Elastic Stack

Brief introduction and what's new

@medcl

About me

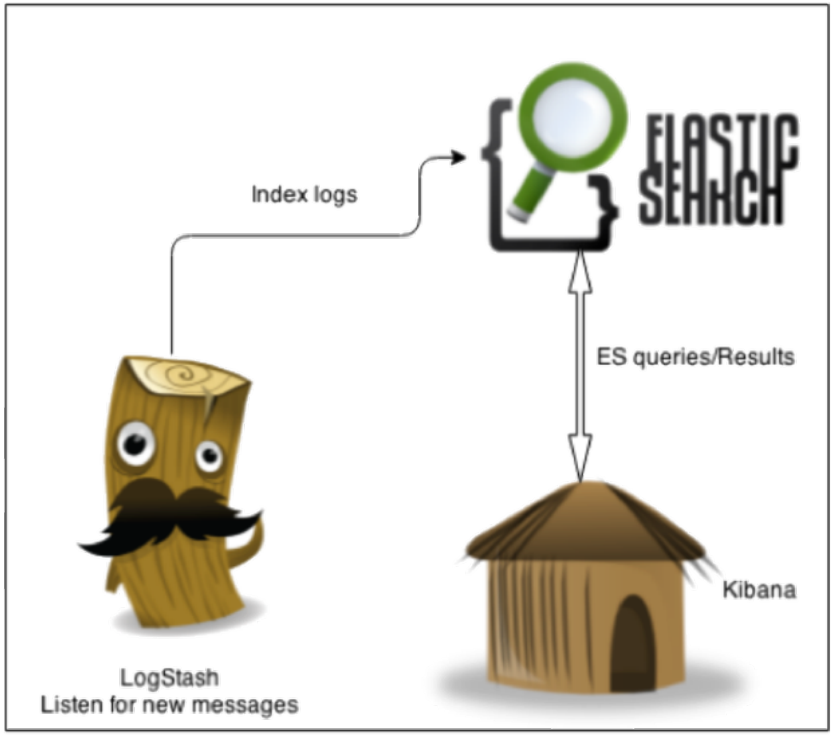
- Medcl，曾勇（Zeng Yong）
- Developer @ Elastic
  - Follow Elasticsearch since v0.5, 2010
  - Joined Elastic since September, 2015
  - Now in Beats team
- @medcl
- medcl@elastic.co
- http://github.com/medcl
- Based in Changsha, Hunan, China

elastic

What's Elastic?

- A distributed startup company, since 2012
  - HQ: Mountain View, CA AND Amsterdam, Netherlands
  - With employees in 27 countries (and counting), spread across 18 time zones, speaking over 30 languages
- We are working on Open Source projects!
  - (Luckily some of them are popular, eg:elasticsearch)
- Offering support Subscription, X-pack, Cloud and Trainings
- Find us on: https://github.com/elastic and https://www.elastic.co

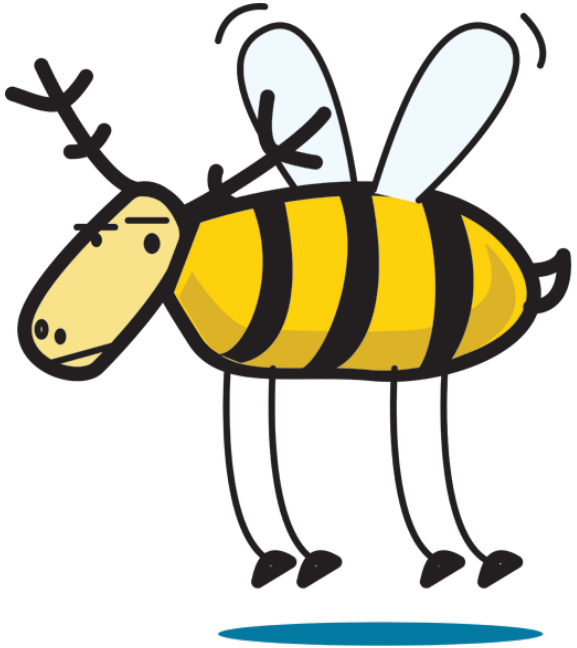
听说过 “ELK” 么?



elastic

ELK is out!

Beats & Packetbeat

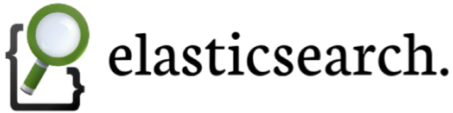


ELKB? BELK? LKBE? BKEL?

elastic

Logo


ELASTIC
SEARCH


elasticsearch.


elasticsearch.


elasticsearch.


Kibana


kibana


kibana


kibana

It’s complicated

es	Nov 5, 2014 **1.4**	May 23, 2015 **1.5**	Jun 9, 2015 **1.6**	Jul 16, 2015 **1.7**
kibana	Feb 19, 2015 **4.0**		Jun 10, 2015 **4.1**	
ls	May 14th, 2015 **1.5**			
beats	May 27th, 2015 **1.0 Beta 1**		July 13th, 2015 **1.0 Beta 2**	Sept 4 th, 2015 **1.0 Beta 3**

elastic

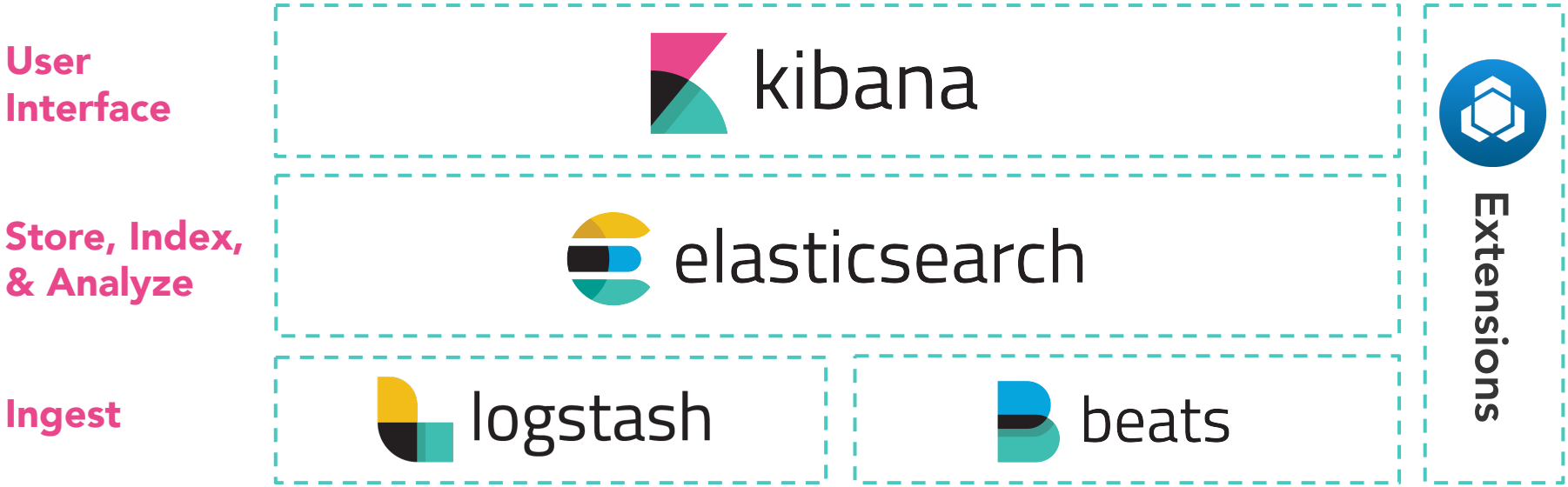
Release Bonanza

	Oct 28th	Nov 21st	Feb 2nd
es	2.0	2.1	2.2
kibana	4.2	4.3	4.4
ls	2.0	2.1	2.2
beats		1.0	1.1

It’s time to unite!

elastic

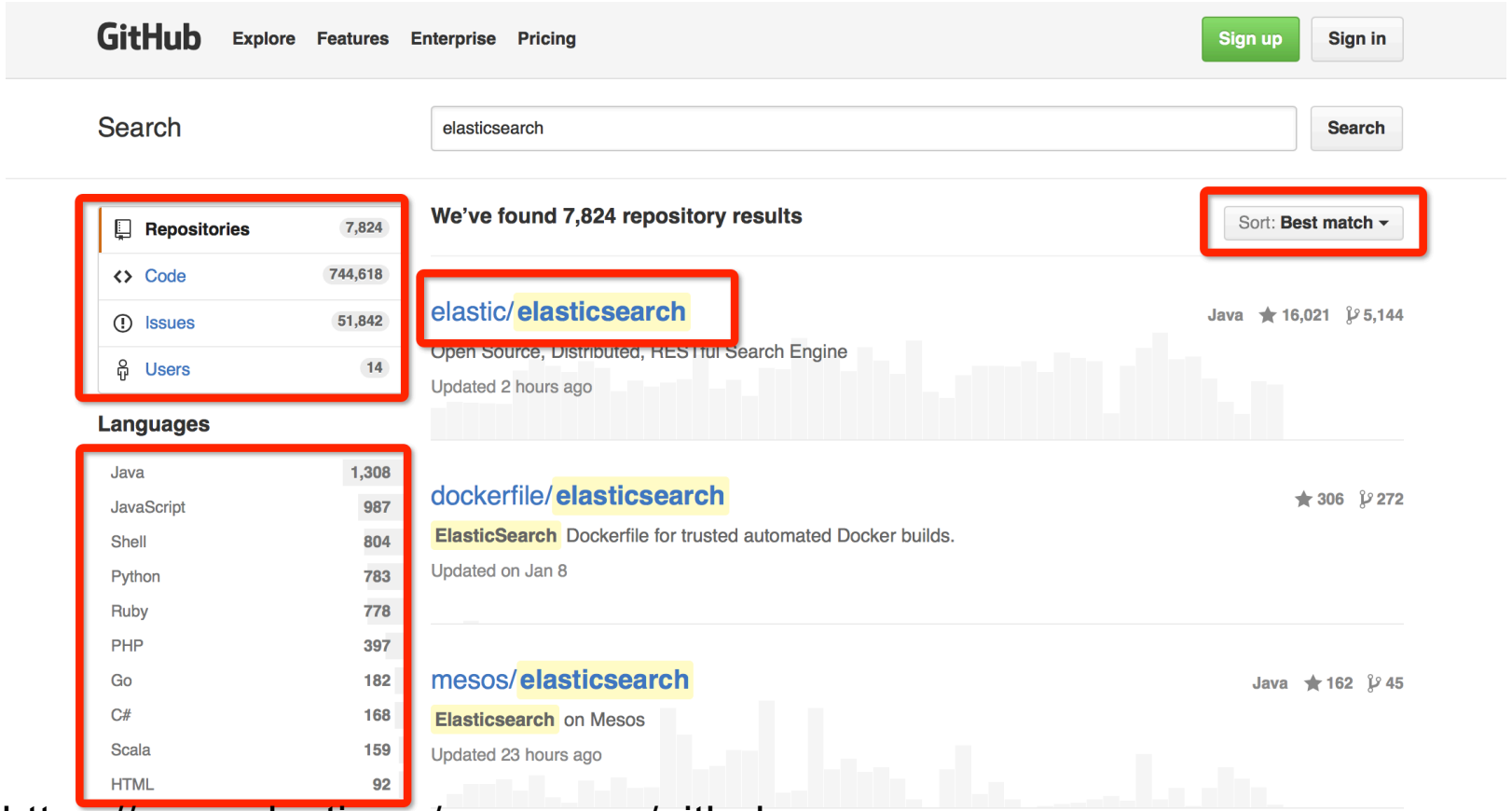
The “Elastic Stack”



elastic

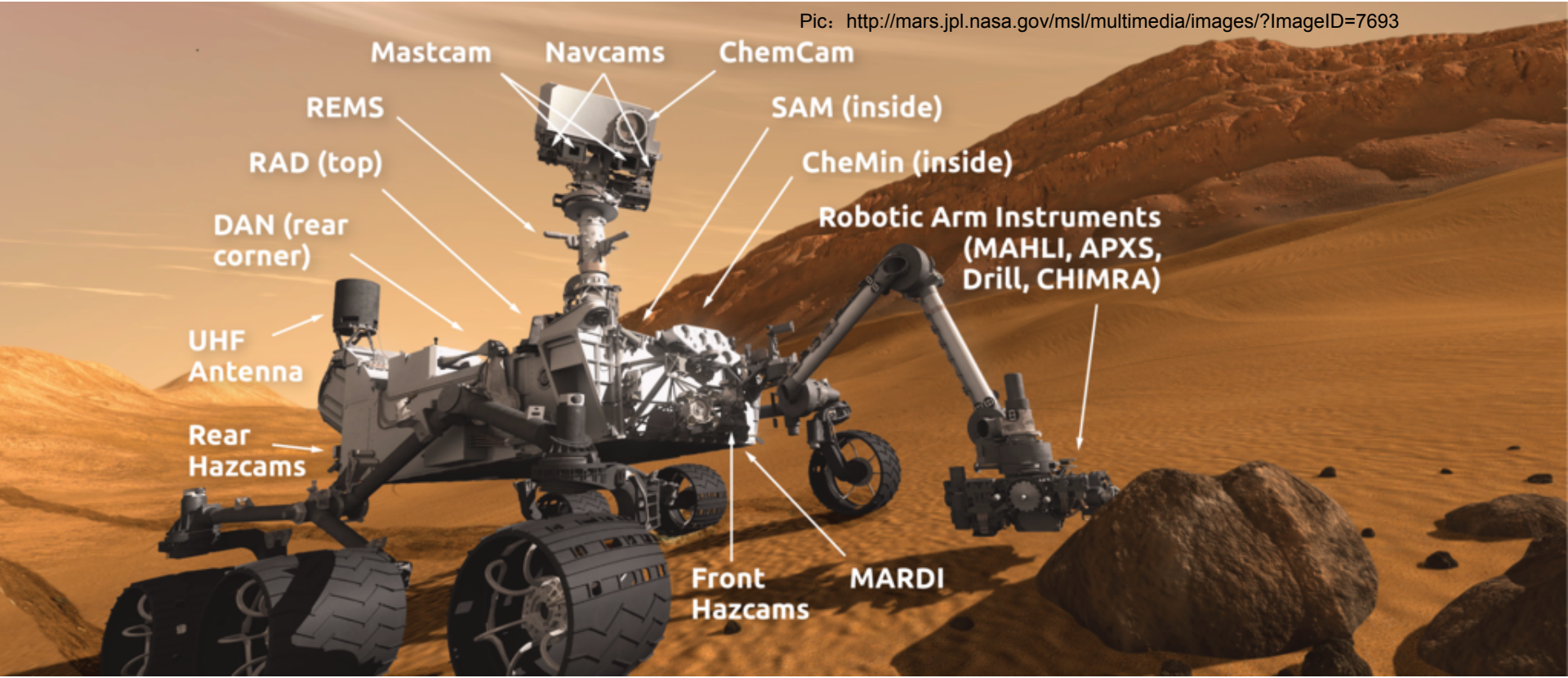
Elastic Stack
能做什么?

Github: Enable Powerful Search For Both End-Users And Developers



https://www.elastic.co/use-cases/github

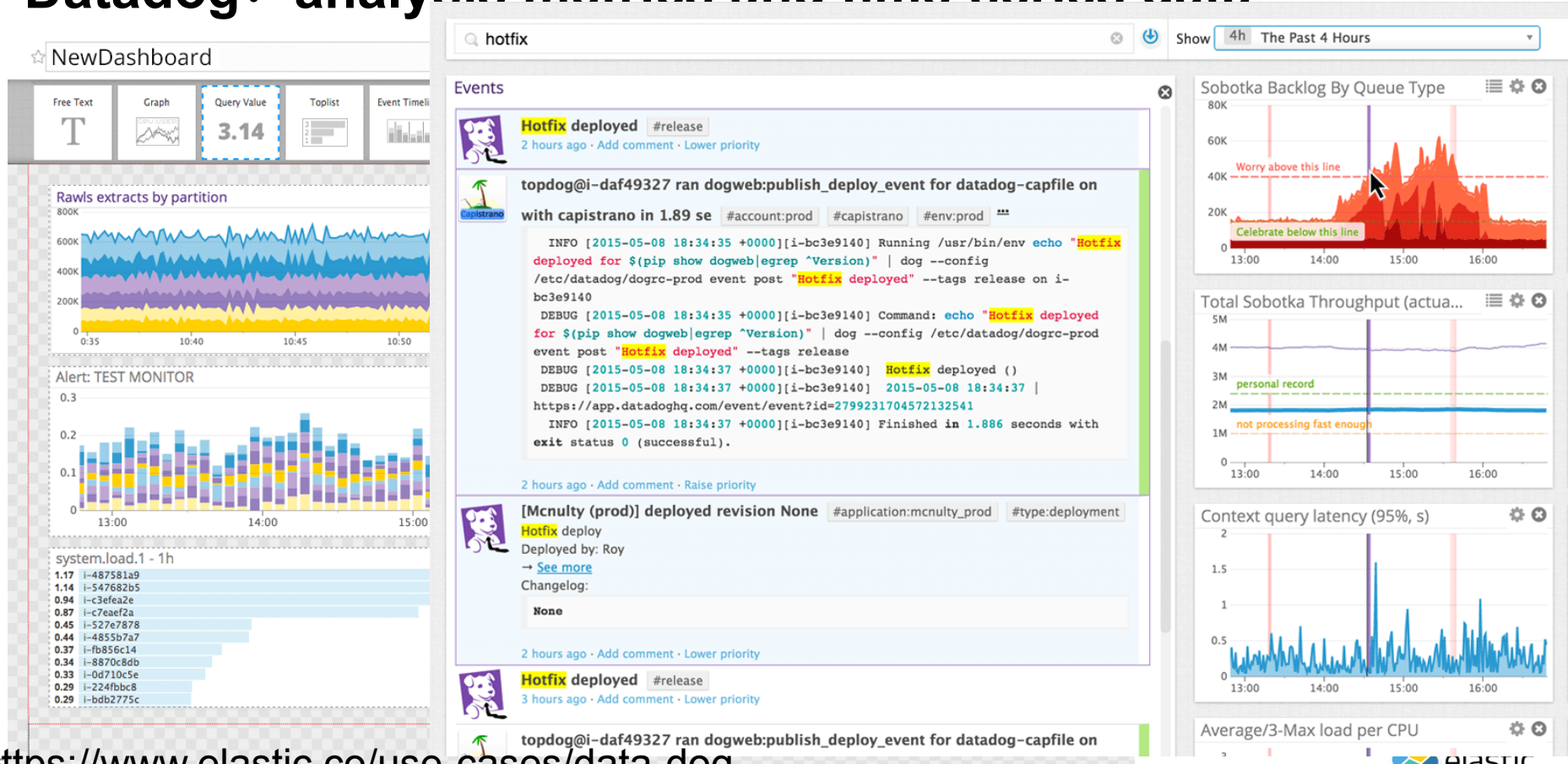
NASA: Unlocking Interplanetary Datasets with Real-Time Search



https://www.elastic.co/elasticon/2015/sf/unlocking-interplanetary-datasets-with-real-time-search

elastic

Datadog：analysis metrics and time-series data

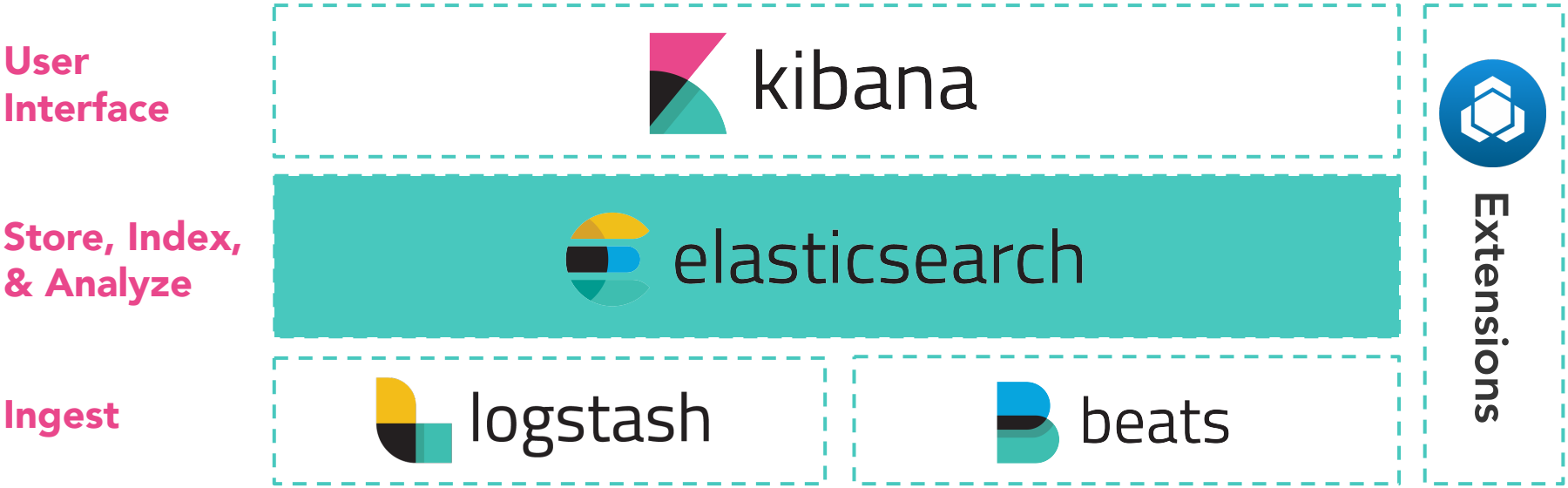


https://www.elastic.co/use-cases/data-dog

更多案例

https://www.elastic.co/use-cases

The “Elastic Stack”



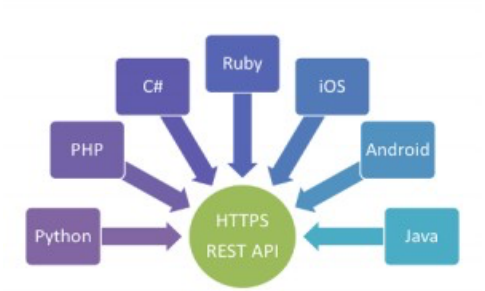
elastic

elasticsearch

Elasticsearch is an open source, distributed, scalable, highly available, document-oriented, RESTful, full text search engine with real-time search and analytics capabilities

Thomson Reuters: “107 clusters ~1747 nodes” @Elastic{ON}16
https://speakerdeck.com/elastic/thomson-reuters-research-journalism-finance-and-elastic

Netflix：”~150 clusters totaling ~3,500 nodes hosting ~1.3 PB of data”
http://techblog.netflix.com/2016/02/evolution-of-netflix-data-pipeline.html?m=1



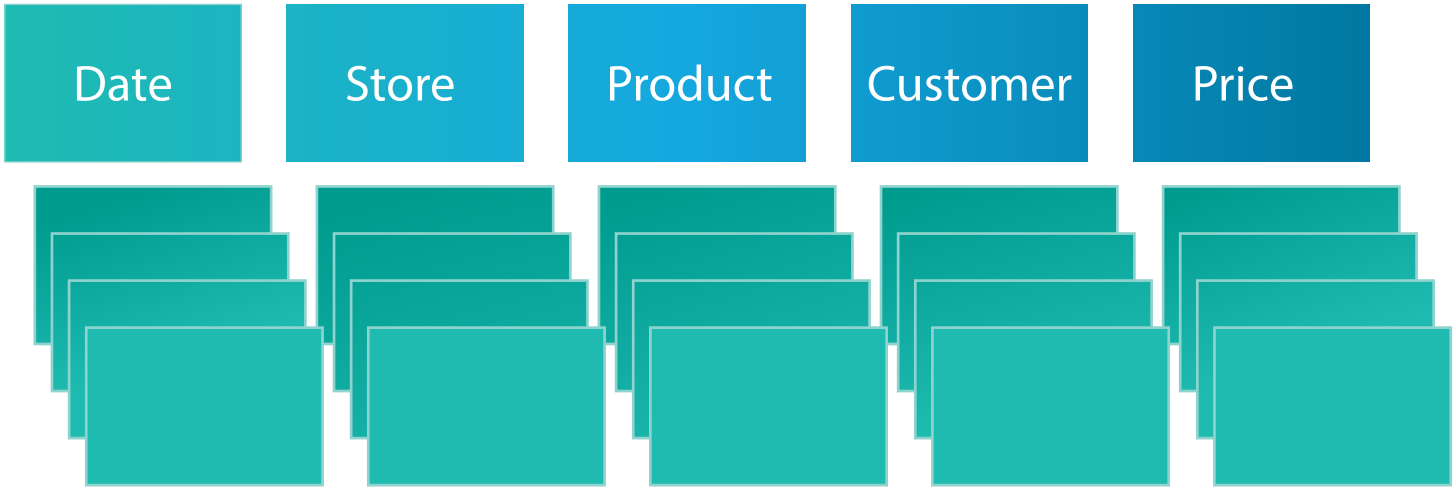
JSON

- Real-time analytics
- Time series data analytics
- Logging analytics
- Security analytics
- Fraud detection
- Prediction modeling
- Recommendations
- …

http://github.com/elastic/elasticsearch

elastic

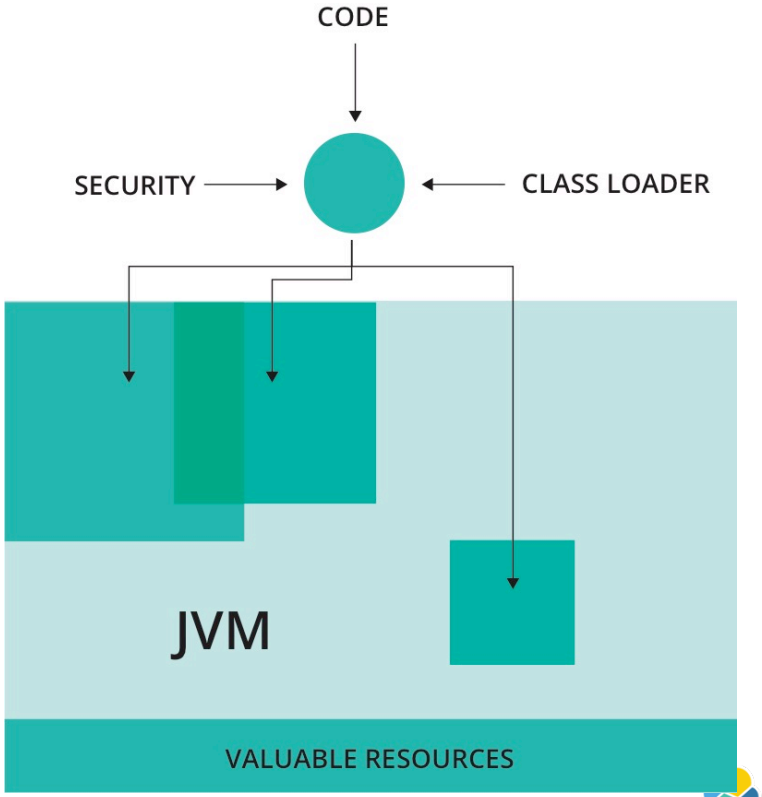
Columnar Store



elastic

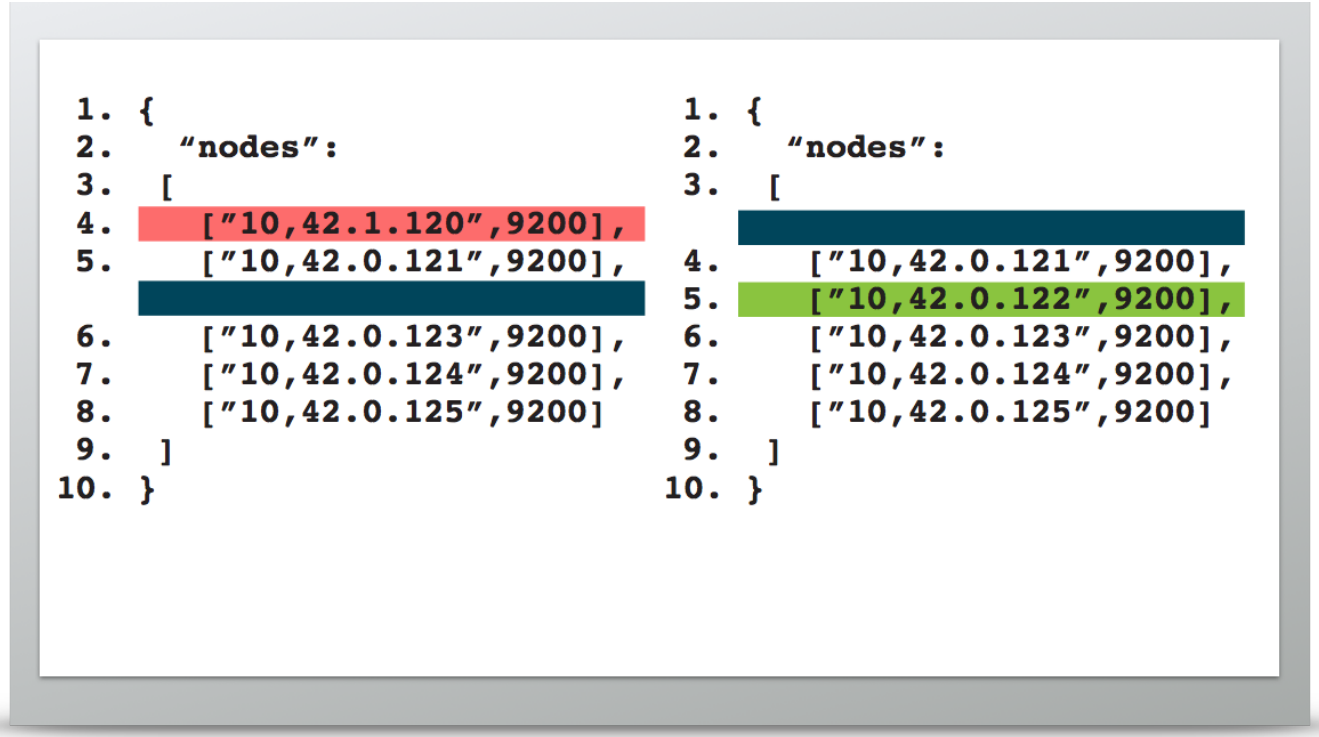
Java Security Manager

- One does not simply **fork a process**



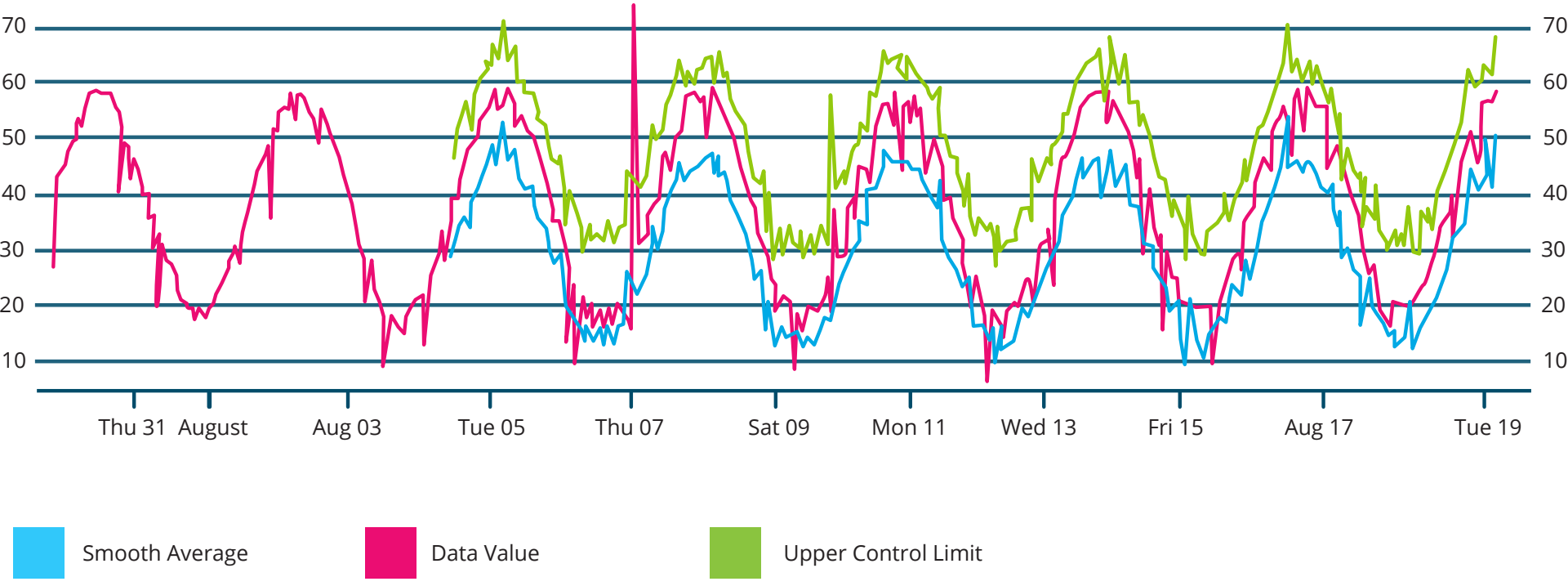
elastic

Cluster State Diffs



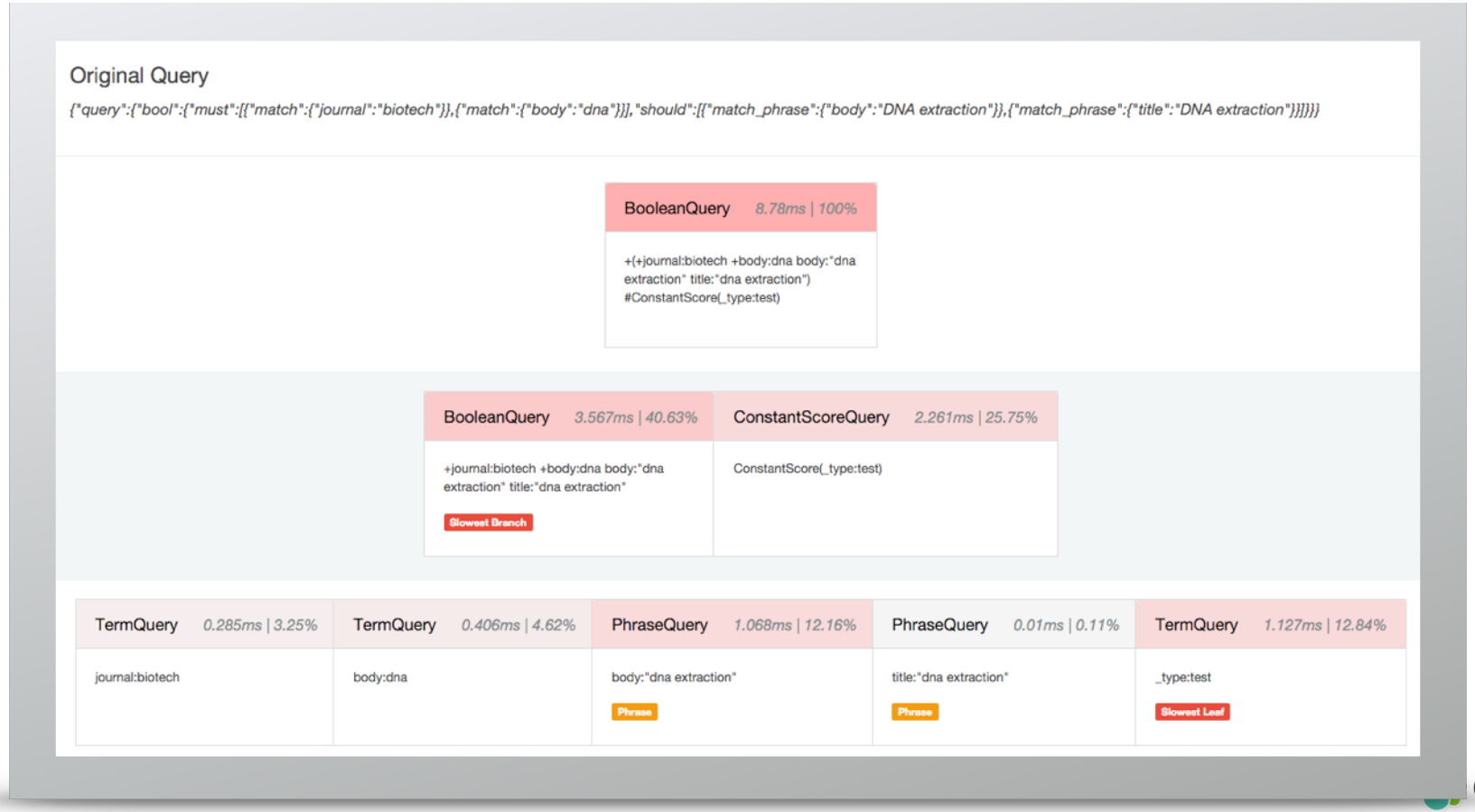
elastic

Pipeline Aggregations



elastic

Profile API



Painless Scripting

- Dynamic/ Static

```
def first = input.doc.first_name.0;
def last  = input.doc.last_name.0;
return first + " " + last;
```

```
String first = (String)((List)((Map)input.get("doc")).get("first_name")).get(0);
String last  = (String)((List)((Map)input.get("doc")).get("last_name")).get(0);
return first + " " + last;
```

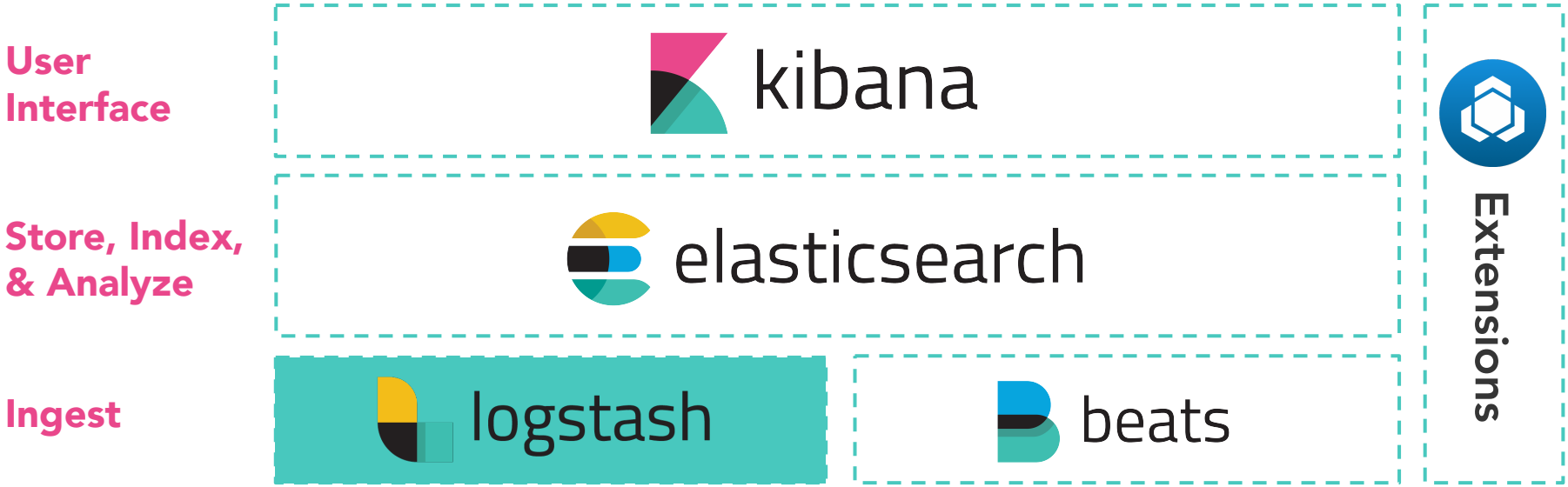
it is **ten times faster!**

- Reindex API
  - The Reindex API makes upgrading Elasticsearch easy
  - Change problematic mappings & upgrade to the latest / greatest
  - An important step towards 5.0 and there is a detailed blog post
- Task Management API
  - Manage long running tasks in Elasticsearch
  - A stepping stone towards future capabilities

What's more

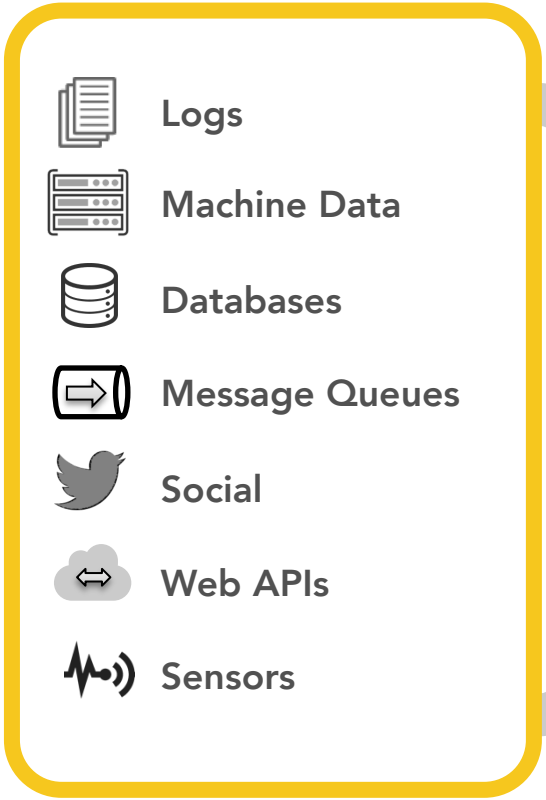
- Plugincommand
  - bin/elasticsearch-plugin
- Lucene 6
  - DimensionalPoints/Multi-dimensionalpoints
    - numeric, date, and geospatial fields will be:50% disk; 50% index time； 75% search time
- Ingest Node
  - grok, split, convert, and date etc.
- Text/Keyword to Replace Strings
- Instant aggregations
  - Date queries(aggregations) now cacheable
- Settings Validation
- Safety in production
- IndexName -> UUID
- Depreated logging

The “Elastic Stack”



elastic

Logstash: Collect from diverse inputs



logstash

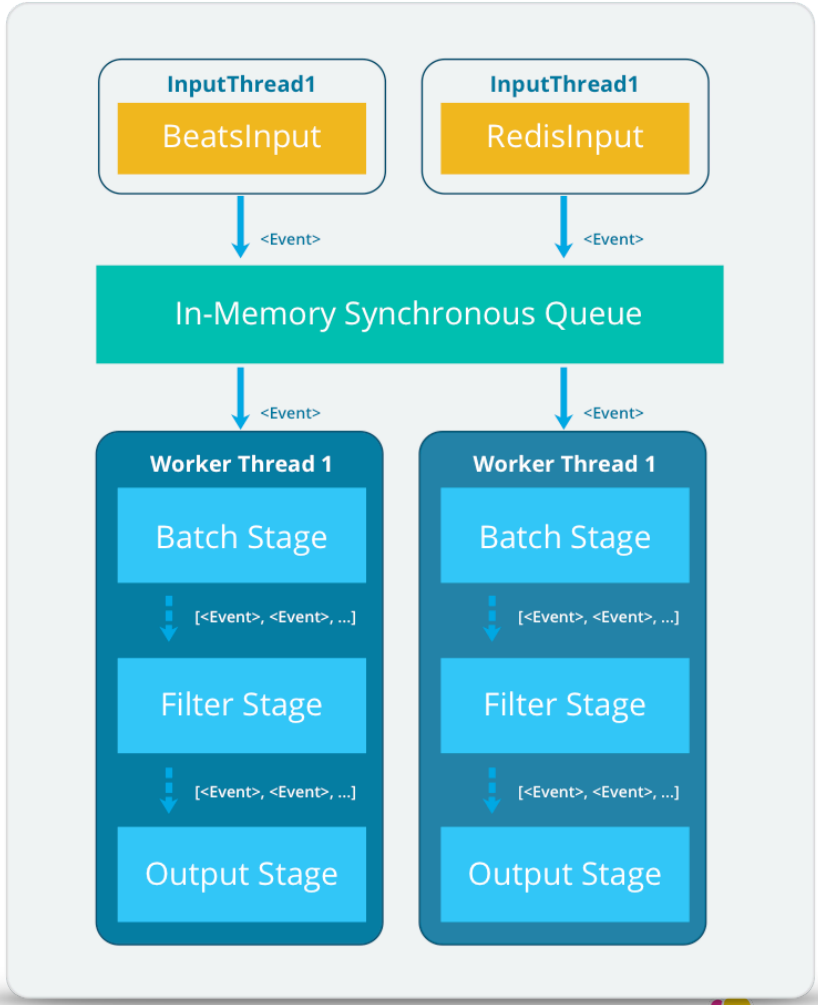
- Collects diverse sources
  - Logs + many others
  - Over 200 plugins
- Connects with live streams
  - Real-Time data
  - Wire / Transaction data
  - Full-Packet Network Capture

http://github.com/elastic/logstash

elastic

New pipeline Architecture

Faster, more reliable pipeline



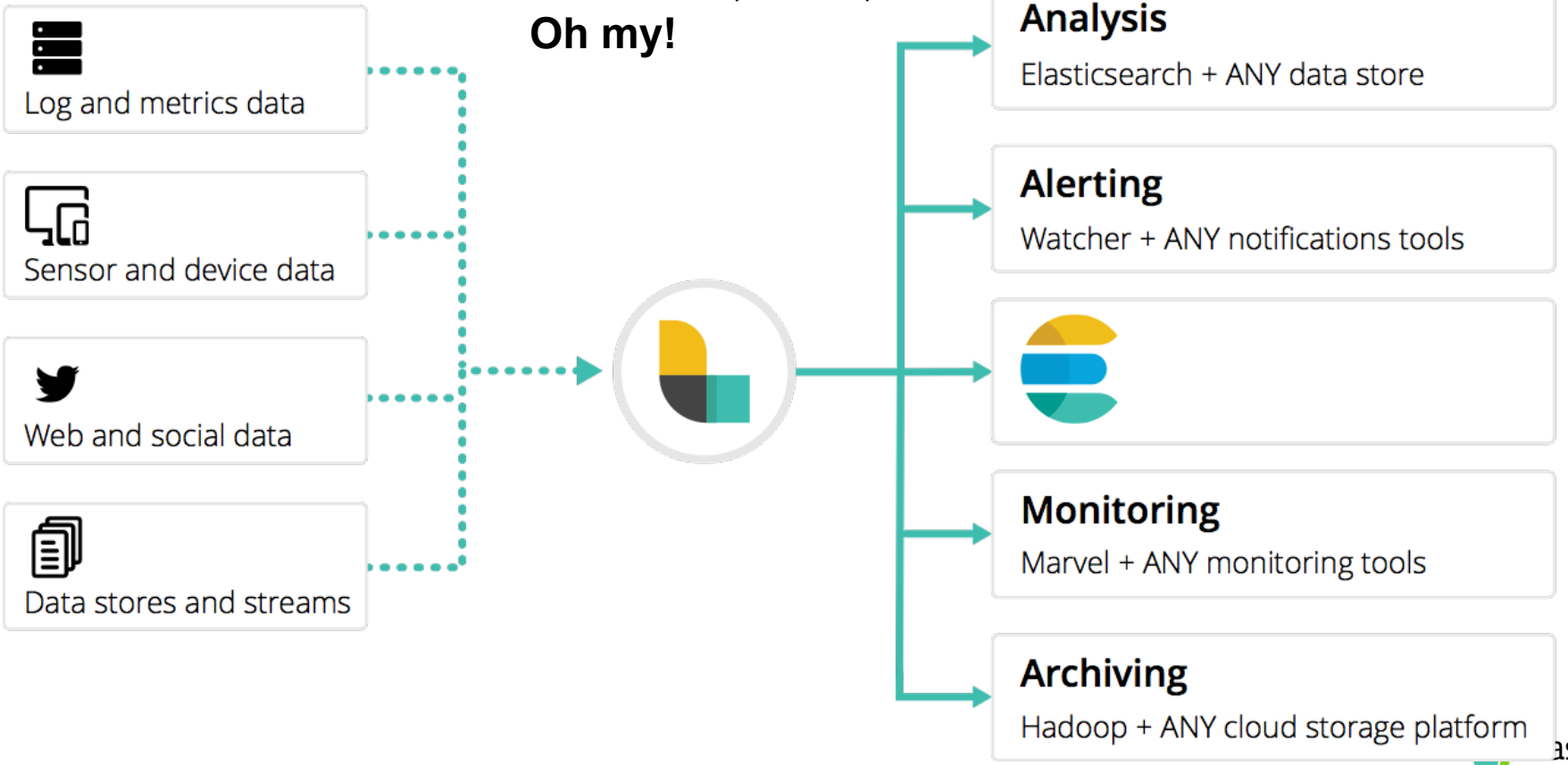
elastic

Config Reload

```
input {
    file {
        path => "/Users/Elastic/Logstash/configs/webapp_logs/app.log"
    }
}

filter {
    if [metadata][logstash_plugin] == "true" {
        mutate {
            add_field => { plugin_type => "%{[metadata][plugin_type]}" }
        }
    }
    else {
        drop {}
    }
}

output {
    stdout { codec => rubydebug }

    elasticsearch {
        hosts => "localhost"
        user => "es_admin"
        password => "logstash+love"
    }
}
```

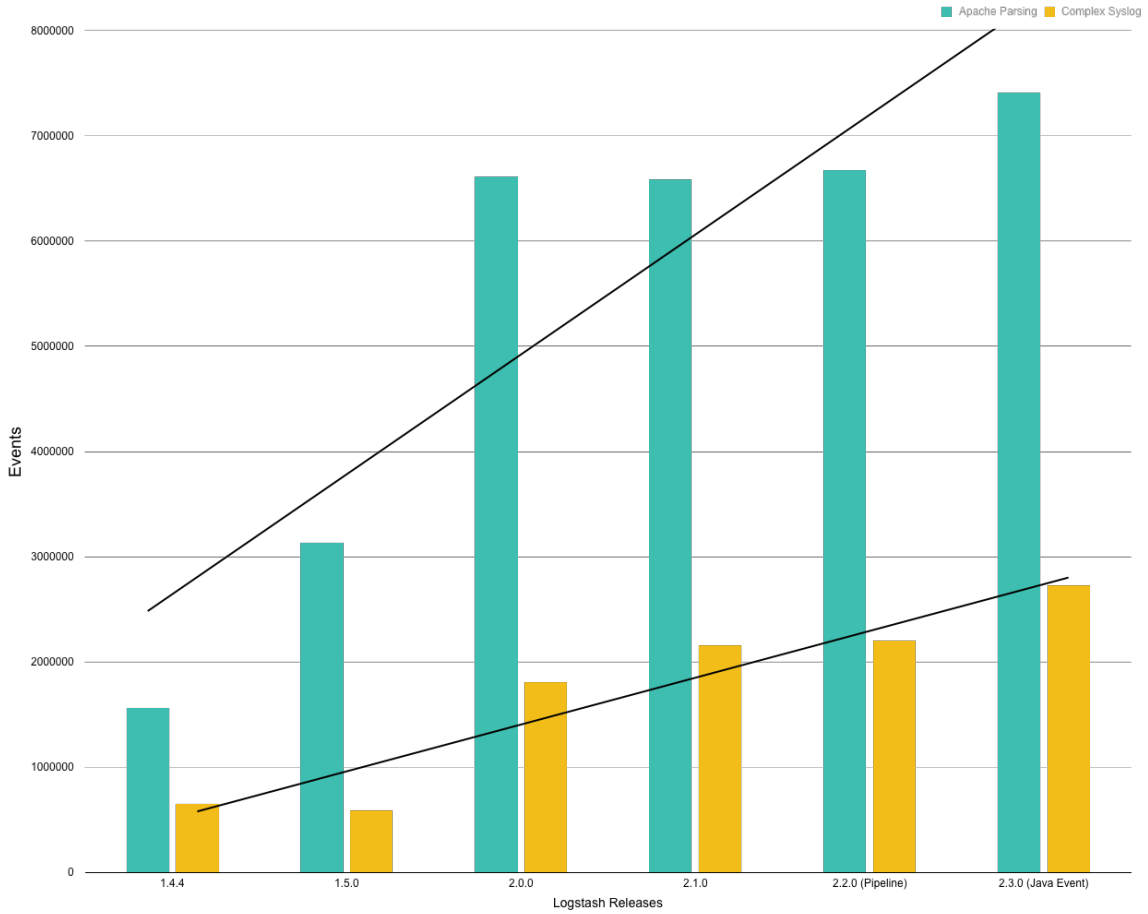
Plugins

Kafka, HDFS,
Salesforce, HTTP,
Oh my!



Performance

Now you can
grok faster



elastic

Monitoring API

```
curl localho
```

```
{
  "events" :
    "in" : 1
    "filtere
    "out" :
  },
  "jvm" : {
    "timesta
    "uptime_
    "mem" :
      "heap_
      "heap_
      ....
```

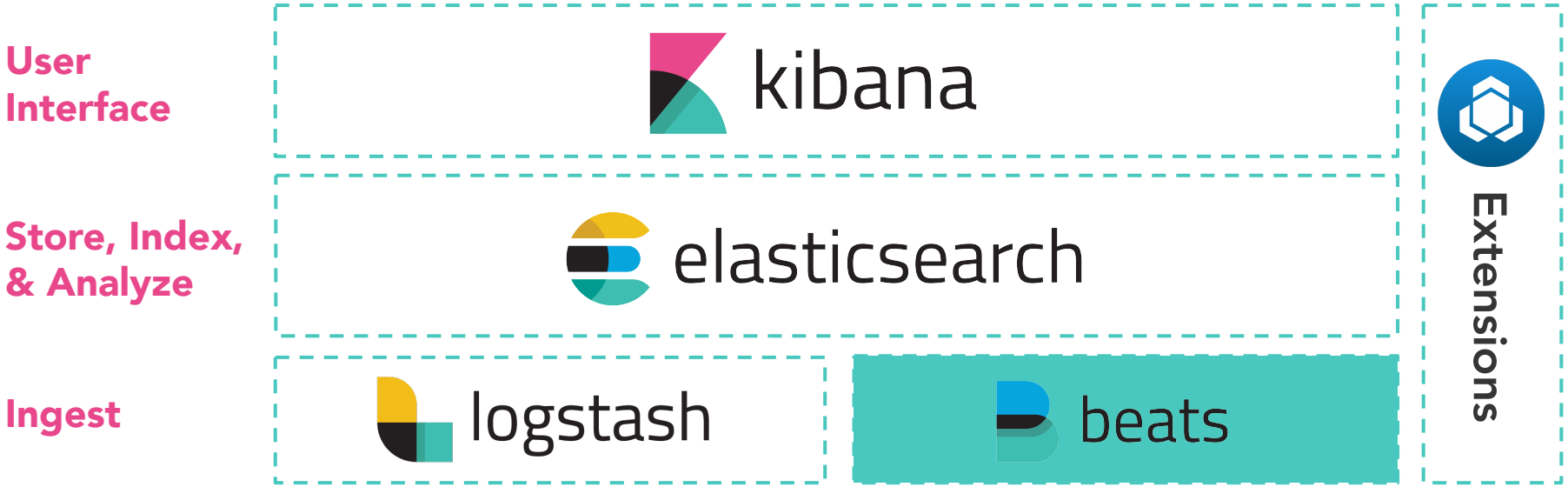
```
curl localhost:9600/_node/hot_threads?human
```

```
Hot threads at 2016-03-30T20:08:22-07:00, busiestThreads=3:
    5.22 % of of cpu usage by waiting thread named '[main]>worker3'
       java.lang.Object.wait(Native Method)
       java.lang.Object.wait(Object.java:460)
       org.jruby.RubyThread$SleepTask.run(RubyThread.java:1050)
       org.jruby.RubyThread.executeBlockingTask(RubyThread.java:1066)
       org.jruby.RubyThread.wait_timeout(RubyThread.java:1414)
       org.jruby.ext.thread.Queue.pop(Queue.java:152)
       org.jruby.ext.thread.Queue.pop(Queue.java:127)
       org.jruby.ext.thread.SizedQueue.pop(SizedQueue.java:111)
       org.jruby.ext.thread.SizedQueue$INVOKER$i$pop.call(SizedQueue$INVOK
       org.jruby.runtime.callsite.CachingCallSite.call(CachingCallSite.jav
    2.44 % of of cpu usage by timed_waiting thread named '[main]-pipeline-
       java.lang.Object.wait(Native Method)
       ....
```

What’s more

- Plugin command
  - bin/logstash-plugin
- Kafka0.9 support
  - Support SSLencryptionandclientauth

The “Elastic Stack”



elastic

beats

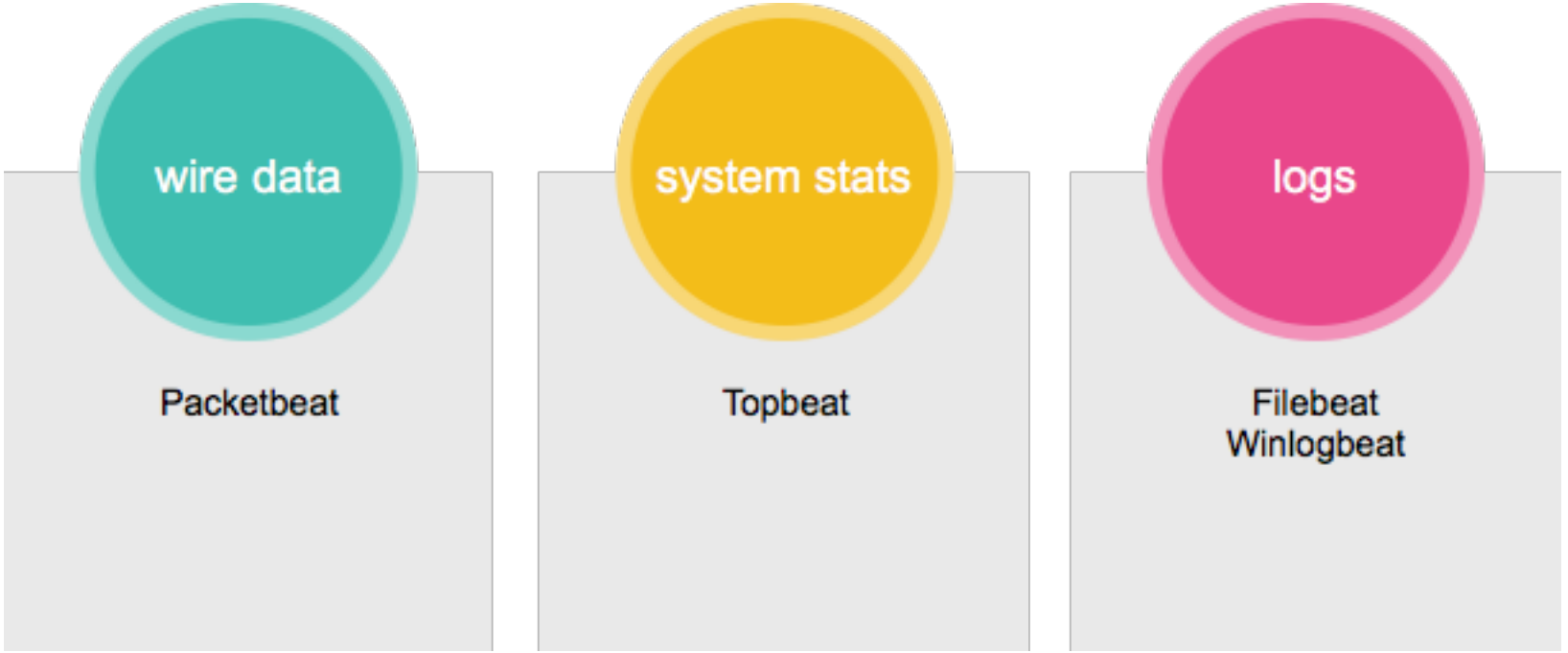
- Beats are lightweight shippers that collect and ship all kinds of **operational data** to Elasticsearch
  - Small application
  - Install as agent on your servers
  - Written in **Golang**
  - No runtime dependencies
  - Single purpose

http://github.com/elastic/beats

https://www.flickr.com/photos/8barbikes/17256970434/

elastic

Examples of operational data



elastic

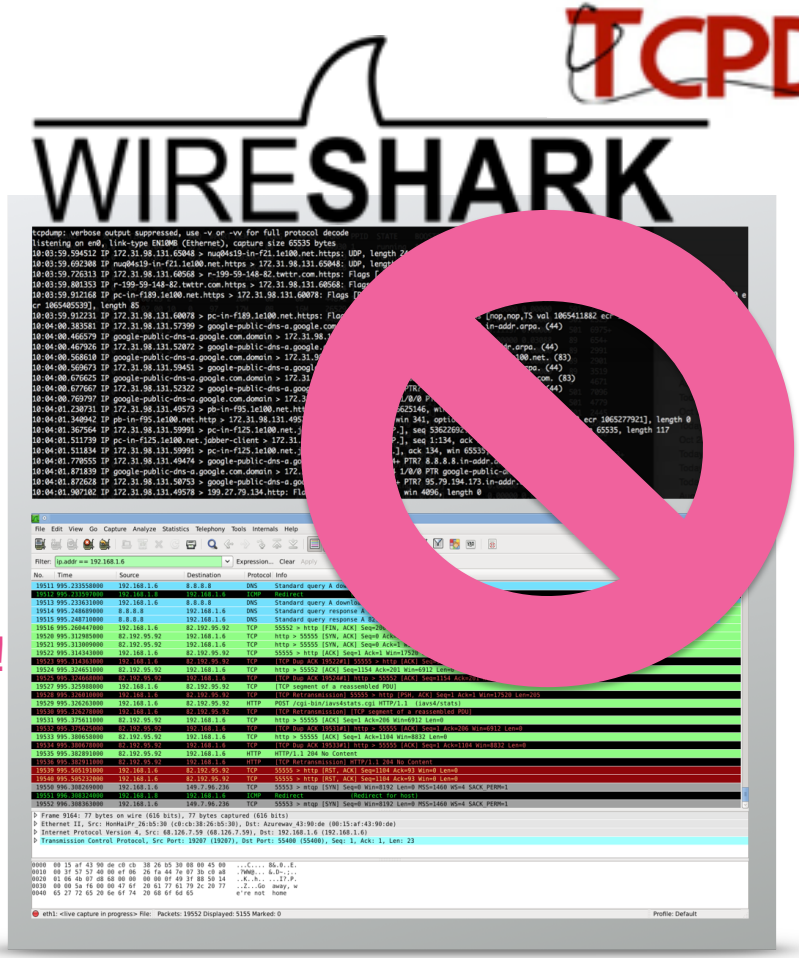
Packetbeat

Sniffs the traffic
between your servers,
parses the **application-level**
protocols on the fly.

Built-in protocols：

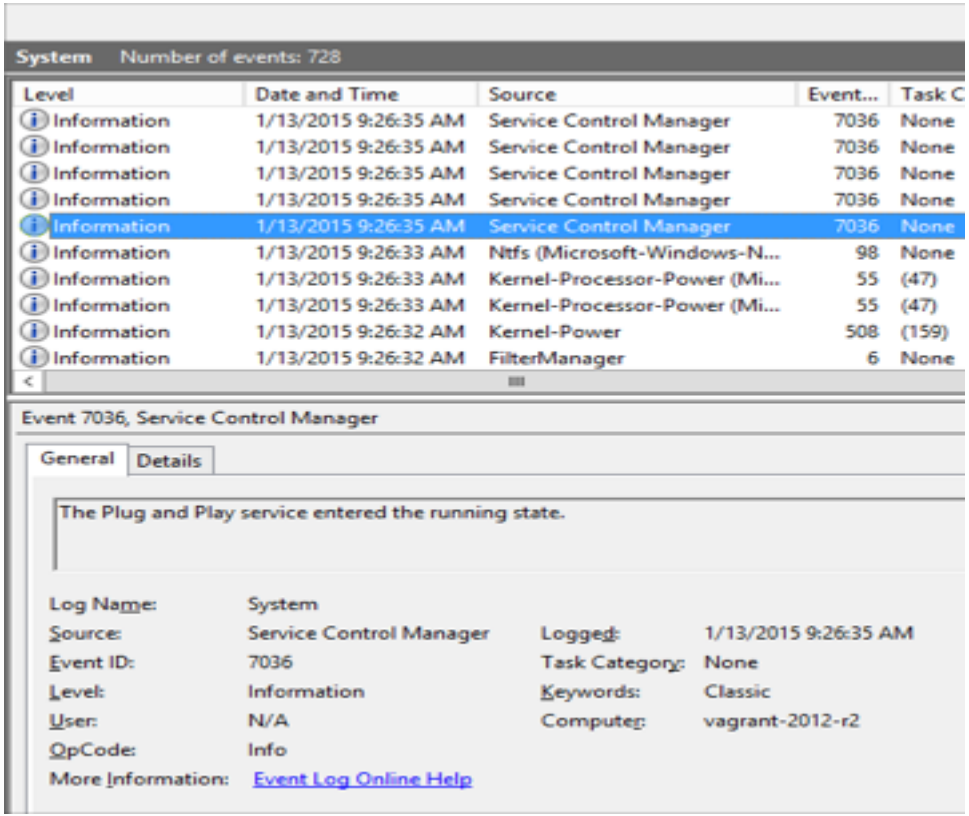
- HTTP
- MySQL
- PostgreSQL
- Redis
- Thrift-RPC
- MongoDB
- DNS
- Memcache
- …

Let’s go realtime!



elastic

winlogbeat!



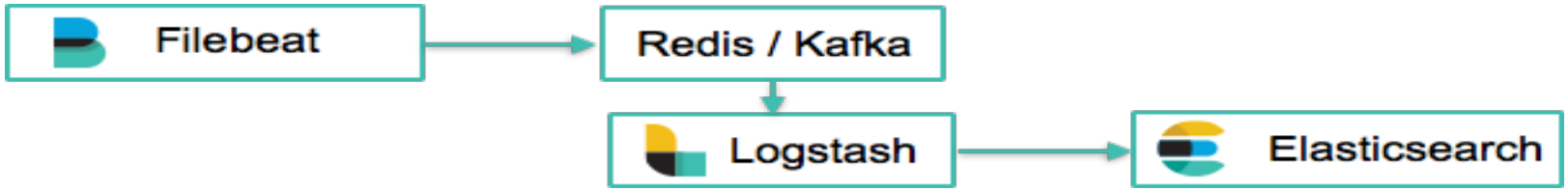
Forwards Windows Event logs to Elasticsearch

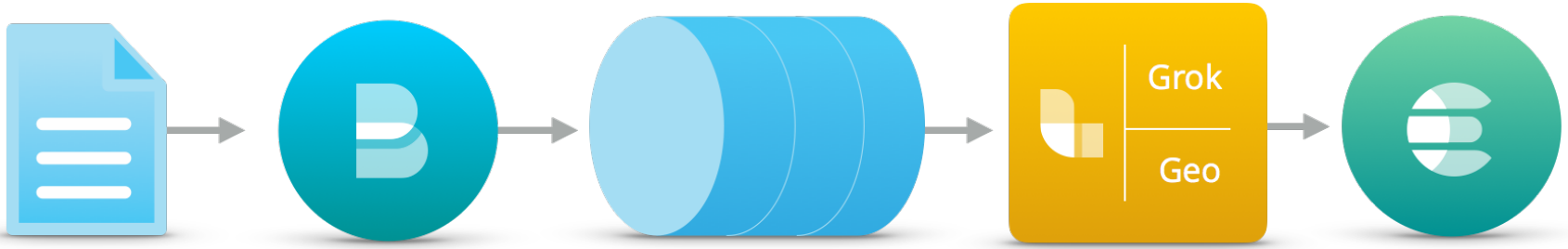
elastic

Filebeat

A more lightweight log shipper

- **Generic filtering**
  Flexibly reduce the amount of data sent of the wire and stored



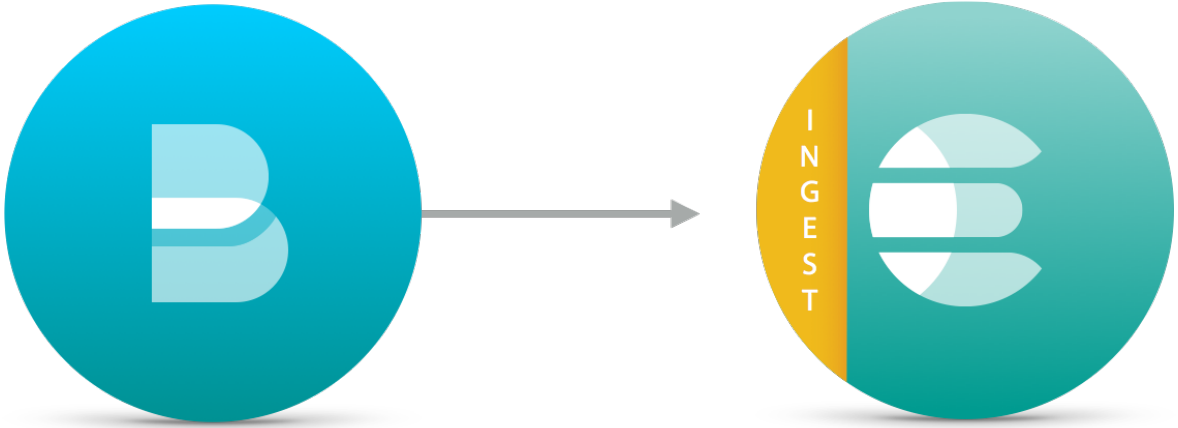


elastic

Simple things **should be simple**

Not
like this

Like *this*



elastic

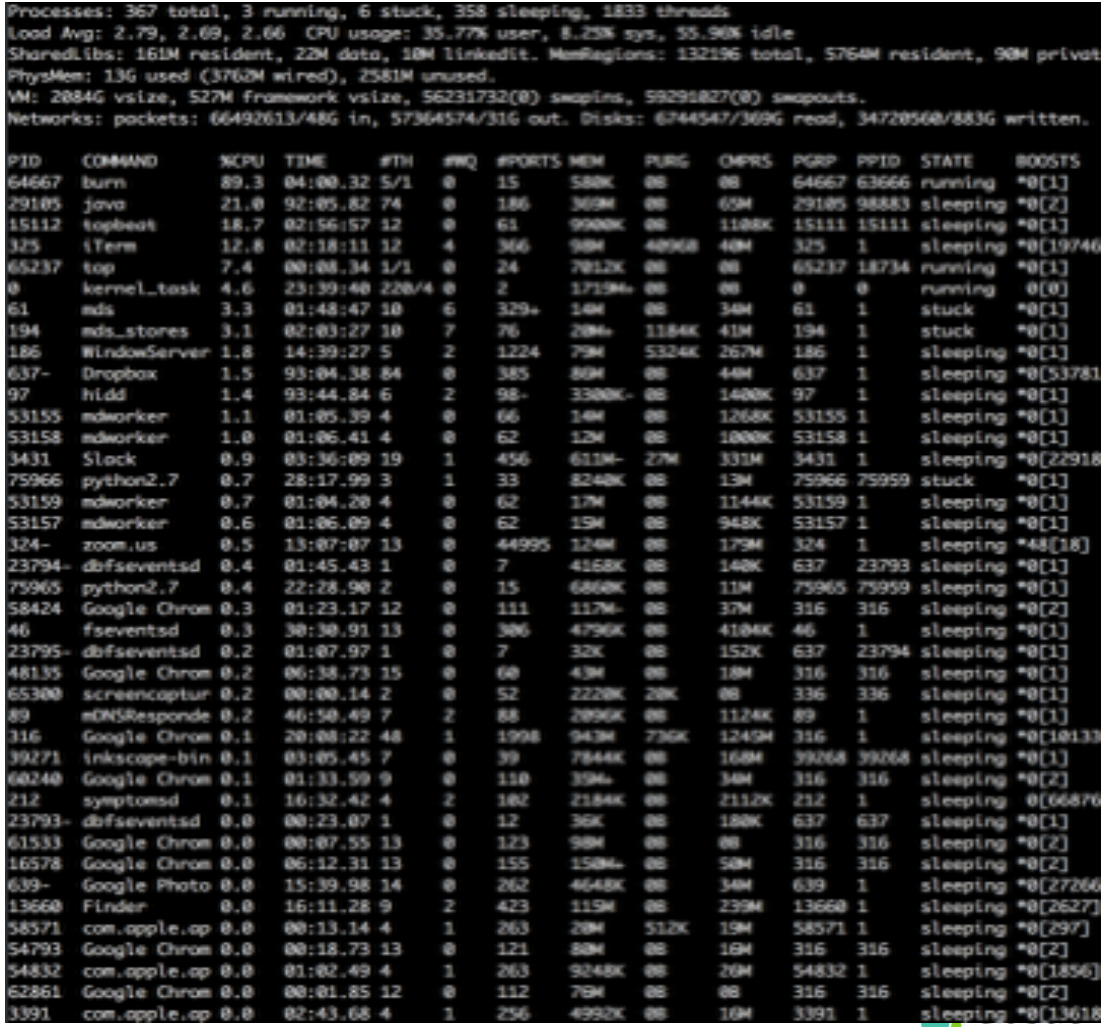
Topbeat

Like the Unix **top** command but sends the output periodically to Elasticsearch. Also works on Windows.

System wide
- system load
- total CPU usage
- …

Per process
- state
- name
- command line
- …

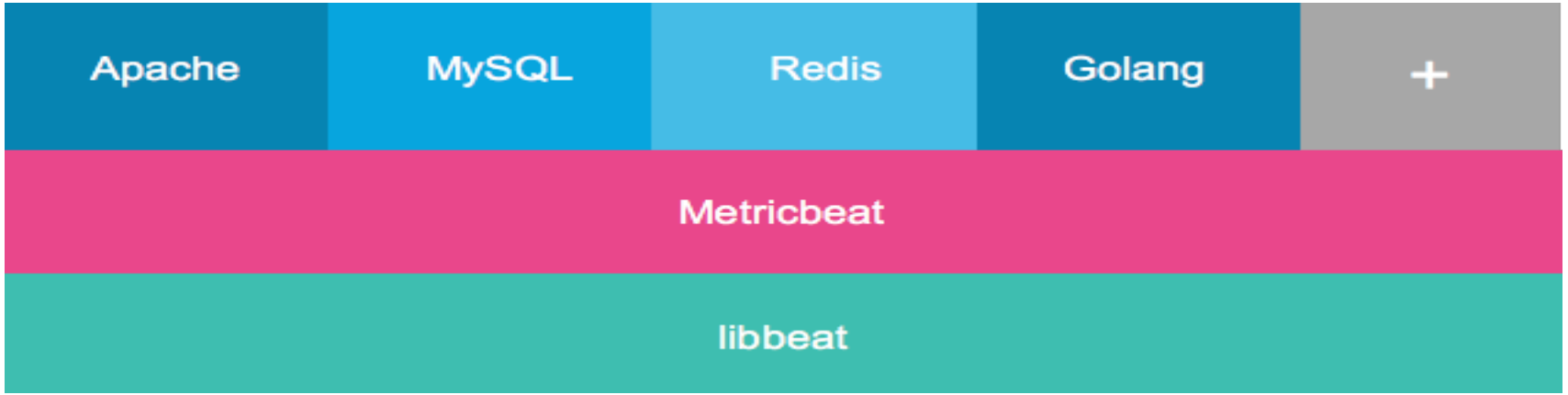
Disk usage
- available disks
- used, free space …



That's More!

Metricbeat: Connecting **Numb3rs**

- Listens to the internal "beat" of systems via APIs.



http://github.com/elastic/beat-generator/

elastic

Custom Fields and generic filtering

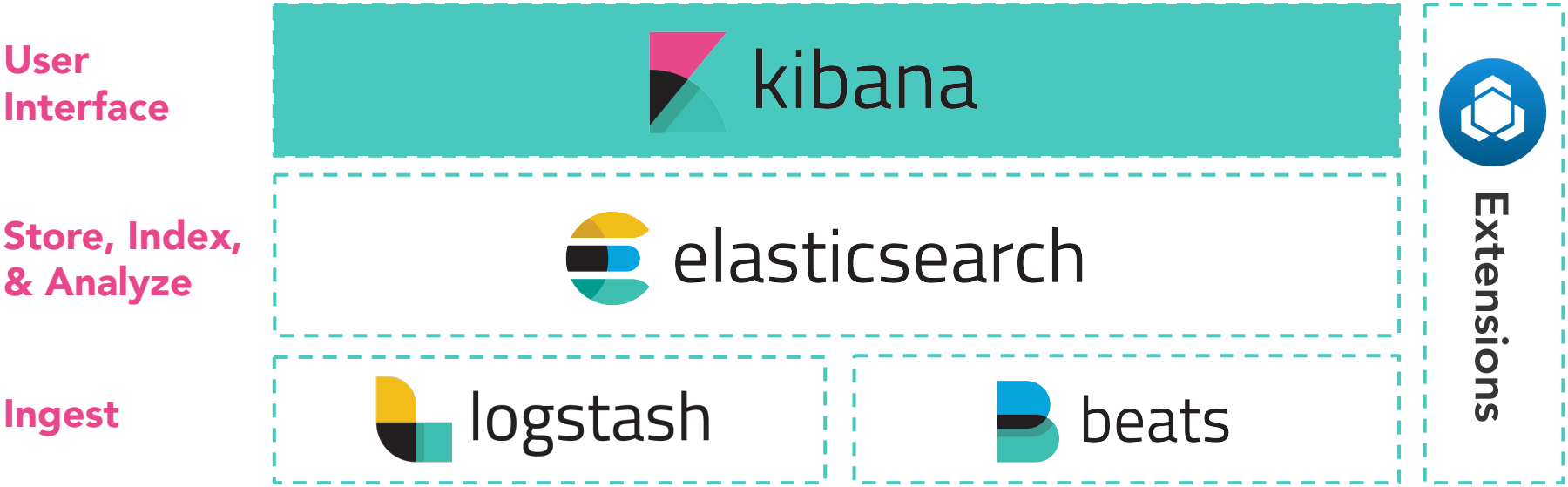
- Filtering the exported data

```
filter:
  - include_fields:
      fields:
        - bytes_in
        - bytes_out
        - ip
        - client_ip
        - dns.question.name
        - dns.question.etld_plus_one
        - dns.response_code
```

What's more

- Decode JSON from log lines
- Kafka output
  - Output to Kafka directly
- Integration with IngestNode
  - set"pipelineparameter" in the Elasticsearch output config
- Support IP/TCP flows
  - Report statistics like packet/byte counts

The “Elastic Stack”



elastic

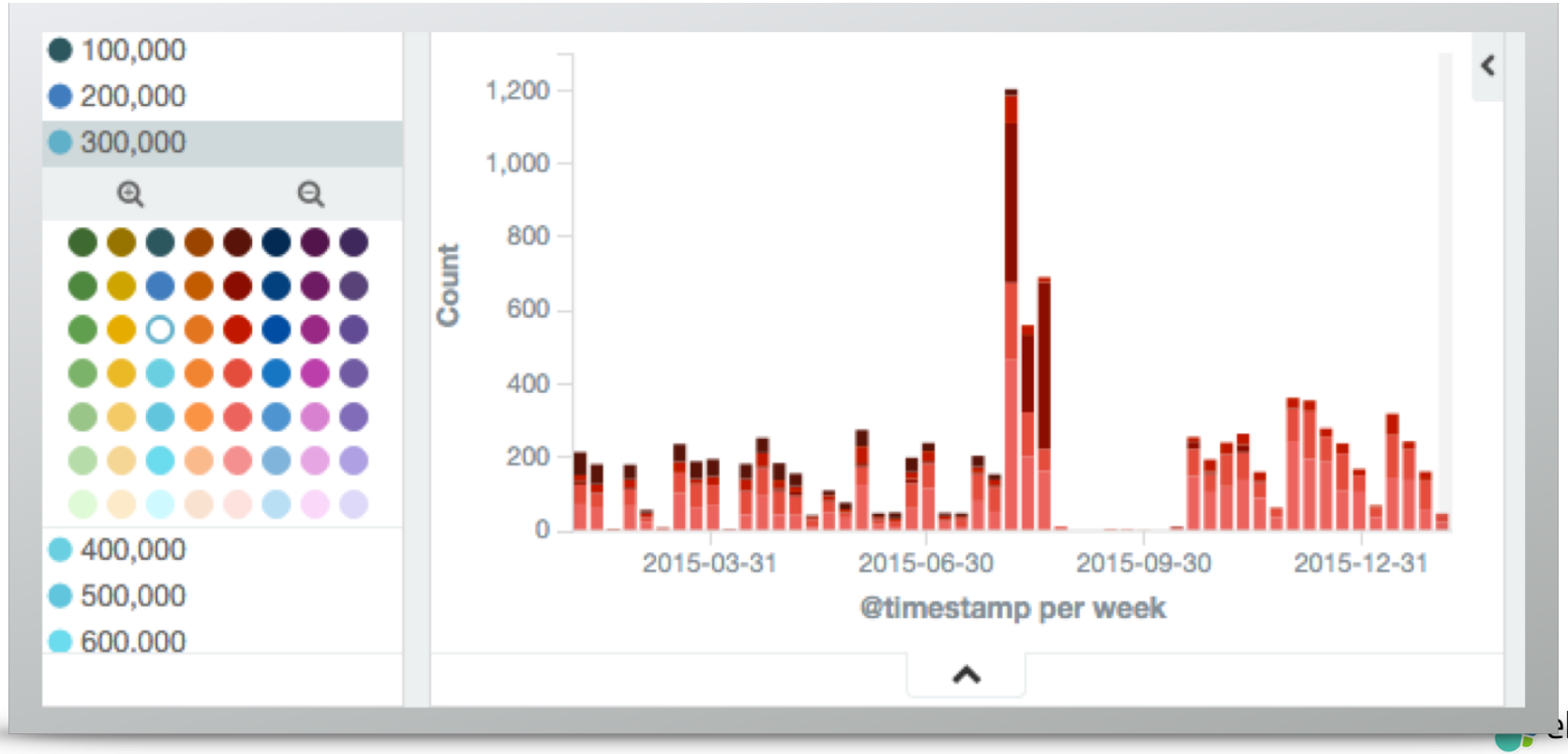
What's Kibana?

Kibana is an
open source analytics
and **visualization platform**
designed to work with
Elasticsearch.

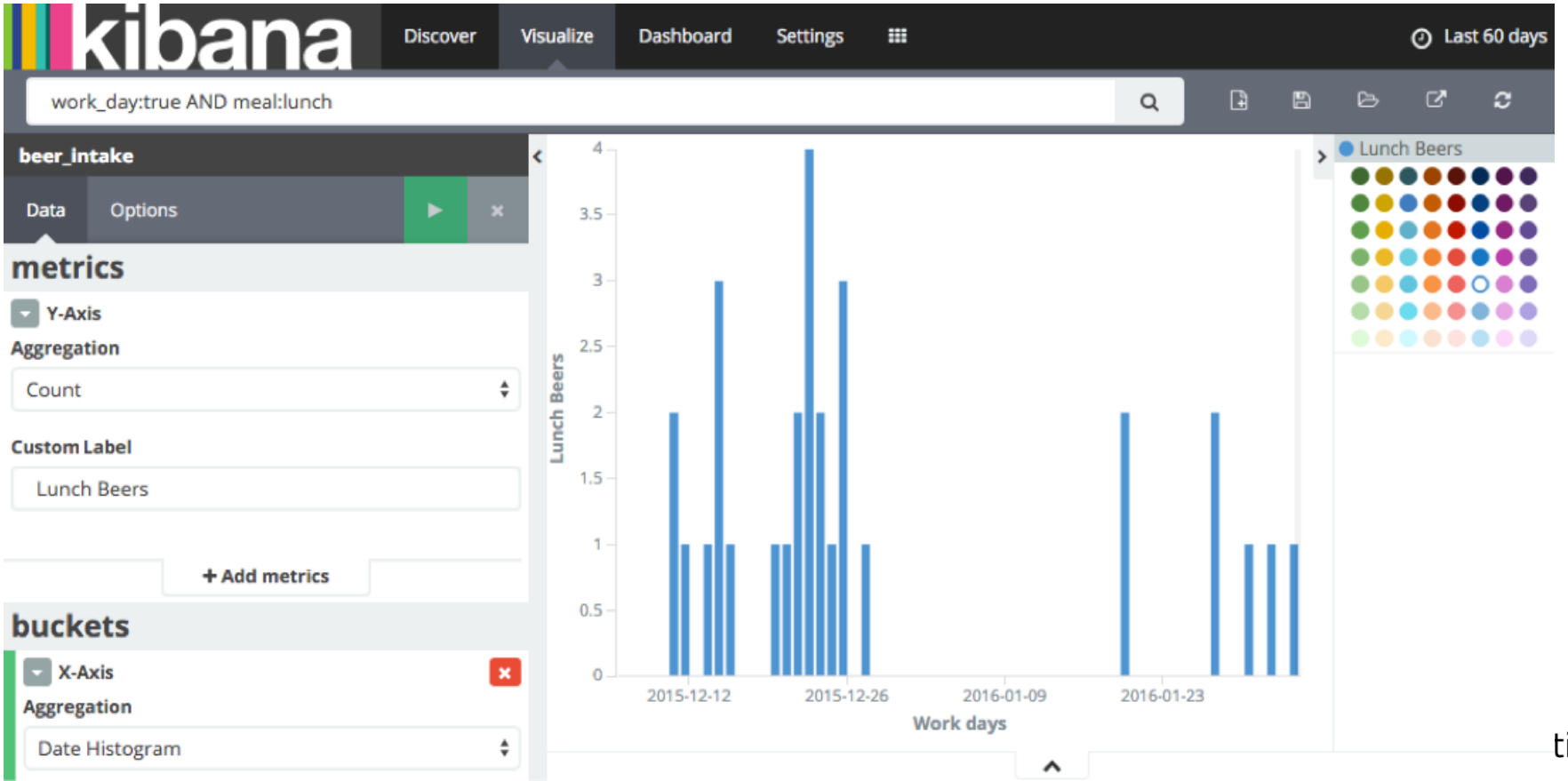
http://github.com/elastic/kibana

elastic

Colour picker

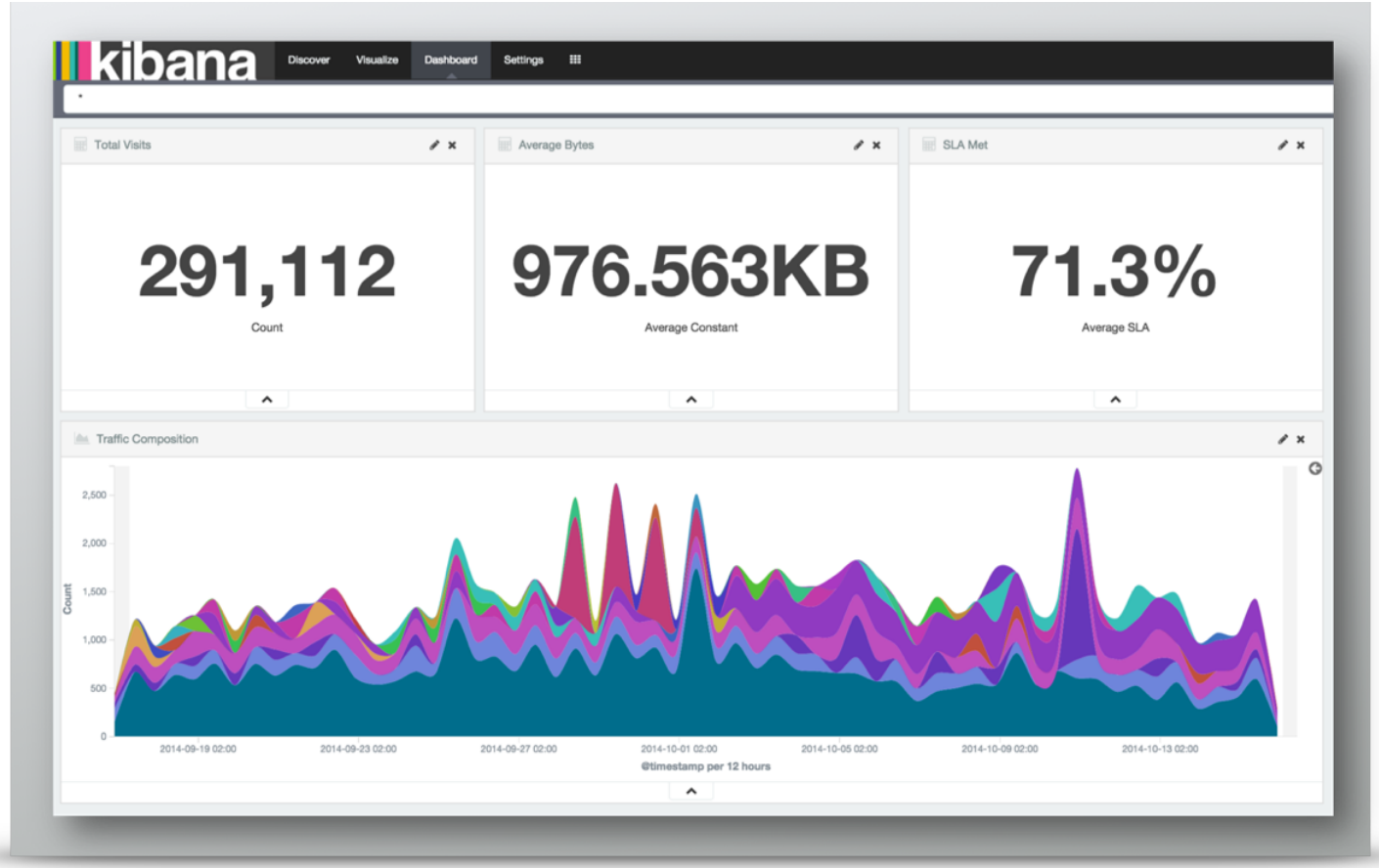


Custom Legends



tic

Field formatters

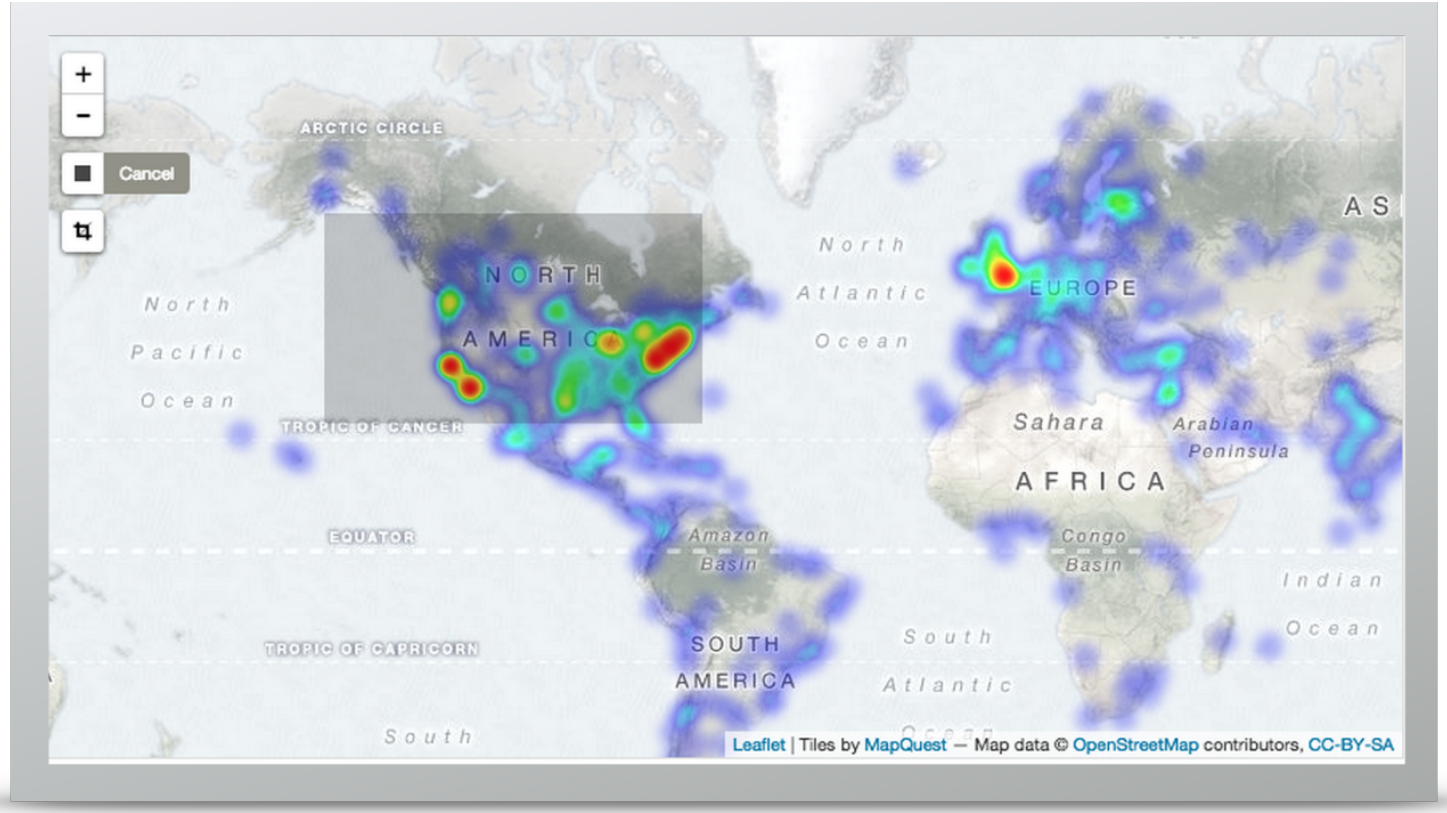


elastic

Black theme

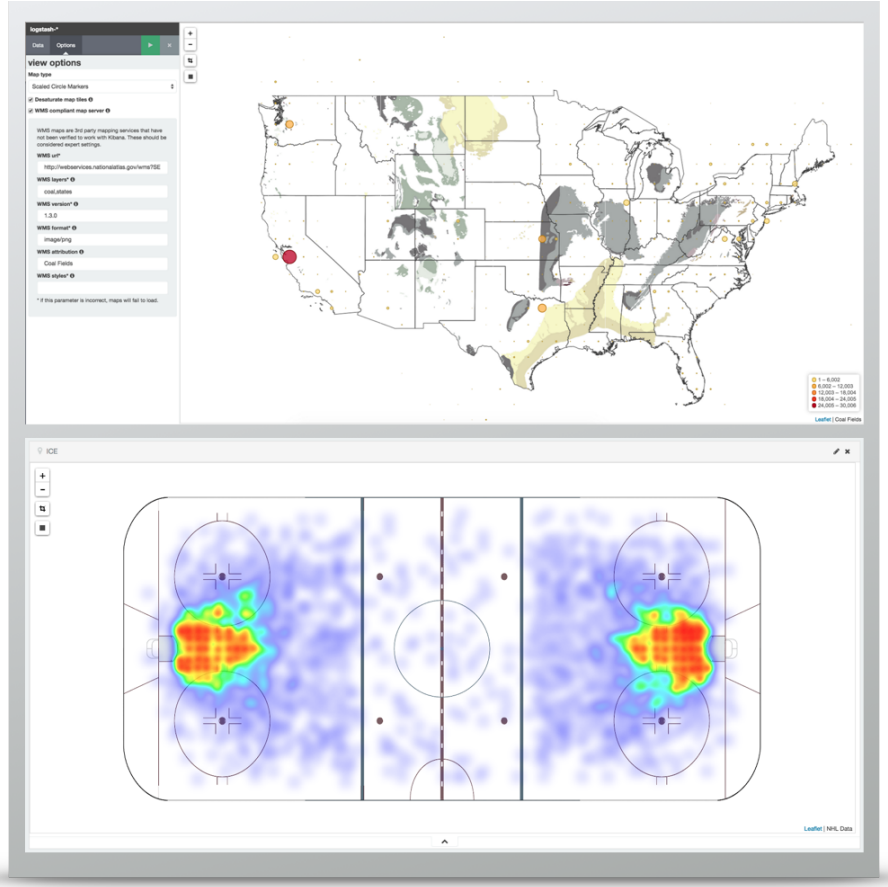
elastic

Heat map



elastic

Pluggable Tile Servers

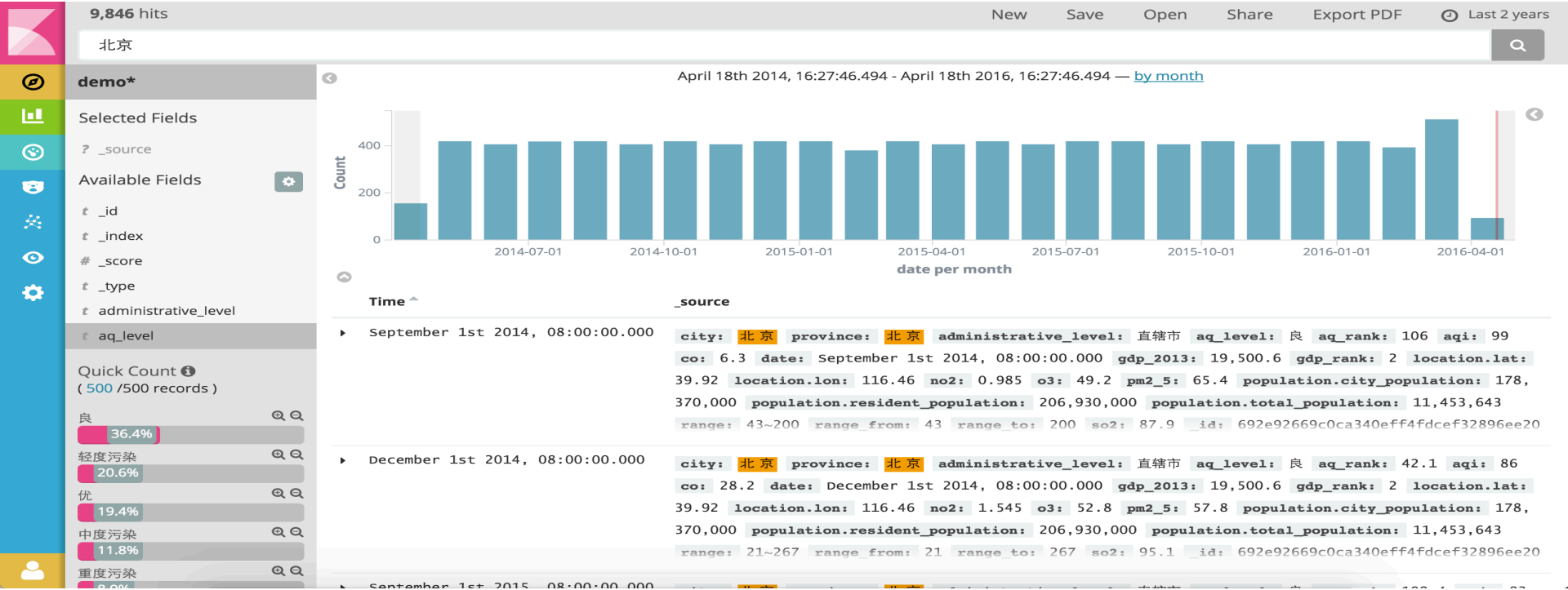


elastic

Global timezone

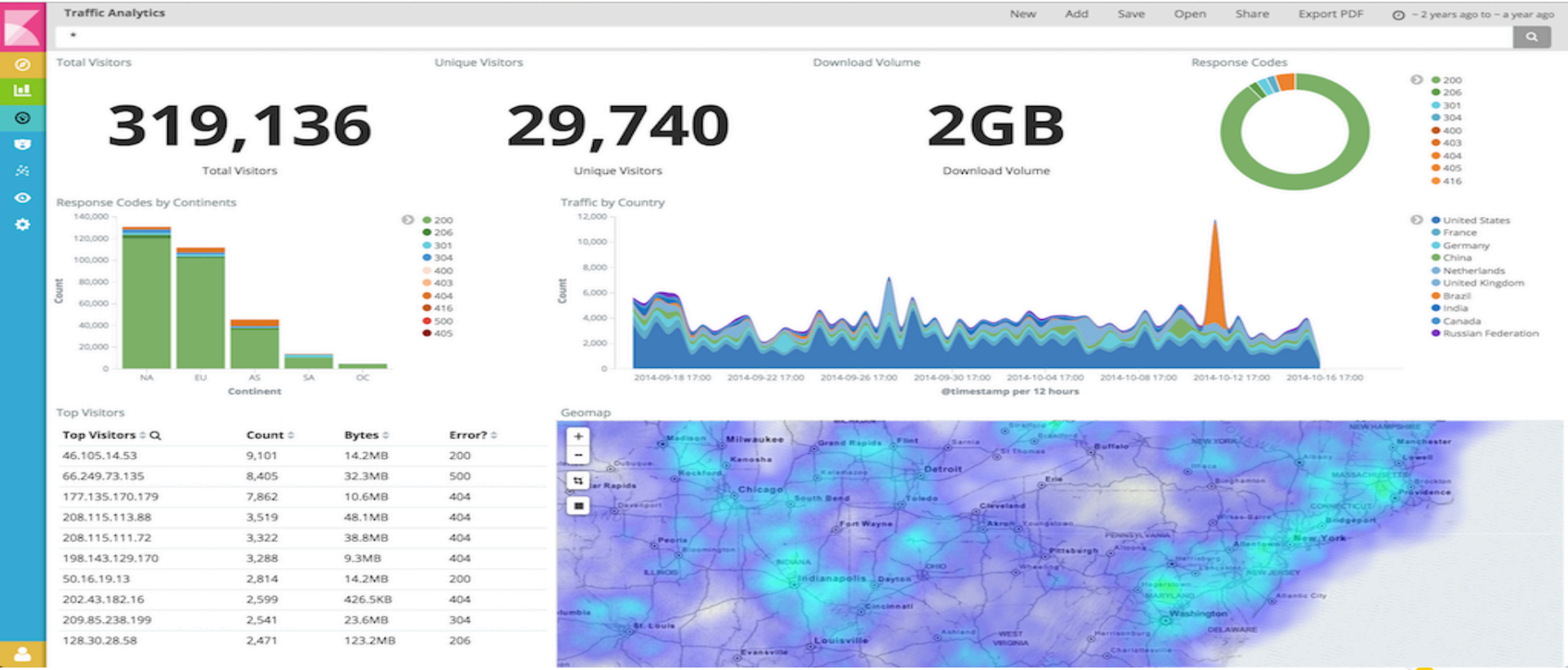
elastic

New design



elastic

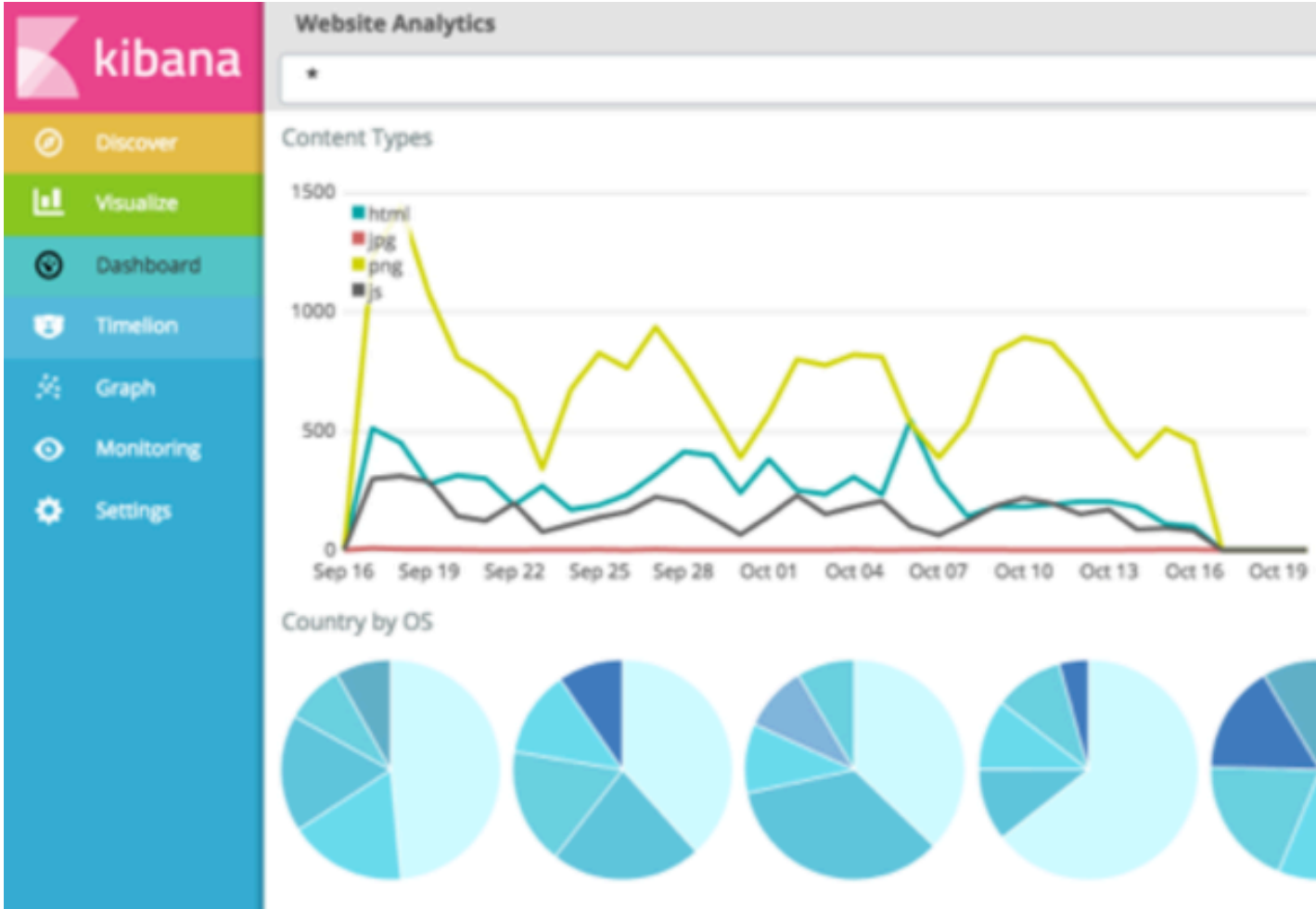
Brings a new focus on your data



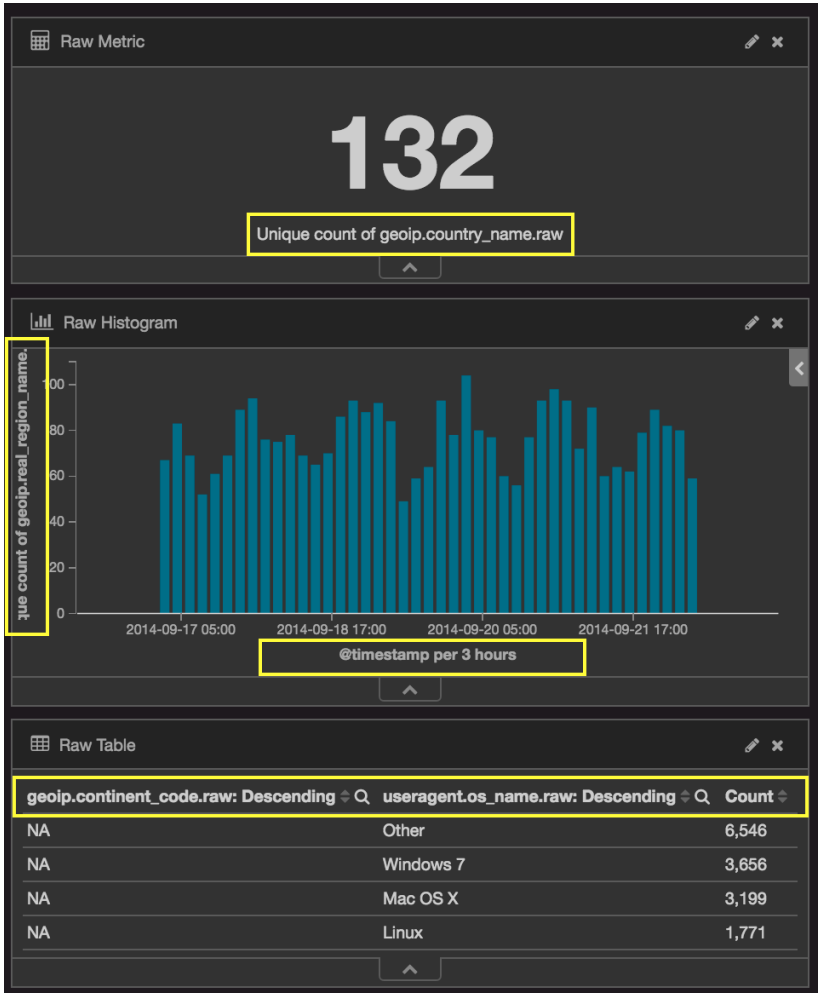
elastic

Applicaton Framework

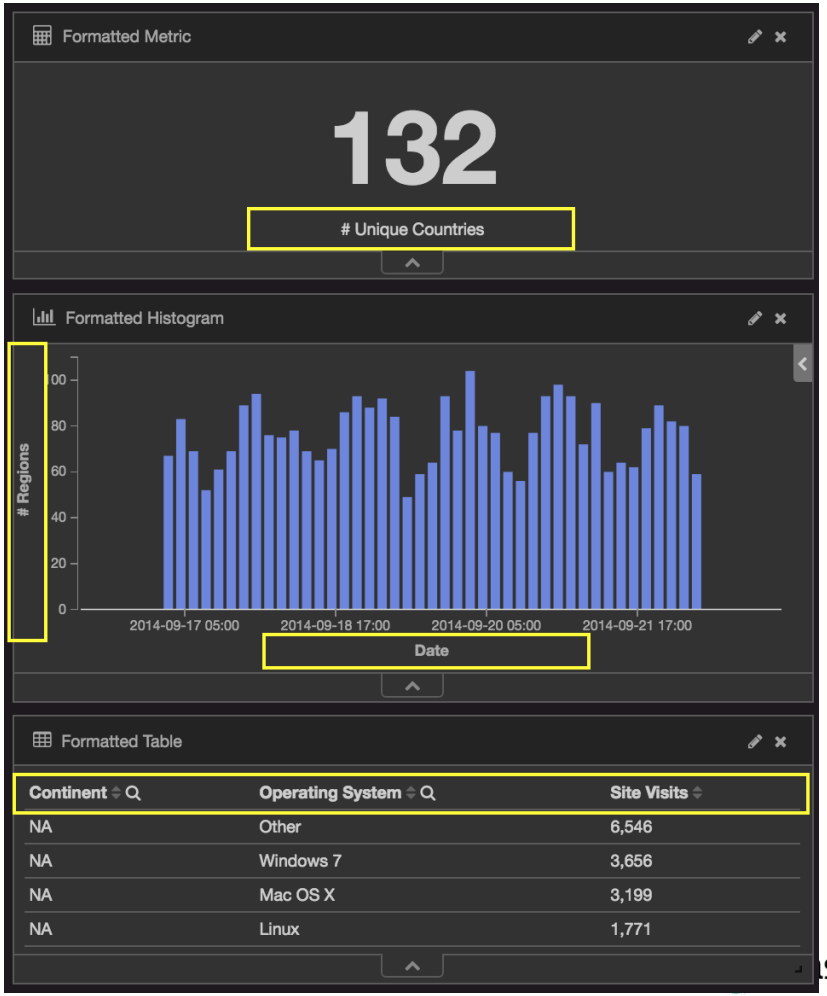
Appear in the
main navigation



System-Generated Labels

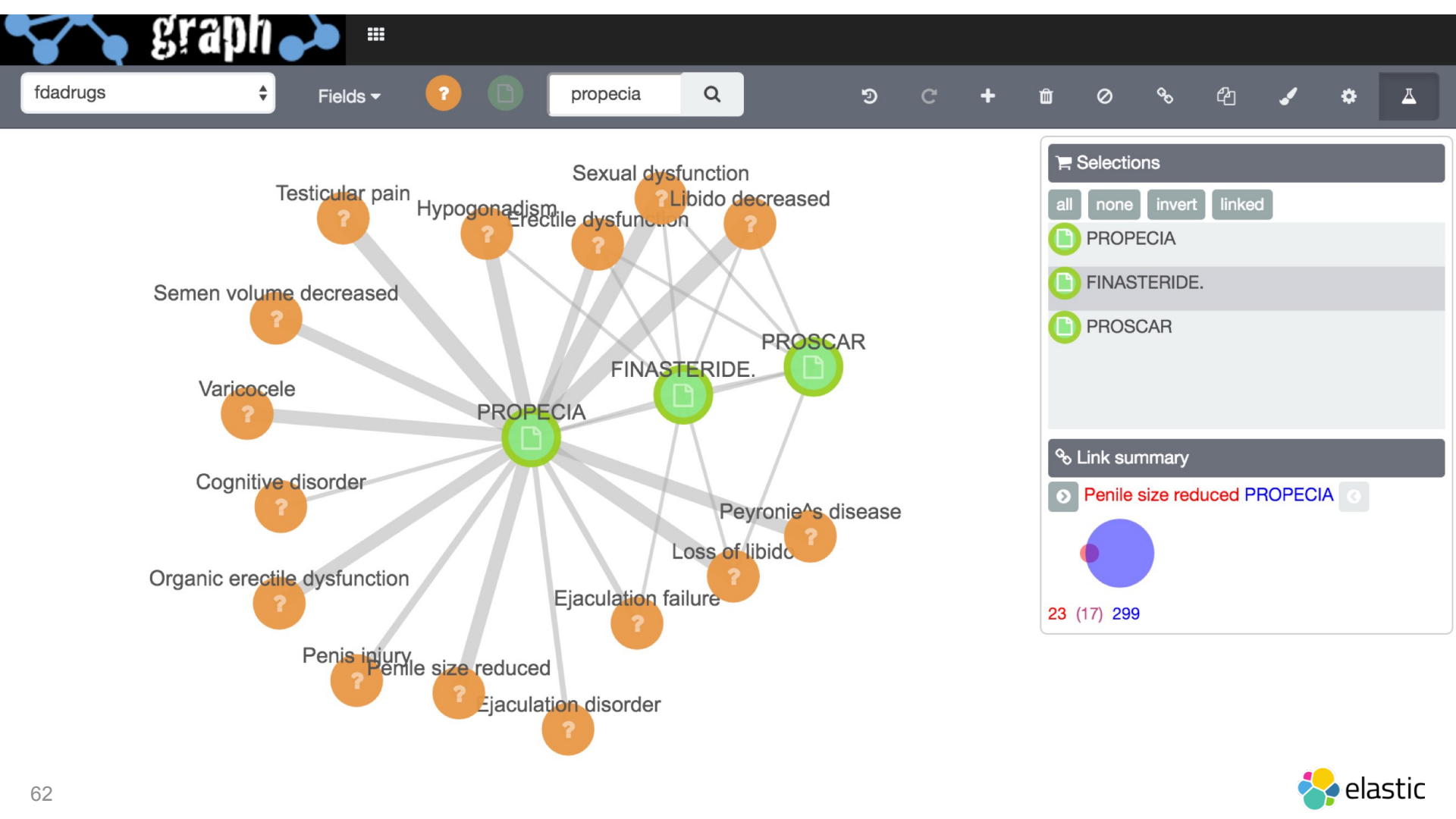


Custom Labels



What’s more

- Plugin command
  - bin/kibana-plugin
- “Sense ”willbe“ Console” (not yet available)
  - Sense plugin will be built into Kibana
- Graph

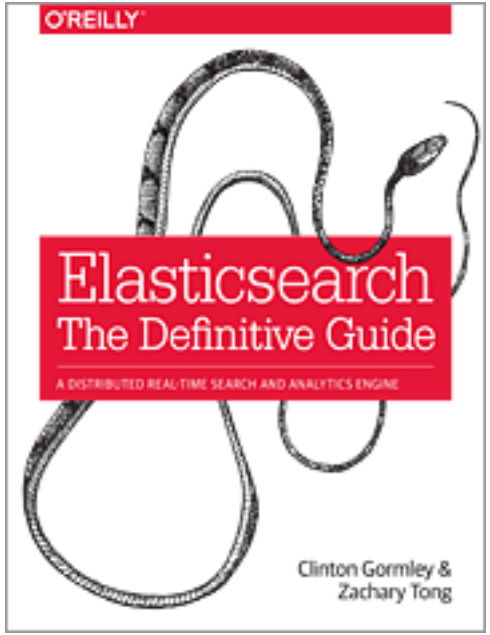

graph
fdadrugs
Fields
propecia
Testicular pain
Hypogonadism
Sexual dysfunction
Libido decreased
Erectile dysfunction
Semen volume decreased
PROSCAR
FINASTERIDE.
Varicocele
PROPECIA
Cognitive disorder
Peyronie^s disease
Loss of libido
Organic erectile dysfunction
Ejaculation failure
Penis injury
Penile size reduced
Ejaculation disorder
Selections
all
none
invert
linked
PROPECIA
FINASTERIDE.
PROSCAR
Link summary
Penile size reduced PROPECIA
23 (17) 299

elastic

Community

- 源码 & Issue: http://github.com/elastic/
- 中文社区: http://elasticsearch.cn
- 官方 QQ 群: 190605846

ES权威指南翻译中，欢迎志愿者加入！
https://github.com/elasticsearch-cn/elasticsearch-definitive-guide



elastic

Thanks!