

# Elasticsearch 1.0

Alexander Reelsen

@spinscale

[alexander.reelsen@elasticsearch.com](mailto:alexander.reelsen@elasticsearch.com)

# Agenda

- Introduction
- Elasticsearch 1.0 features
  - Aggregations
  - Snapshot/Restore
  - Distributed/Scalable percolator
  - cat API
- Elasticsearch 1.1 features
- Q & A

# about

- Me

Interested in metrics, ops and the web

Likes the JVM

Working with elasticsearch since 2011

- Elasticsearch, founded in 2012

Products: Elasticsearch, Logstash, Kibana, Marvel

Professional services: Support & development subscriptions

Trainings

# Introduction

# Unstructured search

GitHub

Explore Features Enterprise Blog

Sign up

Sign in

Search

elasticsearch

Search

|                |        |
|----------------|--------|
| 📁 Repositories | 317    |
| 🔗 Code         | 17,981 |
| 🔔 Issues       | 2,008  |
| 👤 Users        | 2      |

Languages

|            |     |
|------------|-----|
| Java       | 167 |
| 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



elasticsearch.

# Structured search

GitHub

Explore Features Enterprise Blog

Sign up

Sign in

Search

elasticsearch

Search

|                |        |
|----------------|--------|
| 📁 Repositories | 317    |
| 🔗 Code         | 17,981 |
| 🔔 Issues       | 2,008  |
| 👤 Users        | 2      |

Languages

|            |     |
|------------|-----|
| Java       | 317 |
| 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

elasticsearch.

# Enrichment

GitHub

Explore Features Enterprise Blog

Sign up

Sign in

Search

elasticsearch

Search

|                |        |
|----------------|--------|
| 📁 Repositories | 317    |
| 🔗 Code         | 17,981 |
| 🔔 Issues       | 2,008  |
| 👤 Users        | 2      |

Languages

|            |     |
|------------|-----|
| Java       | 167 |
| 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

# Sorting

GitHub

Explore Features Enterprise Blog

Sign up

Sign in

Search

elasticsearch

Search

Sort: Best match

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

Languages

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

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.

# Pagination

GitHub

Explore Features Enterprise Blog

Sign up

Sign in

Search

elasticsearch

Search

|    |              |        |
|----|--------------|--------|
| 📁  | 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 FST Suggester

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

GitHub

Explore Features Enterprise Blog

Sign up

Sign in

Search

elasticsearch

Search

|                |       |
|----------------|-------|
| 📁 Repositories | 317   |
| 🔗 Code         | 7,981 |
| 🔔 Issues       | 1,008 |
| 👤 Users        | 2     |

Languages

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

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

elasticsearch.

# Suggestions



**GitHub** This repository:

**elasticsearch**

**Sign up** **Sign in**

★ **Star** 4,683 **Fork** 1,097

**Browse Issues** **Everyone's Issues** **New Issue**

**Labels**

- Lucene 4.5 Upgrade
- breaking
- bug
- enhancement
- feature
- non-issue

Search elasticsearch/elasticsearch for 'debian'

Search GitHub for 'debian'

1 2 3 ... 19

**Forms** #3702

**reproducible** #3701

**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 in 10 seconds

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

# Installation & first steps

# Zero configuration

```
$ wget https://download.elasticsearch.org/...  
$ tar -xf elasticsearch-1.1.0.tar.gz  
$ ./elasticsearch-1.1.0/bin/elasticsearch  
...  
[2014-01-19 14:53:11,508][INFO ][node] [Scanner] started  
...
```

# Is it alive?

```
» curl localhost:9200
```

```
{  
  "status" : 200,  
  "name" : "Scanner",  
  "version" : {  
    "number" : "1.1.0",  
    "build_hash" : "e018cda7e7a32643d59e0ac3cdb412ccc239af04",  
    "build_timestamp" : "2014-03-25T15:11:47Z",  
    "build_snapshot" : true,  
    "lucene_version" : "4.7.0"  
  },  
  "tagline" : "You Know, for Search"  
}
```

# 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
      }
    } ]
  }
}
```

# 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
        }
      }
    }
  }
}'
```

# Scalability

# Distributed & scalable

- Replication
  - Read scalability
  - Removing SPOF
- Sharding
  - Split logical data over several machines
  - Write scalability
  - Control data flows

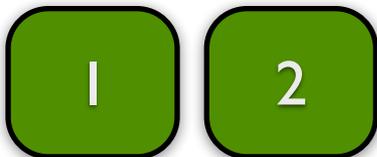
# Distributed & scalable

## node 1 (m)

### orders



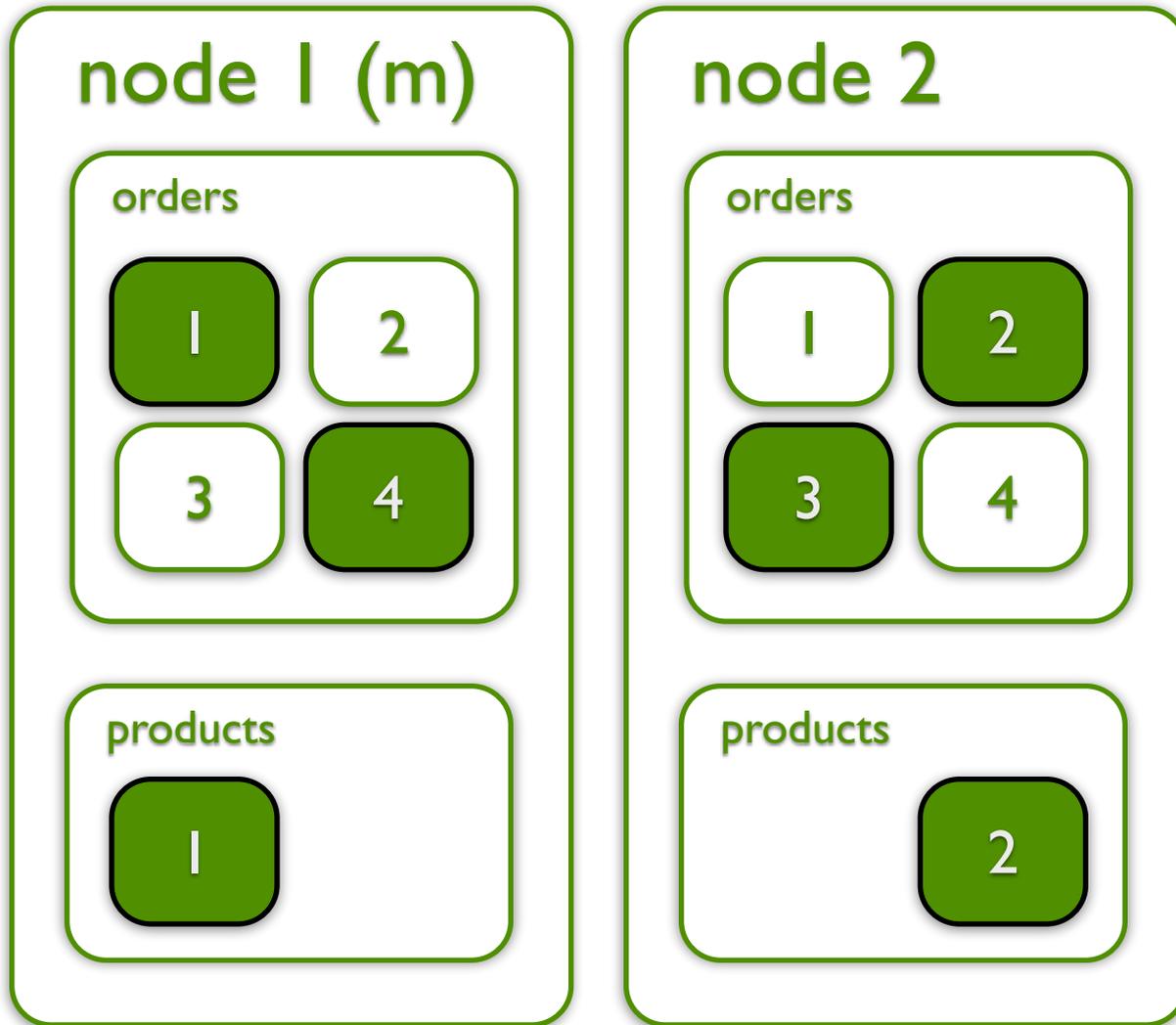
### products



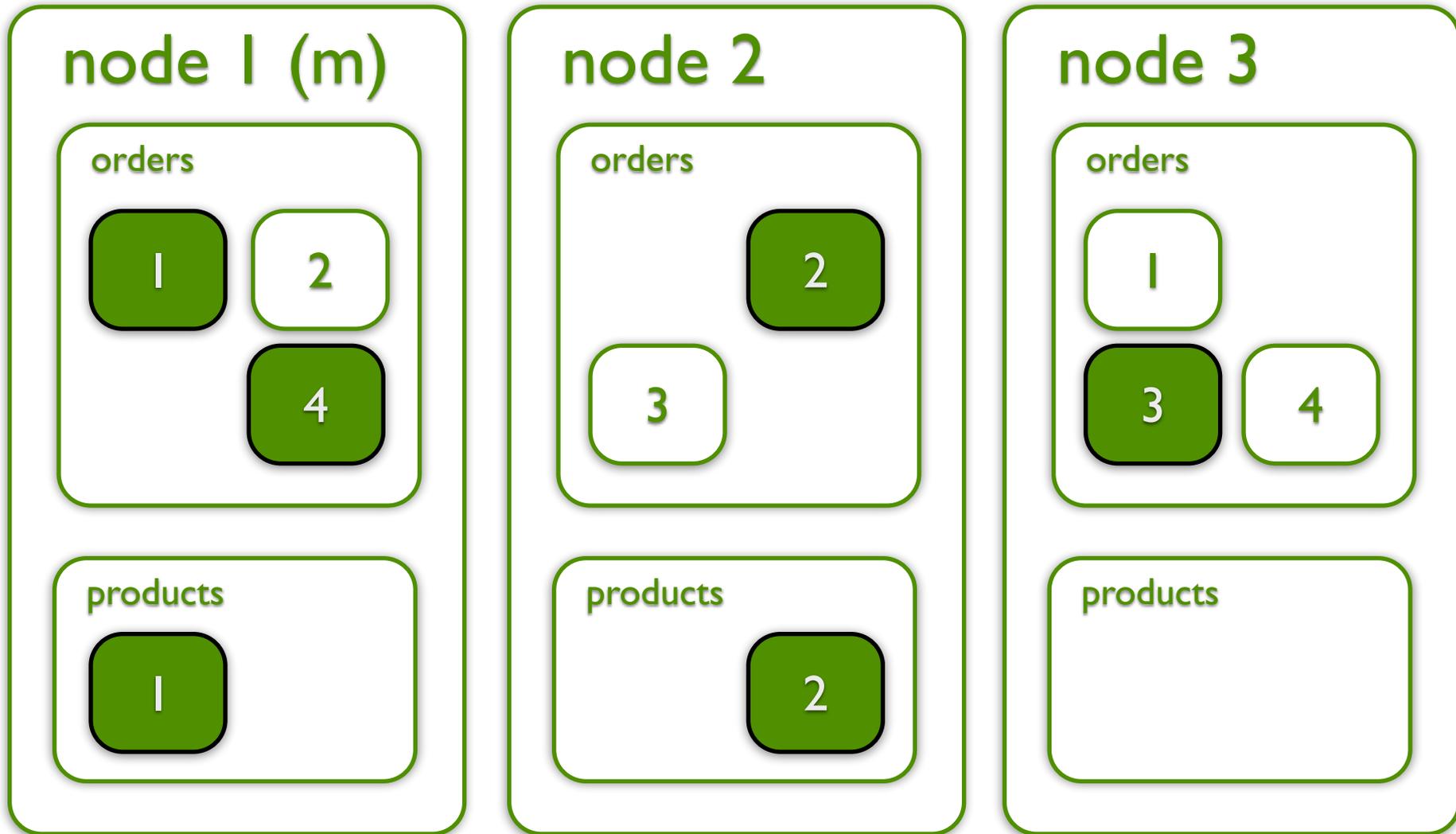
```
curl -X PUT localhost:9200/orders -d '{  
  "settings.index.number_of_shards" : 4  
  "settings.index.number_of_replicas" : 1  
}'
```

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

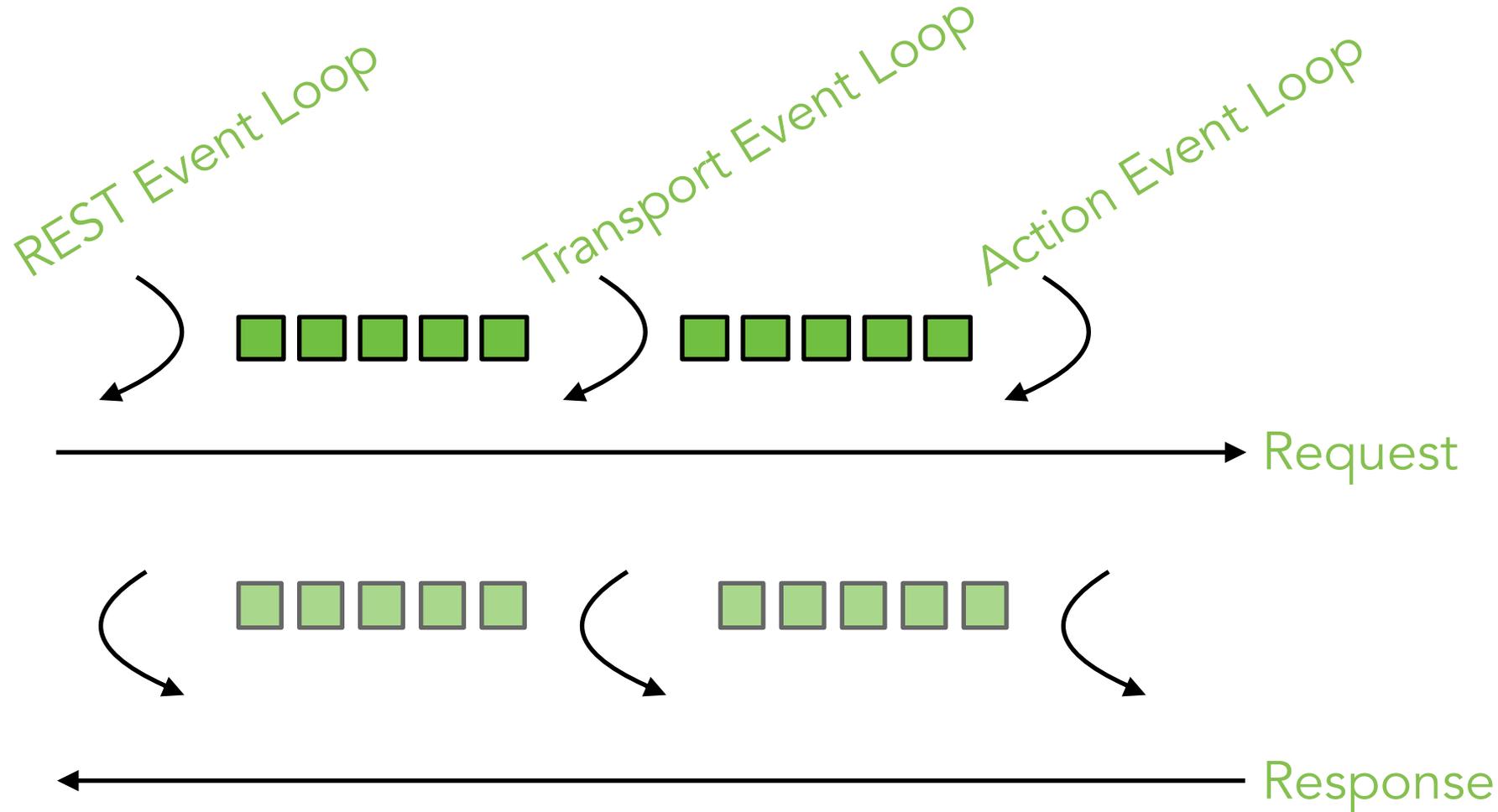
# Distributed and scalable



# Distributed & scalable



# A request under the hood



# Think async!

- Enforces event driven architecture
- Support for non-blocking model
- Enforce loose coupling
- Prefers push over pull
- Callback based concurrency
- Helps to avoid contention on resources / threads

# Elasticsearch 1.0/1.1

# Elasticsearch

## 1.0

- Aggregations
- Snapshot/Restore
- Distributed percolator
- Cat API
- ... and more

## 1.1

- Cardinality Agg
- Percentiles Agg
- Significant Terms Agg
- Search Templates
- Cross fields search

# Road to 1.1

- v0.4.0 - Feb 8, 2010
- v0.5.0 - Mar 5, 2010
- ...
- v0.19.0 - Mar 1, 2012
- v0.20.0 - Dec 7, 2012
- v0.90.0 - Apr 29, 2013
- v1.0.0 - Feb 12, 2014
- v1.1.0 - Mar 25, 2014

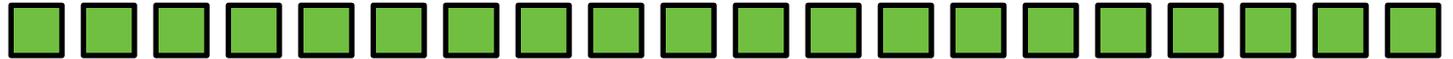
# Aggregations

# Aggregations

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

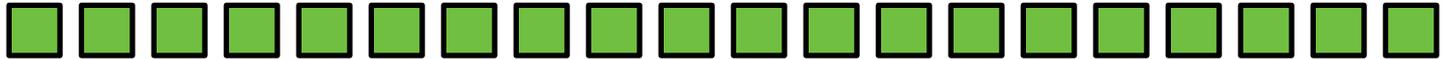
# Aggregations

**Documents**

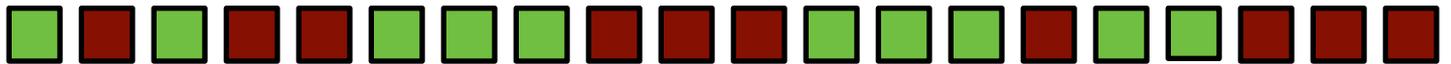


# Aggregations

Documents

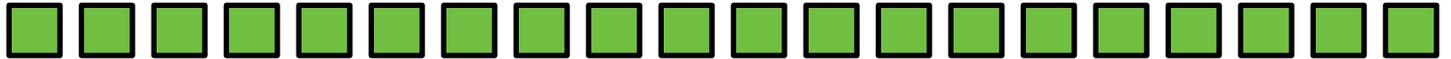


Query

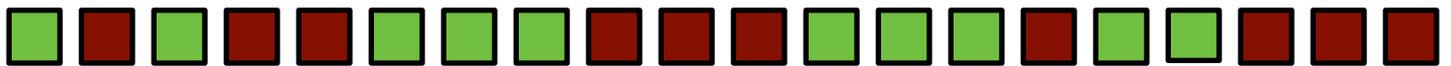


# Aggregations

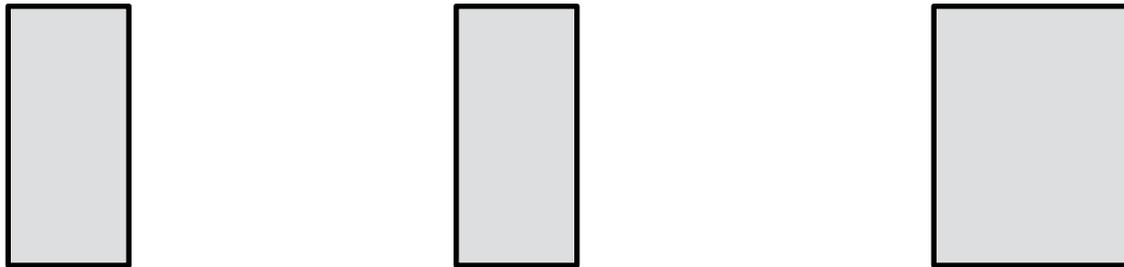
Documents



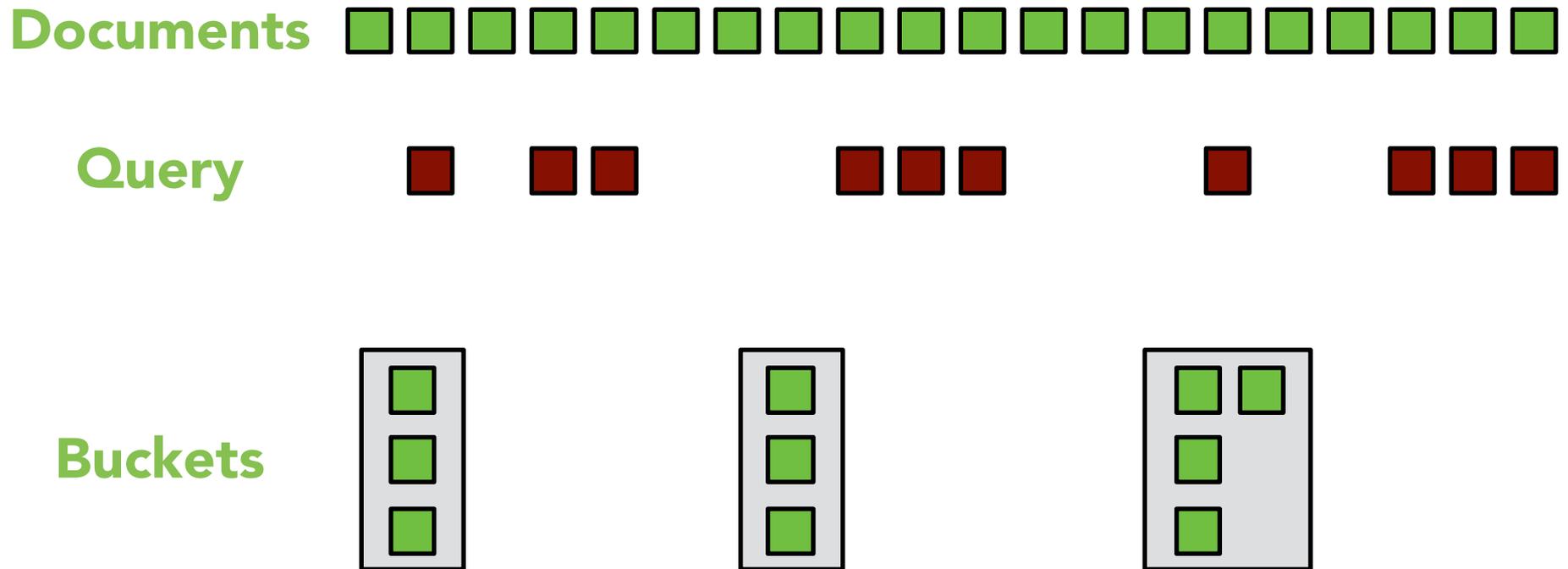
Query



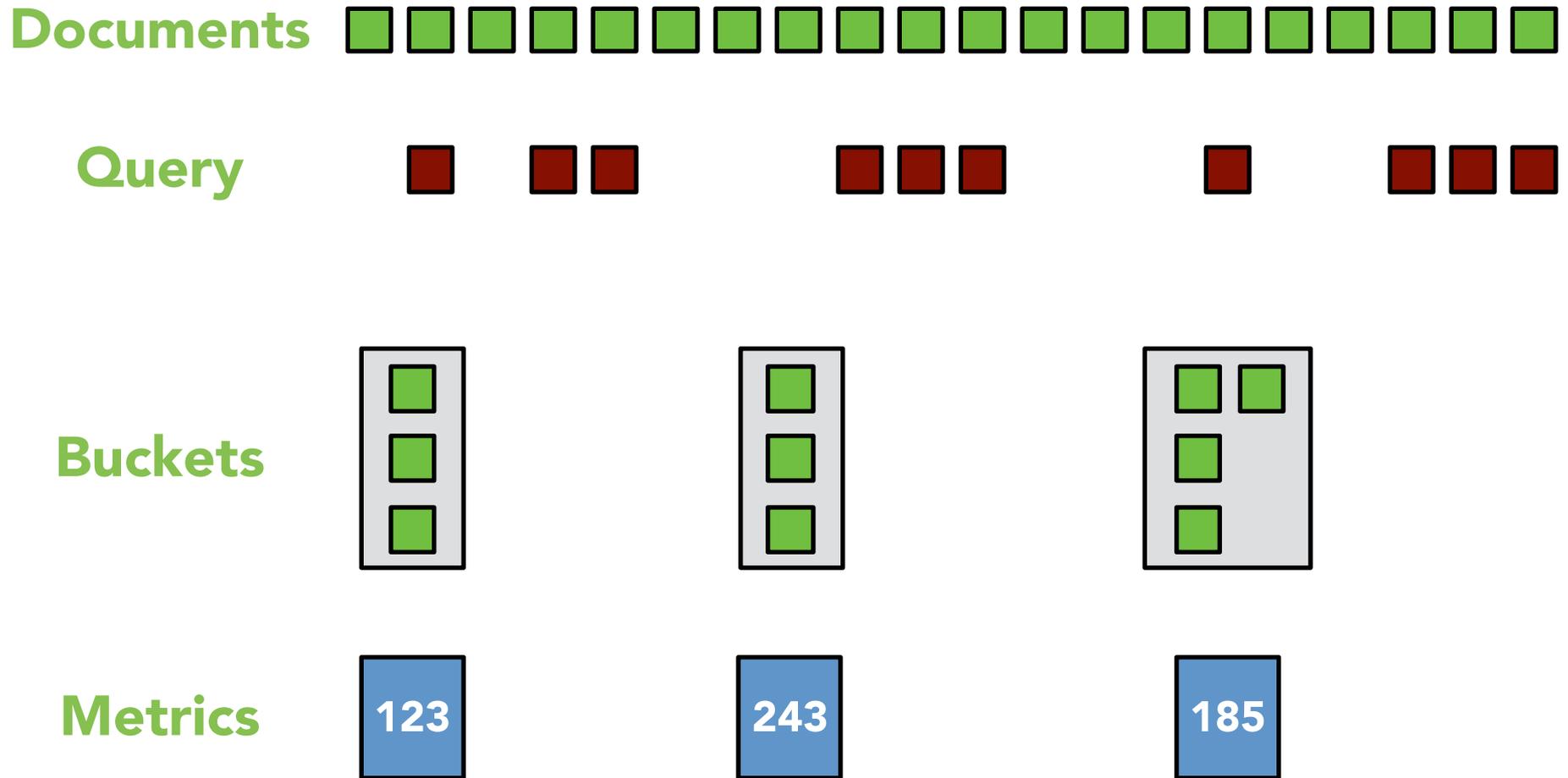
Buckets



# Aggregations



# Aggregations



# bucket aggregators

- global
- filter
- missing
- terms
- range
- date range
- ip range
- histogram
- date histogram
- geo distance
- nested

# metrics aggregators

- count
- stats
- extended stats
- avg
- max
- min
- sum

# Order average

```
» curl -XGET 'localhost:9200/orders/order/_search' -d '{
  "aggs": {
    "average_order_size" : {
      "avg" : { "field" : "total" }
    }
  }
}
```

```
...
  "aggregations": {
    "average_order_size" : {
      "value" : 658.369
    }
  }
...

```

# Order average - filters

```
{
  "aggs": {
    "average_order_size_january": {
      "filter": {
        "range": { "created_at": { "gte": "2014-01-01", "lt": "2014-02-01" } } },
      "aggs": {
        "avg": { "field": "total" }
      }
    }
  }
}
```

```
...
  "aggregations": {
    "average_order_size_january": {
      "doc_count": 8,
      "value": 540.89754
    }
  }
  ...
```

# Order average - by day

```
{
  "aggs": {
    "by_day": {
      "filter": {
        "range": {
          "created_at": {
            "gte": "2014-01-01", "lt": "2014-02-01"
          }
        }
      },
      "aggs": {
        "daily_filter": {
          "date_histogram": {
            "field": "created_at",
            "interval": "day",
            "format": "yyyy-MM-dd"
          },
          "aggs": {
            "average_order_size": { "avg" : { "field" : "total" } }
          }
        }
      }
    }
  }
}
```

# Order average - by day

```
...
  "aggregations": {
    "by_day" : {
      "doc_count" : 32422,
      "daily_filter" : [ {
        "key_as_string" : "2014-01-01",
        "key" : 1388534400000
        "doc_count" : 423,
        "average_order_size" : {
          "value" : 380.0
        }
      }, {
        "key_as_string" : "2014-01-02",
        "key" : 1388534400000
        "doc_count" : 543,
        "average_order_size" : {
          "value" : 323.432
        }
      }, {
        ...
      ]
    }
  }
  ...
}
```

# Order average - by hour

```
{
  "aggs": {
    "by_day": {
      "filter": {
        "range": {
          "created_at": {
            "gte": "2014-01-01", "lt": "2014-02-01"
          }
        }
      },
      "aggs": {
        "hourly_filter": {
          "histogram": {
            "script": "doc[\u0027created_at\u0027].date.hourOfDay",
            "interval": 1
          },
          "aggs": {
            "average_order_size": { "avg": { "field": "total" } }
          }
        }
      }
    }
  }
}
```

# Order average - by hour

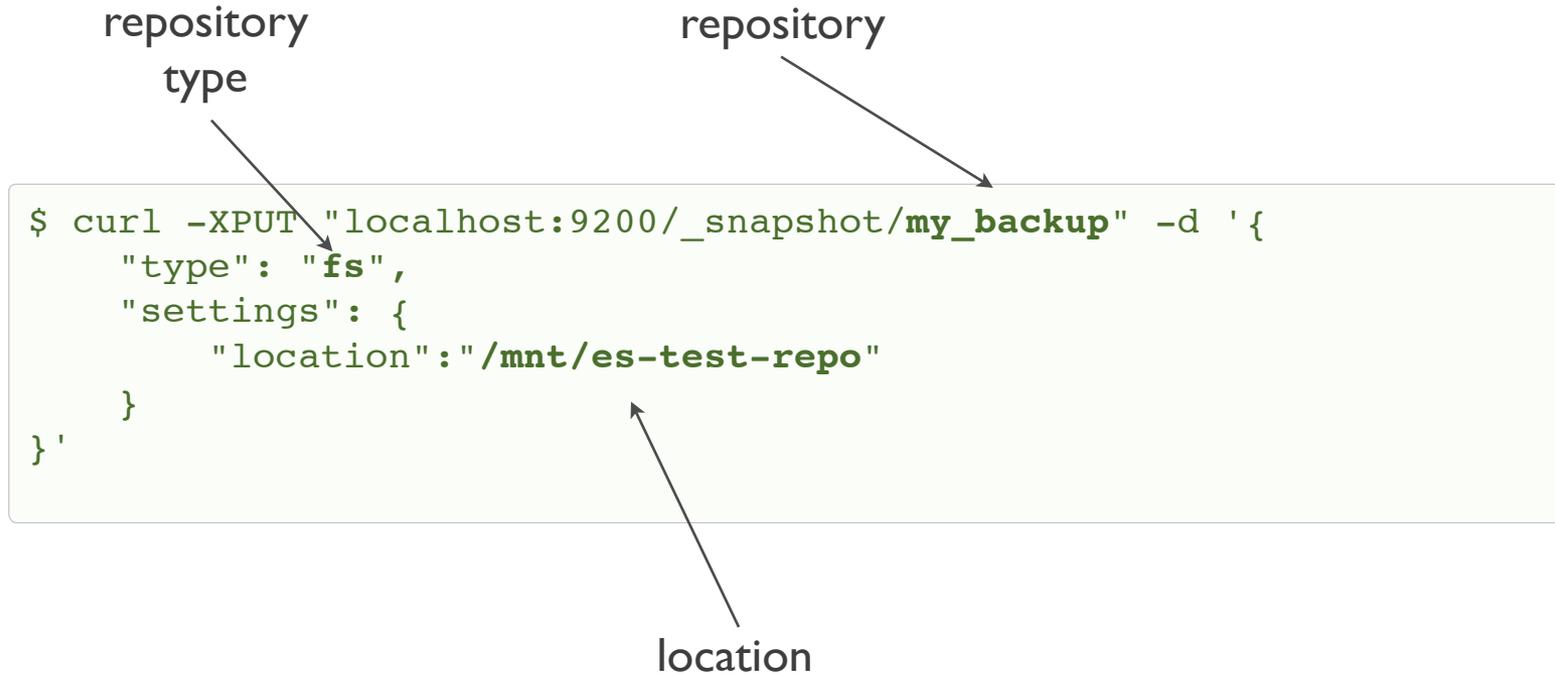
```
...
  "aggregations": {
    "by_day" : {
      "doc_count" : 32422,
      "daily_filter" : [ {
        "key" : "11",
        "doc_count" : 1534,
        "average_order_size" : {
          "value" : 380.0
        }
      }, {
        "key" : "18",
        "doc_count" : 8923,
        "average_order_size" : {
          "value" : 485.4323
        }
      }, {
        ...
      ]
    }
  }
  ...
}
```

# Snapshot/Restore

<http://www.elasticsearch.org/blog/introducing-snapshot-restore/>

# Backup made easy

- Several shell commands + login were needed for pre 1.0 backups, but not via API



# Start snapshot

repository

```
$ curl -XPUT "localhost:9200/_snapshot/my_backup/snapshot_20131010" -d '{  
  "indices": "+test_*,-test_4"  
}'
```

index list  
(optional)

snapshot  
name

# Restore snapshot

close all indices  
that start with test\_

```
$ curl -XPOST "localhost:9200/test_*/_close"
```

repository  
name

snapshot  
name

```
$ curl -XPOST "localhost:9200/_snapshot/my_backup/snapshot_20131010" -d  
'{  
  "indices": "test_*"  
'
```

index  
list

# Distributed & scalable Percolator

<http://www.elasticsearch.org/blog/percolator-redesign-blog-post/>

# percolator

- reverse search
- alerts
- updatable search results

# registering percolator in 0.90

target  
index

query id

```
$ curl -XPUT "localhost:9200/_percolator/tweeter/es-tweets" -d '{
  "query": {
    "match": { "text": "elasticsearch" }
  }
}'
```

# document percolation in 0.90

target  
index

percolation  
end point

```
$ curl -XGET "localhost:9200/twitter/tweet/_percolate" -d '{
  "doc": {
    "text": "#elasticsearch is awesome"
    "nick": "@imotov"
    "name": "Igor Motov"
    "date": "2013-11-03"
  }
}'
```

document  
to be percolated

```
{
  "ok": true
  "matches": ["es-tweets"]
}
```

matching  
queries

# how does it work in 0.90?

- all queries are stored in special `_percolate` index
- `_percolate` index has 1 primary shard which is replicated to every node
- each percolated document is indexed in memory
- all queries are executed against this document sequentially
- **execution time is linear to number of queries!**

# registering percolator in 1.0

reserved percolator  
type

query id

```
$ curl -XPUT "localhost:9200/some_index/.percolator/es-tweets" -d '{
  "query": {
    "match": { "body": "elasticsearch" }
  }
}'
```

any index with as  
many shards as you  
need

# multi index support

```
$ curl -XGET "localhost:9200/twitter,facebook/_percolate" -d '{
  "doc": {
    "body": "#elasticsearch is awesome"
    "nick": "@imotov"
    "name": "Igor Motov"
    "date": "2013-11-03"
  }
}'
```

document  
to be percolated



# other features

- percolation of existing document
- percolate count api
- filter support (in addition to queries in 0.90)
- highlighting, scoring
- multi-index, aliases support
- multi percolate (bulk percolation)

# Cat API

<http://www.elasticsearch.org/blog/introducing-cat-api/>

# Helping sysadmins

- 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 one 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? (v0.90)

```
$ curl "localhost:9200/_cluster/state?pretty&filter_metadata=true&filter_routing_table=true"
{
  "cluster_name" : "elasticsearch",
  "master_node" : "GNf0hEX1TfaBvQXKBF300A",
  "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" : { }
    },
    "GNf0hEX1TfaBvQXKBF300A" : {
      "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
```

# /cat/count

```
$ curl localhost:9200/_cat/count  
1383501234301 12:53:54 3344067
```

count



# `_cat/*` api

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

And more...

# And more...

- Disk-based fielddata

<http://www.elasticsearch.org/blog/disk-based-field-data-a-k-a-doc-values/>

- Fielddata circuit breaker
- Federated search

And 1.1...

# And 1.1...

- Aggregations
  - Percentile
  - Significant terms
  - Cardinality
- Improved multi field search
- Search Templates
- Create index/template: Support for aliases

Thanks for listening

# Q & A

P.S. We're hiring  
<http://elasticsearch.com/about/jobs>  
<http://elasticsearch.com/support>

Alexander Reelsen  
@spinscale  
alexander.reelsen@elasticsearch.com