

Elasticsearch

Search made easy

Alexander Reelsen

@spinscale

alexander.reelsen@elasticsearch.com

Agenda

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

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

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

Search - by term



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 ▾

 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

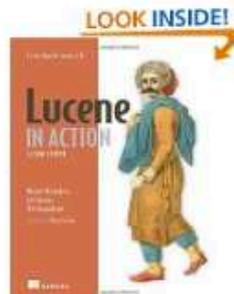
- ★★★★☆ & Up
- ★★★★☆ & Up
- ★★★★☆ & Up
- ★★★★☆ & Up

"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)

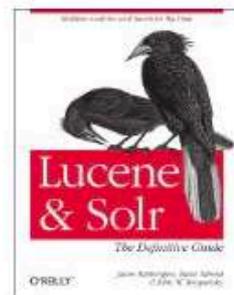
\$25.99 used (26 offers)

★★★★☆ (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

Search - by ID

amazon
Join Prime

Your Amazon.com | Today's Deals | Gift Cards | Sell | Help



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

Search - by attribute



Warenkorb (0)
0,00 €



[DAMEN](#) | [HERREN](#) | [KINDER](#) | [NEWS & STYLE](#)

[Schuhe](#) | [Bekleidung](#) | [Sports](#) | [Accessoires](#) | [Premium](#) | [Wohnen](#) | [Marken](#) | [SALE %](#) | [Lounge](#)

Alle Filter zurücksetzen

Kategorie

- Damen (0)
- Herren (596)
- Kinder (38)
- Wohnen (0)

Marke

- 3 Pommes
- 55 DSL
- 98-86
- adidas Performance
- Armani Jeans

Unsere Topseller

NEU



Napapijri
GONICA - Hemd - rot/wei&blau
119,95 €

NEU



New Zealand Auckland
Hemd - rot
79,95 €



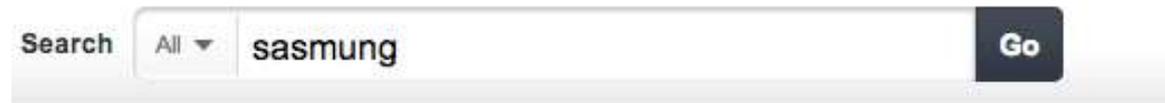
ESPRIT Collection
Hemd - chili red
39,95 €

NEU



Mustang
Hemd - dark shadow
39,95 €

Search - Suggestions



"samsung"

Showing results for [samsung](#). Search instead for [samsung](#).

Related Searches: [samsung galaxy](#), [samsung galaxy s3](#), [samsung tv](#).

Search - Highlighting

github

Explore GitHub

Search

Features

Blog

Sign up for free

Sign in

Search

elasticsearch

Search

Repositories

1,325

Code

47,938

Issues

2,967

Users

4

Languages

Ruby

287

Java

252

JavaScript

202

Python

135

PHP

76

Shell

46

We've found 1,325 repository results

Sort: Best match



elasticsearch/elasticsearch

Java ★ 3,721 🌐 764

Open Source, Distributed, RESTful Search Engine

Last updated 16 hours ago



mobz/elasticsearch-head

JavaScript ★ 384 🌐 115

A web front end for an elastic search cluster

Last updated 8 days ago



nervetattoo/elasticsearch

PHP ★ 123 🌐 30

Simple PHP client for **ElasticSearch**

Last updated a day ago

elasticsearch.

Search - Analytics



Custom

memcache AND NOT program:sshd

Search

Reset

7,679 hits

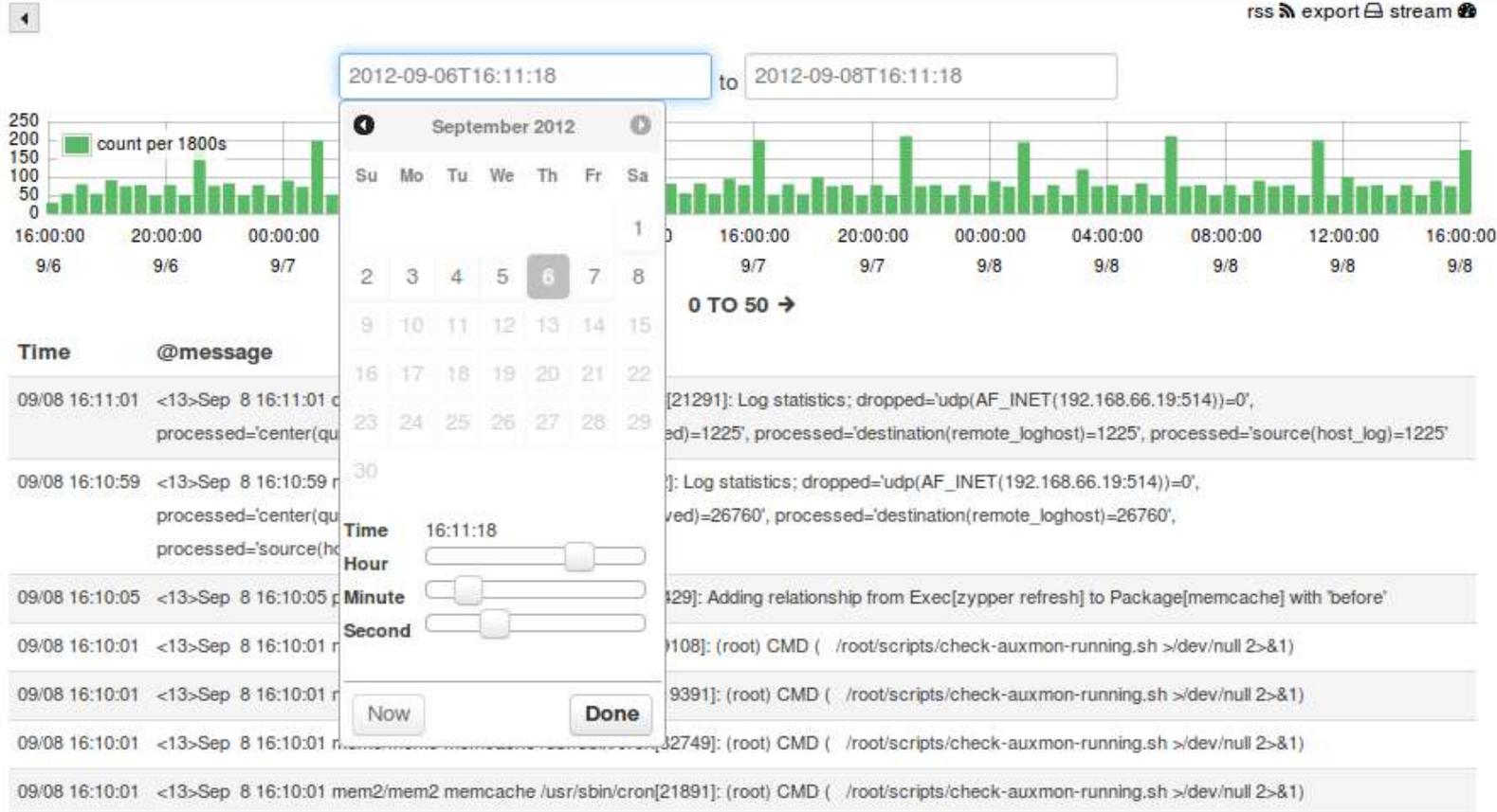
rss export stream

Columns

- @message
- @source
- @source_host
- @source_path
- @tags
- @timestamp
- @type pid program
- role source_host

Analysis

- @message
- @source
- @source_host
- @source_path
- @tags
- @timestamp
- @type
- pid
- program
- role
- source_host



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

Installation & setup

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" } } }'
```

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`

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]
```

Sharding & Replication

Sharding & Replication

- **Replication:** Share same data over several machines

Increasing throughput due to concurrency

Allow outage of nodes without data loss

- **Sharding:** Index partitioning

Split logical data into physically smaller parts

Control data flows

Sharding

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

The screenshot shows the ElasticSearch web interface. At the top, the URL is `http://localhost:9200/` and the cluster name is `Skyhawk` with a status of `cluster health: green (1, 5)`. Below the navigation tabs (Overview, Browser, Structured Query [+], Any Request [+]), the **Cluster Overview** section is active, featuring a `New Index` button. The `products` index is highlighted, showing a size of 430b (430b) and 0 documents. Below the index name, there are `Info` and `Actions` dropdown menus. At the bottom, the node `Skyhawk` is shown with ID `HZwbFDZBSfG_-PZEKwozKA` and IP `inet[/192.168.178.20:9200]`, also with `Info` and `Actions` dropdowns. To the right of the node information, five green boxes represent the shards, numbered 0 through 4, indicating that all five shards are present and healthy.

Replication

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

The screenshot shows the ElasticSearch web interface. At the top, the 'ElasticSearch' logo is on the left, followed by a search bar containing 'http://localhost:9200/' and a 'Connect' button. To the right, the cluster name 'Skyhawk' is displayed, along with a green status bar indicating 'cluster health: green (2, 1)'. Below this, there are navigation tabs for 'Overview', 'Browser', 'Structured Query [+]', and 'Any Request [+]', with 'Overview' selected. The main heading is 'Cluster Overview', with a 'New Index' button to its right. The primary focus is on the 'products' index, which has a size of 66b (132b) and 0 documents. Below the index name are 'Info' and 'Actions' buttons. A table lists the nodes in the cluster:

| Node Name | IP Address | Replicas |
|--------------|--|----------|
| Abner Little | pAVI8rIEQHSVNHZbjpVedg inet[/192.168.178.20:9201] | 0 |
| Skyhawk | HZwbFDZBSfG_-PZEKwozKA inet[/192.168.178.20:9200] | 0 |

Sharding & Replication

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

ElasticSearch **Skyhawk** **cluster health: green (4, 5)**

Overview

Cluster Overview

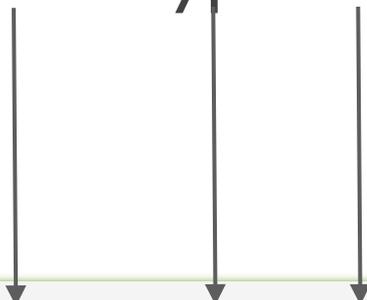
products
size: 430b (760b)
docs: 0 (0)

| | | | | | |
|--|---|---|---|---|---|
| Abner Little pAVI8rIEQHSVNHzbjpvEdg inet[/192.168.178.20:9201] <input type="button" value="Info"/> <input type="button" value="Actions"/> | 0 | 1 | 2 | | |
| Hugh Jones KvQ04bQ8SDGSySUIYC7abA inet[/192.168.178.20:9203] <input type="button" value="Info"/> <input type="button" value="Actions"/> | | 1 | | 3 | |
| Misfit _vSwO9WMQjy5Zhss2-j4nA inet[/192.168.178.20:9202] <input type="button" value="Info"/> <input type="button" value="Actions"/> | 0 | | | | 4 |
| Skyhawk HZwbFDZBSfG_-PZEKwozKA inet[/192.168.178.20:9200] <input type="button" value="Info"/> <input type="button" value="Actions"/> | | | 2 | 3 | 4 |

Indexing

Importing data

index type id



```
# 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
}'
```

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)

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"  
      }  
    }  
  }  
}
```

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 |

Searching

Different ways of searching

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

Searching

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

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
        }
      }
    ]
  }
}
```

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
        }
      }
    ]
  }
}
```

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
        }
      }
    ]
  }
}
```

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
        }
      }
    ]
  }
}
```

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
        }
      }
    ]
  }
}
```

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
        }
      }
    ]
  }
}
```

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

Faceting

Filter by:

▼ Price (per night)

- € 0 - € 49
- € 50 - € 99
- € 100 - € 149
- € 150 - € 199
- € 200 +

▼ Deals

- Smart Deal

▼ Star rating

- 1 star (1)
- 2 stars (18)
- 3 stars (119)
- 4 stars (78)
- 5 stars (9)
- Unrated (33)

▼ Hotel type

- Hotel (221)
- Apartment (13)
- Hostel (5)
- Guest house (19)

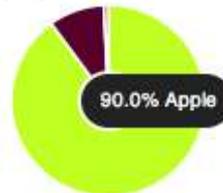
Articles published between 1/1/2011 and 24/1/2011



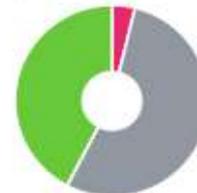
Categories



Topics



Sentiment



Top 5 Authors

- neznámé : 331 (2.03%)
- Zpravy : 122 (0.75%)
- tomasprss : 100 (0.61%)
- Daniel Dočekal : 93 (0.57%)
- MacBlog.sk : 84 (0.51%)

Hide Dashboard

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
      }
    }
  }
}'
```

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 }
      ]
    }
  }
}
```

Scripting

- Apply custom scoring logic before returning results
- Apply math operations with data from fields to change score
- Scripting languages: MVEL, javascript, groovy, python

Plugins & clients

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

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

Roadmap

Roadmap

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

Thanks!

<http://www.elasticsearch.org>

<http://groups.google.com/group/elasticsearch>

Alexander Reelsen

alexander.reelsen@elasticsearch.com

@spinscale

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/>