

elasticsearch in 15 minutes

David Pilato
@dadoonet



Elasticsearch.com



- Créée en 2012 par les auteurs d'Elasticsearch
- **Formation** (publique et intra)
- **Consulting** (support dév)
- Abonnement annuel **support pour la production** avec 3 niveaux de SLA

Plug & Play

Installation

```
$ wget https://download.elasticsearch.org/...  
$ tar -xf elasticsearch-0.90.2.tar.gz  
$ ./elasticsearch-0.90.2/bin/elasticsearch -f  
... [INFO ][node][Ghost Maker] {0.90.2}[5645]: initializing ...
```

Index a document...

```
$ curl -XPUT localhost:9200/products/product/1 -d '{  
  "title" : "Welcome!"  
}'
```

Update a document...

```
$ curl -XPUT localhost:9200/products/product/1 -d '{  
  "title" : "Welcome to the Elasticsearch meetup!"  
}'
```

Documents as JSON

Data structure with basic types, arrays and deep hierarchies

```
{
  "id"          : "abc123",
  "title"       : "A JSON Document",
  "body"        : "A JSON document is a ...",
  "published_on" : "2013/06/27 10:00:00",
  "featured"    : true,
  "tags"        : ["search", "json"],
  "author"      : {
    "first_name" : "Clara",
    "last_name"  : "Rice",
    "email"      : "clara@rice.org"
  }
}
```

Search for documents....

```
$ curl localhost:9200/products/_search?q=welcome
```

Add a node...

```
$ ./elasticsearch-0.90.2/bin/elasticsearch -f -D es.node.name=Node2  
...[cluster.service] [Node2] detected_master [Node1] ...
```

Node1 [KgDkQwLHRQyodn9f3v4Gyw]

+	-----+	+	-----+
	products	0	■
	products	1	■
	products	2	■
	products	3	■
	products	4	■

+ -----+ | + | -----+ |

Unassigned: 5 shards

+	-----+	+	-----+
	products	0	□
	products	1	□
	products	2	□
	products	3	□
	products	4	□

+ -----+ | + | -----+ |

Add a node...

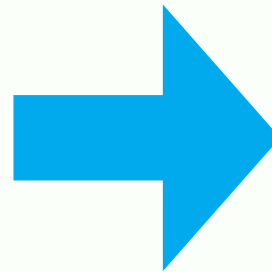
```
$ ./elasticsearch-0.90.2/bin/elasticsearch -f -D es.node.name=Node2  
...[cluster.service] [Node2] detected_master [Node1] ...
```

Node1 [KgDkQwLHRQyodn9f3v4Gyw]

products	0	■
products	1	■
products	2	■
products	3	■
products	4	■

Unassigned: 5 shards

products	0	□
products	1	□
products	2	□
products	3	□
products	4	□



Node1 [KgDkQwLHRQyodn9f3v4Gyw]

products	0	■
products	1	■
products	2	■
products	3	■
products	4	■

Node2 [xzgjTuScRVyJ3cJEyUhPqQ]

products	0	□
products	1	□
products	2	□
products	3	□
products	4	□

Add another node...

```
$ ./elasticsearch-0.90.2/bin/elasticsearch -f -D es.node.name=Node3  
...[cluster.service] [Node3] detected_master [Node1] ...
```

Node1 [KgDkQwLHRQyodn9f3v4Gyw]

+	-----+	+	-----+
	products	0	■
	products	1	■
	products	2	■
	products	3	■
	products	4	■

+

Node2 [xzgjTuScRVyJ3cJEyUhPqQ]

+	-----+	+	-----+
	products	0	□
	products	1	□
	products	2	□
	products	3	□
	products	4	□

+

Add another node...

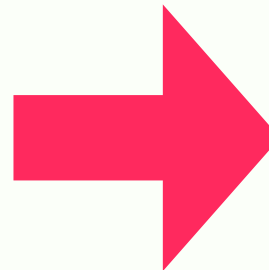
```
$ ./elasticsearch-0.90.2/bin/elasticsearch -f -D es.node.name=Node3  
...[cluster.service] [Node3] detected_master [Node1] ...
```

Node1 [KgDkQwLHRQyodn9f3v4Gyw]

products	0	■
products	1	■
products	2	■
products	3	■
products	4	■

Node2 [xzgjTuScRVyJ3cJEyUhPqQ]

products	0	□
products	1	□
products	2	□
products	3	□
products	4	□



Node3 [rUPD5n0dQkuR32FbZeEg4w]

products	0	■
products	2	□
products	4	■

Node2 [zjHOMsjcTtmeBWqRmY6H7A]

products	0	□
products	1	□
products	3	□
products	4	□

Node1 [KgDkQwLHRQyodn9f3v4Gyw]

products	1	■
products	2	■
products	3	■

Add another node...

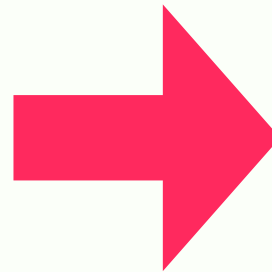
```
$ ./elasticsearch-0.90.2/bin/elasticsearch -f -D es.node.name=Node3  
...[cluster.service] [Node3] detected_master [Node1] ...
```

Node1 [KgDkQwLHRQyodn9f3v4Gyw]

products	0	■
products	1	■
products	2	■
products	3	■
products	4	■

Node2 [xzgjTuScRVyJ3cJEyUhPqQ]

products	0	□
products	1	□
products	2	□
products	3	□
products	4	□



Node3 [rUPD5n0dQkuR32FbZeEg4w]

products	0	■
products	2	□
products	4	■

Node2 [zjHOMsjcTtmeBWqRmY6H7A]

products	0	□
products	1	□
products	3	□
products	4	□

Node1 [KgDkQwLHRQyodn9f3v4Gyw]

products	1	■
products	2	■
products	3	■

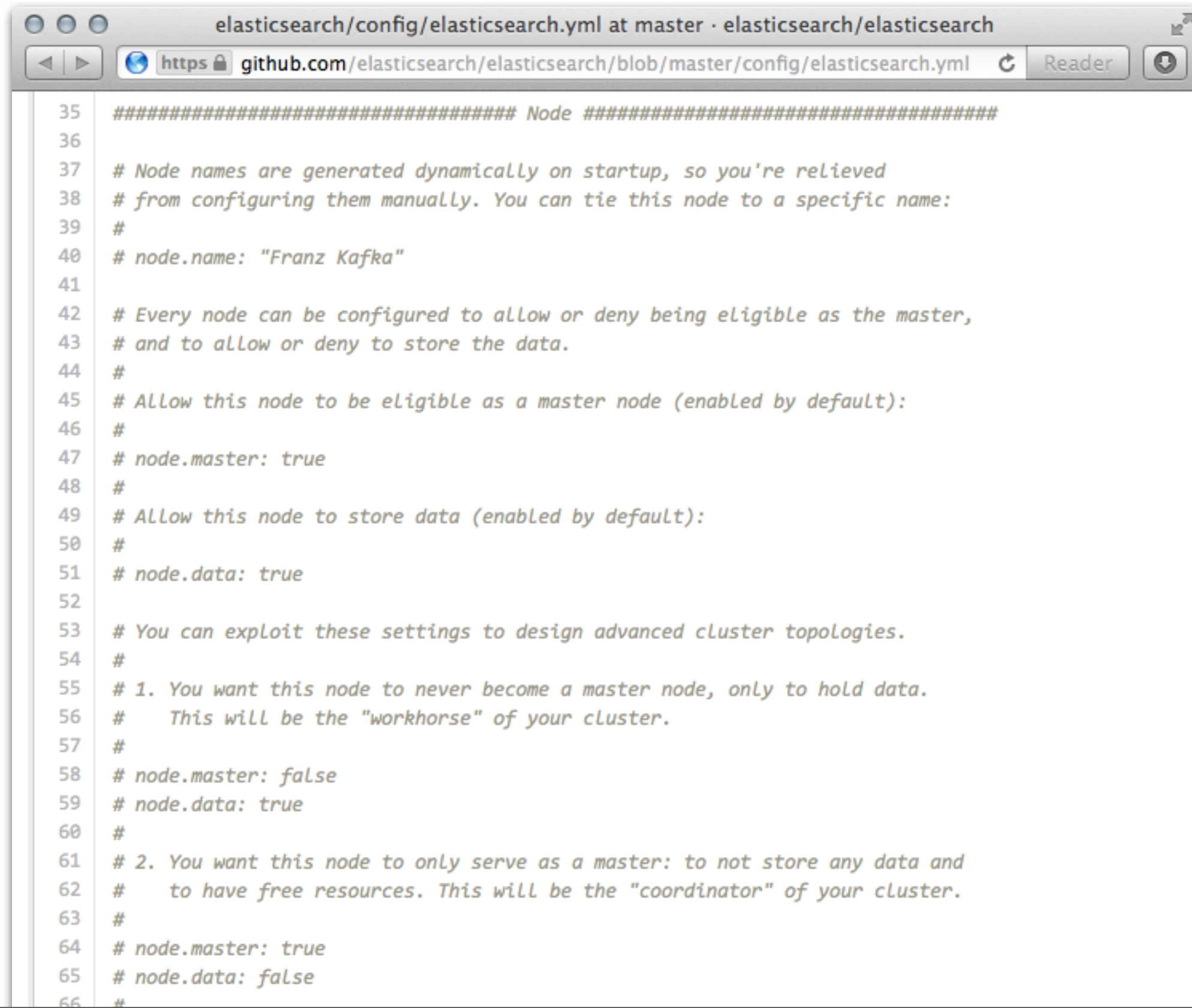
"index.routing.allocation.exclude.name" : "Node1"

"cluster.routing.allocation.exclude.name" : "Node3"



**KEEP
CALM
AND
DEFAULT
CONFIG**

Until you know what to tweak...

A screenshot of a web browser window displaying the Elasticsearch configuration file (elasticsearch.yml) on the GitHub repository. The browser's address bar shows the URL: https://github.com/elasticsearch/elasticsearch/blob/master/config/elasticsearch.yml. The page title is "elasticsearch/config/elasticsearch.yml at master · elasticsearch/elasticsearch". The configuration file content is shown in a monospaced font with line numbers on the left. The visible lines are 35 through 66. The configuration includes comments and settings for node names, master eligibility, and data storage.

```
35 ##### Node #####
36
37 # Node names are generated dynamically on startup, so you're relieved
38 # from configuring them manually. You can tie this node to a specific name:
39 #
40 # node.name: "Franz Kafka"
41
42 # Every node can be configured to allow or deny being eligible as the master,
43 # and to allow or deny to store the data.
44 #
45 # Allow this node to be eligible as a master node (enabled by default):
46 #
47 # node.master: true
48 #
49 # Allow this node to store data (enabled by default):
50 #
51 # node.data: true
52
53 # You can exploit these settings to design advanced cluster topologies.
54 #
55 # 1. You want this node to never become a master node, only to hold data.
56 #    This will be the "workhorse" of your cluster.
57 #
58 # node.master: false
59 # node.data: true
60 #
61 # 2. You want this node to only serve as a master: to not store any data and
62 #    to have free resources. This will be the "coordinator" of your cluster.
63 #
64 # node.master: true
65 # node.data: false
66 #
```

Search & Find

```
$ curl -XGET "http://localhost:9200/_search?q=<YOUR QUERY>"
```

Terms	apple apple iphone
Phrases	"apple iphone"
Proximity	"apple safari"~5
Fuzzy	apple~0.8
Wildcards	app* *pp*
Boosting	apple^10 safari
Range	[2011/05/01 TO 2011/05/31] [java TO json]
Boolean	apple AND NOT iphone +apple -iphone (apple OR iphone) AND NOT review
Fields	title:iphone^15 OR body:iphone published_on:[2011/05/01 TO "2011/05/27 10:00:00"]

http://lucene.apache.org/java/3_1_0/queryparsersyntax.html

JSON-based Query DSL

```
curl -XGET localhost:9200/articles/_search -d '{
  "query" : {
    "filtered" : {
      "query" : {
        "bool" : {

          "must" : {
            "match" : {
              "author.first_name" : {
                "query" : "claire",
                "fuzziness" : 0.1
              }
            }
          },

          "must" : {
            "multi_match" : {
              "query" : "elasticsearch",
              "fields" : ["title^10", "body"]
            }
          }
        }
      },

      "filter": {
        "and" : [
          { "terms" : { "tags" : ["search"] } },
          { "range" : { "published_on": { "from": "2013" } } },
          { "term" : { "featured" : true } }
        ]
      }
    }
  }
}'
```

JSON-based Query DSL

```
curl -XGET localhost:9200/_search?pretty
{
  "query" : {
    "filtered" : {
      "query" : {
        "bool" : {
          "must" : {
            "match" : {
              "author.first_name" : "John",
              "query" : "john",
              "fuzziness" : 0.1
            }
          }
        }
      },
      "must" : {
        "multi_match" : {
          "query" : "elasticsearch",
          "fields" : ["title^10", "body"]
        }
      }
    },
    "filter" : {
      "and" : [
        { "terms" : { "tags" : ["search"] } },
        { "range" : { "published_on" : { "from" : "2013" } } },
        { "term" : { "featured" : true } }
      ]
    }
  }
}
```



JSON-based Query DSL

```
curl -XGET localhost:9200/articles/_search -d '{
  "query" : {
    "filtered" : {
      "query" : {
        "bool" : {
          "must" : {
            "match" : {
              "author.first_name" : {
                "query" : "claire",
                "fuzziness" : 0.1
              }
            }
          },
          "must" : {
            "multi_match" : {
              "query" : "claire",
              "fields" : ["author.first_name", "author.last_name"]
            }
          }
        }
      },
      "filter" : {
        "and" : [
          { "terms" : { "tags" : ["search"] } },
          { "range" : { "published_on" : { "from" : "2013" } } },
          { "term" : { "featured" : true } }
        ]
      }
    }
  }
}'
```

```
"must" : {
  "match" : {
    "author.first_name" : {
      "query" : "claire",
      "fuzziness" : 0.1
    }
  },
  "must" : {
    "multi_match" : {
      "query" : "claire",
      "fields" : ["author.first_name", "author.last_name"]
    }
  }
}
```

JSON-based Query DSL

```
curl -XGET localhost:9200/articles/_search -d '{
  "query" : {
    "filtered" : {
      "query" : {
        "bool" : {

          "must" : {
            "match" : {
              "author.first_name" : {
                "query" : "claire",
                "fuzziness" : 0.1
              }
            }
          },

          "must" : {
            "multi_match" : {
              "query" : "elas
              "fields" : ["tit

            }
          }
        }
      },

      "filter": {
        "and" : [
          { "terms" : { "tags"
          { "range" : { "publisk
          { "term" : { "featured" : true } }
        ]
      }
    }
  }
}'
```

```
},
"must" : {
  "multi_match" : {
    "query" : "elasticsearch",
    "fields" : ["title^10", "body"]
  }
}
```

JSON-based Query DSL

```
curl -XGET localhost:9200/articles/_search -d '{
  "query" : {
    "filtered" : {
      "query" : {
        "bool" : {

          "must" : {
            "match" : {
              "author.first_name" : {
                "query" : "claire",
                "fuzziness" : 0.1
              }
            }
          },

          "must" : {
            "multi_match" : {
              "query" : "elasticsearch",
              "fields" : ["title^10", "body"]
            }
          }
        }
      },

      "filter" : {
        "and" : [
          { "terms" : { "tags" : ["search"] } },
          { "range" : { "published_on" : { "from" : "2013" } } },
          { "term" : { "featured" : true } }
        ]
      }
    }
  }
}
```

```
    "filter": {
      "and" : [
        { "terms" : { "tags" : ["search"] } },
        { "range" : { "published_on": { "from": "2013" } } },
        { "term" : { "featured" : true } }
      ]
    }
  }
}
```

Full-text Search

"Find all articles with 'search' in their title or body, give matches in titles higher score"

Structured Search

"Find all articles from year 2013 tagged 'search'"

Combined Search

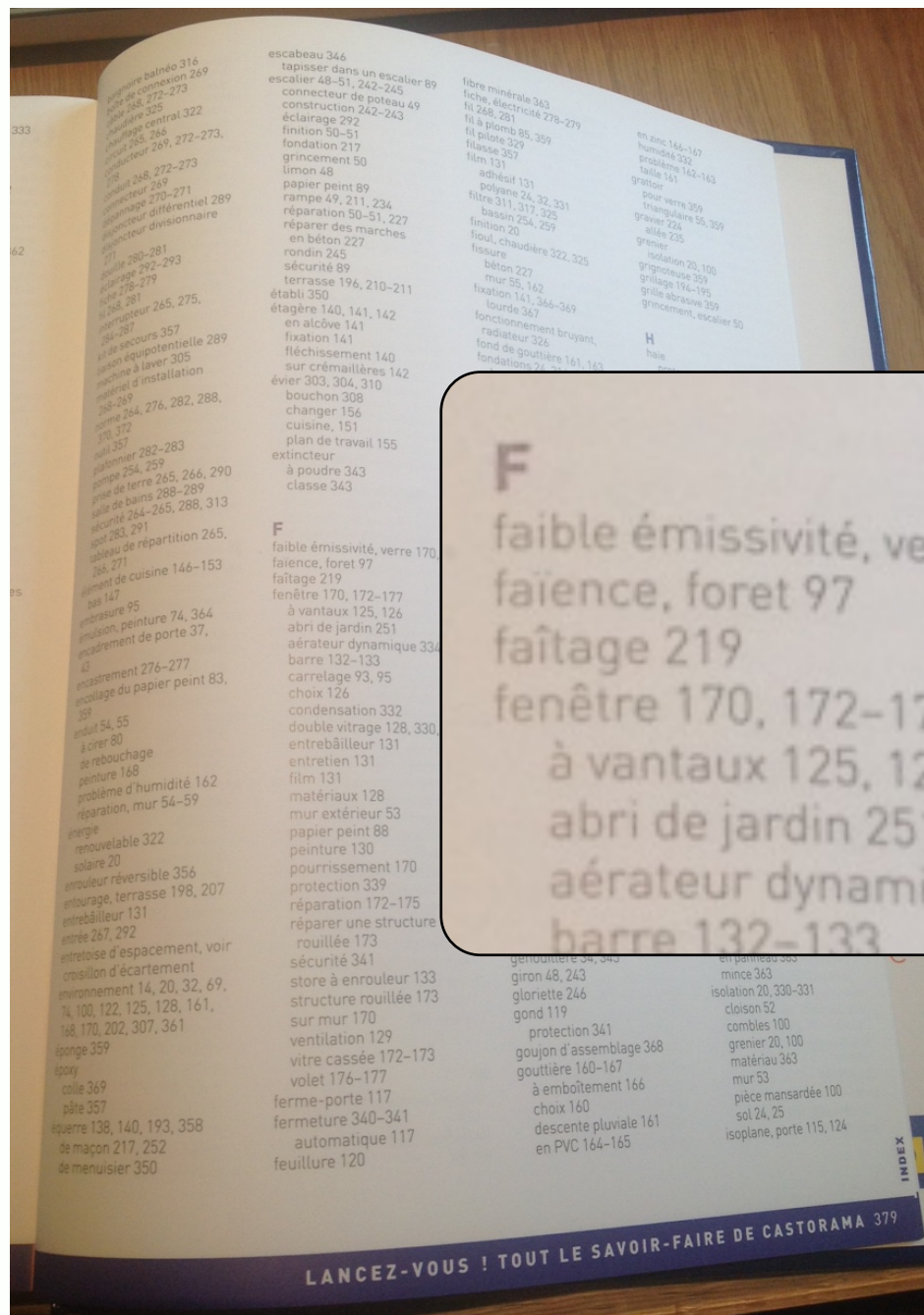
"Find all articles with 'search' in their title or body, give matches in titles higher score and filter articles from year 2013 tagged 'search' "

Custom Scoring

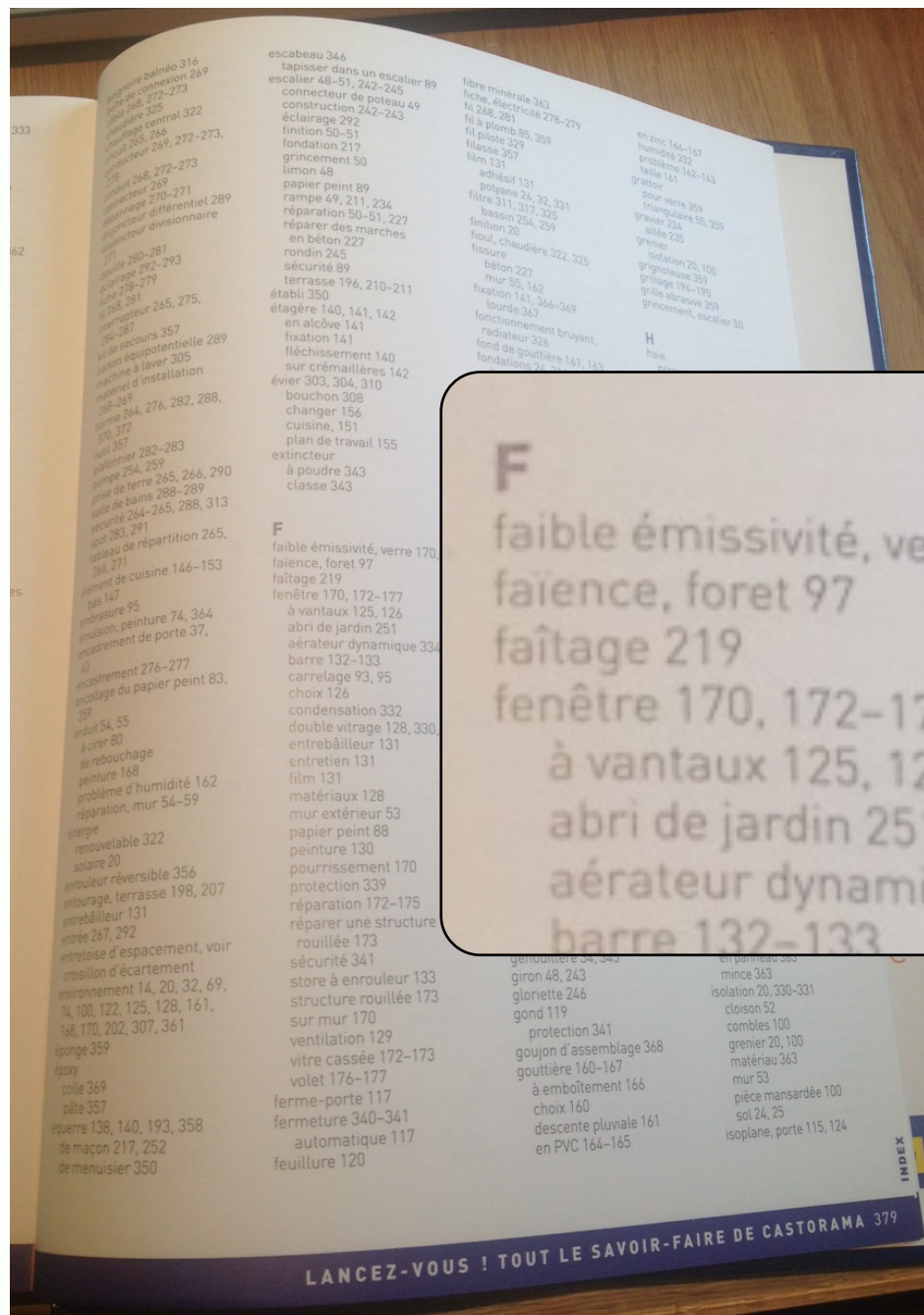
See custom_score and custom_filters_score queries

What is really a search engine?

What is really a search engine?



What is really a search engine?



Fetch document field → Pick configured analyzer → Parse text into tokens → Apply token filters → Store into index

The `_analyze` API

`_analyze?text=...&tokenizer=X&filters=A,B,C`

The `_analyze` API

`_analyze?text=...&tokenizer=X&filters=A,B,C`

```
curl 'localhost:9200/_analyze?pretty&tokenizer=keyword&filters=lowercase' \  
-d 'This is a Test'
```

The `_analyze` API

`_analyze?text=...&tokenizer=X&filters=A,B,C`

```
curl 'localhost:9200/_analyze?pretty&tokenizer=keyword&filters=lowercase' \
-d 'This is a Test'
```

```
{
  "tokens" : [ {
    "token" : "this is a test",
    "start_offset" : 0, "end_offset" : 14,
    "type" : "word", "position" : 1
  } ]
}
```

The `_analyze` API

`_analyze?text=...&tokenizer=X&filters=A,B,C`

```
curl 'localhost:9200/_analyze?pretty&tokenizer=keyword&filters=lowercase' \
-d 'This is a Test'
```

```
{
  "tokens" : [ {
    "token" : "this is a test",
    "start_offset" : 0, "end_offset" : 14,
    "type" : "word", "position" : 1
  } ]
}
```

```
curl 'localhost:9200/_analyze?pretty&analyzer=standard' -d 'This is a Test'
```

The _analyze API

`_analyze?text=...&tokenizer=X&filters=A,B,C`

```
curl 'localhost:9200/_analyze?pretty&tokenizer=keyword&filters=lowercase' \
-d 'This is a Test'
```

```
{
  "tokens" : [ {
    "token" : "this is a test",
    "start_offset" : 0, "end_offset" : 14,
    "type" : "word", "position" : 1
  } ]
}
```

```
curl 'localhost:9200/_analyze?pretty&analyzer=standard' -d 'This is a Test'
```

```
{
  "tokens" : [ {
    "token" : "test",
    "start_offset" : 10, "end_offset" : 14,
    "type" : "<ALPHANUM>", "position" : 4
  } ]
}
```

Mapping

Configuring document properties for the search engine

```
curl -XPUT localhost:9200/articles/_mapping -d '{
  "article" : {
    "properties" : {
      "title" : {
        "type"      : "string",
        "analyzer"  : "french"
      }
    }
  }
}
```

Slice & Dice



Search

☐ Keep refine selections**Search**[Show more...](#)Sort by: [Relationship](#)View: [Basic](#)33,778 results for **linkedin** [\[Save this search\]](#)Refine By BETA

Location

- ☒ All Locations
- ☐ United States (20762)
- ☐ San Francisco Bay Area (2444)
- ☐ India (2303)
- ☐ United Kingdom (2101)
- ☐ Greater New York City Area (2079)

[Show more...](#)

Relationship

- ☒ All LinkedIn Members
- ☐ 1st Connections (192)
- ☐ 2nd Connections (2706)
- ☐ Group Members (860)
- ☐ 3rd + Everyone Else (30390)

Industry

- ☒ All Industries
- ☐ Information Technology and Services (4219)
- ☐ Marketing and Advertising (2709)
- ☐ Staffing and Recruiting (2301)
- ☐ Computer Software (1947)
- ☐ Internet (1782)

[Show more...](#)Current Company +Past Company +School +Groups +**Reid Hoffman** 1st

Partner at Greylock

San Francisco Bay Area | Internet

In Common: [120 shared connections](#) [2 shared groups](#)**David Hahn** 1st

Director of Product Management, LinkedIn

San Francisco Bay Area | Internet

In Common: [83 shared connections](#)**Candy Chastain Mielke** 1st

Consultant at LinkedIn

Colorado Springs, Colorado Area | Internet

In Common: [128 shared connections](#) [2 shared groups](#)**Wade Burgess** 1st

Business Leader

Greater Omaha Area | Internet

In Common: [44 shared connections](#) [2 shared groups](#)**Alex Vauthey** 1st

Director of Engineering, Monetization at LinkedIn

San Francisco Bay Area | Information Technology and Services

In Common: [103 shared connections](#)**Ruslan Belkin** 1st

Sr. Director of Engineering at LinkedIn

San Francisco Bay Area | Computer Software

In Common: [95 shared connections](#)**Elizabeth Reaves** 1st

Product Manager at LinkedIn

San Francisco Bay Area | Internet

In Common: [92 shared connections](#) [4 shared groups](#)**Riccardo Ferretti** 1st

Principal Engineer at LinkedIn

San Francisco Bay Area | Computer Software

In Common: [121 shared connections](#)**Shirley Xu** 1st

Data Scientist at LinkedIn

Search tools

[Advanced search](#)[Saved Searches](#)

Ads by LinkedIn Members

[S&OP Summit](#)

group

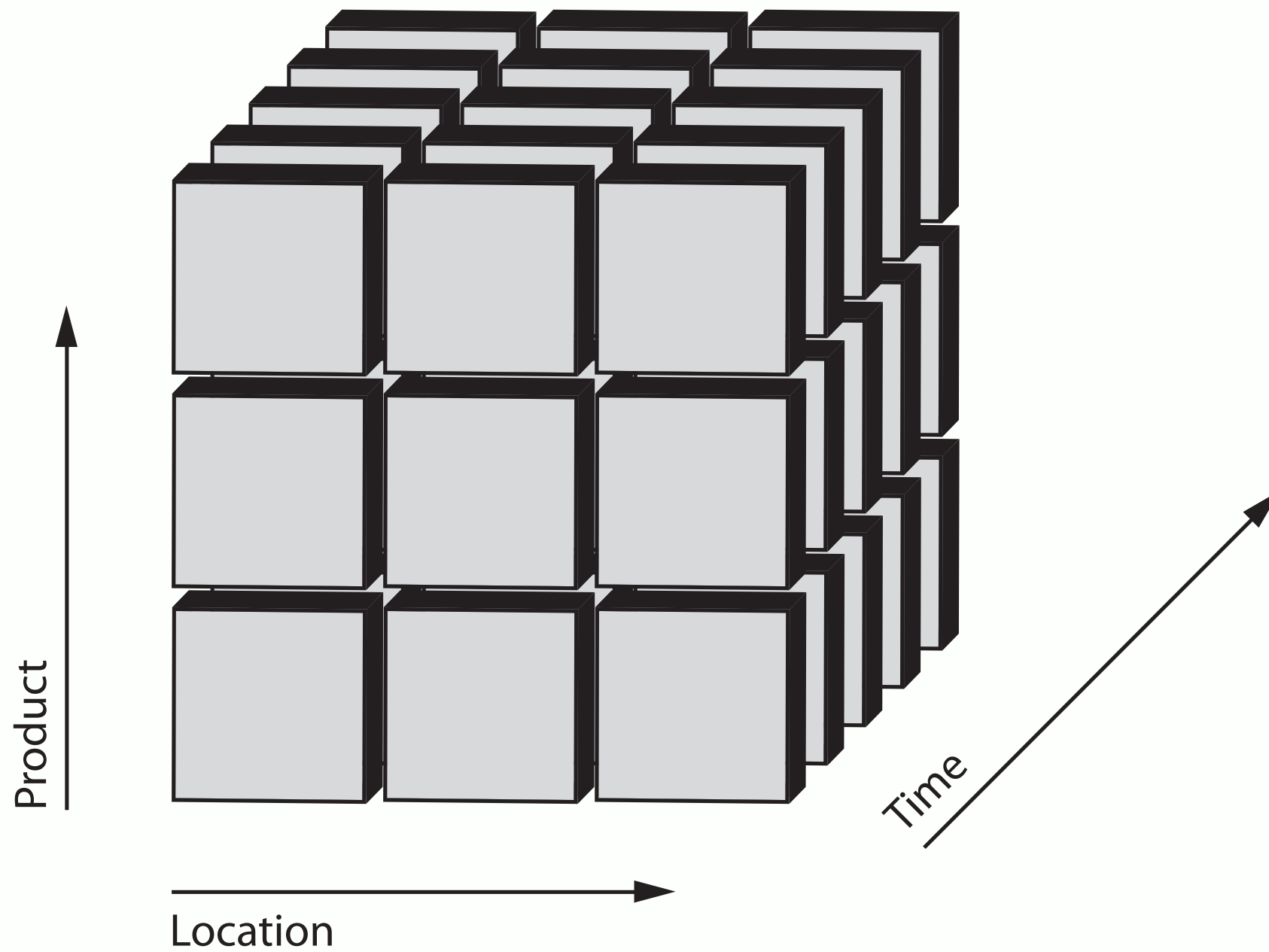
January 20/21 2010,
The Palms Hotel Sales
& Operations Planning
Innovat[events.linkedin.com](#)From: [Richard Bracchi](#)[Channel Not
Performing?](#)

relayware

On demand solutions
for channel
performance
optimization.[www.relayware.com](#)From: [Mike Morgan](#)[What's this?](#)

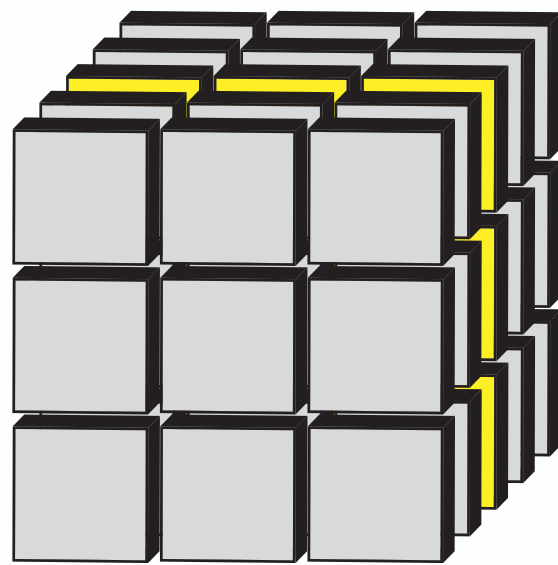
Query

Facets

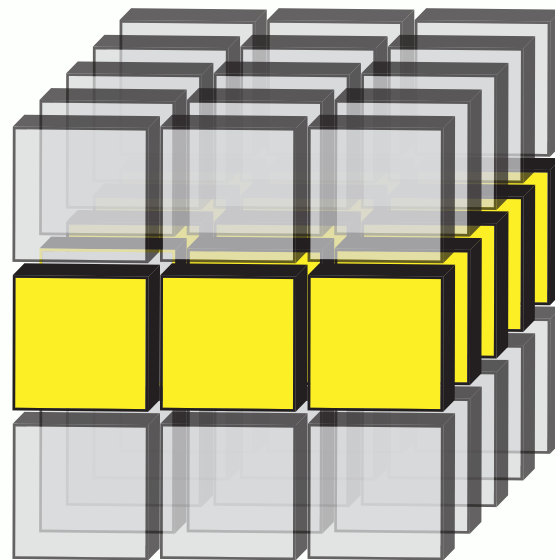


OLAP Cube

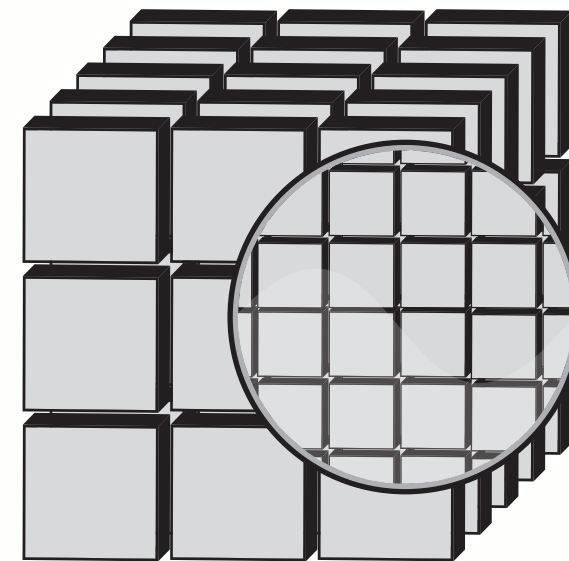
Dimensions, measures, aggregations



Slice



Dice



Drill Down / Roll Up

Show me sales numbers for all products across all locations in year 2013

Show me product A sales numbers across all locations over all years

Show me products sales numbers in location X over all years

"Tag Cloud" With the terms Facet

Simplest "map/reduce" aggregation: document count per tag

```
curl -XPOST 'localhost:9200/articles/_search?search_type=count&pretty' -d '{
  "facets": {
    "tag-cloug": {
      "terms": {
        "field": "tags"
      }
    }
  }
}'
```

```
"facets": {
  "tag-cloug": {
    "terms": [ {
      "term": "ruby",
      "count": 3
    }, {
      "term": "java",
      "count": 2
    },
    ...
  ]
}
```

Statistics on Student Scores With the terms_stats Facet

Aggregating statistics per subject

```
curl -XGET 'localhost:9200/scores/_search/?search_type=count&pretty' -d '{
  "facets": {
    "scores-per-subject" : {
      "terms_stats" : {
        "key_field" : "subject",
        "value_field" : "score"
      }
    }
  }
}'
```

```
"facets": {
  "scores-per-subject": {
    "_type": "terms_stats",
    "missing": 0,
    "terms": [ {
      "term": "math",
      "count": 4,
      "total_count": 4,
      "min": 25.0,
      "max": 92.0,
      "total": 267.0,
      "mean": 66.75
    }, ... ]
  }
}
```

Statistics on Student Scores With the terms_stats Facet

Realtime filtering with queries and filters

```
curl -X GET 'localhost:9200/demo-scores/_search/?search_type=count&pretty'  
{  
  "query" : {  
    "match" : {  
      "student" : "john"  
    }  
  },  
  "facets" : {  
    "scores-per-subject" : {  
      "terms_stats" : {  
        "key_field" : "subject",  
        "value_field" : "score"  
      }  
    }  
  }  
}
```

```
"facets" : {  
  "scores-per-subject" : {  
    "_type" : "terms_stats",  
    "missing" : 0,  
    "terms" : [ {  
      "term" : "math",  
      "count" : 1,  
      "total_count" : 1,  
      "min" : 85.0,  
      "max" : 85.0,  
      "total" : 85.0,  
      "mean" : 85.0  
    }, ... ]  
  }  
}
```

Facets

Terms

Terms Stats

Statistical

Range

Histogram

Date Histogram

Filter

Query

Geo Distance

Above & Beyond

Above & Beyond

Bulk operations *(For indexing and search operations)*

Percolator *("reversed search" — alerts, classification, ...)*

Suggesters *("Did you mean ...?")*

Index aliases *(Grouping, filtering or "renaming" of indices)*

Index templates *(Automatic index configuration)*

Monitoring API *(Amount of memory used, number of operations, ...)*

...

Above & Beyond

Bulk operations *(For indexing and search operations)*

Percolator *("reversed search" — alerts, classification, ...)*

Suggesters *("Did you mean ...?")*

Index aliases *(Grouping, filtering or "renaming" of indices)*

Index templates *(Automatic index configuration)*

Monitoring API *(Amount of memory used, number of operations, ...)*

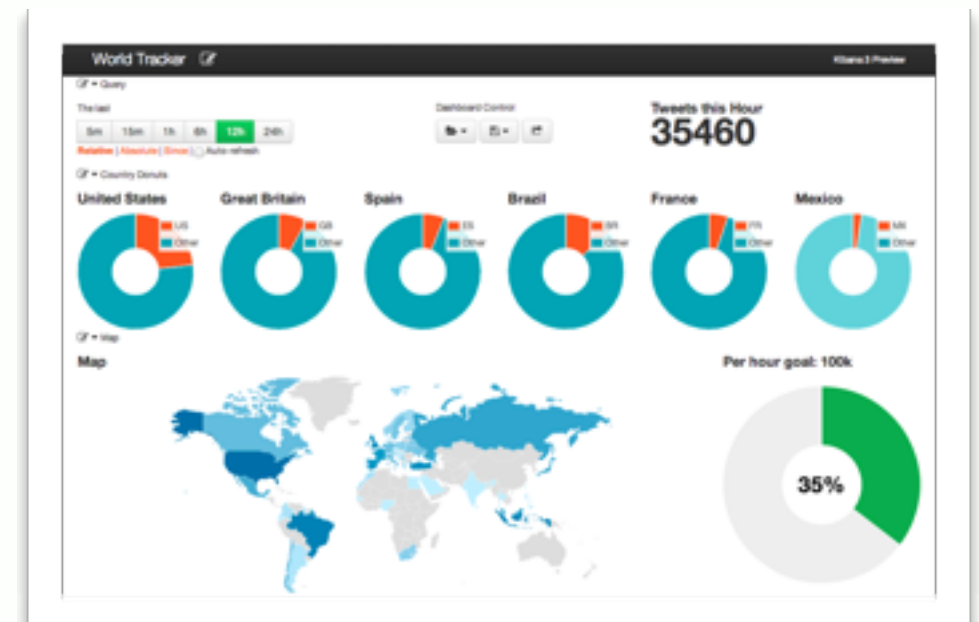
...

GUI?



Give Kibana a try

<http://three.kibana.org/>





thanks!