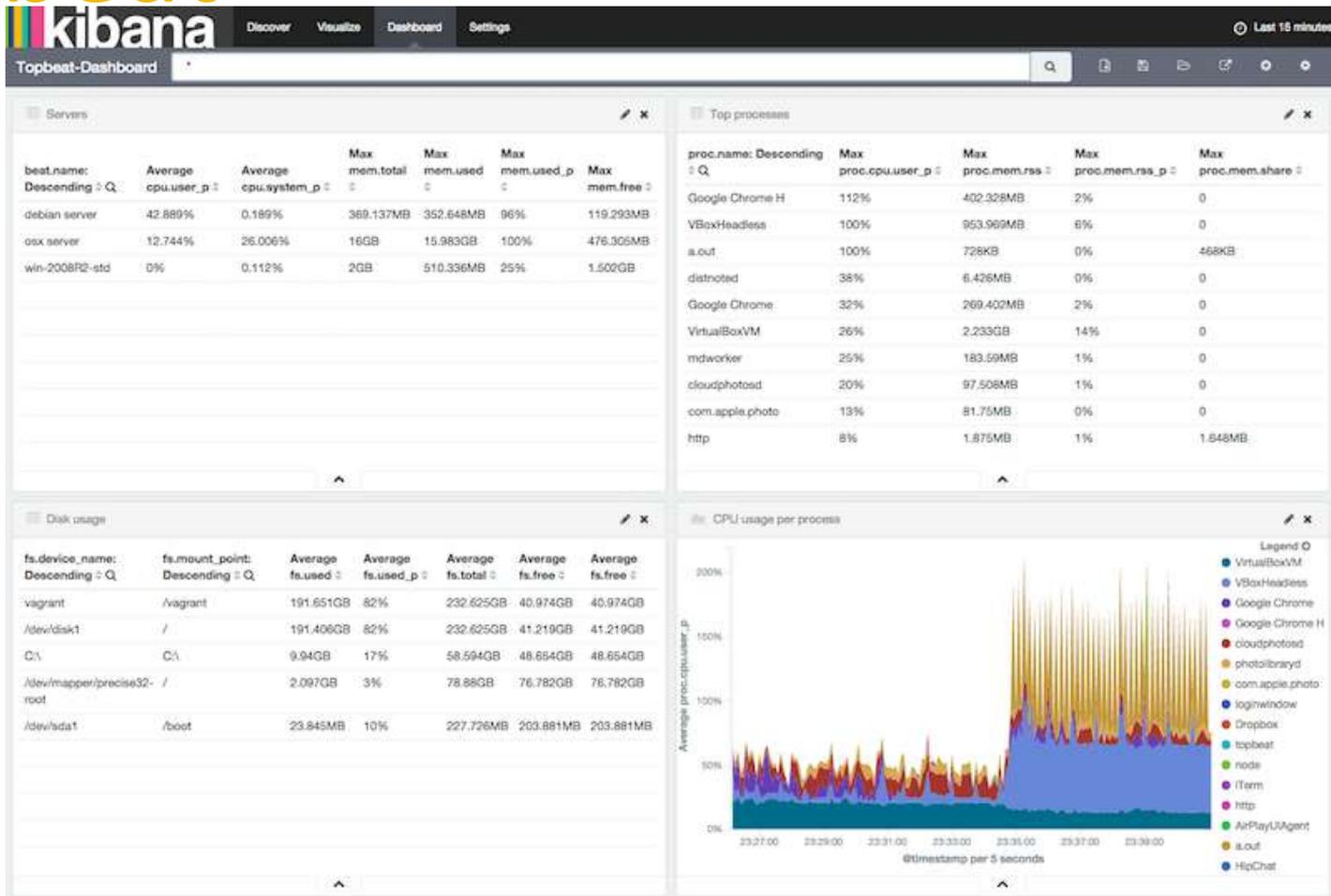
Output

Elasticsearch, Logstash, File, console

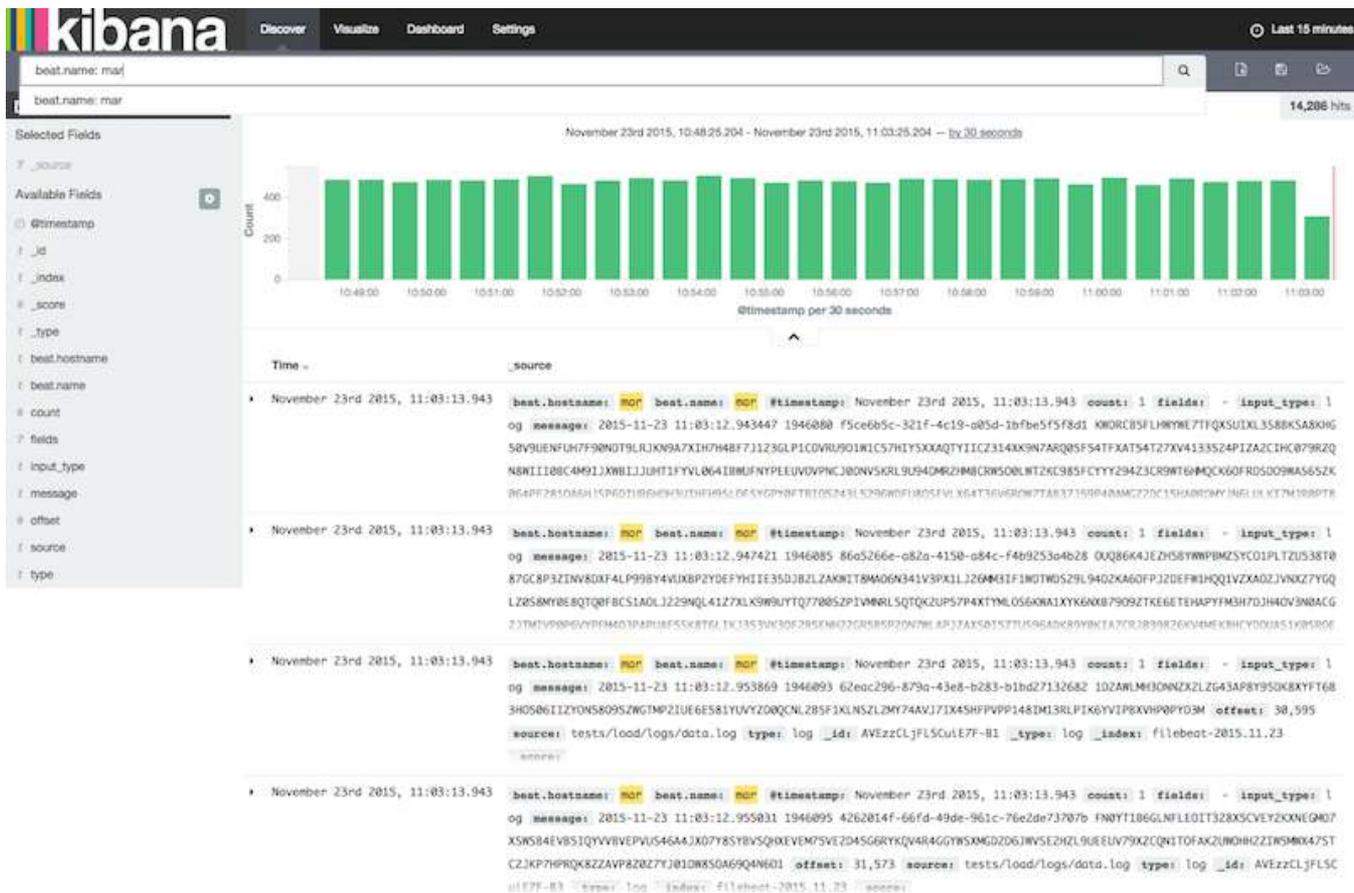
Extensibility

protocols can be added easily

Topbeat



Filebeat - logstash forwarder as beat



Beats



Metricbeat, Winlogbeat



Beat stats



Multiline filtering

Getting up and running - easy!

```
tar zxvf filebeat-1.X.Y-darwin.tgz ; cd filebeat-1.Y.Z
```

```
./filebeat -c filebeat.yml
```

```
./topbeat -c topbeat.yml
```

```
./packetbeat -c packetbeat.yml
```

Community Beats

apachebeat

elasticbeat

httpbeat

dockerbeat

redisbeat

phpfpmbat

uwsgibeat

execbeat

unifiedbeat

pingbeat

nginxbeat

factbeat

hsbeat

nagioscheckbeat

Beats: Further info

<https://www.elastic.co/guide/en/beats/libbeat/current/getting-started.html>

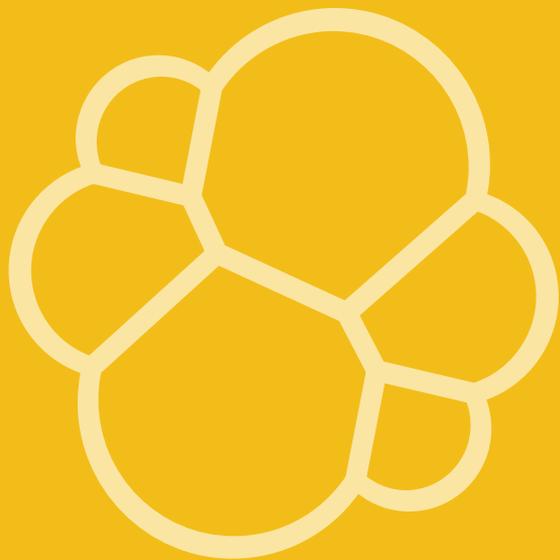
<https://www.elastic.co/guide/en/beats/packetbeat/current/index.html>

<https://www.elastic.co/guide/en/beats/filebeat/current/index.html>

<https://www.elastic.co/guide/en/beats/topbeat/current/index.html>

<https://www.elastic.co/guide/en/beats/winlogbeat/current/index.html>

<https://speakerdeck.com/tsg/get-real-time-insights-from-your-application-with-packetbeat-and-elasticsearch>



Elasticsearch

Ingest node

Document enrichment before indexing

Simple document editing

Processors

set, append, remove, rename, convert, gsub, join, split, lowercase, uppercase, trim, grok, date, fail

Dead letter queue

failure handlers to change field or destination index

Ingest node

master only

Document enrichment before indexing

Simple document editing

Processors

set, append, remove, rename, convert, gsub, join, split, lowercase, uppercase, trim, grok, date, fail

Dead letter queue

failure handlers to change field or destination index

Ingest node - Configure pipeline

master only

```
PUT/_ingest/pipeline/access-log-pipeline
```

```
{  
  "description" : "Apache Logs Pipeline",  
  "processors" : [  
    { "grok" : { ... } },  
    { "convert" : { ... } },  
    { "convert" : { ... } },  
    { "date" : { ... } },  
    { "geoip" : { ... } },  
  ]  
}
```

Ingest node - Grok Processor

master only

```
...  
  
  {  
    "grok" : {  
      "field" : "message",  
      "pattern" : "%{COMBINEDAPACHELOG}"  
    }  
  },  
...  
...
```

Ingest node - Convert Processor

master only

```
...  
  {  
    "convert" : {  
      "field": "response",  
      "type": "integer"  
    }  
  },  
...
```

Ingest node - Convert Processor

master only

...

```
{
```

```
  "convert" : {
```

```
    "field": "bytes",
```

```
    "type": "integer"
```

```
  }
```

```
},
```

...

Ingest node - Date Processor

master only

```
...  
  {  
    "date" : {  
      "match_field": "timestamp",  
      "match_formats" : [ "dd/MMM/YYYY:HH:mm:ss Z" ]  
    }  
  },  
...
```

Ingest node - GeoIP Processor

master only

...

```
{
```

```
  "geoip" : {
```

```
    "source_field" : "clientip"
```

```
  }
```

```
}
```

...

Ingest node - Index document

master only

```
POST logs/log?pipeline=access-log-pipeline
```

```
{
```

```
  "message" : "70.193.17.92 - - [08/Sep/2014:02:54:42  
+0000] \"GET /presentations/logstash-scale11x/images/  
ahhh__rage_face_by_samusmmx-d5g5zap.png HTTP/1.1\" 200  
175208 \"http://mobile.rivals.com/board_posts.asp?  
SID=880&mid=198829575&fid=2208&tid=198829575&Team=&TeamId  
=&SiteId=\" \"Mozilla/5.0 (Linux; Android 4.2.2; VS980 4G  
Build/JDQ39B) AppleWebKit/537.36 (KHTML, like Gecko)  
Chrome/33.0.1750.135 Mobile Safari/537.36\""
```

```
}
```

Ingest node - Indexed document

master only

```
{
  "_index": "logs", "_type": "log", "_id": "AVKiNsYu-Si4Nc0nCP5b",
  "_version": 1, "found": true,
  "_source": {
    "request": "/presentations/logstash-scale11x/images/
ahhh__rage_face_by_samusmmx-d5g5zap.png",
    "agent": "\"Mozilla/5.0 (Linux; Android 4.2.2; VS980 4G Build/JDQ39B)
AppleWebKit/537.36 (KHTML, like Gecko) Chrome/33.0.1750.135 Mobile
Safari/537.36\"",
    "geoip": {
      "continent_name": "North America",
      "city_name": "Charlotte",
      "country_iso_code": "US",
      "region_name": "North Carolina",
      "location": { "lon": -80.8431, "lat": 35.2271 }
    }
  },
}
```

Ingest node - Indexed document

master only

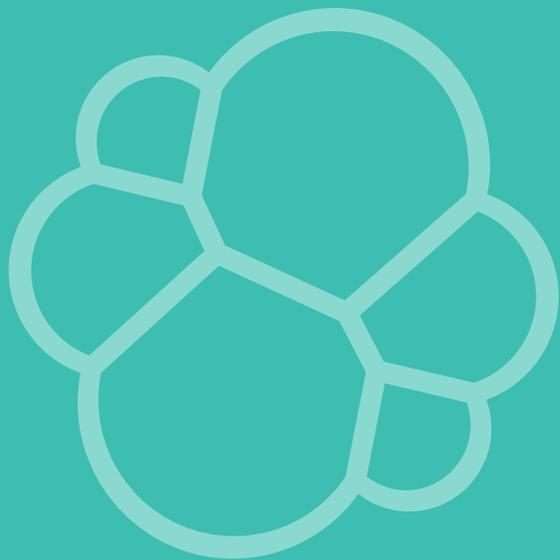
...

```
"auth": "-", "ident": "-", "verb": "GET", "httpversion": "1.1",
  message: "70.193.17.92 - - [08/Sep/2014:02:54:42 +0000] \"GET /
presentations/logstash-scale11x/images/ahhh__rage_face_by_samusmx-
d5g5zap.png HTTP/1.1\" 200 175208 \"http://mobile.rivals.com/
board_posts.asp?
SID=880&mid=198829575&fid=2208&tid=198829575&Team=&TeamId=&SiteId=\"
\"Mozilla/5.0 (Linux; Android 4.2.2; VS980 4G Build/JDQ39B) AppleWebKit/
537.36 (KHTML, like Gecko) Chrome/33.0.1750.135 Mobile Safari/537.36\"",
  "referrer": "\"http://mobile.rivals.com/board_posts.asp?
SID=880&mid=198829575&fid=2208&tid=198829575&Team=&TeamId=&SiteId=\"",
  "response": 200, bytes: 175208,
  "clientip": "70.193.17.92",
  "rawrequest": null,
  "@timestamp": "2014-09-08T02:54:42.000Z"
}
```

All pluggable

master only

```
bin/plugin install ingest-geoip
```



Community & Documentation

The Elastic Community



40,000

Community
members



1,500

Global subscription
customers



35,000

Commits against
Elastic stack to-date



ElasticSearch Meetups

Find out what's happening in ElasticSearch Meetup groups around the world and start meeting up with the ones near you.



Groups
127

Members
44,785

Interested
2,440

Cities
86

Countries
31

Docs

Elasticsearch: Store, Search, and Analyze

- [Elasticsearch Reference \[2.1\]](#) — [other versions](#)
- [Elasticsearch - The Definitive Guide](#)
- [Resiliency status](#)
- [Plugins and Integrations \[2.1\]](#) — [other versions](#)
- [Elasticsearch Clients](#)
- [Sense Editor](#)

Marvel: Monitoring for Elasticsearch

- [Marvel Reference \[2.1\]](#) — [other versions](#)

Shield: Security for Elasticsearch

- [Shield Reference \[2.1\]](#) — [other versions](#)

Watcher: Alerts for Elasticsearch

- [Watcher Reference \[2.1\]](#) — [other versions](#)

Found: Elasticsearch as a Service

- [Found Reference](#)

Kibana: Explore, Visualize, and Share

- [Kibana Reference \[4.3\]](#) — [other versions](#)

Logstash: Collect, Enrich, and Transport

- [Logstash Reference \[2.1\]](#) — [other versions](#)
- [Logstash Roadmap](#)

Beats: Collect, Parse, and Ship

- [Beats Platform Reference \[1.0.0\]](#) — [other versions](#)
- [Packetbeat Reference \[1.0.0\]](#) — [other versions](#)
- [Topbeat Reference \[1.0.0\]](#) — [other versions](#)
- [Filebeat Reference \[1.0.0\]](#) — [other versions](#)

Elasticsearch for Apache Hadoop

- [es-hadoop reference \[2.1\]](#) — [other versions](#)

Utilities

- [Curator Index Management \[3.4\]](#) — [other versions](#)
- [Rivers \[2.0\]](#) — [other versions](#)

O'REILLY



Elasticsearch

The Definitive Guide

A DISTRIBUTED REAL-TIME SEARCH AND ANALYTICS ENGINE

Clinton Gormley &
Zachary Tong

[all categories ▾](#)
[Categories](#)
[Latest](#)
[New \(139\)](#)
[Unread \(1\)](#)
[Top](#)
[+ New Topic](#)


Category	Latest	Topics
Announcements 1 new Release announcements, end of life notifications and other bits about Elastic products that we think will be useful to everyone. ■ Community Plugins	[ANNOUNCEMENT] - esBench 0.0.2 released • new 4h [ANN] Elasticsearch Mapper Attachment plugin 3.1.0 released 14d [ANN] Elasticsearch Mapper Attachment plugin 3.0.3 released 14d	1 / day 1 / week
Beats 6 new Any questions regarding Beats, forwarders for various types of data: PacketBeat for network metrics... and more to come! ■ libbeat ■ Packetbeat ■ Topbeat ■ Filebeat ■ Winlogbeat	✘ Beats from the Open Source Community 6d Filebeat from a Windows Network Share • new 2h How to Setup FileBeat with Basic Auth for LogStash Output? • new 2h	4 / day 14 / week
Elasticsearch 53 new Any questions related to Elasticsearch, including specific features, language clients and plugins.	✘ Connectivity issues with a new/upgraded 2.X cluster? Read here first 🙄 14d BulkIndexing is ~10X slower in 2.1 when index is dropped and recreated 8m How update with groovy script will work if changer refresh interval • new 15m	26 / day 147 / week
Found 15 new The Found forum is dedicated to all questions related to Found as a Service, Elastic's hosted Elasticsearch service.	NodeNotConnectedException after migrating to Shield 2h The configuration page says "A change is already being applied to this cluster" about 1 hour • new 10h Unsuccessful Authentication Error • new 13h	5 / day 28 / week
Hadoop and Elasticsearch 2 new Questions about Elasticsearch and all things Hadoop (Map/Reduce, Hive, Pig, Cascading, Spark and friends)	✘ Elasticsearch for Apache Hadoop 2.1.1 and 2.2.0-m1 released Aug 29 Error while loading RDDOperationScope, Missing dependency 'bad symbolic reference' • new 11h Spark-sql does not seem to read from a nested schema 1d	1 / day 4 / week
Logstash 44 new Everything related to your favorite centralized logging platform, including	✘ Logstash Plugins Community Maintainers 11d How to define fields in kibana • new 1h	19 / day

https://discuss.elastic.co/

Shield 3 new

Security for your Elastic stack.

 Where does not work - no certificates checked • new 39m

 Migrating from access control to shield / Make deny the default • new 19h

 Downloading shield • new 1d

2 / day

4 / week



This repository Search

Pull requests Issues Gist



elastic / elasticsearch

Unwatch 1,422

Star 13,961

Fork 4,632

Code

Issues 1,029

Pull requests 169

Pulse

Graphs

Settings

Filters

is:issue is:open

Labels

Milestones

New Issue

1,029 Open 8,075 Closed

Author

Labels

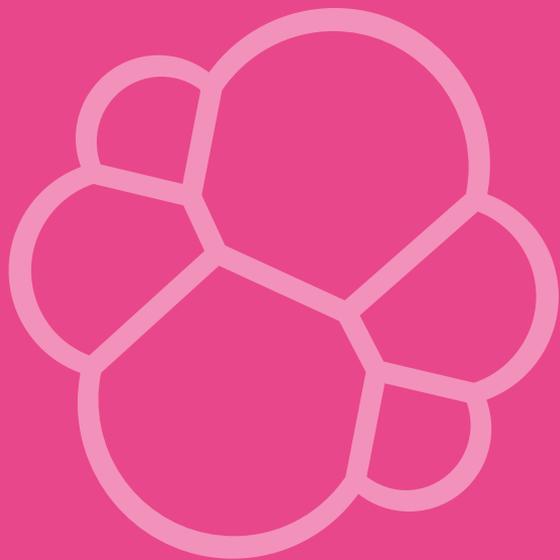
Milestones

Assignee

Sort

- Document need to upgrade plugins during rolling/ full cluster restarts** `adoptme` `docs`
#15389 opened 2 hours ago by clintongormley
- Remove the object notation for string fields** `:Mapping` `breaking`
#15388 opened 2 hours ago by jpountz
- Documentation about translog fsync is a bit confusing** `docs`
#15387 opened 2 hours ago by jpountz
- Error on Elasticsearch 1.0.3 EndWithValue**
#15384 opened 3 hours ago by elliswood
- Change docs on "node client" to not use an in-memory node** `docs` `v2.2.0` `v3.0.0`
#15383 opened 9 hours ago by rjernst
- Config `index.mapper.dynamic: false` is not honored.** `:Mapping` `adoptme` `bug` `low hanging frull`
#15381 opened 15 hours ago by zamblauskas
- Dynamic mappings fail when a single document generates inconsistent mapping updates** `:Mapping` `bug`
#15377 opened 18 hours ago by jpountz
- XContentBuilder throws NumberFormatException for Date field**
#15375 opened 20 hours ago by drmaas

github



Customers

Global Customer Base

Hi-Tech



Finance



Telco



Retail



!!



Elasticsearch is the backbone across all of Wikimedia's sites, powering billions of real-time user prefix and full-text searches every day. !!

Use Case	Search, Logging, Analytics
Products	Elasticsearch, Logstash, Kibana

Chad Horohoe
Software Engineering

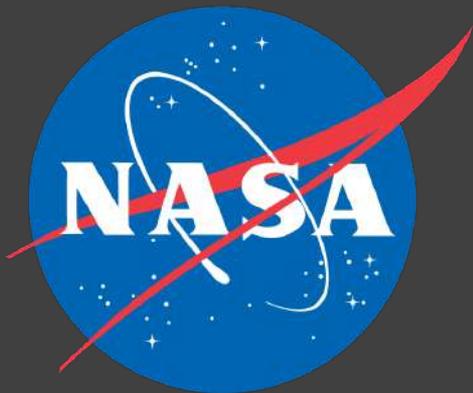
mozilla

”

Elasticsearch, Logstash, and Kibana allow for real-time indexing, search, and analytics for over 300 million events per day. This protects our network, services, and systems from security threats. ”

Use Case	Search, Logging, Analytics, Security
Products	Elasticsearch, Logstash, Kibana

Jeff Bryner
Security Engineer



”

With the Elastic Stack, we log more than 30K messages and 100K documents four times every day from the Mars Rover to optimize our space missions.”

Use Case	Search, Logging, Analytics
Products	Elasticsearch, Logstash, Kibana

Dan Isla
Data Scientist

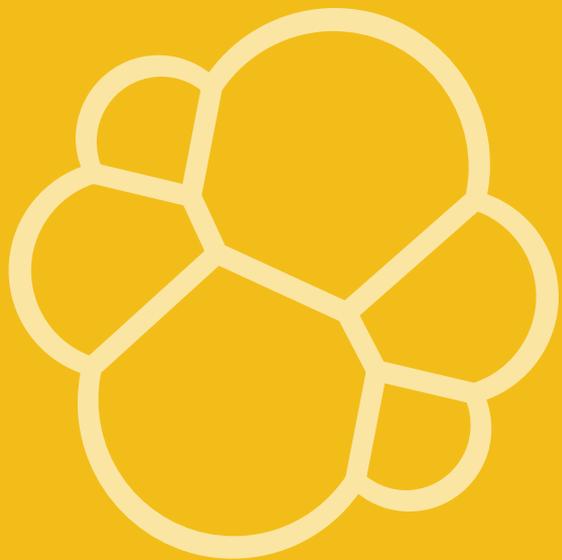


”

Using Elasticsearch, we index more than 500 billion documents for real-time logging and analytics for our mission critical applications.”

Use Case	Logging, Analytics
Products	Elasticsearch, Logstash

Bhaskar Karambelkar
Sr. Security Data Scientist



Roundup



Ease-of-use



Minimal dependencies



Extensibility



Consistency



Flexibility

The Elastic Stack

 Elastic Stack

Plugins

Monitoring

Security

Alerting

User Interface

Kibana

Store, Index,
& Analyze

Elasticsearch

Ingest

Logstash

Beats

Hosted Service

Found: Elasticsearch as a Service



Thank You

We're hiring

<https://www.elastic.co/about/careers>

We're helping

<https://www.elastic.co/subscriptions>