Elasticsearch

Search made easy

Alexander Reelsen
@spinscale
alexander.reelsen@elasticsearch.com

Copyright Elasticsearch 2013. Copying, publishing and/or distributing without written permission is strictly prohibited

elasticsearch.

Agenda

- Why is search complex?
- Replication & Sharding
- Installation & initial setup
- Importing data
- Searching data
- Plugin-based architecture
- Clients & integrations
- Roadmap

Copyright Elasticsearch 2013. Copying, publishing and/or distributing without written permission is strictly prohibited

Elasticsearch - The company

- Founded in 2012

  By the people behind the Elasticsearch project

  http://www.elasticsearch.com

- Professional services

  Training (public & onsite)

  Consultancy (development support)

  Production support subscription

  - targeting production
  - 3 levels of SLAs
  - differing in response times and availability

Copyright Elasticsearch 2013. Copying, publishing and/or distributing without written permission is strictly prohibited

Search is hard

- Functional requirements

  Find the right data (effectivity/relevance)
- Non-functional requirements

  Find the data right (efficiency/speed)

- Speed is useless without relevance
- Biggest problem: Search is highly subjective

Copyright Elasticsearch 2013. Copying, publishing and/or distributing without written permission is strictly prohibited

Search - by term

amazon Join Prime | Your Amazon.com | Today's Deals | Gift Cards | Sell | Help | kindle From $69

Shop by Department | Search | All | lucene | Go | Hello. Sign in Your Account | Join Prime | 0 Cart | Wish List

Departments
Books
- Computers & Technology
- Web Development
- Java Programming
- Data Mining
- Online Internet Searching
- + See more...

Kindle Store
- Computers & Technology

+ See All 2 Departments

International Shipping (What's this?)
- AmazonGlobal Eligible

Shipping Option (What's this?)
- Free Super Saver Shipping

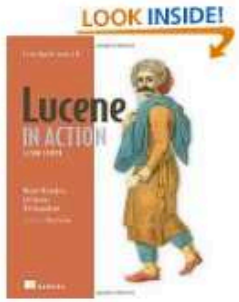
Book Format
- Paperback
- Kindle Edition

Avg. Customer Review
- ★★★★☆ & Up
- ★★★☆☆ & Up
- ★★☆☆☆ & Up
- ★☆☆☆☆ & Up

"lucene"

Related Searches: lucene 4, lucene net, solr.

Showing 1 - 18 of 147 Results | Choose a Department to enable sorting



Lucene in Action, Second Edition: Covers Apache Lucene 3.0 by Michael McCandless, Erik Hatcher and Otis Gospodnetic (Jul 28, 2010)

~~$49.99~~ **$31.49** Paperback

Order in the next **11 hours** and get it by **Friday, Mar 15.**

Only 12 left in stock - order soon.

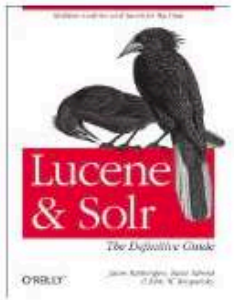
More Buying Choices - Paperback
$26.10 new (56 offers)
$25.99 used (26 offers)

★★★★½ (30)

Eligible for **FREE** Super Saver Shipping.

Sell this back for an Amazon.com Gift Card

Books: See all 142 items



Lucene and Solr: The Definitive Guide: The comprehensive guide to Lucene and Solr for realtime Big Data by Jason Rutherglen, Ryan Tabora and Jack Krupansky (Jul 22, 2013)

~~$49.99~~ **$34.60** Paperback

Available for Pre-order. This item will be released on July 22, 2013.

Eligible for **FREE** Super Saver Shipping and 1 more promotion

Books: See all 142 items

Copyright Elasticsearch 2013. Copying, publishing and/or distributing without written permission is strictly prohibited

elasticsearch.

Search - by ID

amazon Join Prime | Your Amazon.com | Today's Deals | Gift Cards | Sell | Help | kindle

Shop by Department | Search | All | 9781849515528 | Go | Hello. Sign in Your Account | Join Prime

Department
Books
Kindle Store

International Shipping (What's this?)
AmazonGlobal Eligible

Shipping Option (What's this?)
Free Super Saver Shipping

Avg. Customer Review
★★★★☆ & Up
★★★☆☆ & Up
★★☆☆☆ & Up
★☆☆☆☆ & Up

Price
Under $25
$25 to $50
$ [ ] to $ [ ] GO

Discount
10% Off or More

Availability
Include Out of Stock

"9781849515528"

Showing 2 Results

Play Framework Cookbook by Alexander Reelsen (Aug 8, 2011)

~~$44.99~~ **$39.99** Paperback

Order in the next **7 hours** and get it by **Friday, Mar 15.**

More Buying Choices - Paperback
$35.99 new (18 offers)
$73.08 used (6 offers)

★★★★☆ (14)

Eligible for **FREE** Super Saver Shipping.

Sell this back for an Amazon.com Gift Card

Books: See all items

Play framework Cookbook by Alexander Reelsen (Aug 8, 2011)

$19.67 Kindle Edition

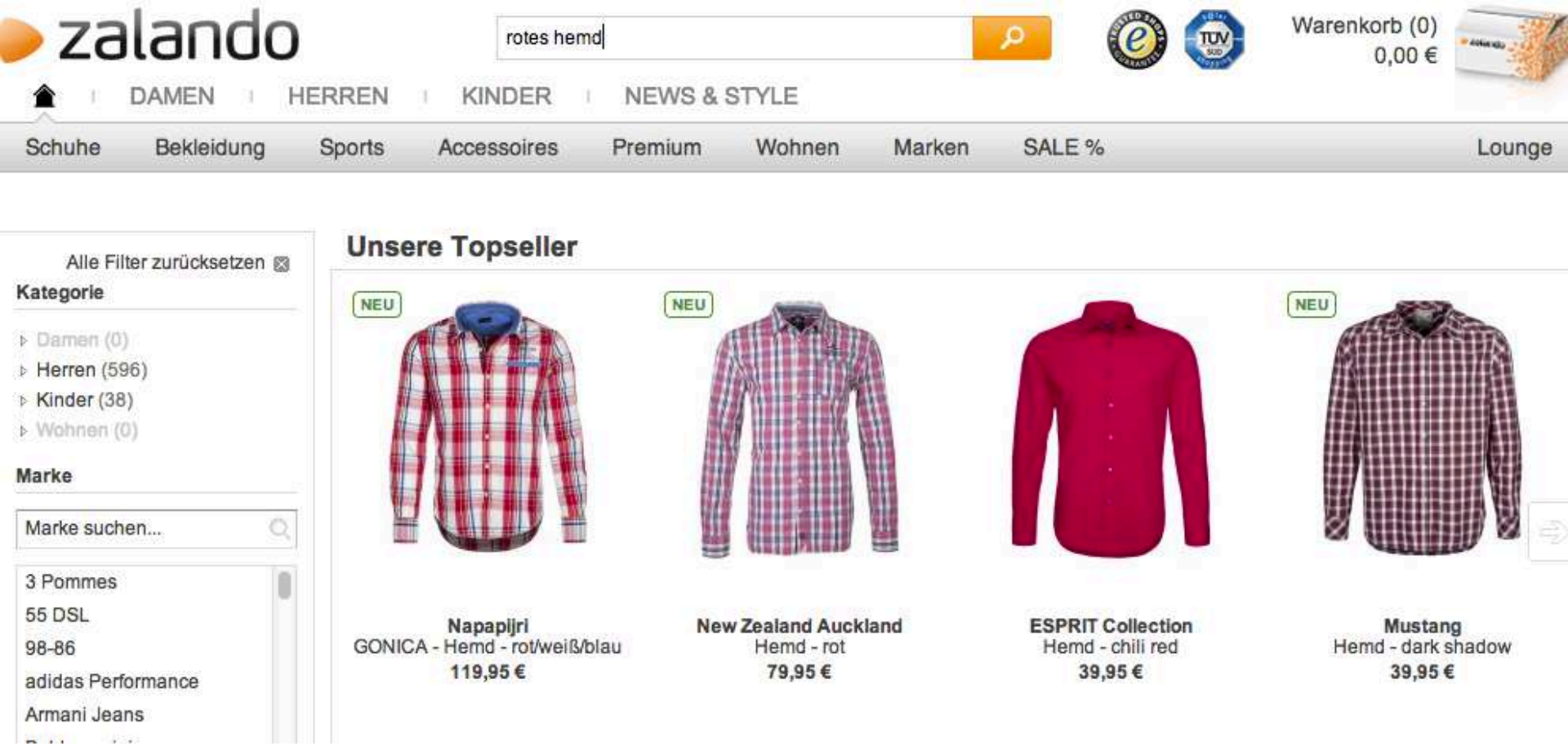
★★★★☆ (14)

Kindle Store: See all items

Copyright Elasticsearch 2013. Copying, publishing and/or distributing without written permission is strictly prohibited

elasticsearch.

Search - by attribute

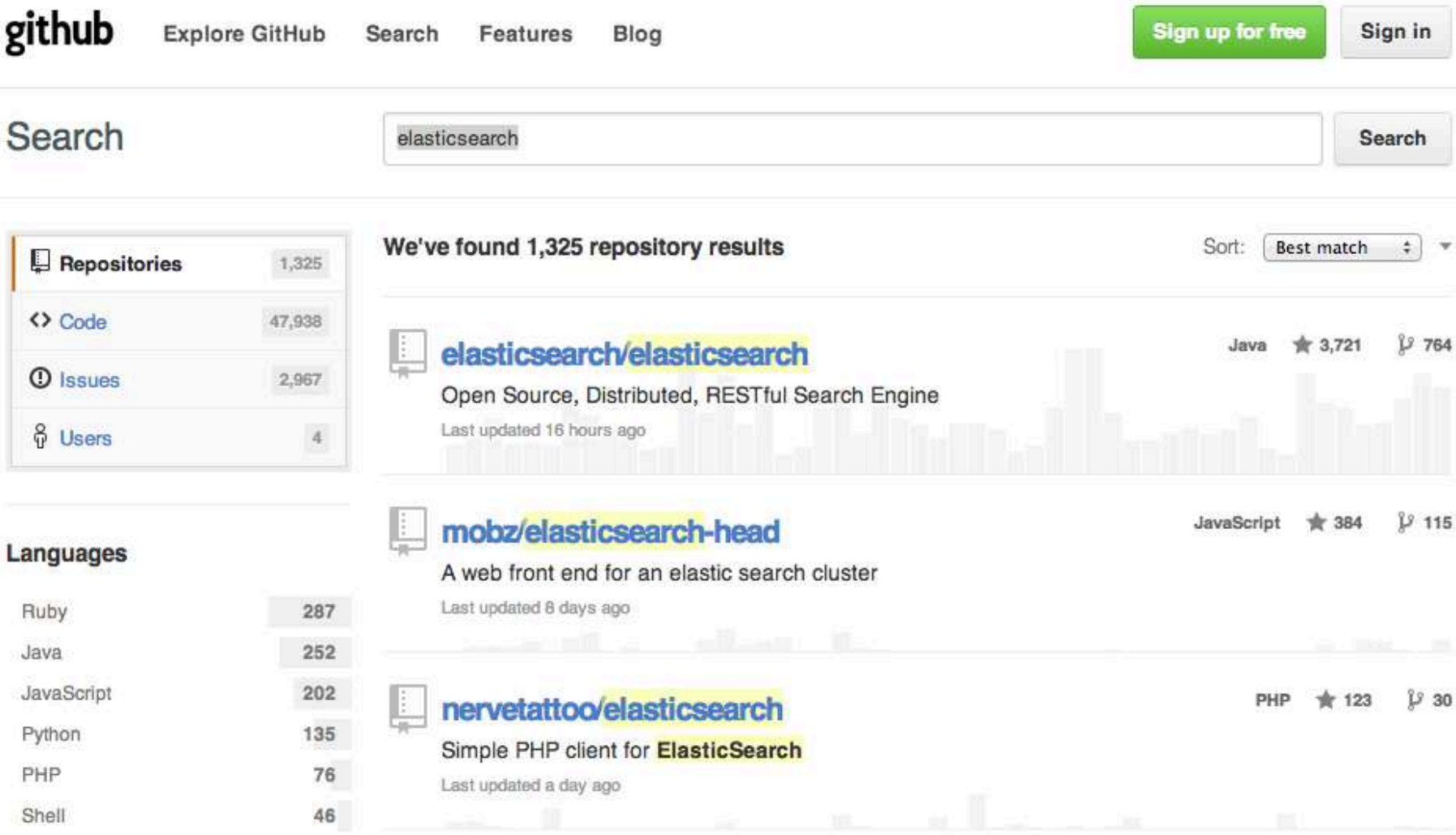


Copyright Elasticsearch 2013. Copying, publishing and/or distributing without written permission is strictly prohibited

Search - Suggestions

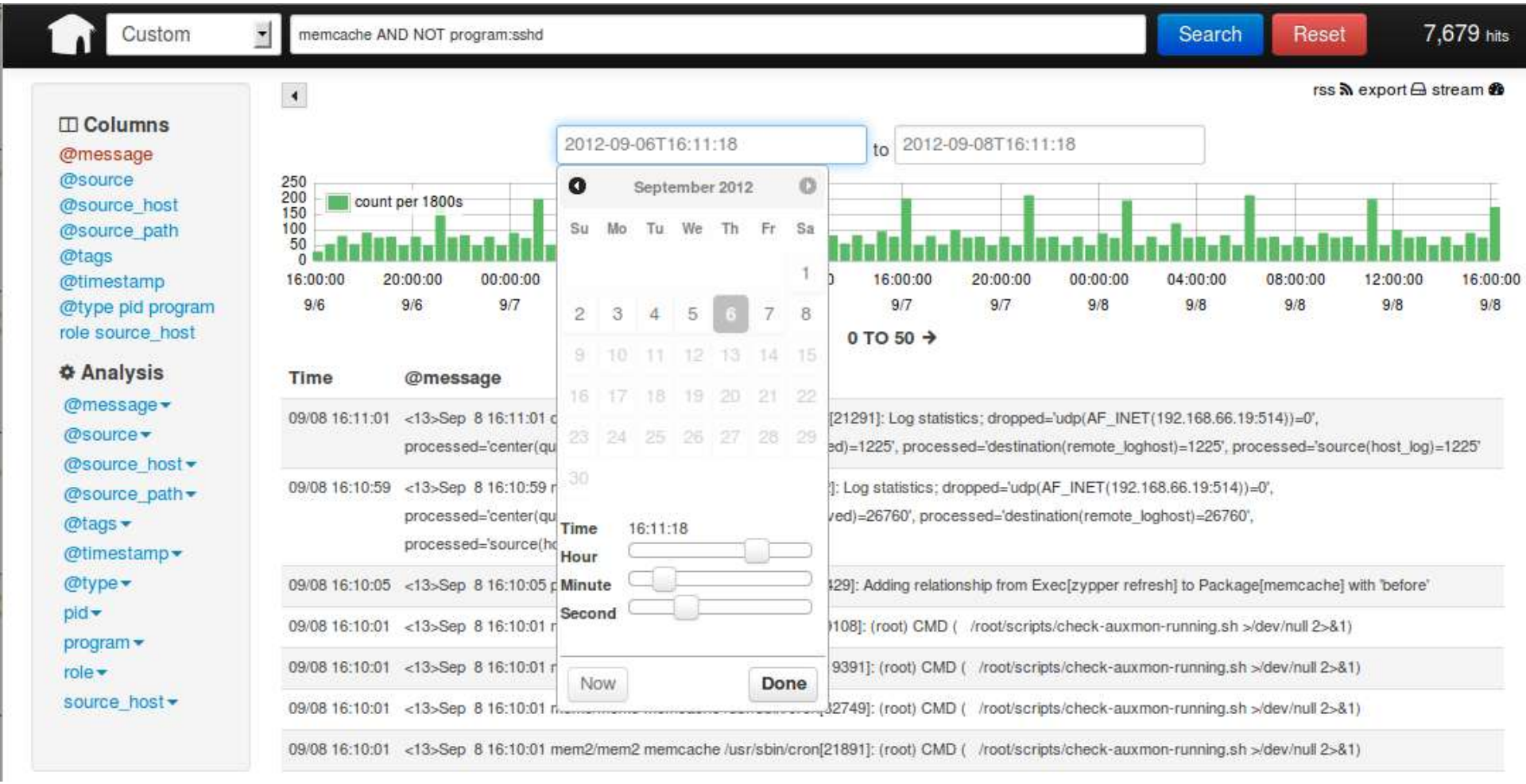
Copyright Elasticsearch 2013. Copying, publishing and/or distributing without written permission is strictly prohibited

Search - Highlighting



Copyright Elasticsearch 2013. Copying, publishing and/or distributing without written permission is strictly prohibited

Search - Analytics



Copyright Elasticsearch 2013. Copying, publishing and/or distributing without written permission is strictly prohibited

elasticsearch.

What is elasticsearch?

- Schema-free, REST & JSON based, distributed document store
- Apache License 2.0
- Language specific drivers
- Zero configuration

- Used by github, soundcloud, stackoverflow, mozilla, klout

Copyright Elasticsearch 2013. Copying, publishing and/or distributing without written permission is strictly prohibited

Installation & setup

Copyright Elasticsearch 2013. Copying, publishing and/or distributing without written permission is strictly prohibited

Zero configuration

```
# wget --no-check-certificate https://download.elasticsearch.org/elasticsearch/
elasticsearch/elasticsearch-0.90.1.zip

# unzip elasticsearch-0.90.1.zip

# cd elasticsearch-0.90.1

# bin/elasticsearch -f
```

```
# curl -X PUT http://localhost:9200/products/product/1 -d '{ "name" : "high
quality search engine" }'

{”ok”:true,”_index”:”products”,”_type”:”product”,”_id”:”1”,”_version”:1}
```

```
# curl -X POST 'http://localhost:9200/products/product/_search?pretty' -d
'{ "query" : { "match" : { "name" : " search"} } }'
```

Copyright Elasticsearch 2013. Copying, publishing and/or distributing without written permission is strictly prohibited

Configuration

- `config/elasticsearch.yml` or `config/elasticsearch.json`
- instance-wide settings (zen discovery, network setup, available analyzers)
- Index default configurations (number of shards, number of replicas)
- Seperate logging configuration (simplified log4j): `config/logging.yml`

Copyright Elasticsearch 2013. Copying, publishing and/or distributing without written permission is strictly prohibited

elasticsearch.yml

```
discovery.zen.multicast.enabled: false

http:
  max_content_length: 100000

index:
  number_of_shards: 1

  analysis:
    analyzer:
      default:
        type: standard

      lowercase_analyzer:
        type: custom
        tokenizer: standard
        filter: [standard, lowercase]
```

Copyright Elasticsearch 2013. Copying, publishing and/or distributing without written permission is strictly prohibited

Sharding & Replication

Copyright Elasticsearch 2013. Copying, publishing and/or distributing without written permission is strictly prohibited

Sharding & Replication

- Replication: Share same data over several machines
  - Increasing throughput due to concurrency
  - Allow outage of nodes without dataloss
- Sharding: Index partitioning
  - Split logical data into physically smaller parts
  - Control data flows

Copyright Elasticsearch 2013. Copying, publishing and/or distributing without written permission is strictly prohibited

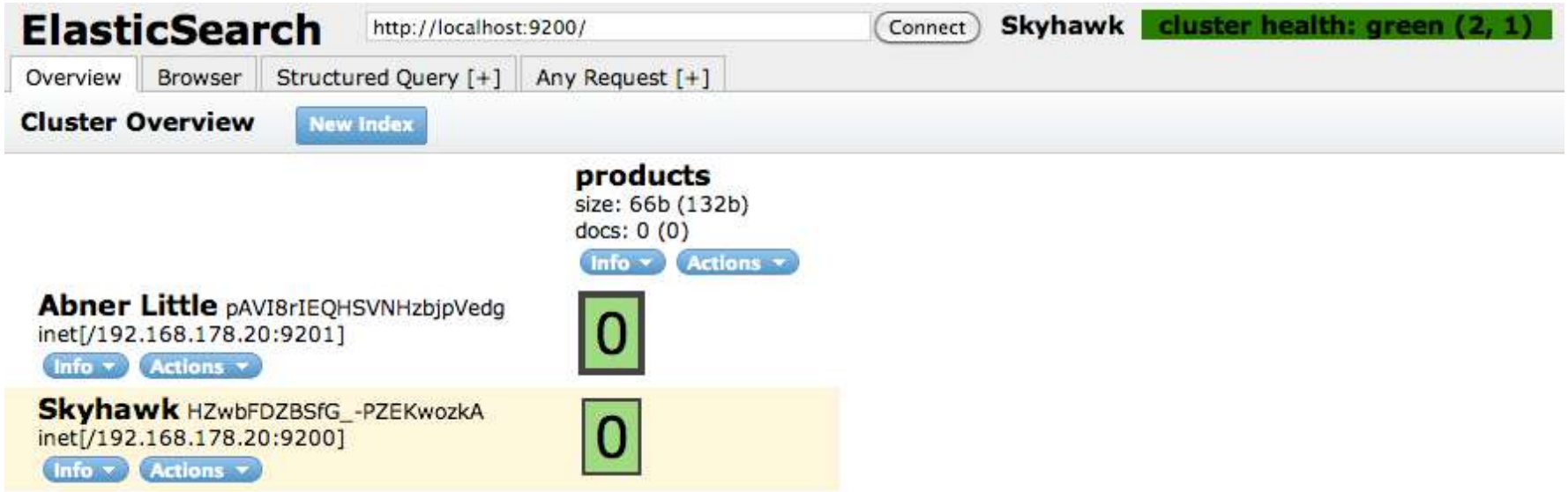
Sharding

```
curl -X PUT http://localhost:9200/products -d '{
  "settings" : {
    "index" : {
      "number_of_shards" : "5",
      "number_of_replicas" : "0"
    }
  }
}'
```

Copyright Elasticsearch 2013. Copying, publishing and/or distributing without written permission is strictly prohibited

Replication

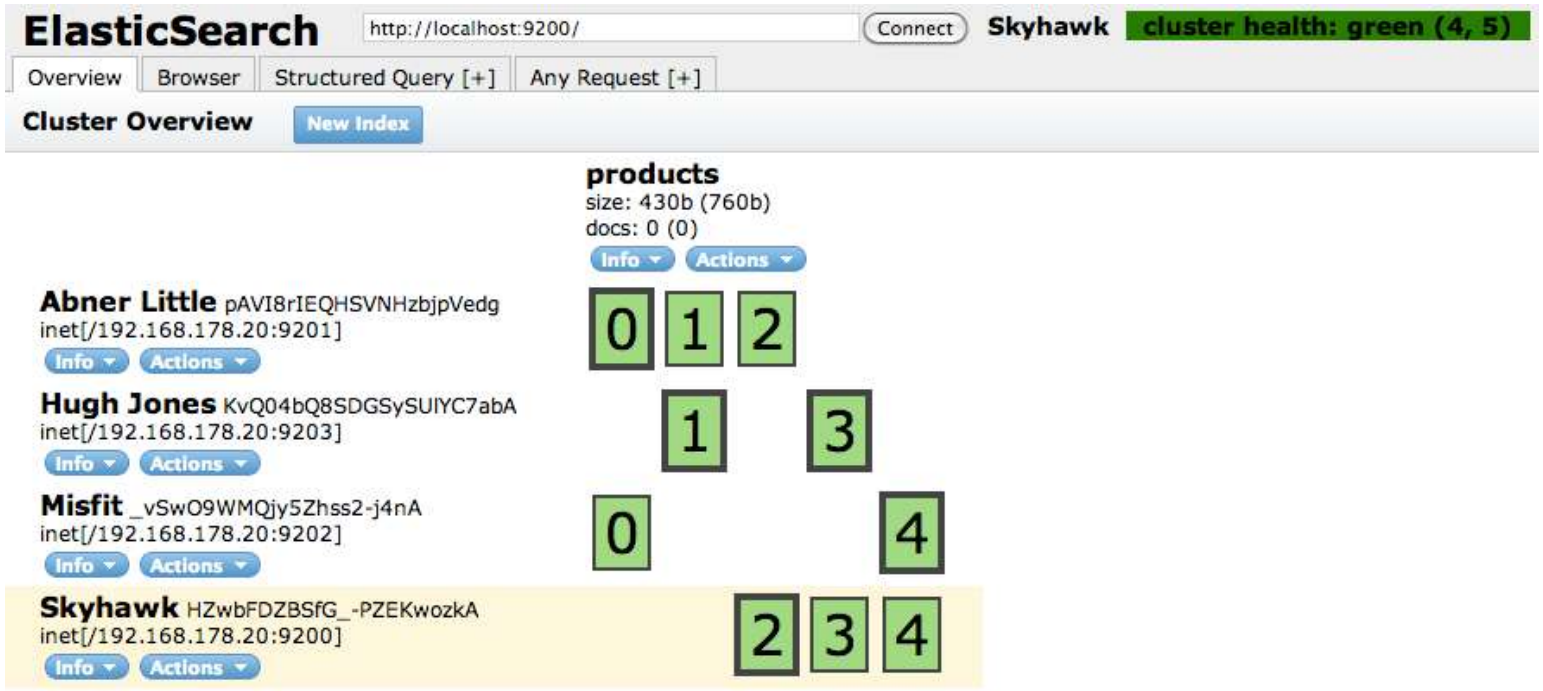
```
curl -X PUT http://localhost:9200/products -d '{
  "settings" : {
    "index" : {
      "number_of_shards" : "1",
      "number_of_replicas" : "1"
    }
  }
}'
```



Copyright Elasticsearch 2013. Copying, publishing and/or distributing without written permission is strictly prohibited

Sharding & Replication

```
curl -X PUT http://localhost:9200/products -d '{
  "settings" : {
    "index" : {
      "number_of_shards" : "5",
      "number_of_replicas" : "1"
} } }'
```

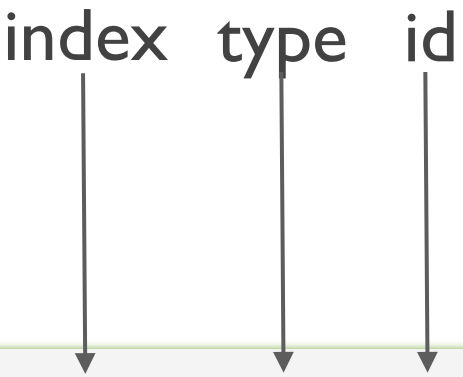


Copyright Elasticsearch 2013. Copying, publishing and/or distributing without written permission is strictly prohibited

Indexing

Copyright Elasticsearch 2013. Copying, publishing and/or distributing without written permission is strictly prohibited

Importing data




```
# curl -X PUT 'http://localhost:9200/articles/article/1' -d '{
   "title"   : "My first article",
   "content" : "... some lengthy article ...",
   "tags"    : [ "news", "sports", "introduction" ],
   "created" : "2013/04/04 16:54:23",
   "viewed"  : 234,
   "cost"    : 0.99
}'
```

Copyright Elasticsearch 2013. Copying, publishing and/or distributing without written permission is strictly prohibited

Mapping

- Matching fields with data types
- Inferred if not configured (dangerous!)
- Types: `float`, `long`, `boolean`, `date` (+formatting), `object`, `nested`
- `String` type can have arbitrary analyzers
- Fields can be split up in more fields (multi field)

Copyright Elasticsearch 2013. Copying, publishing and/or distributing without written permission is strictly prohibited

Sample mapping

```
# curl 'localhost:9200/articles/article/_mapping?pretty=1'
{
  "article" : {
    "properties" : {
      "content" : { "type" : "string" },
      "title"   : { "type" : "string" },
      "tags"    : { "type" : "string" },
      "viewed"  : { "type" : "long"   },
      "cost"    : { "type" : "double" },
      "created" : {
        "type" : "date", "format" : "yyyy/MM/dd HH:mm:ss||yyyy/MM/dd"
      }
    }
  }
}
```

Copyright Elasticsearch 2013. Copying, publishing and/or distributing without written permission is strictly prohibited

Analyzers

Waxolutionists - The smart blip experience, unopened, only 20$!!

Standard Analyzers	Analyzed Text
standard	waxolutionists smart blip experience unopened only 20
simple	waxolutionists the smart blip experience unopened only
whitespace	Waxolutionists - The smart blip experience, unopened, only 20$!!
stop	waxolutionists smart blip experience unopened only
keyword	Waxolutionists - The smart blip experience, unopened, only 20$!!
pattern	waxolutionists smart blip experience unopened only
snowball	waxolutionist smart blip experi unopen onli 20

Copyright Elasticsearch 2013. Copying, publishing and/or distributing without written permission is strictly prohibited

Searching

Copyright Elasticsearch 2013. Copying, publishing and/or distributing without written permission is strictly prohibited

Different ways of searching

- Search queries

  `match`, `term`, `prefix`, `id`, `fuzzy`
- Counting only, Geo-based queries
- More like this, Highlighting
- Faceting, Percolation, Scripting
- Suggestions

Copyright Elasticsearch 2013. Copying, publishing and/or distributing without written permission is strictly prohibited

Searching

```
curl -X POST http://localhost:9200/articles/article/_search?
pretty=1 -d '
{
  "from" : 0,
  "size" : 10,
  "query" : {
    "match" : {
      "title" : "first"
    }
  }
}'
```

Copyright Elasticsearch 2013. Copying, publishing and/or distributing without written permission is strictly prohibited

Searching


```
{
  "took": 2,
  "timed_out": false,
  "_shards": { "total": 15, "successful": 15, "failed": 0 },
  "hits": {
    "total": 1,
    "max_score": 0.15342641,
    "hits": [
      {
        "_index": "articles", "_type": "article", "_id": "1",
        "_score": 0.15342641,
        "_source": {
          "title": "My first article",
          "content": "... some lengthy article ...",
          "tags": [ "news", "sports", "introduction" ],
          "created": "2013/04/04 16:54:23",
          "viewed"  : 234,
          "cost"    : 0.99
        }
      }
    ]
  }
}
```


Copyright Elasticsearch 2013. Copying, publishing and/or distributing without written permission is strictly prohibited

Searching


```
{
  "took": 2,
  "timed_out": false,
  "_shards": { "total": 15, "successful": 15, "failed": 0 },
  "hits": {
    "total": 1,
    "max_score": 0.15342641,
    "hits": [
      {
        "_index": "articles", "_type": "article", "_id": "1",
        "_score": 0.15342641,
        "_source": {
          "title": "My first article",
          "content": "... some lengthy article ...",
          "tags": [ "news", "sports", "introduction" ],
          "created": "2013/04/04 16:54:23",
          "viewed"  : 234,
          "cost"    : 0.99
        }
      }
    ]
  }
}
```


Copyright Elasticsearch 2013. Copying, publishing and/or distributing without written permission is strictly prohibited

Searching

```
{
  "took": 2,
  "timed_out": false,
  "_shards": { "total": 15, "successful": 15, "failed": 0 },
  "hits": {
    "total": 1,
    "max_score": 0.15342641,
    "hits": [
      {
        "_index": "articles", "_type": "article", "_id": "1",
        "_score": 0.15342641,
        "_source": {
          "title": "My first article",
          "content": "... some lengthy article ...",
          "tags": [ "news", "sports", "introduction" ],
          "created": "2013/04/04 16:54:23",
          "viewed"  : 234,
          "cost"    : 0.99
        }
      }
    ]
  }
}
```

Copyright Elasticsearch 2013. Copying, publishing and/or distributing without written permission is strictly prohibited

Searching

```
{
  "took": 2,
  "timed_out": false,
  "_shards": { "total": 15, "successful": 15, "failed": 0 },
  "hits": {
    "total": 1,
    "max_score": 0.15342641,
    "hits": [
      {
        "_index": "articles", "_type": "article", "_id": "1",
        "_score": 0.15342641,
        "_source": {
          "title": "My first article",
          "content": "... some lengthy article ...",
          "tags": [ "news", "sports", "introduction" ],
          "created": "2013/04/04 16:54:23",
          "viewed"  : 234,
          "cost"    : 0.99
        }
      }
    ]
  }
}
```

Copyright Elasticsearch 2013. Copying, publishing and/or distributing without written permission is strictly prohibited

Searching

```
{
  "took": 2,
  "timed_out": false,
  "_shards": { "total": 15, "successful": 15, "failed": 0 },
  "hits": {
    "total": 1,
    "max_score": 0.15342641,
    "hits": [
      {
        "_index": "articles", "_type": "article", "_id": "1",
        "_score": 0.15342641,
        "_source": {
          "title": "My first article",
          "content": "... some lengthy article ...",
          "tags": [ "news", "sports", "introduction" ],
          "created": "2013/04/04 16:54:23",
          "viewed"  : 234,
          "cost"    : 0.99
        }
      }
    ]
  }
}
```

Copyright Elasticsearch 2013. Copying, publishing and/or distributing without written permission is strictly prohibited

Searching

```
{
  "took": 2,
  "timed_out": false,
  "_shards": { "total": 15, "successful": 15, "failed": 0 },
  "hits": {
    "total": 1,
    "max_score": 0.15342641,
    "hits": [
      {
        "_index": "articles", "_type": "article", "_id": "1",
        "_score": 0.15342641,
        "_source": {
          "title": "My first article",
          "content": "... some lengthy article ...",
          "tags": [ "news", "sports", "introduction" ],
          "created": "2013/04/04 16:54:23",
          "viewed"  : 234,
          "cost"    : 0.99
        }
      }
    ]
  }
}
```

Copyright Elasticsearch 2013. Copying, publishing and/or distributing without written permission is strictly prohibited

Faceting

- Faceting allows aggregation of search results

- Term: Group results by a term
- Range: Group by price or date ranges
- Histogram: Group results in equally sized buckets, also as date histogram
- Statistical: Include statistical data like min, max, sum, avg & some more

Copyright Elasticsearch 2013. Copying, publishing and/or distributing without written permission is strictly prohibited

Faceting

Copyright Elasticsearch 2013. Copying, publishing and/or distributing without written permission is strictly prohibited

elasticsearch.

Faceting - Query

```
curl -X POST http://localhost:9200/articles/article/_search?pretty=1 -d '
{
  "from" : 0,
  "size" : 10,
  "query" : {
    "match" : {
      "title" : "first"
    }
  },
  "facets" : {
    "tagsFacet" : {
      "terms" : {
        "field" : "tags",
        "size" : 10
      }
    }
  }
}'
```

Copyright Elasticsearch 2013. Copying, publishing and/or distributing without written permission is strictly prohibited

Faceting response

```
{
  "took" : 154,
  "timed_out" : false,
  "_shards" : { ... },
  "hits" : { ... },

  "facets" : {
    "tagsFacet" : {
      "_type" : "terms",
      "missing" : 0,
      "total" : 3,
      "other" : 0,
      "terms" : [
        { "term" : "sports", "count" : 201 },
        { "term" : "news", "count" : 160 },
        { "term" : "introduction", "count" : 1 }
      ]
    }
  }
}
```

Copyright Elasticsearch 2013. Copying, publishing and/or distributing without written permission is strictly prohibited

Scripting

- Apply custom scoring logic before returning results

- Apply math operations with data from fields to change score
- Scripting languages: MVEL, javascript, groovy, python

Copyright Elasticsearch 2013. Copying, publishing and/or distributing without written permission is strictly prohibited

Plugins & clients

Copyright Elasticsearch 2013. Copying, publishing and/or distributing without written permission is strictly prohibited

Pluggable architecture

- Modularized architecture
- Plugins are simple zip files with a predefined layout
- Different plugin use-cases

  Lucene features

  Monitoring

  Scripting languages

  Rivers

  Transport & Discovery

  Field types, facet types

Copyright Elasticsearch 2013. Copying, publishing and/or distributing without written permission is strictly prohibited

Clients & integrations

- Tons of languages supported already

  Perl, Python, Ruby, PHP, JavaScript, .NET, Scala, Clojure, Erlang

- Lots integrations available

  Grails, Play Framework (1,2), spring & spring-data

  Django, Haystack, Catalyst, Node, Mongoose

  Wordpress, Drupal, Symfony2, CakePHP

  Nagios, Munin, collectd, MCollective, chef

Copyright Elasticsearch 2013. Copying, publishing and/or distributing without written permission is strictly prohibited

Roadmap

Copyright Elasticsearch 2013. Copying, publishing and/or distributing without written permission is strictly prohibited

Roadmap

- Current stable version: Elasticsearch 0.90.1
- On our way to 1.0!

- Documentation

Copyright Elasticsearch 2013. Copying, publishing and/or distributing without written permission is strictly prohibited

Thanks!

http://www.elasticsearch.org
http://groups.google.com/group/elasticsearch

Alexander Reelsen
alexander.reelsen@elasticsearch.com
@spinscale

Copyright Elasticsearch 2013. Copying, publishing and/or distributing without written permission is strictly prohibited

Resources

- Introduction: Getting down and dirty with elasticsearch (Clinton Gormley)

  http://www.slideshare.net/clintongormley/down-and-dirty-with-elasticsearch

- Document relations (Martijn v. Groningen)

  http://www.berlinbuzzwords.de/sites/berlinbuzzwords.de/files/slides/document-relations-bbuz-2013.pdf

- The state of open source logging (Rashid Khan & Shay Banon)

  http://www.berlinbuzzwords.de/sites/

Copyright Elasticsearch 2013. Copying, publishing and/or distributing without written permission is strictly prohibited