Elasticsearch, Logstash & Kibana

Kevin Kluge
@kevinkluge
kevin.kluge@elasticsearch.com

elasticsearch.

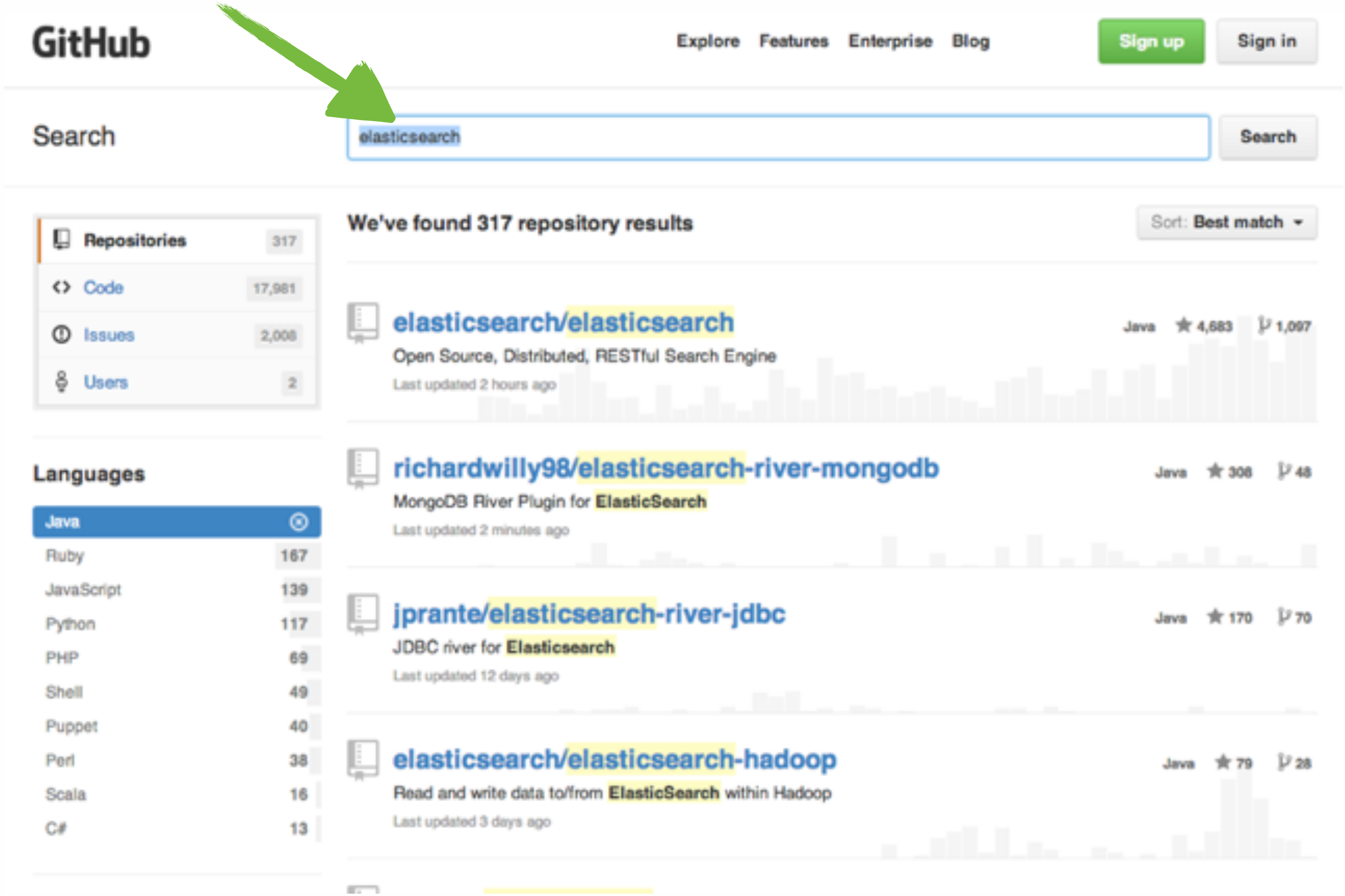
Saturday, February 22, 14

Elasticsearch in 10 seconds

- Schema-free, REST & JSON based document store
- Distributed and horizontally scalable
- Open Source: Apache License 2.0
- Zero configuration
- Written in Java, extensible

elasticsearch.

Unstructured search



elasticsearch

Repositories	317
Code	17,981
Issues	2,008
Users	2

Languages

Java	
Ruby	167
JavaScript	139
Python	117
PHP	69
Shell	49
Puppet	40
Perl	38
Scala	16
C#	13

We've found 317 repository results

elasticsearch/elasticsearch — Java ★ 4,683 ⑂ 1,097
Open Source, Distributed, RESTful Search Engine
Last updated 2 hours ago

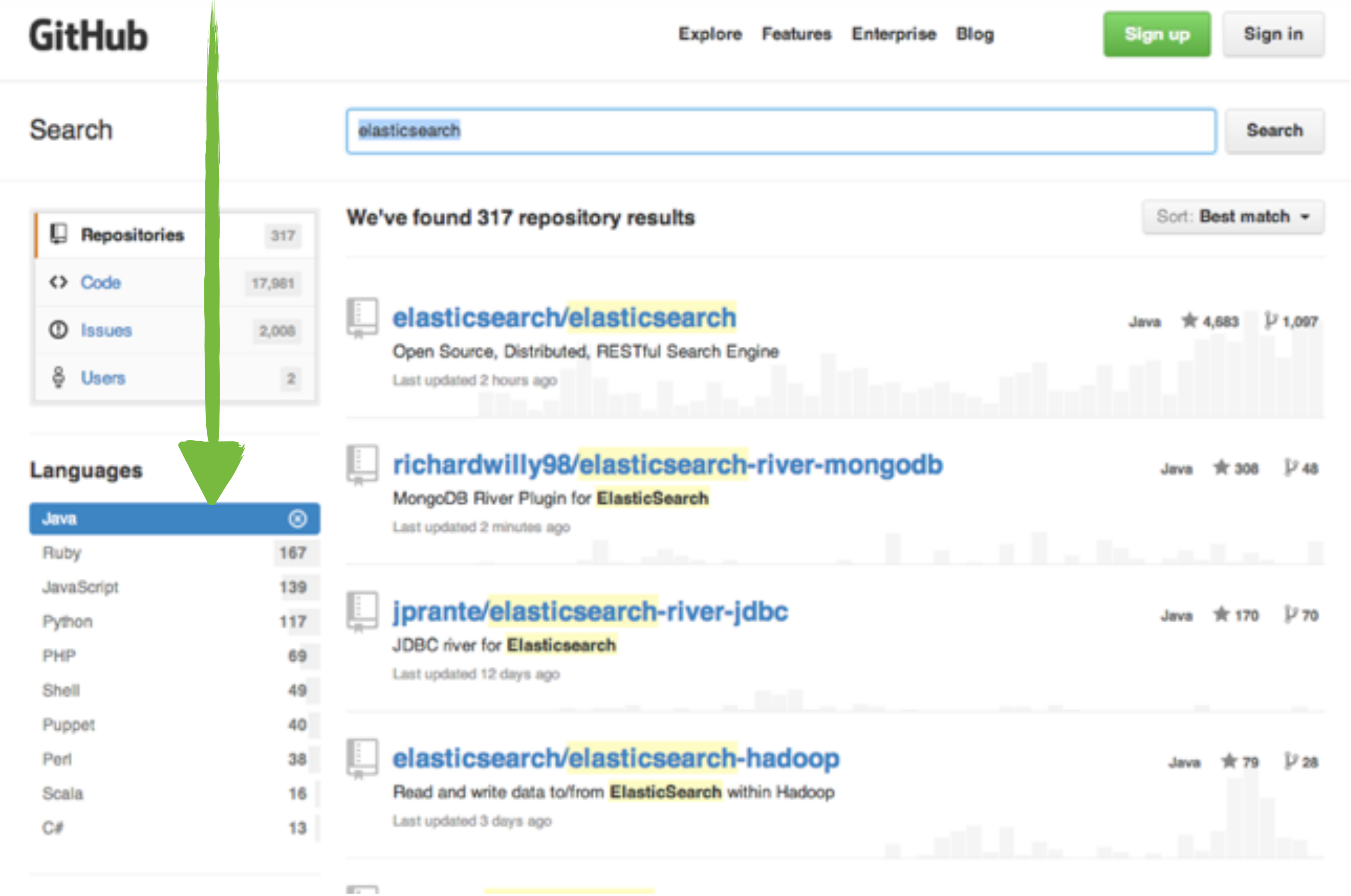
richardwilly98/elasticsearch-river-mongodb — Java ★ 308 ⑂ 48
MongoDB River Plugin for ElasticSearch
Last updated 2 minutes ago

jprante/elasticsearch-river-jdbc — Java ★ 170 ⑂ 70
JDBC river for Elasticsearch
Last updated 12 days ago

elasticsearch/elasticsearch-hadoop — Java ★ 79 ⑂ 28
Read and write data to/from ElasticSearch within Hadoop
Last updated 3 days ago

elasticsearch.

Structured search



elasticsearch

Repositories	317
Code	17,981
Issues	2,008
Users	2

Languages

Java	
Ruby	167
JavaScript	139
Python	117
PHP	69
Shell	49
Puppet	40
Perl	38
Scala	16
C#	13

We've found 317 repository results

Sort: Best match

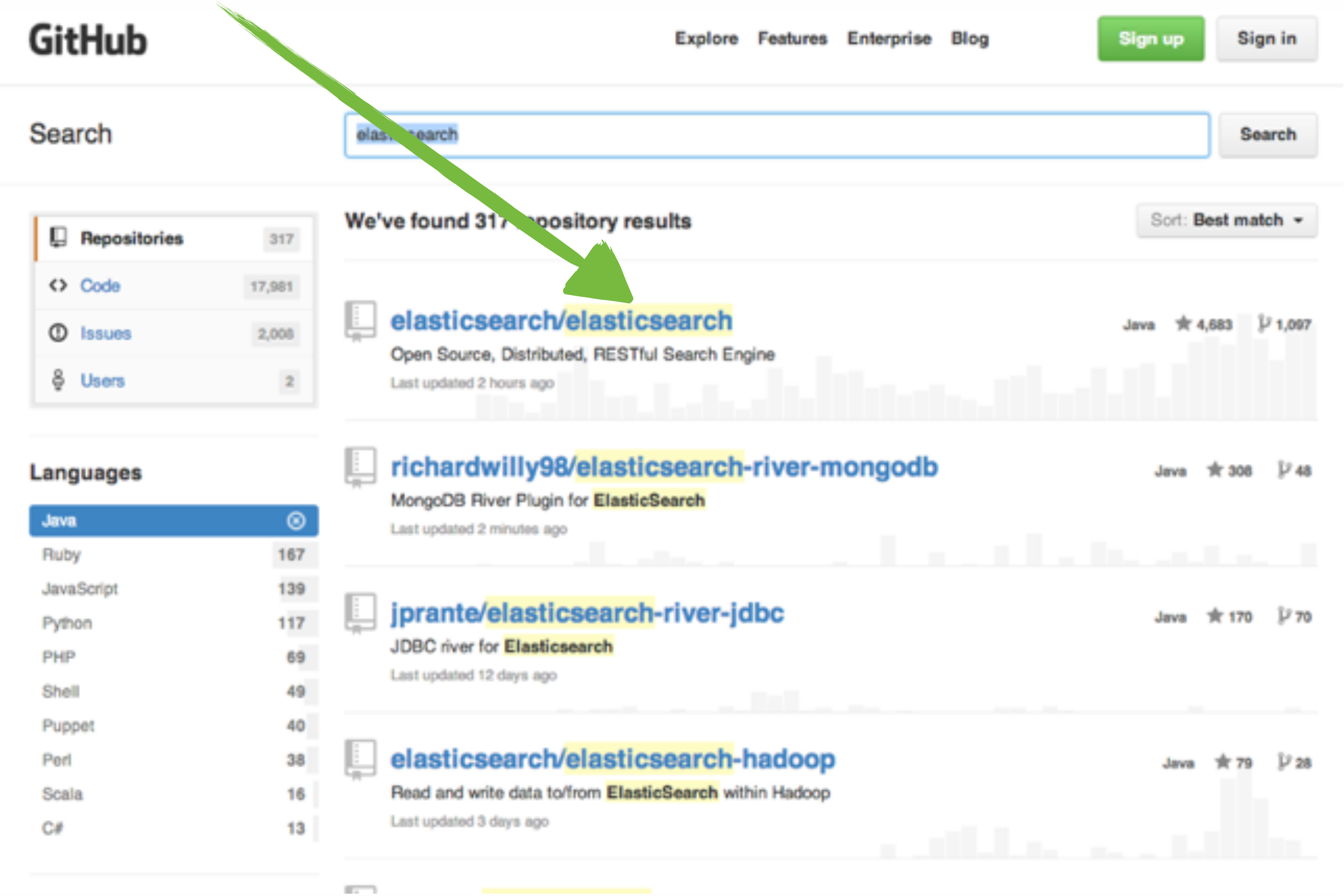
elasticsearch/elasticsearch — Java ★ 4,683 1,097
Open Source, Distributed, RESTful Search Engine
Last updated 2 hours ago

richardwilly98/elasticsearch-river-mongodb — Java ★ 308 48
MongoDB River Plugin for ElasticSearch
Last updated 2 minutes ago

jprante/elasticsearch-river-jdbc — Java ★ 170 70
JDBC river for Elasticsearch
Last updated 12 days ago

elasticsearch/elasticsearch-hadoop — Java ★ 79 28
Read and write data to/from ElasticSearch within Hadoop
Last updated 3 days ago

Enrichment



Search: elasticsearch

We've found 317 repository results

Sort: Best match

Repositories	317
Code	17,981
Issues	2,008
Users	2

Languages

Java	
Ruby	167
JavaScript	139
Python	117
PHP	69
Shell	49
Puppet	40
Perl	38
Scala	16
C#	13

elasticsearch/elasticsearch — Java ★ 4,683 ⑂ 1,097
Open Source, Distributed, RESTful Search Engine
Last updated 2 hours ago

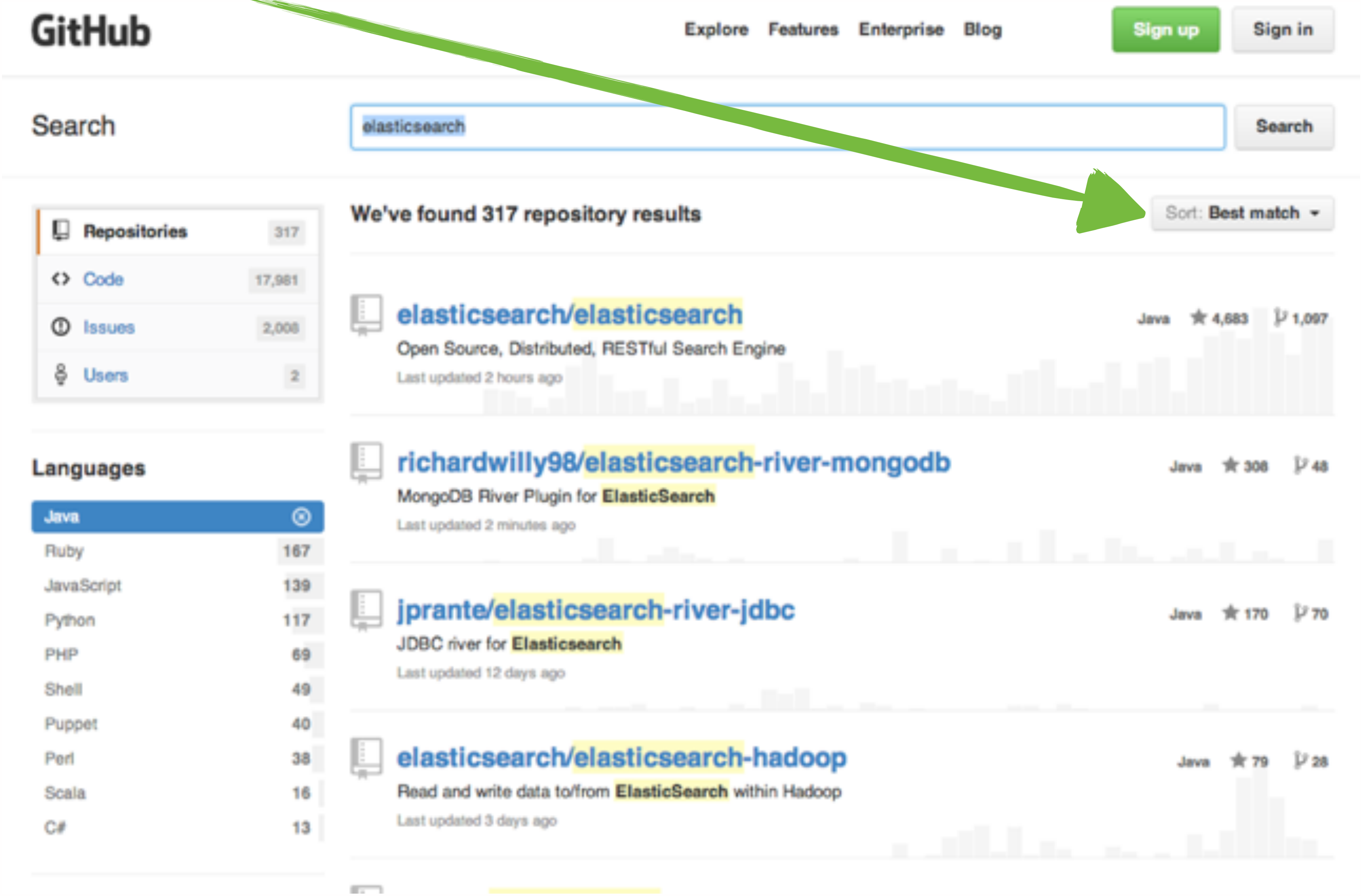
richardwilly98/elasticsearch-river-mongodb — Java ★ 308 ⑂ 48
MongoDB River Plugin for ElasticSearch
Last updated 2 minutes ago

jprante/elasticsearch-river-jdbc — Java ★ 170 ⑂ 70
JDBC river for Elasticsearch
Last updated 12 days ago

elasticsearch/elasticsearch-hadoop — Java ★ 79 ⑂ 28
Read and write data to/from ElasticSearch within Hadoop
Last updated 3 days ago

elasticsearch.

Sorting



We've found 317 repository results

elasticsearch/elasticsearch — Java ★ 4,683 ⑂ 1,097
Open Source, Distributed, RESTful Search Engine
Last updated 2 hours ago

richardwilly98/elasticsearch-river-mongodb — Java ★ 308 ⑂ 48
MongoDB River Plugin for ElasticSearch
Last updated 2 minutes ago

jprante/elasticsearch-river-jdbc — Java ★ 170 ⑂ 70
JDBC river for Elasticsearch
Last updated 12 days ago

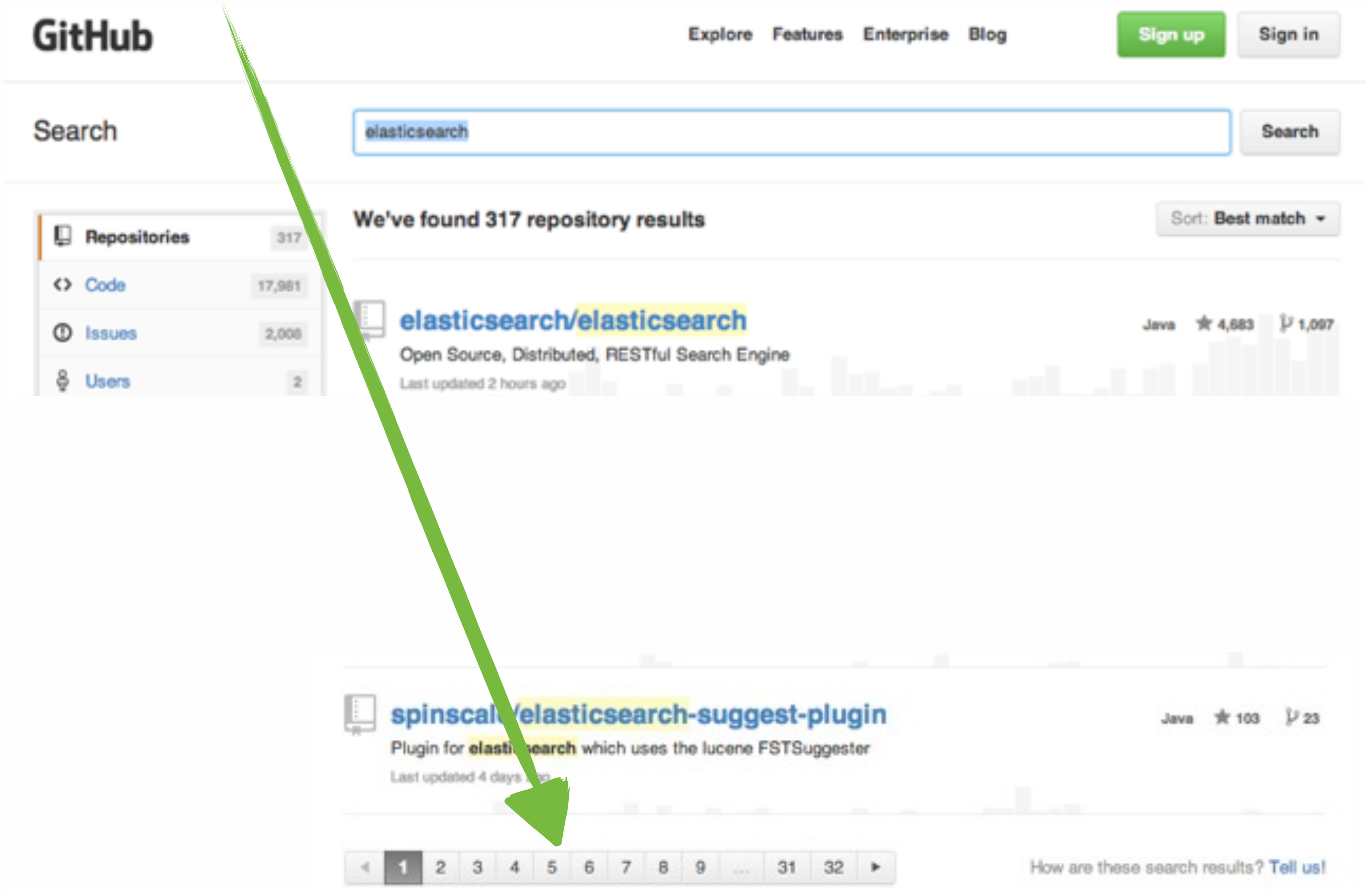
elasticsearch/elasticsearch-hadoop — Java ★ 79 ⑂ 28
Read and write data to/from ElasticSearch within Hadoop
Last updated 3 days ago

Languages

Java	
Ruby	167
JavaScript	139
Python	117
PHP	69
Shell	49
Puppet	40
Perl	38
Scala	16
C#	13

elasticsearch.

Pagination



elasticsearch

Repositories	317
Code	17,981
Issues	2,008
Users	2

We've found 317 repository results

Sort: Best match

elasticsearch/elasticsearch — Java ★ 4,683 1,097
Open Source, Distributed, RESTful Search Engine
Last updated 2 hours ago

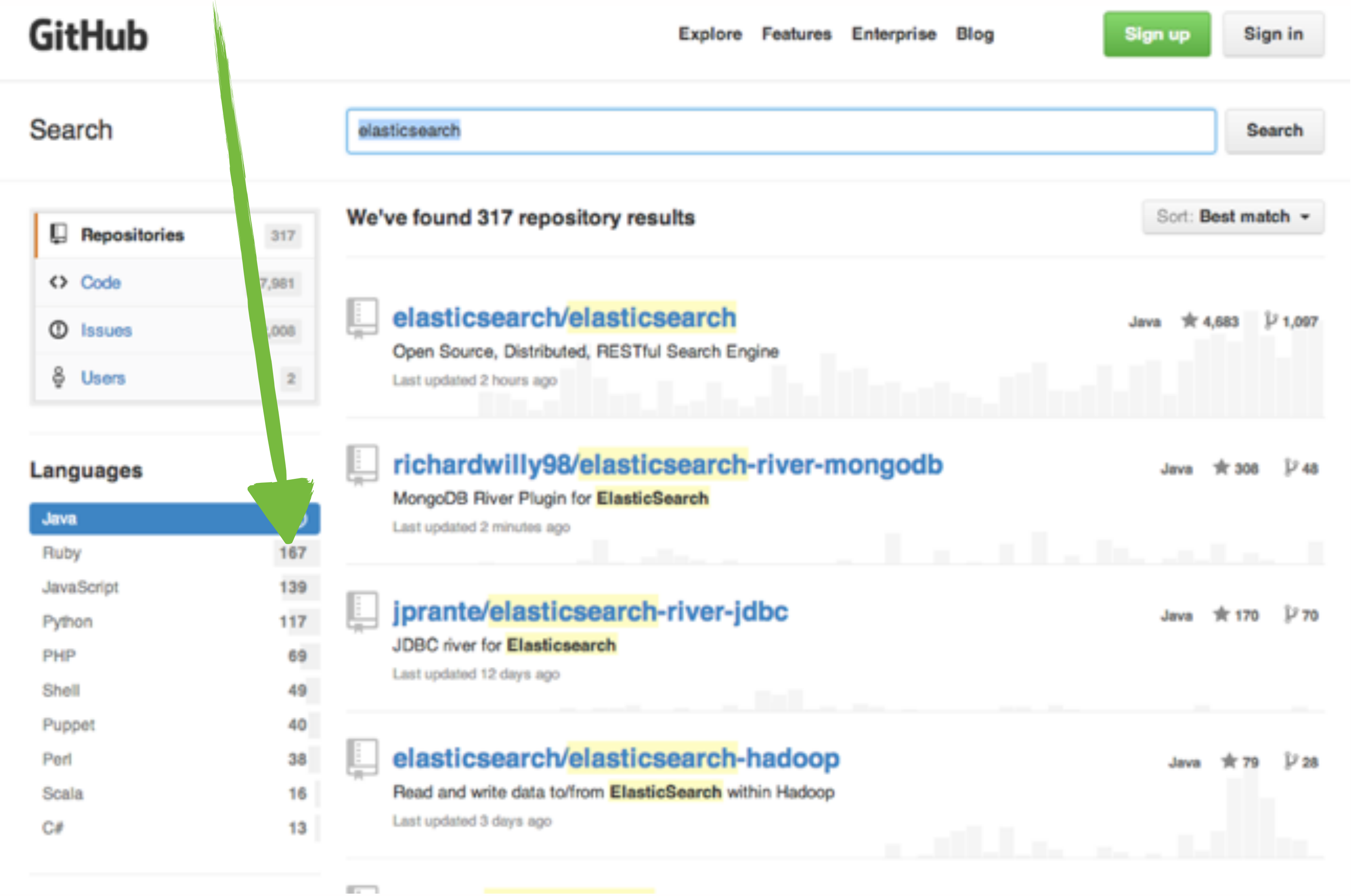
spinscale/elasticsearch-suggest-plugin — Java ★ 103 23
Plugin for elasticsearch which uses the lucene FSTSuggester
Last updated 4 days ago

◂ 1 2 3 4 5 6 7 8 9 … 31 32 ▸

How are these search results? Tell us!

elasticsearch.

Aggregation



elasticsearch

We've found 317 repository results

Repositories	317
Code	7,981
Issues	,008
Users	2

Languages

Java	
Ruby	167
JavaScript	139
Python	117
PHP	69
Shell	49
Puppet	40
Perl	38
Scala	16
C#	13

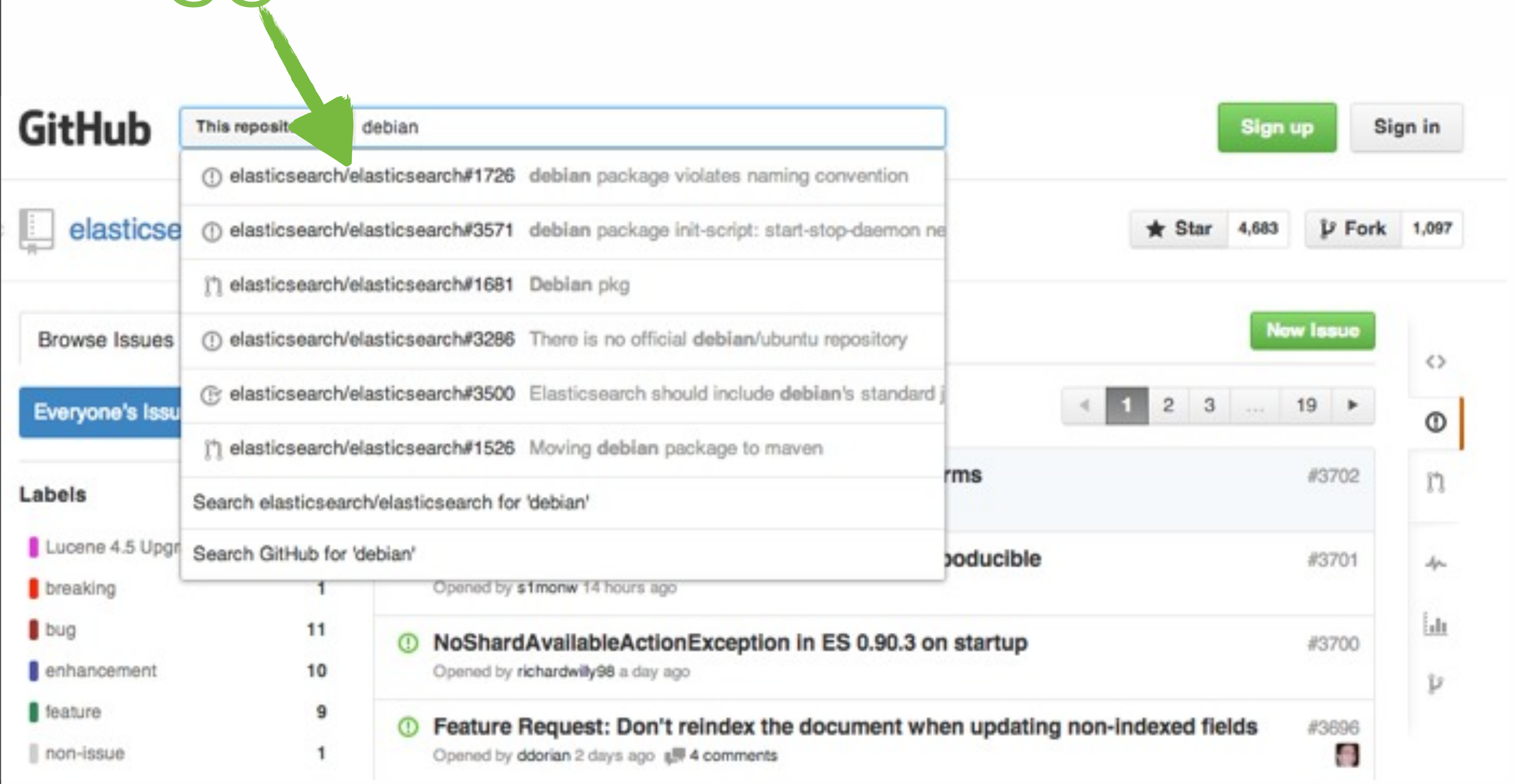
elasticsearch/elasticsearch Java ★ 4,683 1,097
Open Source, Distributed, RESTful Search Engine
Last updated 2 hours ago

richardwilly98/elasticsearch-river-mongodb Java ★ 308 48
MongoDB River Plugin for ElasticSearch
Last updated 2 minutes ago

jprante/elasticsearch-river-jdbc Java ★ 170 70
JDBC river for Elasticsearch
Last updated 12 days ago

elasticsearch/elasticsearch-hadoop Java ★ 79 28
Read and write data to/from ElasticSearch within Hadoop
Last updated 3 days ago

Suggestions



GitHub

This reposit… debian

- elasticsearch/elasticsearch#1726 debian package violates naming convention
- elasticsearch/elasticsearch#3571 debian package init-script: start-stop-daemon ne
- elasticsearch/elasticsearch#1681 Debian pkg
- elasticsearch/elasticsearch#3286 There is no official debian/ubuntu repository
- elasticsearch/elasticsearch#3500 Elasticsearch should include debian's standard j
- elasticsearch/elasticsearch#1526 Moving debian package to maven

Search elasticsearch/elasticsearch for 'debian'

Search GitHub for 'debian'

elasticse

Labels

Lucene 4.5 Upgr	
breaking	1
bug	11
enhancement	10
feature	9
non-issue	1

rms #3702

oducible #3701
Opened by s1monw 14 hours ago

NoShardAvailableActionException in ES 0.90.3 on startup #3700
Opened by richardwilly98 a day ago

Feature Request: Don't reindex the document when updating non-indexed fields #3696
Opened by ddorian 2 days ago 4 comments

elasticsearch.

Installation & first steps

2 minutes to live

```
$ wget https://download.elasticsearch.org/...
$ tar -xf elasticsearch-1.0.0.tar.gz
$ ./elasticsearch-1.0.0/bin/elasticsearch
...
[2014-01-19 14:53:11,508][INFO ][node] [Scanner] started
...
```

Also puppet modules and RPM/DEB

Is it alive?

```
» curl localhost:9200
{
  "status" : 200,
  "name" : "Scanner",
  "version" : {
    "number" : “1.0.0",
    "build_hash" : "e018cda7e7a32643d59e0ac3cdb412ccc239af04",
    "build_timestamp" : "2014-01-17T15:11:47Z",
    "build_snapshot" : true,
    "lucene_version" : “4.6.1"
  },
  "tagline" : "You Know, for Search"
}
```

elasticsearch.

Create...

```
» curl -XPUT localhost:9200/books/book/1 -d '
{
  "title" : "Elasticsearch - The definitive guide",
  "authors" : "Clinton Gormley",
  "started" : "2013-02-04",
  "pages" : 230
}'
```

Update...

```
» curl -XPUT localhost:9200/books/book/1 -d '
{
  "title" : "Elasticsearch - The definitive guide",
  "authors" : [ "Clinton Gormley", "Zachary Tong" ],
  "started" : "2013-02-04",
  "pages" : 230
}'
```

Delete…

```
» curl –X DELETE localhost:9200/books/book/1
```

Realtime GET…

```
» curl –X GET localhost:9200/books/book/1
» curl –X GET localhost:9200/books/book/1/_source
```

Search

```
» curl -XGET localhost:9200/books/_search?q=elasticsearch
```

```
{
  "took" : 2, "timed_out" : false,
  "_shards" : { "total" : 5, "successful" : 5, "failed" : 0 },
  "hits" : {
    "total" : 1, "max_score" : 0.076713204,
    "hits" : [ {
      "_index" : “books", "_type" : “book", "_id" : "1",
      "_score" : 0.076713204, "_source" : {
        "title" : "Elasticsearch - The definitive guide",
        "authors" : [ "Clinton Gormley", "Zachary Tong" ],
        "started" : “2013-02-04", "pages" : 230
      }
    } ]
  }
}
```

elasticsearch.

Search - Query DSL

```
» curl –XGET ‘localhost:9200/books/book/_search' –d '{
    "query": {
        "filtered" : {
            "query" : {
                "match": {
                    "text" :  {
                        "query" : “To Be Or Not To Be",
                        "cutoff_frequency" : 0.01
                    }
                }
            },
            "filter" : {
                "range": {
                  "price": {
                    "gte": 20.0
                    "lte": 50.0
       ...
    }
}'
```

Distributed and scalable

Basic terms

- Index

  Logical collection of data; might be time based
  Analogous to a database

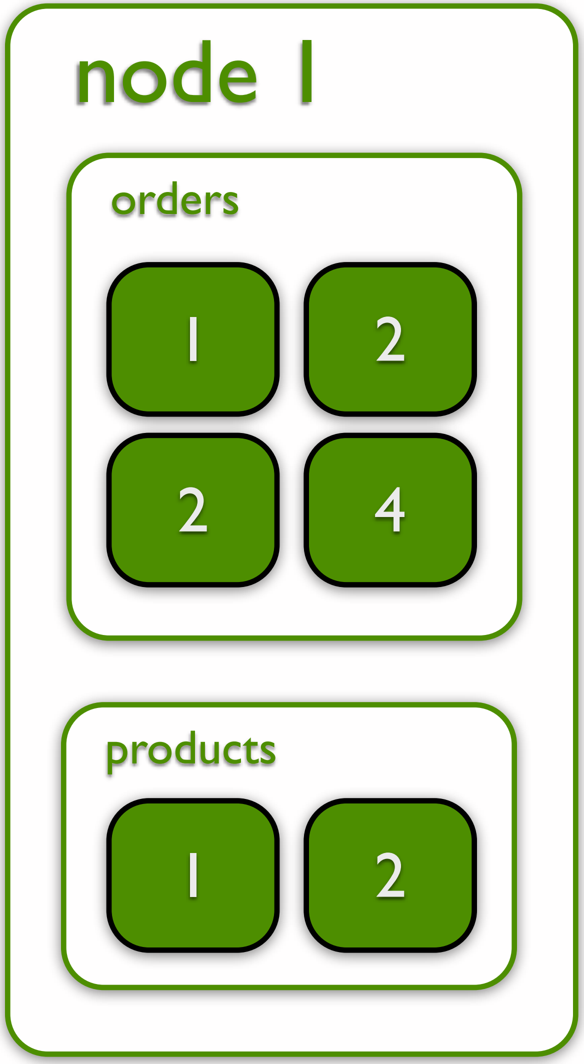
- Replication

  Read scalability
  Removing SPOF

- Sharding

  Split logical data over several machines
  Write scalability
  Control data flows

Shards and replicas

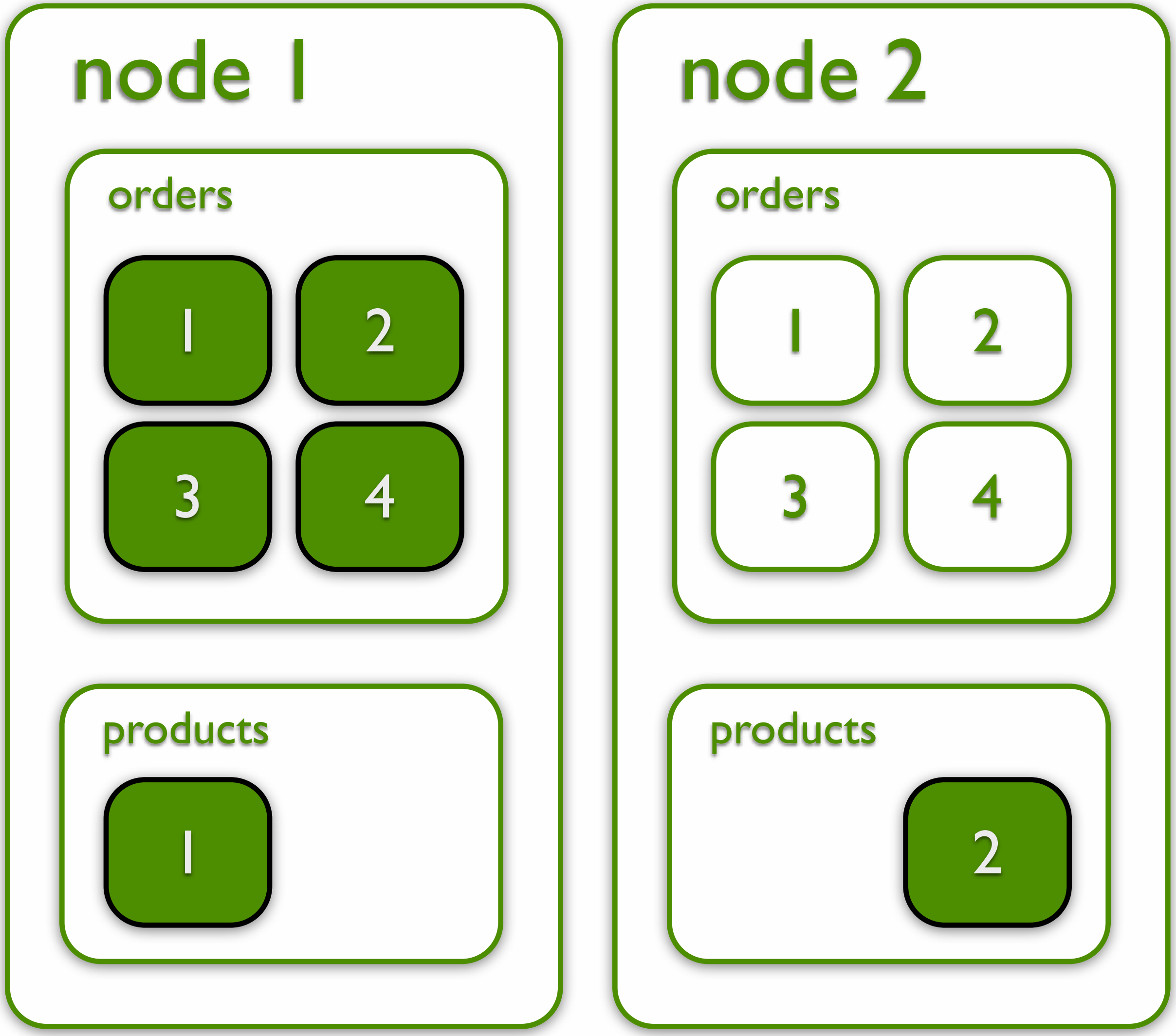


```
curl -X PUT localhost:9200/orders -d '{
  "settings.index.number_of_shards" : 4
  "settings.index.number_of_replicas" : 1
}'
```

```
curl -X PUT localhost:9200/products -d '{
  "settings.index.number_of_shards" : 2
  "settings.index.number_of_replicas" : 0
}'
```

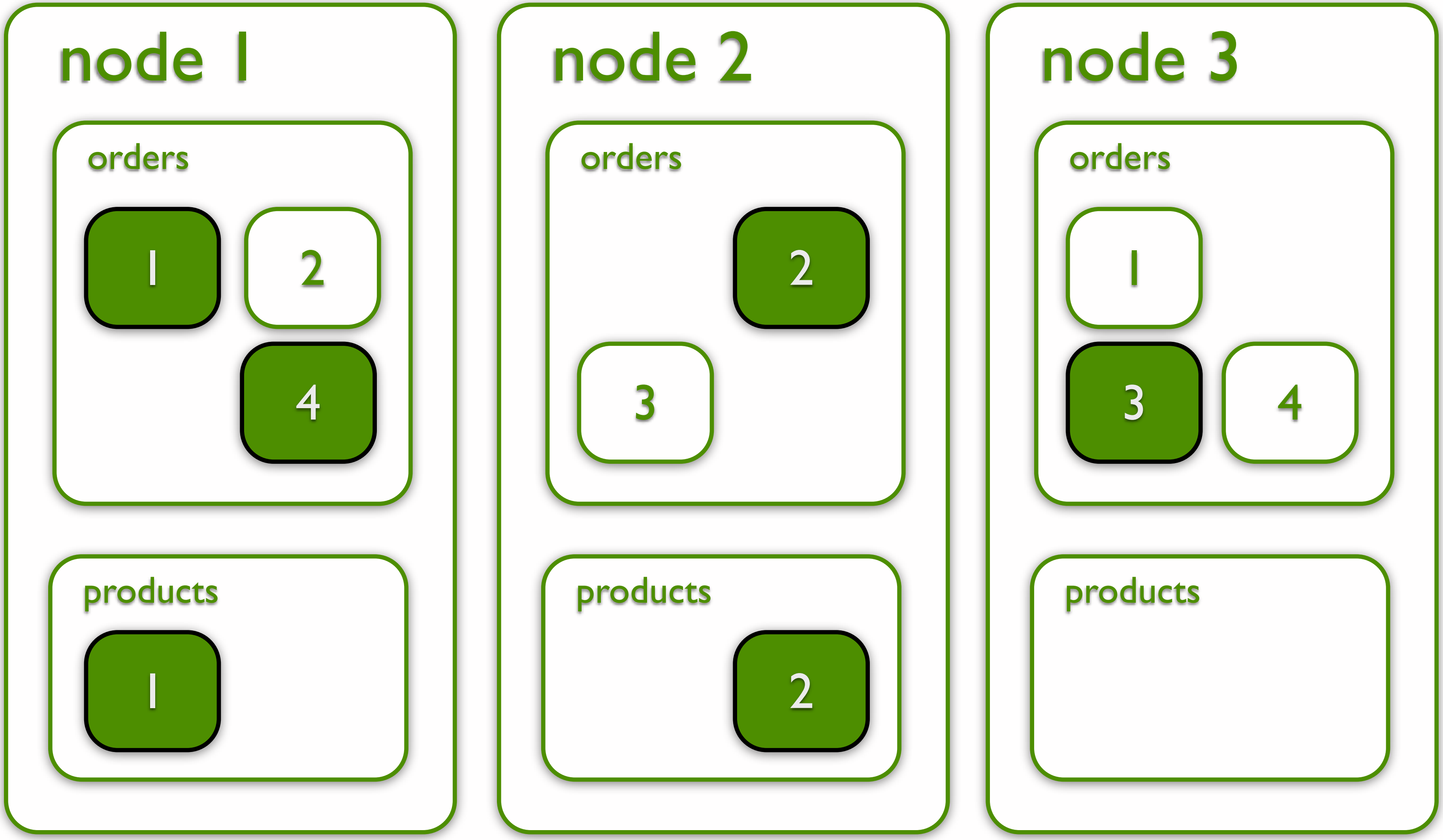
elasticsearch.

Shards and replicas



elasticsearch.

Automatic leveling



elasticsearch.

Cluster management

- Single master at any point in time
- Multicast based discovery (optionally unicast)
- Configuration is required here

  Tell each node the name of the cluster to join

  Set minimum master nodes
- Tip: reserve 3 nodes for master role and do not put data on them

Sizing a cluster or node

- Data and operation dependent

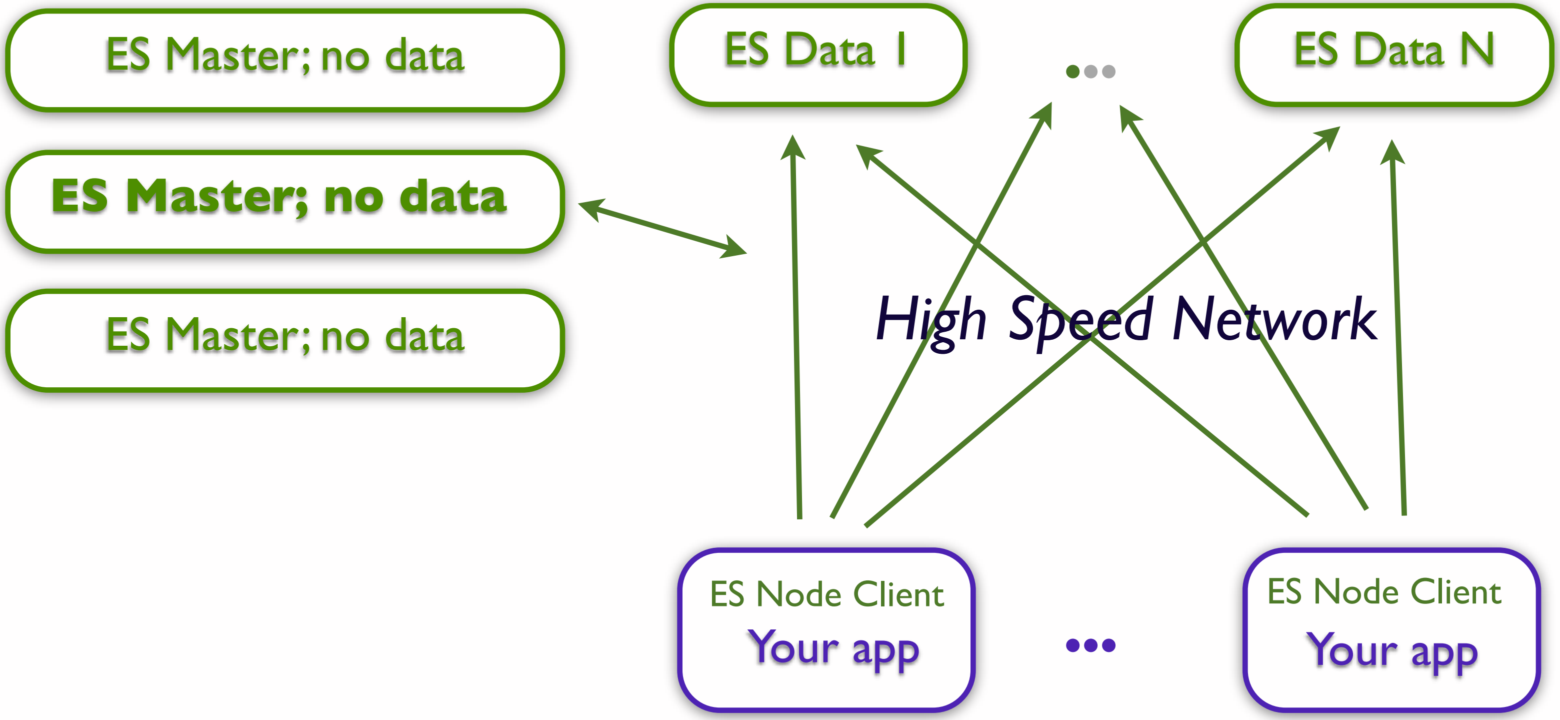
  How big are your documents? How many fields in them?
  What is your query rate?
  Do you do facets/aggregations, sorting, custom scoring?
  What is your write rate?
  Do you delete documents? Update them?
  Is the data time-based?

- Test on one node, no replicas

  Look at shard size, JVM heap usage and GC frequency, number of shards/node, docs per shard, CPU util, disk util, index pattern

- Tip: 30 GB heap

Deployment architecture



- Above shows local disk; SAN OK
- Tip: clusters spanning high latency WANs are not recommended. Cross-zone in EC2 is OK.

elasticsearch.

Elasticsearch use-cases

What is data?

- Whatever provides value for your business

- Domain data

  Internal: Orders, products

  External: Social media streams, email

- Application data

  Log files

  Metrics

Use case: Product search engine

Product search engine

- Just index all your products and be happy?

  Search is not that easy

- Synonyms, Suggestions, Faceting, Custom scoring, Analytics, Decompounding, Query optimization, beyond search

- User your domain knowledge

Scoring

- Is full-text search relevancy really your preferred scoring algorithm?

- Possible influential factors

  Age of the product, been ordered in last 24h
  In Stock?
  No shipping costs
  Special offer
  Rating (product or seller)

  http://www.elasticsearch.org/guide/en/elasticsearch/reference/current/query-dsl-function-score-query.html

Faceting & user exploration

- Products grouped by

  Category
  Material
  Brand

- Allowing to filter

  All of the facets
  Price range
  Color
  Seller
  Ratings (hard!)

Notification with percolation

- Customer: If a product matches name *X* and costs below price *Y*, is color *Z*, then I want to get a mail

  More likely: Notify customer, when it is back in stock

- Enter percolation!

  Not: Index a document and fire a query

  But: Index a query and check a document for a match

https://speakerdeck.com/javanna/whats-new-in-percolator

Use-case:
Analytics

elasticsearch.

Analytics

- Aggregation of information
- Facets are one dimensional

  Categories/brands/material of all results of this query
- Questions are multidimensional

  Average revenue per category id per day
- Elasticsearch 1.0 has aggregations

  Nested faceting

elasticsearch.

Create knowledge from data

- Orders

  How many orders were created every day in the last month?
  How many orders were created per state in the last month?

- Money

  What is the average revenue per shopping cart?
  What is the average shopping cart size per order per hour?

- Product portfolio

  Take the location of people into account for special offers?
  Analyse page views: Premium or low budget ecommerce site?

Ecosystem

- Plugins

  Many third party plugins available

- Clients for many languages

  Ruby, python, php, perl, javascript, (.NET coming)
  Scala, clojure, go

- Kibana

- Logstash

- Hadoop integration

elasticsearch.

XING
K
stackoverflow
SOUNDCLOUD
Mc Graw Hill
Fog Creek SOFTWARE
foursquare
LiveChat.com
loggly
NETWORKED INSIGHTS

elasticsearch.

Tools for sys admins

REST-based management

- Elasticsearch is full of monitoring APIs

  Everything is returned as JSON

- Humans are not the world's best JSON parsers

- What if elasticsearch had an easy to use interface from the commandline?

Which node is the master?

```
$ curl "localhost:9200/_cluster/state?pretty&filter_metadata=true&
filter_routing_table=true"
{
  "cluster_name" : "elasticsearch",
  "master_node" : "GNf0hEXlTfaBvQXKBF300A",
  "blocks" : { },
  "nodes" : {
    "ObdRqLHGQ6CMI5rOEstA5A" : {
      "name" : "Triton",
      "transport_address" : “inet[/10.0.1.11:9300]”,
      "attributes" : { }
    },
    "4C7pKbfhTvu0slcSy_G4_w" : {
      "name" : "Kid Colt",
      "transport_address" : "inet[/10.0.1.12:9300]",
      "attributes" : { }
    },
    "GNf0hEXlTfaBvQXKBF300A" : {
      "name" : "Lang, Steven",
      "transport_address" : "inet[/10.0.1.13:9300]",
      "attributes" : { }
    }
  }
}
```

Which one is the master? (v1.0)

```
$ curl localhost:9200/_cat/master
GNf0hEXlTfaBvQXKBF300A 10.0.1.13 Lang, Steven
```

elasticsearch.

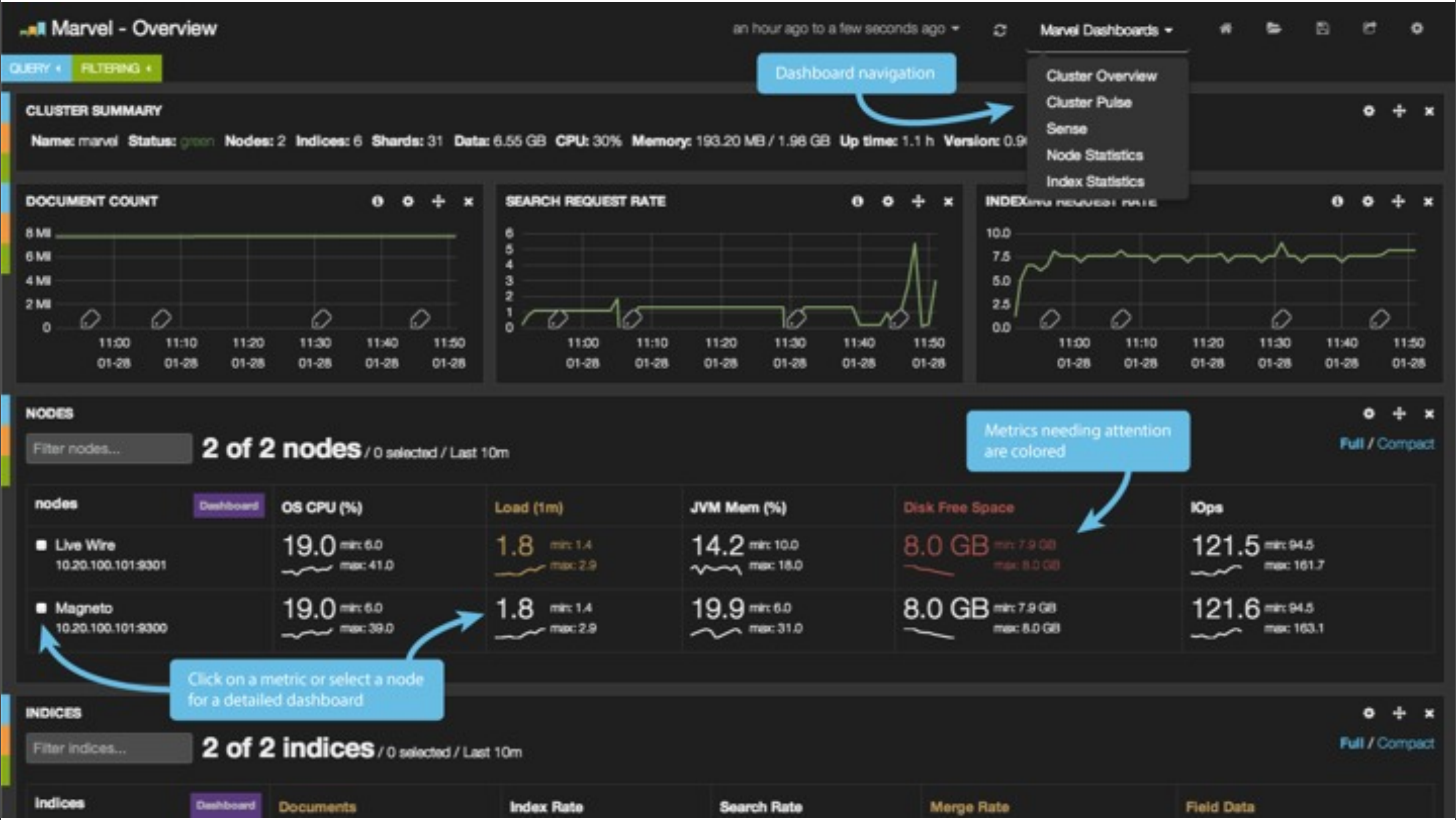
_cat/* api

- /_cat/allocation
- /_cat/count
- /_cat/health
- /_cat/master
- /_cat/aliases
- /_cat/nodes
- /_cat/recovery
- /_cat/shards
- /_cat/indices
- /_cat/thread_pool

Monitor your cluster with Marvel

- Point in time views are a start
- Marvel shows historical trends
- Visualize cluster behavior, act before problems
- Free for development, $500/year for up to 5 nodes

Overview



Dashboard navigation

CLUSTER SUMMARY

Name: marvel **Status:** green **Nodes:** 2 **Indices:** 6 **Shards:** 31 **Data:** 6.55 GB **CPU:** 30% **Memory:** 193.20 MB / 1.98 GB **Up time:** 1.1 h **Version:** 0.9

NODES

2 of 2 nodes / 0 selected / Last 10m

Metrics needing attention are colored

nodes Dashboard	OS CPU (%)	Load (1m)	JVM Mem (%)	Disk Free Space	IOps
Live Wire 10.20.100.101:9301	19.0 min: 6.0 max: 41.0	1.8 min: 1.4 max: 2.9	14.2 min: 10.0 max: 18.0	8.0 GB min: 7.9 GB max: 8.0 GB	121.5 min: 94.5 max: 161.7
Magneto 10.20.100.101:9300	19.0 min: 6.0 max: 39.0	1.8 min: 1.4 max: 2.9	19.9 min: 6.0 max: 31.0	8.0 GB min: 7.9 GB max: 8.0 GB	121.6 min: 94.5 max: 163.1

Click on a metric or select a node for a detailed dashboard

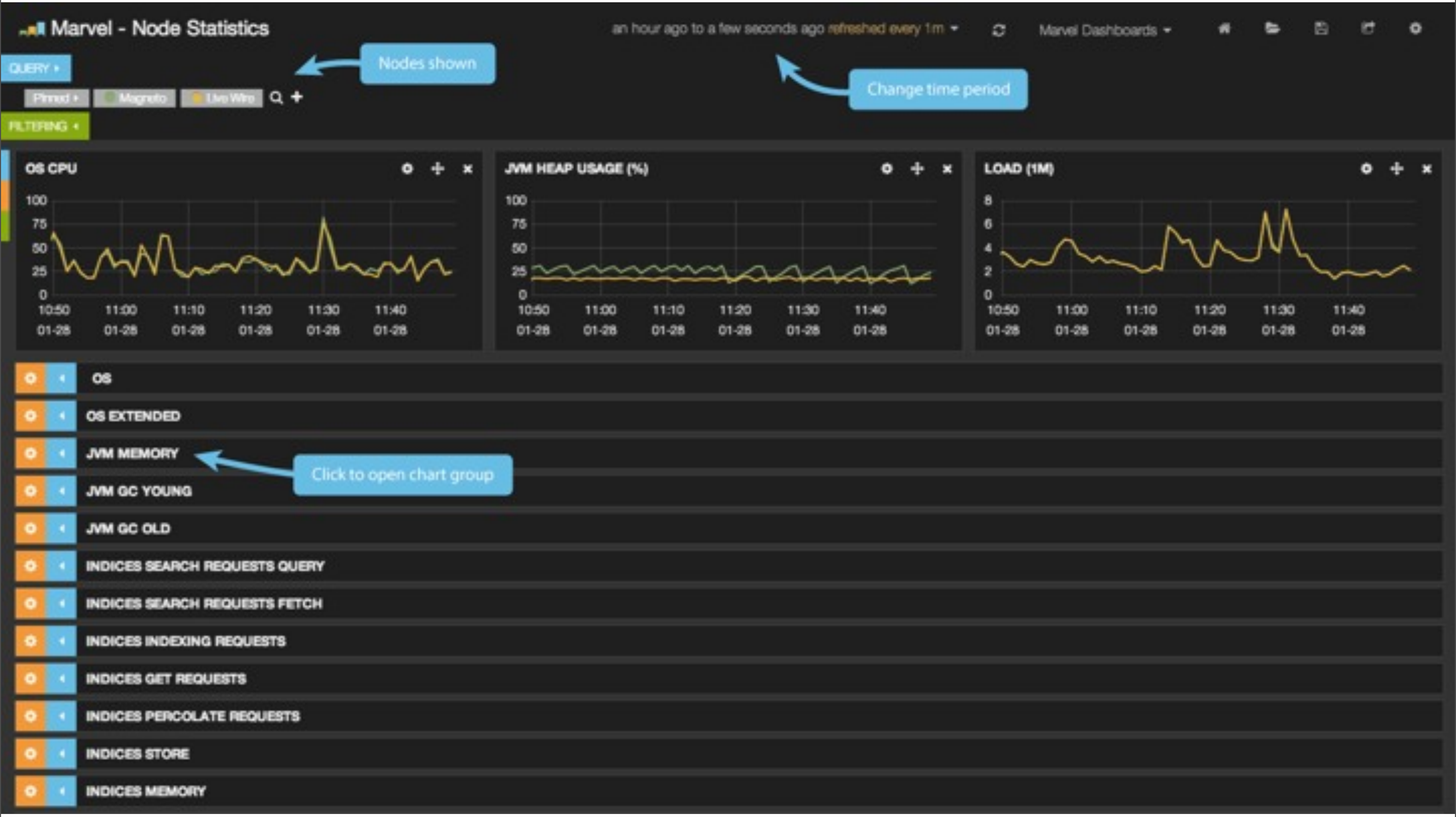
INDICES

2 of 2 indices / 0 selected / Last 10m

Indices Dashboard	Documents	Index Rate	Search Rate	Merge Rate	Field Data

elasticsearch.

Node statistics

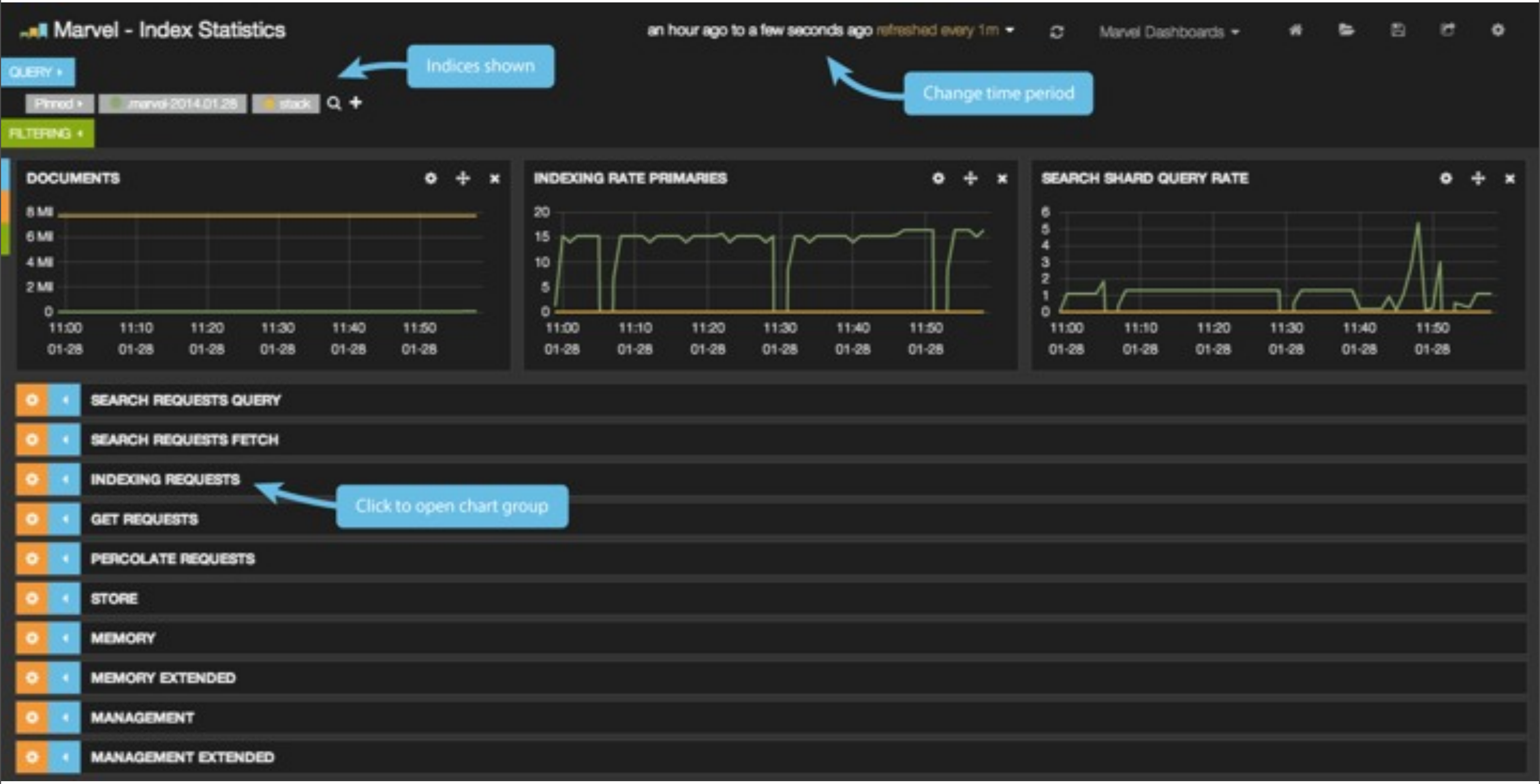


Nodes shown

Change time period

Click to open chart group

Index statistics

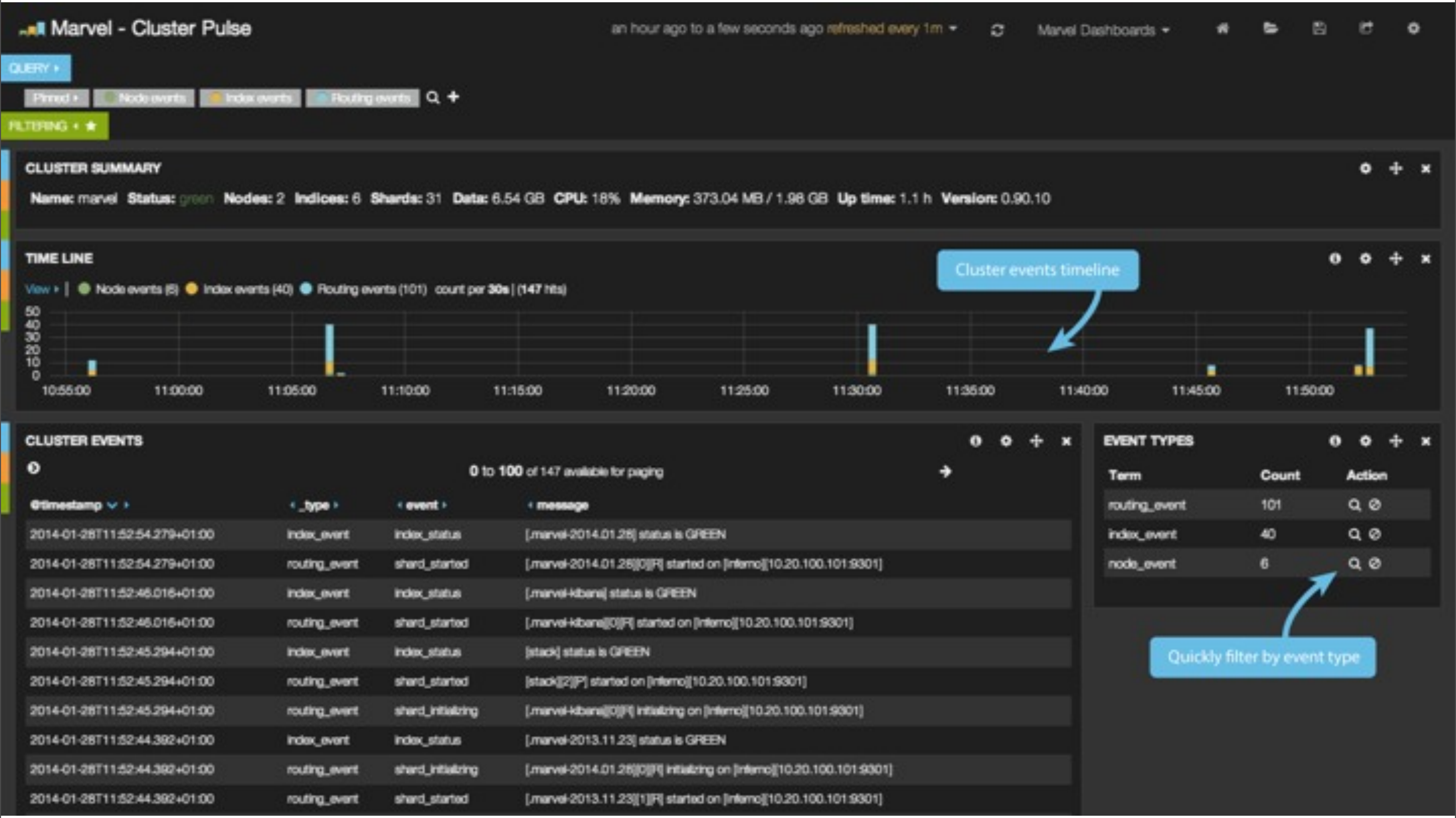


Indices shown

Change time period

Click to open chart group

Cluster Pulse



CLUSTER SUMMARY

Name: marvel **Status:** green **Nodes:** 2 **Indices:** 6 **Shards:** 31 **Data:** 6.54 GB **CPU:** 18% **Memory:** 373.04 MB / 1.98 GB **Up time:** 1.1 h **Version:** 0.90.10

TIME LINE

Cluster events timeline

CLUSTER EVENTS

0 to **100** of 147 available for paging

@timestamp	_type	event	message
2014-01-28T11:52:54.279+01:00	index_event	index_status	[.marvel-2014.01.28] status is GREEN
2014-01-28T11:52:54.279+01:00	routing_event	shard_started	[.marvel-2014.01.28][0][R] started on [Inferno][10.20.100.101:9301]
2014-01-28T11:52:46.016+01:00	index_event	index_status	[.marvel-kibana] status is GREEN
2014-01-28T11:52:46.016+01:00	routing_event	shard_started	[.marvel-kibana][0][R] started on [Inferno][10.20.100.101:9301]
2014-01-28T11:52:45.294+01:00	index_event	index_status	[stack] status is GREEN
2014-01-28T11:52:45.294+01:00	routing_event	shard_started	[stack][2][P] started on [Inferno][10.20.100.101:9301]
2014-01-28T11:52:45.294+01:00	routing_event	shard_initializing	[.marvel-kibana][0][R] initializing on [Inferno][10.20.100.101:9301]
2014-01-28T11:52:44.392+01:00	index_event	index_status	[.marvel-2013.11.23] status is GREEN
2014-01-28T11:52:44.392+01:00	routing_event	shard_initializing	[.marvel-2014.01.28][0][R] initializing on [Inferno][10.20.100.101:9301]
2014-01-28T11:52:44.392+01:00	routing_event	shard_started	[.marvel-2013.11.23][1][R] started on [Inferno][10.20.100.101:9301]

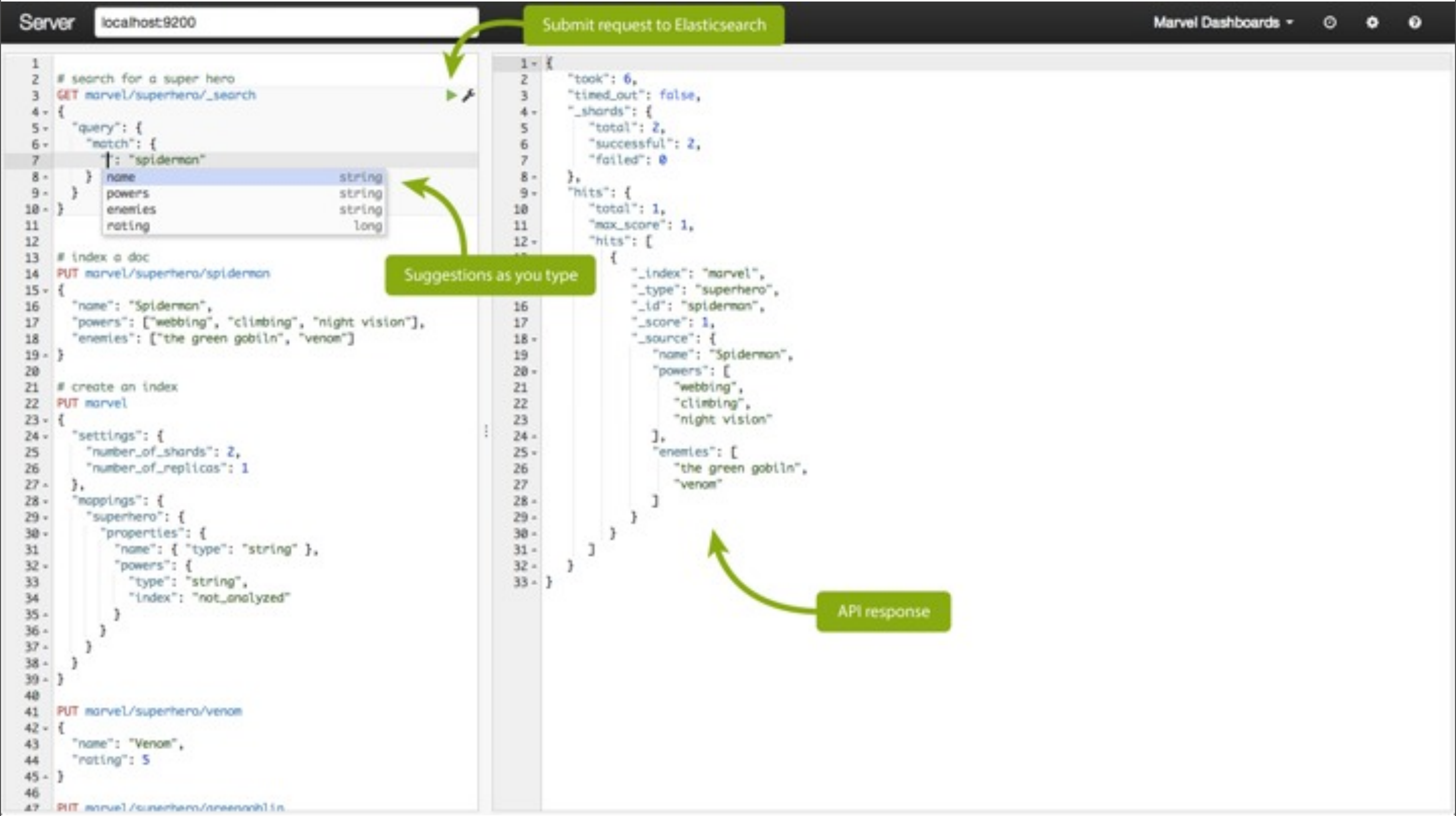
EVENT TYPES

Term	Count	Action
routing_event	101	🔍 ⊘
index_event	40	🔍 ⊘
node_event	6	🔍 ⊘

Quickly filter by event type

elasticsearch.

Sense



Server: localhost:9200

Submit request to Elasticsearch

Suggestions as you type

API response

```

# search for a super hero
GET marvel/superhero/_search
{
  "query": {
    "match": {
      "": "spiderman"
    }
  }
}


# index a doc
PUT marvel/superhero/spiderman
{
  "name": "Spiderman",
  "powers": ["webbing", "climbing", "night vision"],
  "enemies": ["the green gobiln", "venom"]
}

# create an index
PUT marvel
{
  "settings": {
    "number_of_shards": 2,
    "number_of_replicas": 1
  },
  "mappings": {
    "superhero": {
      "properties": {
        "name": { "type": "string" },
        "powers": {
          "type": "string",
          "index": "not_analyzed"
        }
      }
    }
  }
}

PUT marvel/superhero/venom
{
  "name": "Venom",
  "rating": 5
}

PUT marvel/superhero/greengoblin
```

```
{
   "took": 6,
   "timed_out": false,
   "_shards": {
      "total": 2,
      "successful": 2,
      "failed": 0
   },
   "hits": {
      "total": 1,
      "max_score": 1,
      "hits": [
         {
            "_index": "marvel",
            "_type": "superhero",
            "_id": "spiderman",
            "_score": 1,
            "_source": {
               "name": "Spiderman",
               "powers": [
                  "webbing",
                  "climbing",
                  "night vision"
               ],
               "enemies": [
                  "the green gobiln",
                  "venom"
               ]
            }
         }
      ]
   }
}
```

Log analysis with Logstash and Kibana

elasticsearch.

Logstash in 10 seconds

- Managing events and logs
- Collect, parse, enrich, store data
- Modular: many, many inputs and outputs
- Apache License 2.0
- Ruby app (JRuby)
- Part of Elasticsearch family

What is a log?

- Time-based data

- This data is everywhere!

  Server logs
  Twitter stream
  Financial transactions
  Metric / monitoring data
  ...

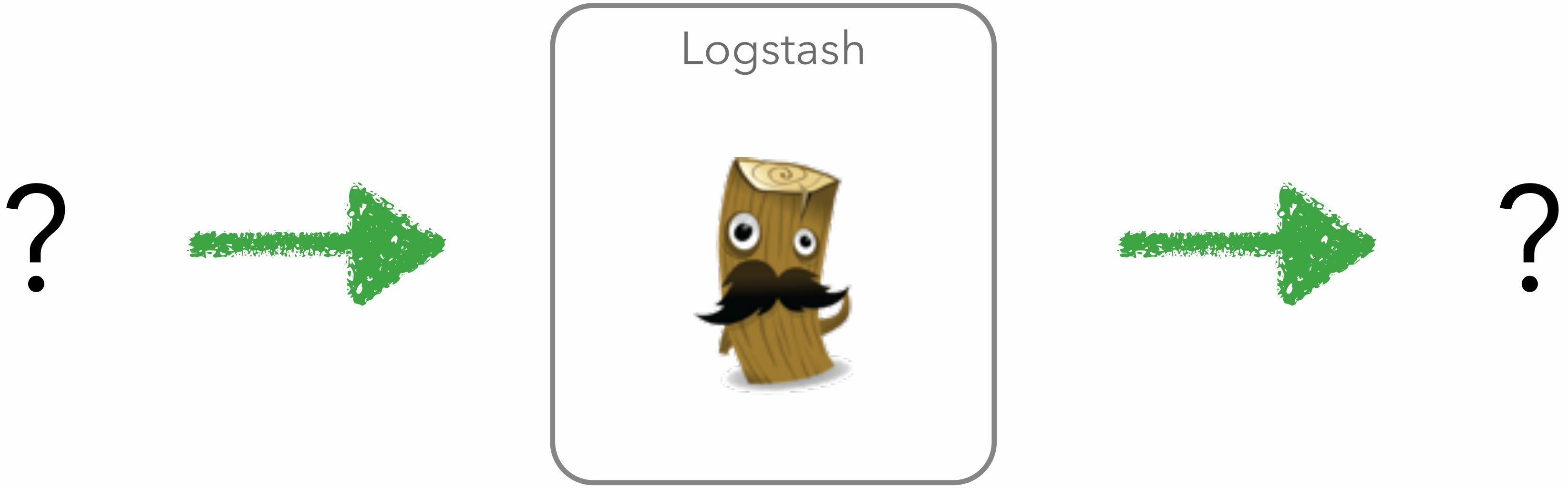
- Log all things

Why collect & centralize logs?

- Access log files without system access
- Shell scripting: Too limited or slow
- Using unique ids for errors, aggregate it across your stack
- Reporting (everyone can create his/her own report)
- Bonus points: Unify your data to make it easily searchable

Logstash architecture

Input	Filter	Output
collect and split	*alter and enrich*	*store and visualize*



? → Logstash → ?

Inputs

- Monitoring: collectd, graphite, ganglia, snmptrap, zenoss
- Datastores: elasticsearch, redis, sqlite, s3
- Queues: rabbitmq, zeromq
- Logging: eventlog, lumberjack, gelf, log4j, relp, syslog, varnish log
- Platforms: drupal_dblog, gemfire, heroku, sqs, s3, twitter
- Local: exec, generator, file, stdin, pipe, unix
- Protocol: imap, irc, stomp, tcp, udp, websocket, wmi, xmpp

Filters

- alter, anonymize, checksum, csv, drop, multiline
- dns, date, extractnumbers, geoip, i18n, kv, noop, ruby, range
- json, urldecode, useragent
- metrics, sleep
- … many, many more …

Outputs

- Store: elasticsearch, gemfire, mongodb, redis, riak, rabbitmq
- Monitoring: ganglia, graphite, graphtastic, nagios, opentsdb, statsd, zabbix
- Notification: email, hipchat, irc, pagerduty, sns
- Protocol: gelf, http, lumberjack, metriccatcher, stomp, tcp, udp, websocket, xmpp
- External Monitoring: boundary, circonus, cloudwatch, datadog, librato
- External service: google big query, google cloud storage, jira, loggly, riemann, s3, sqs, syslog, zeromq
- Local: csv, exec, file, pipe, stdout, null

Installation

- Ruby application, but Java required (JRuby)
- Download single tgz, deb, RPM (also repositories)

  No gem/dependency nightmares!
- Puppet module

Simple example

- Download, create config and run

```
input {
    stdin {}
}

output {
    stdout { debug => true }
}
```

simple.conf

```
echo foo | java -jar logstash-1.3.3-flatjar.jar agent -f simple.conf
{
       "message" => "foo",
      "@version" => "1",
    "@timestamp" => "2014-01-20T13:30:59.648Z",
          "host" => "kryptic.fritz.box"
}
```

Simple filter with grok

```
input {
    stdin {}
}

filter {
  grok {
    match => [ "message", "%{WORD:firstname} %{WORD:lastname} %
{NUMBER:age}" ]
  }
}

output {
    stdout { debug => true }
}
```

Simple filter with grok

```
echo "Alexander Reelsen 30" | java -jar
logstash-1.3.3-flatjar.jar agent -f sample-2.conf
{
       "message" => "Alexander Reelsen 30",
      "@version" => "1",
    "@timestamp" => "2014-01-21T16:56:02.502Z",
          "host" => "kryptic",
     "firstname" => "Alexander",
      "lastname" => "Reelsen",
           "age" => "30"
}
```

elasticsearch.

Syslog example with grok

```
input { stdin {} }

filter {
  grok {
    match => { "message" => "%
{SYSLOGTIMESTAMP:syslog_timestamp} %
{SYSLOGHOST:syslog_hostname} %{DATA:syslog_program}(?:\[%
{POSINT:syslog_pid}\])?: %{GREEDYDATA:syslog_message}" }
  }
  date {
    match => [ "syslog_timestamp",
               "MMM  d HH:mm:ss", "MMM dd HH:mm:ss" ]
  }
}

output { stdout { debug => true } }
```

Syslog example with grok

```
Jun 10 04:04:01 lvps109-104-93-171 postfix/smtpd[11105]:
connect from mail-we0-f196.google.com[74.125.82.196]
```

```
{
             "message" => "Jun 10 04:04:01
lvps109-104-93-171 postfix/smtpd[11105]: connect from
mail-we0-f196.google.com[74.125.82.196]",
            "@version" => "1",
          "@timestamp" => "2014-06-10T04:04:01.000+02:00",
                "host" => "kryptic.local",
    "syslog_timestamp" => "Jun 10 04:04:01",
     "syslog_hostname" => "lvps109-104-93-171",
      "syslog_program" => "postfix/smtpd",
          "syslog_pid" => "11105",
      "syslog_message" => "connect from mail-we0-
f196.google.com[74.125.82.196]"
}
```

CLF log files

```
{
        "message" => "193.99.144.85 - - [23/Jan/2014:17:11:55 +0000]
\"GET / HTTP/1.1\" 200 140 \"-\" \"Mozilla/5.0 (Windows NT 6.1; WOW64)
AppleWebKit/535.19 (KHTML, like Gecko) Chrome/18.0.1025.5 Safari/
535.19\"",
       "@version" => "1",
     "@timestamp" => "2014-01-24T07:56:02.460Z",
           "host" => "kryptic.local",
       "clientip" => "193.99.144.85",
          "ident" => "-",
           "auth" => "-",
      "timestamp" => "23/Jan/2014:17:11:55 +0000",
           "verb" => "GET",
        "request" => "/",
    "httpversion" => "1.1",
       "response" => "200",
          "bytes" => "140",
       "referrer" => "\"-\"",
          "agent" => "\"Mozilla/5.0 (Windows NT 6.1; WOW64)
AppleWebKit/535.19 (KHTML, like Gecko) Chrome/18.0.1025.5 Safari/
535.19\""
}
```

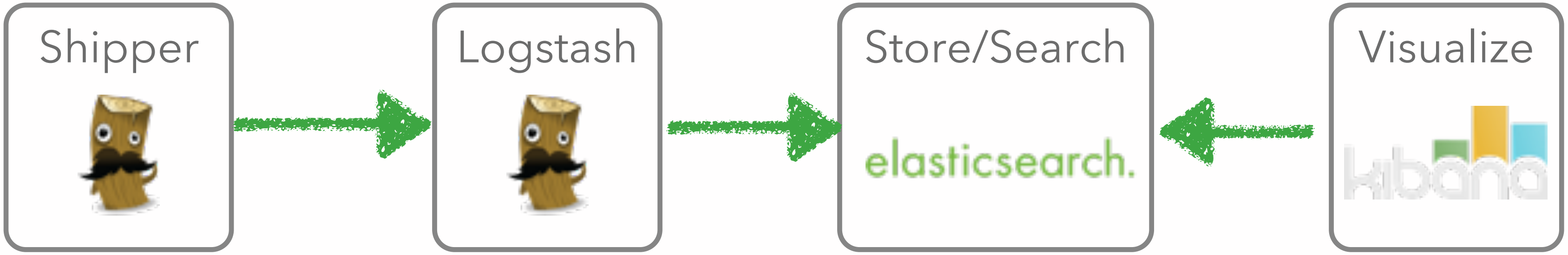
Write to elasticsearch

```
input { stdin {} }

filter {
  grok {
    match => [ message, "%{COMBINEDAPACHELOG}" ]
  }
}

output {
  elasticsearch_http {}
}
```

Deploying ELK for scale



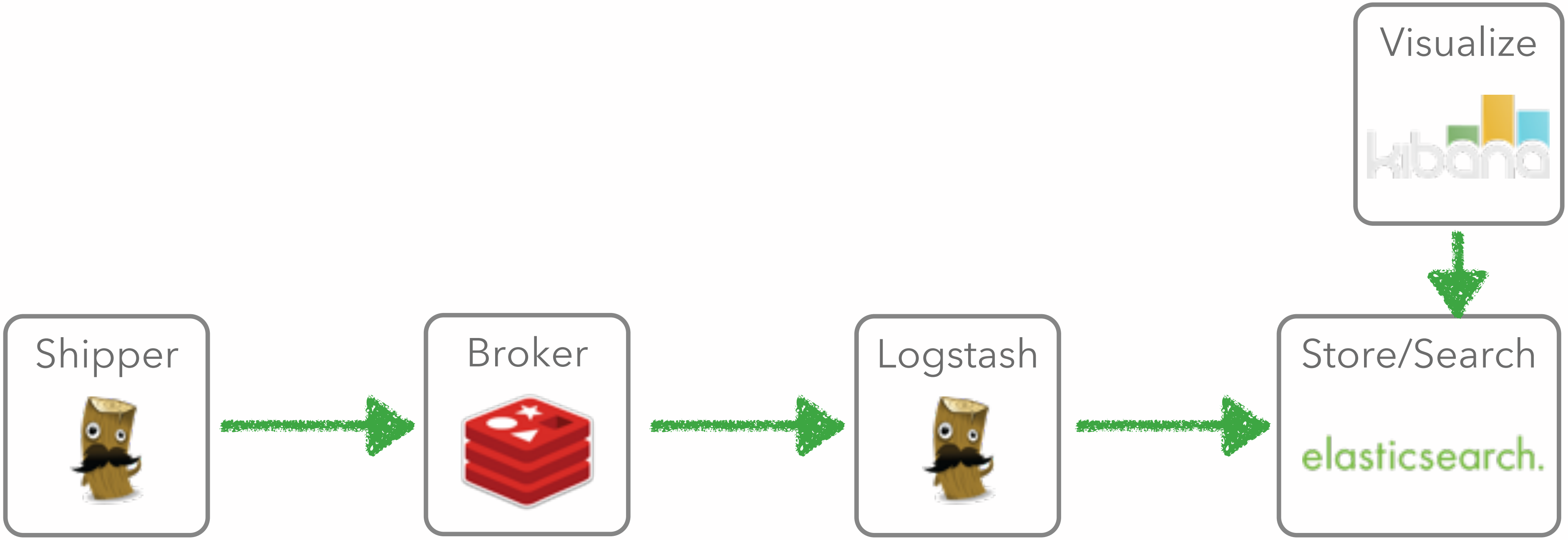
Shipper

Logstash

Store/Search

Visualize

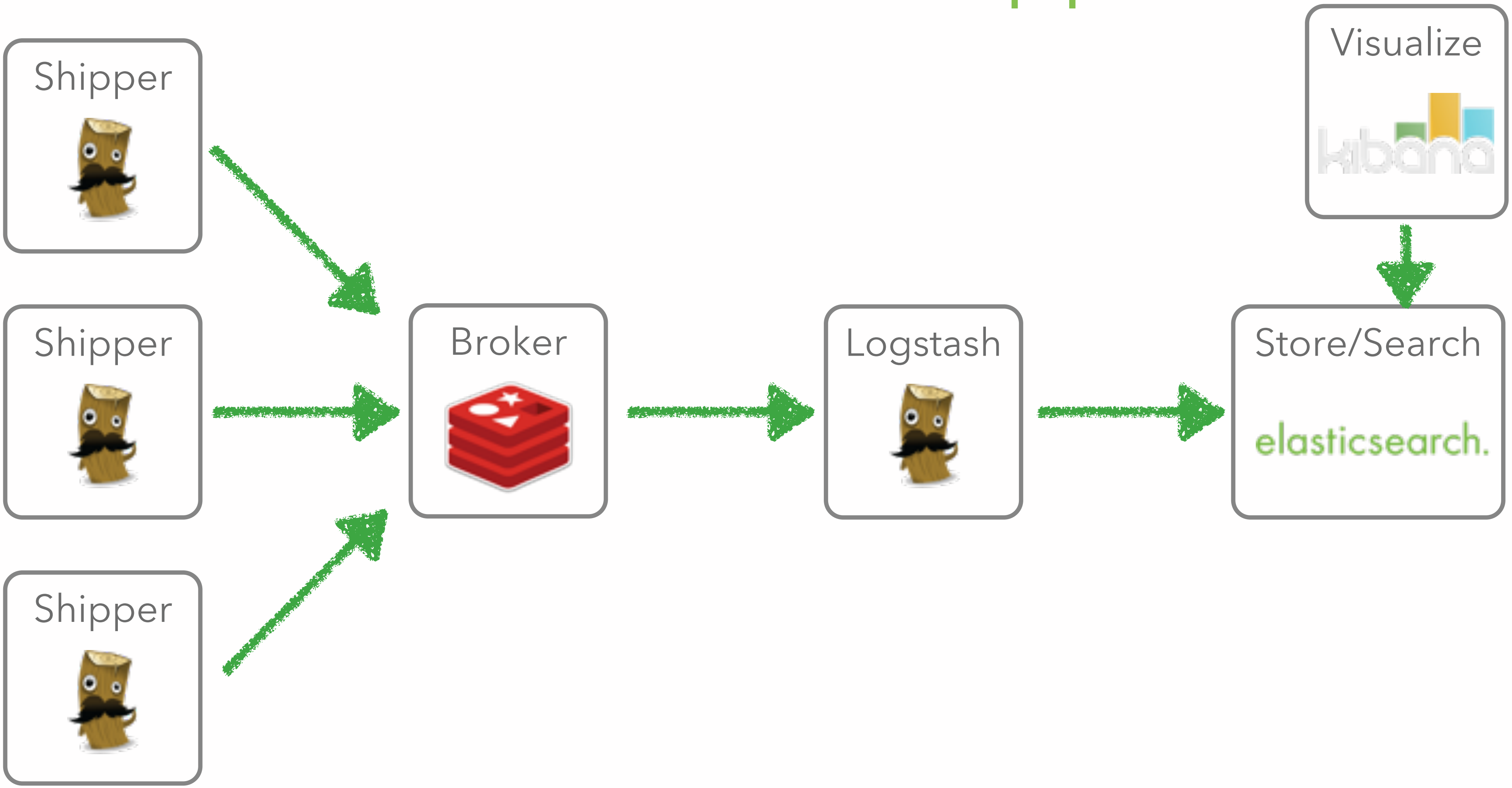
Add a broker



Brokers help with scale and stability by buffering the input and protecting against output downtime.

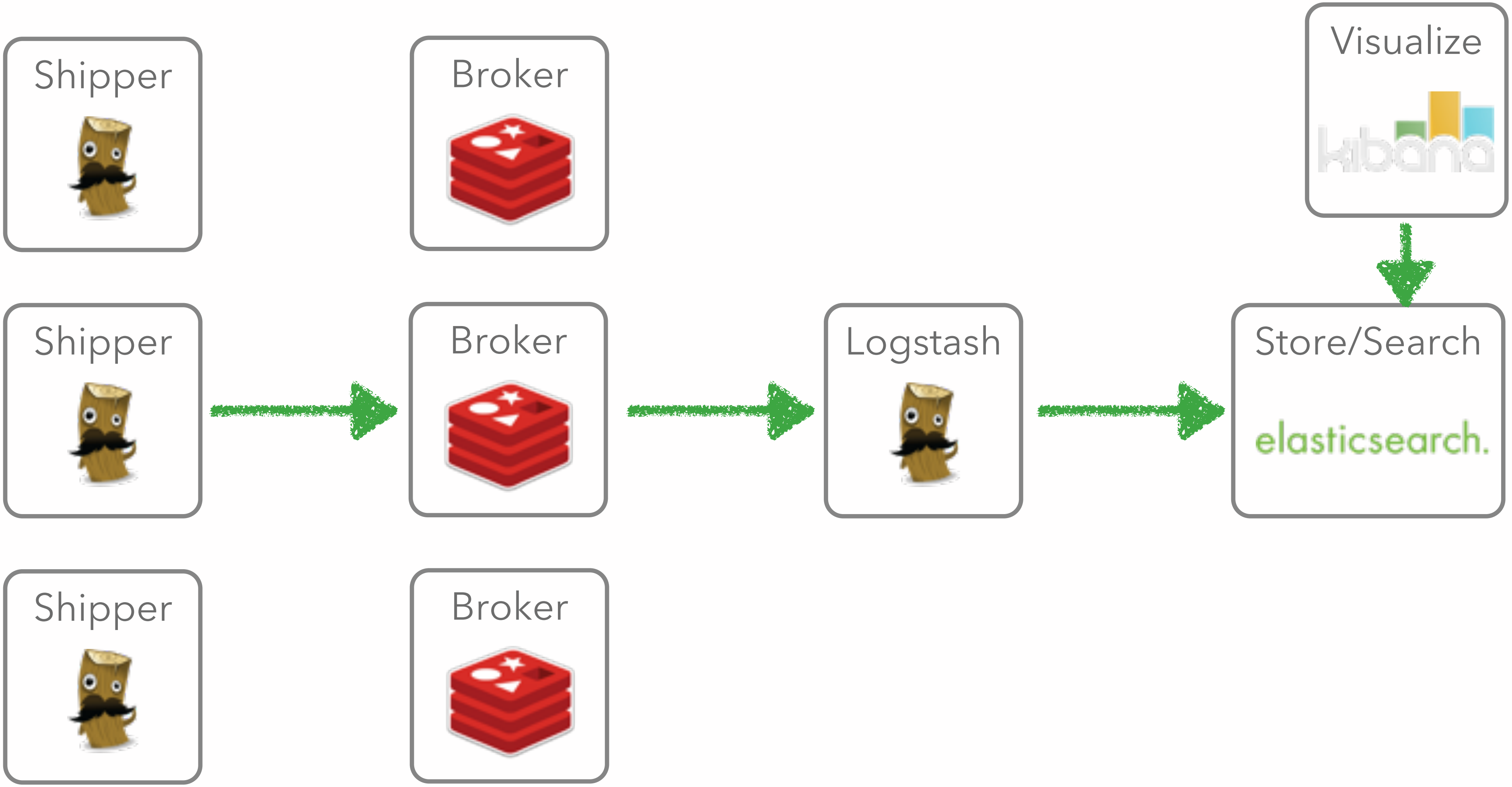
Tip: set limits on broker queue to push back on source as well.

Scale out the shipper



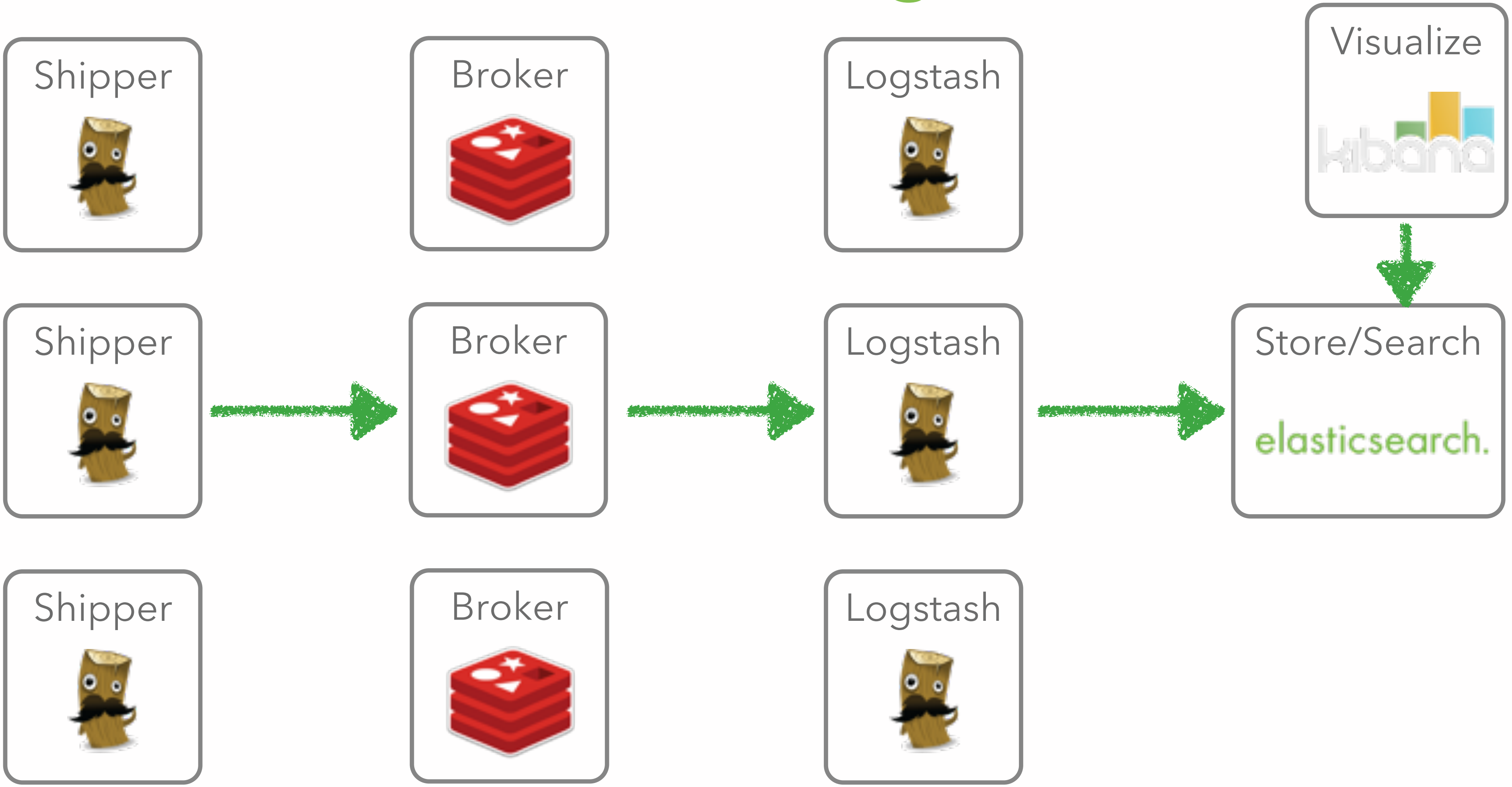
elasticsearch.

Scale out the broker



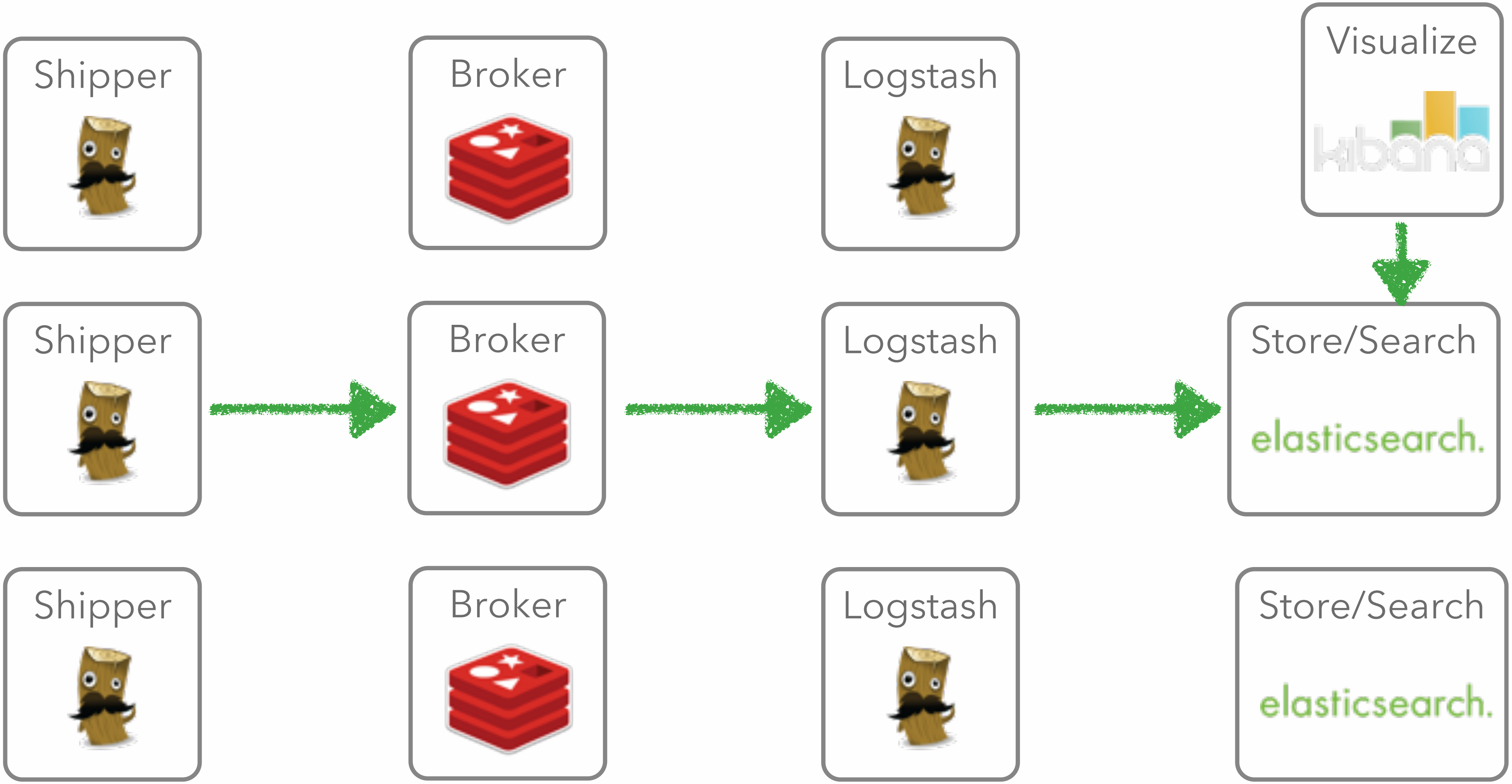
elasticsearch.

Scale out Logstash



elasticsearch.

Scale out Elasticsearch



elasticsearch.

Saturday, February 22, 14

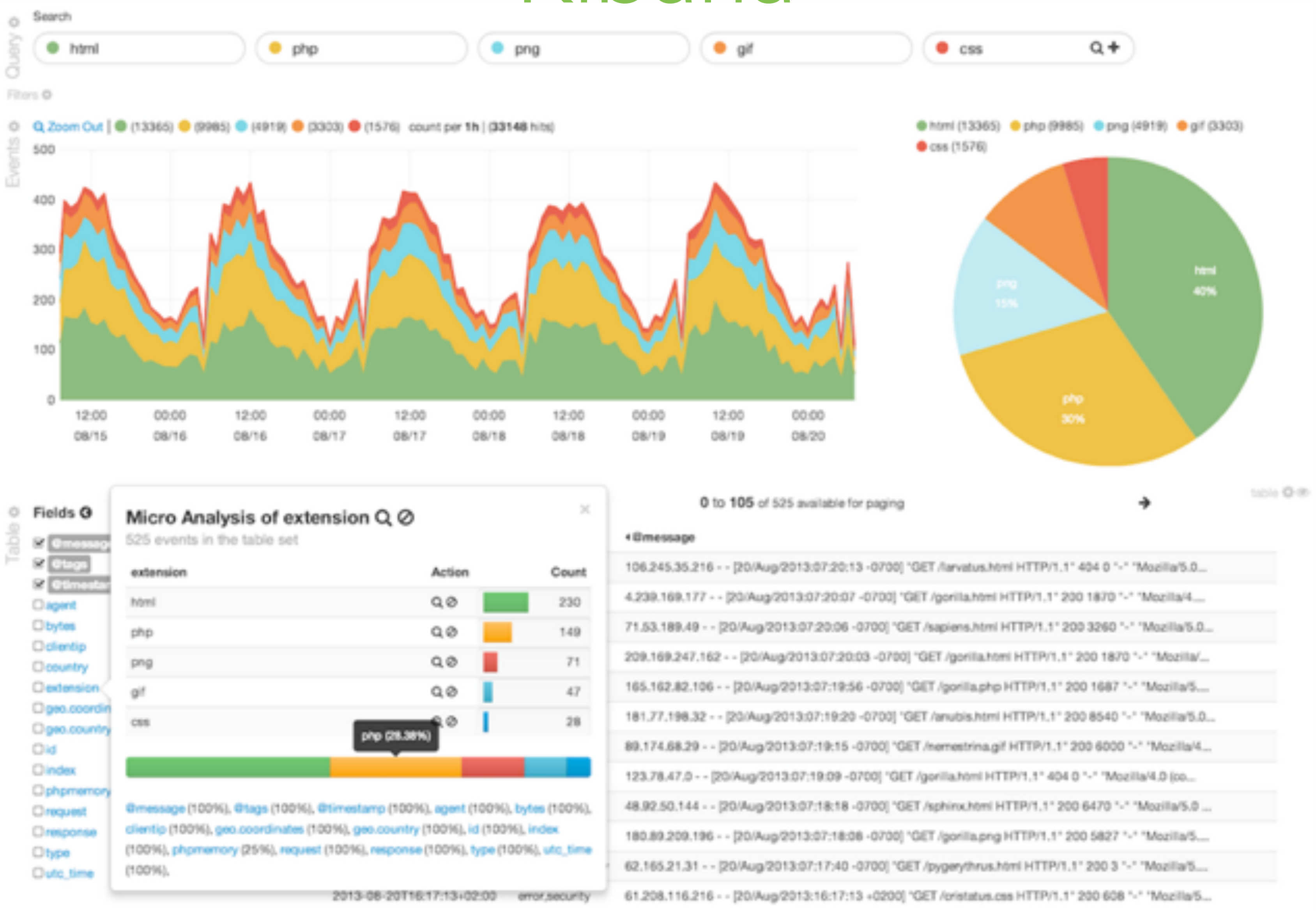
Logstash scaling

- Events get passed via Ruby SizedQueue
- input/worker/output threads, can be configured
- Each input is one thread, unless explicitly configured
- One worker thread by default, use -w to change
- Output is a single thread (some outputs have their own queueing thread)

http://logstash.net/docs/1.3.3/life-of-an-event

Visualize with Kibana

Kibana



Micro Analysis of extension

525 events in the table set

extension	Action	Count
html		230
php		149
png		71
gif		47
css		28

php (28.38%)

@message (100%), @tags (100%), @timestamp (100%), agent (100%), bytes (100%), clientip (100%), geo.coordinates (100%), geo.country (100%), id (100%), index (100%), phpmemory (25%), request (100%), response (100%), type (100%), utc_time (100%).

2013-08-20T16:17:13+02:00 error,security

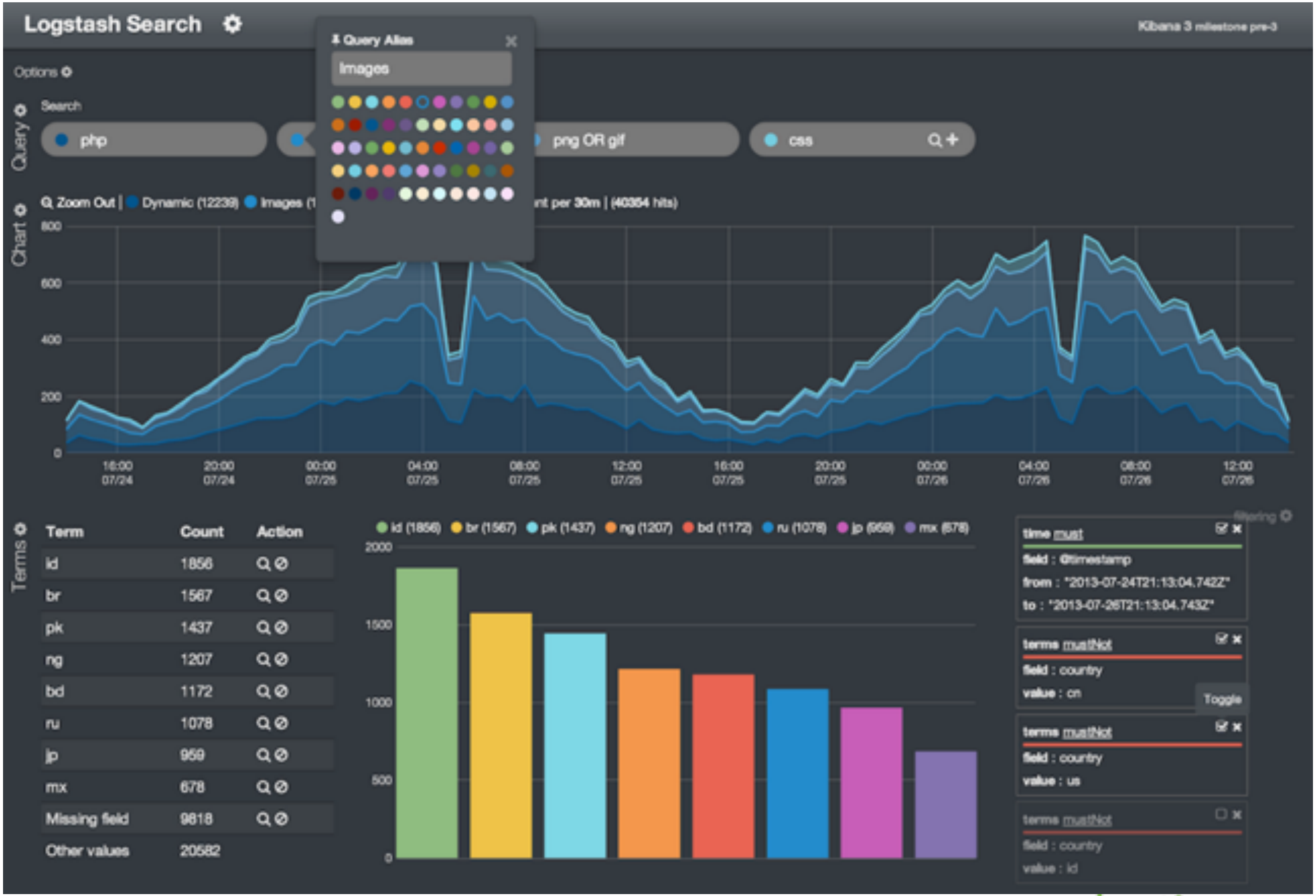
0 to 105 of 525 available for paging

@message

```
106.245.35.216 - - [20/Aug/2013:07:20:13 -0700] "GET /larvatus.html HTTP/1.1" 404 0 "-" "Mozilla/5.0...
4.239.169.177 - - [20/Aug/2013:07:20:07 -0700] "GET /gorilla.html HTTP/1.1" 200 1870 "-" "Mozilla/4....
71.53.189.49 - - [20/Aug/2013:07:20:06 -0700] "GET /sapiens.html HTTP/1.1" 200 3260 "-" "Mozilla/5.0...
209.169.247.162 - - [20/Aug/2013:07:20:03 -0700] "GET /gorilla.html HTTP/1.1" 200 1870 "-" "Mozilla/...
165.162.82.106 - - [20/Aug/2013:07:19:56 -0700] "GET /gorilla.php HTTP/1.1" 200 1687 "-" "Mozilla/5....
181.77.198.32 - - [20/Aug/2013:07:19:20 -0700] "GET /anubis.html HTTP/1.1" 200 8540 "-" "Mozilla/5.0...
89.174.68.29 - - [20/Aug/2013:07:19:15 -0700] "GET /nemestrina.gif HTTP/1.1" 200 6000 "-" "Mozilla/4...
123.78.47.0 - - [20/Aug/2013:07:19:09 -0700] "GET /gorilla.html HTTP/1.1" 404 0 "-" "Mozilla/4.0 (co...
48.92.50.144 - - [20/Aug/2013:07:18:18 -0700] "GET /sphinx.html HTTP/1.1" 200 6470 "-" "Mozilla/5.0 ...
180.89.209.196 - - [20/Aug/2013:07:18:08 -0700] "GET /gorilla.png HTTP/1.1" 200 5827 "-" "Mozilla/5....
62.165.21.31 - - [20/Aug/2013:07:17:40 -0700] "GET /pygerythrus.html HTTP/1.1" 200 3 "-" "Mozilla/5....
61.208.116.216 - - [20/Aug/2013:16:17:13 +0200] "GET /cristatus.css HTTP/1.1" 200 608 "-" "Mozilla/5...
```

elasticsearch.

Kibana



Term	Count	Action
id	1856	🔍 ⊘
br	1567	🔍 ⊘
pk	1437	🔍 ⊘
ng	1207	🔍 ⊘
bd	1172	🔍 ⊘
ru	1078	🔍 ⊘
jp	959	🔍 ⊘
mx	678	🔍 ⊘
Missing field	9818	🔍 ⊘
Other values	20582	

time must ☑ ✕
field : @timestamp
from : "2013-07-24T21:13:04.742Z"
to : "2013-07-26T21:13:04.743Z"

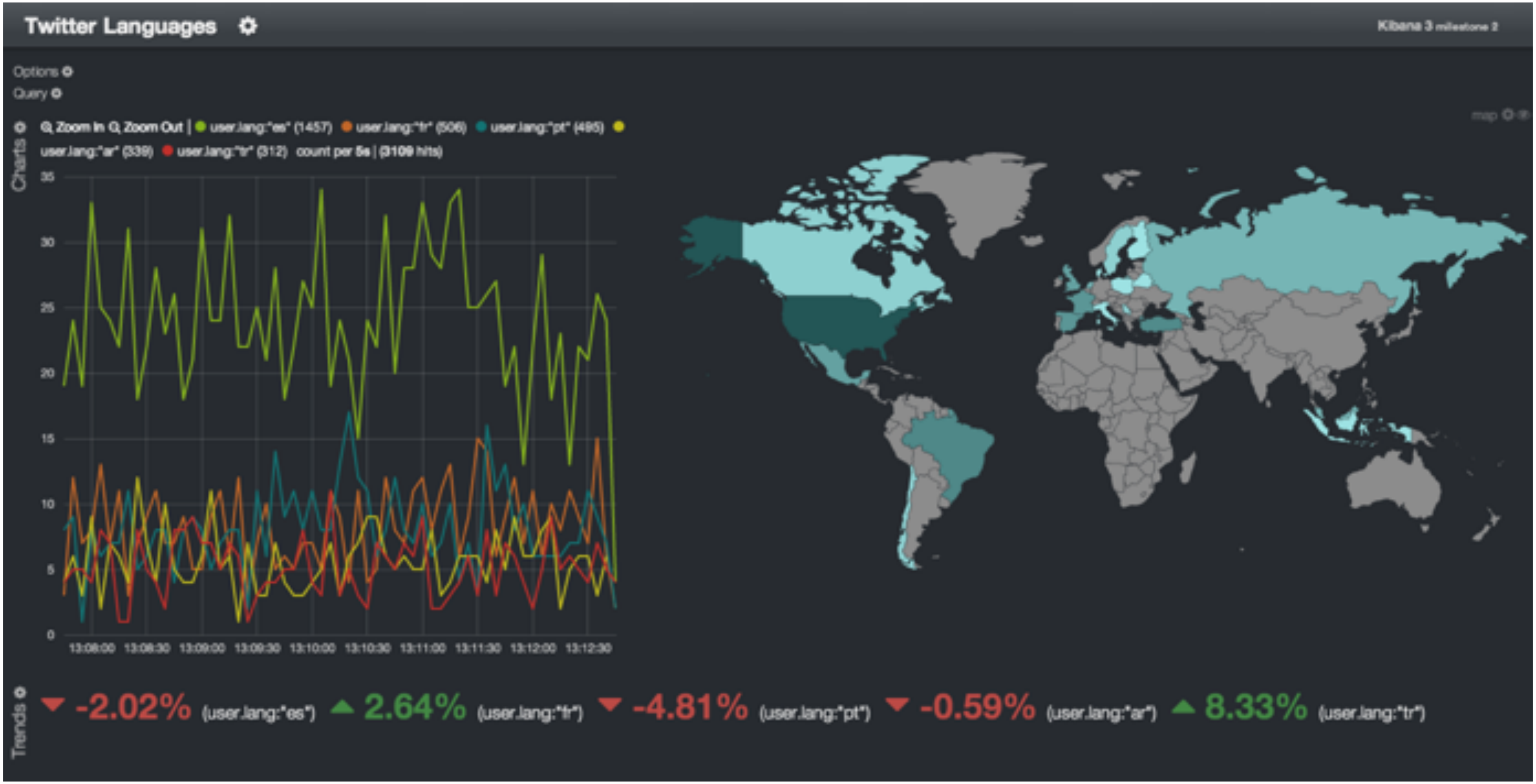
terms mustNot ☑ ✕
field : country
value : cn
Toggle

terms mustNot ☑ ✕
field : country
value : us

terms mustNot □ ✕
field : country
value : id

Saturday, February 22, 14

Kibana


Twitter Languages
Kibana 3 milestone 2
Options
Query
Charts
Zoom In
Zoom Out
● user.lang:"es" (1457)
● user.lang:"fr" (506)
● user.lang:"pt" (495)
● user.lang:"ar" (339)
● user.lang:"tr" (312)
count per 5s | (3109 hits)
35
30
25
20
15
10
5
0
13:08:00
13:08:30
13:09:00
13:09:30
13:10:00
13:10:30
13:11:00
13:11:30
13:12:00
13:12:30
map
Trends
▼ -2.02% (user.lang:"es")
▲ 2.64% (user.lang:"fr")
▼ -4.81% (user.lang:"pt")
▼ -0.59% (user.lang:"ar")
▲ 8.33% (user.lang:"tr")

Useful helpers

- Curator: index management
  http://www.elasticsearch.org/blog/curator-tending-your-time-series-indices/
- Puppet module
  https://github.com/elasticsearch/puppet-logstash
- logstash forwarder: low overhead collector
  https://github.com/elasticsearch/logstash-forwarder
- Logstash cookbook
  http://cookbook.logstash.net/

More info

- Github: https://github.com/elasticsearch

  Code, issues there

  Except Logstash issues at https://logstash.jira.com

- Mailing lists

  Google groups, logstash-users and elasticsearch

- IRC channels

  #logstash and #elasticsearch on freenode

- We're hiring!

  jobs@elasticsearch.com