

Buscas rápidas com Elasticsearch



Leonardo Antunes

Github: [antunesleo](#)

Email: antunesleo4@gmail.com

Twitter: [antunesleo_](#)

O que é elasticsearch

Engine de buscas em tempo real

Orientado a documentos (NoSQL)

Arquitetura baseada em Restful

Distribuído, escalável e alta disponibilidade

Construído em cima do apache lucene

O que é elasticsearch

Engine de buscas em tempo real

Orientado a documentos (NoSQL)

Arquitetura baseada em Restful

Distribuído, escalável e alta disponibilidade

Construído em cima do apache lucene

O que é elasticsearch

Engine de buscas em tempo real

Orientado a documentos (NoSQL)

Arquitetura baseada em Restful

Distribuído, escalável e alta disponibilidade

Construído em cima do apache lucene

O que é elasticsearch

Engine de buscas em tempo real

Orientado a documentos (NoSQL)

Arquitetura baseada em Restful

Distribuído, escalável e alta disponibilidade

Construído em cima do apache lucene

O que é elasticsearch

Engine de buscas em tempo real

Orientado a documentos (NoSQL)

Arquitetura baseada em Restful

Distribuído, escalável e alta disponibilidade

Construído em cima do apache lucene

Empresas que utilizam o ES

NETFLIX

ebay

tinder

 GitHub



mercado
livre



globo
.com

BILZARD[®]
ENTERTAINMENT

vimeo

citibank

Exemplo de uso em produção

The screenshot shows the GitHub search interface. At the top, the navigation bar includes links for Features, Business, Explore, Marketplace, and Pricing. A search bar at the top right contains the text 'elas' and a magnifying glass icon. Below the search bar, a dropdown menu shows 'elas' with a magnifying glass icon and a button labeled 'All GitHub'. The main content area displays search results for 'Python'. On the left, a sidebar lists various categories: Repositories (751K), Code (124M), Commits (12M), Issues (2M), Marketplace (5), Topics (1K), Wikis (222K), and Users (39K). Below this, a 'Languages' section lists various programming languages and their repository counts. The main search results area shows '751,994 repository results' with a 'Sort: Best match' dropdown. The first two results are for 'geekcomputers/Python' and 'TheAlgorithms/Python', both with a Python logo and star counts. The third result is 'Show-Me-the-Code/python', with an HTML logo and star count. The 'geekcomputers/Python' result includes a description 'My Python Examples', 'MIT license', and 'Updated 7 days ago'. The 'TheAlgorithms/Python' result includes a description 'All Algorithms implemented in Python', a list of tags 'sorting-algorithms', 'python', 'education', and 'algorithm', 'MIT license', and 'Updated 8 days ago'. The 'Show-Me-the-Code/python' result includes a description 'Show Me the Code Python version.'.

Repositories 751K

Code 124M

Commits 12M

Issues 2M

Marketplace 5

Topics 1K

Wikis 222K

Users 39K

Languages

| | |
|------------------|---------|
| Python | 501,939 |
| Jupyter Notebook | 37,619 |
| HTML | 17,175 |
| JavaScript | 11,295 |
| Shell | 7,933 |
| C | 5,489 |
| C++ | 5,474 |
| CSS | 4,303 |

Python

Python is a dynamically typed programming language.

[See topic](#)

751,994 repository results Sort: **Best match** ▾

geekcomputers/Python Python ★ 6.3k

My Python Examples

MIT license Updated 7 days ago

TheAlgorithms/Python Python ★ 5.7k

All Algorithms implemented in Python

[sorting-algorithms](#) [python](#) [education](#) [algorithm](#)

MIT license Updated 8 days ago

Show-Me-the-Code/python HTML ★ 1.9k

Show Me the Code Python version.

Benchmark da Caelum

Foram inseridos nos dois bancos 1.466.485 registros aleatórios de pessoas.

Pessoa (Tabela/Index)

- ID.
- Data de criação.
- Nome.
- Tipo de pessoa.

Fonte: <http://blog.caelum.com.br/buscas-eficientes-com-elasticsearch>

SQL - Postgres

```
SELECT * FROM pessoa  
WHERE nome LIKE '%a%'  
ORDER BY nome, tipopessoa;
```

Resultados: **602.187**

Tempo: **1m 6s**

Query DSL - Elasticsearch

```
GET pessoas_idx/_search
{
  "query": {
    "match": { "nome": "a" }
  },
  "sort": [
    { "nome": "asc", "tipopessoa": "asc" }
  ]
}
```

Resultados: 602.187

Tempo: 1.2s no primeiro acesso, média de 10 ms nos demais.

55x mais rápido

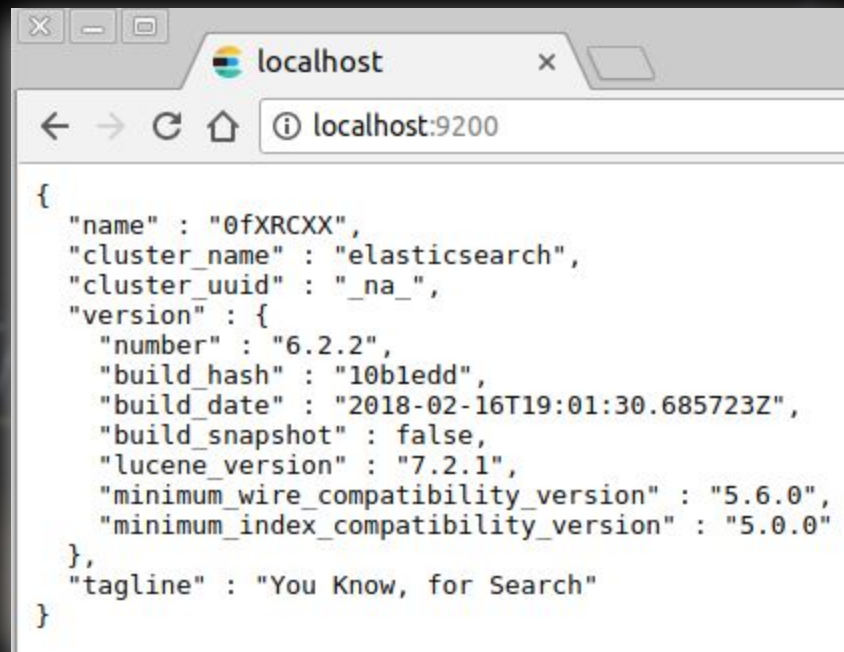
ES ou SQL Databases?



SQL

Como o ES funciona

\$ sudo systemctl start elasticsearch

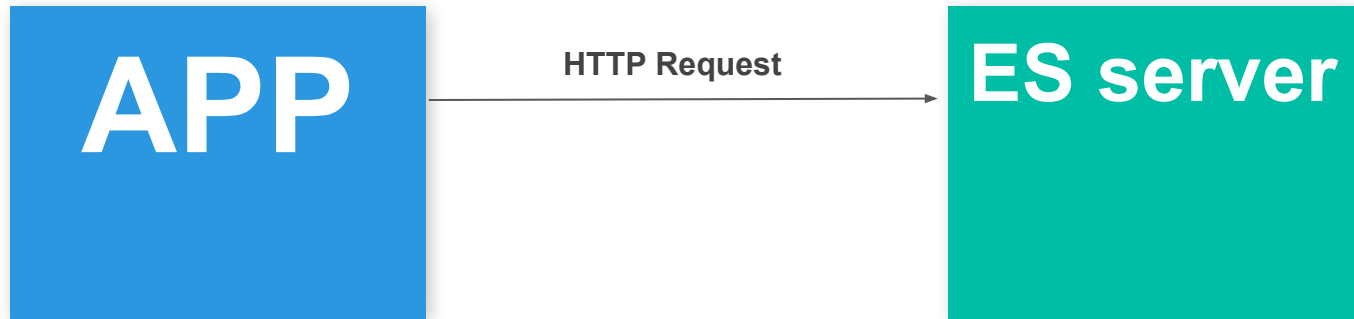
A screenshot of a web browser window. The address bar shows 'localhost:9200'. The page content is a JSON object representing the Elasticsearch cluster status.

```
{
  "name" : "0fXRCXX",
  "cluster_name" : "elasticsearch",
  "cluster_uuid" : "_na_",
  "version" : {
    "number" : "6.2.2",
    "build_hash" : "10b1edd",
    "build_date" : "2018-02-16T19:01:30.685723Z",
    "build_snapshot" : false,
    "lucene_version" : "7.2.1",
    "minimum_wire_compatibility_version" : "5.6.0",
    "minimum_index_compatibility_version" : "5.0.0"
  },
  "tagline" : "You Know, for Search"
}
```

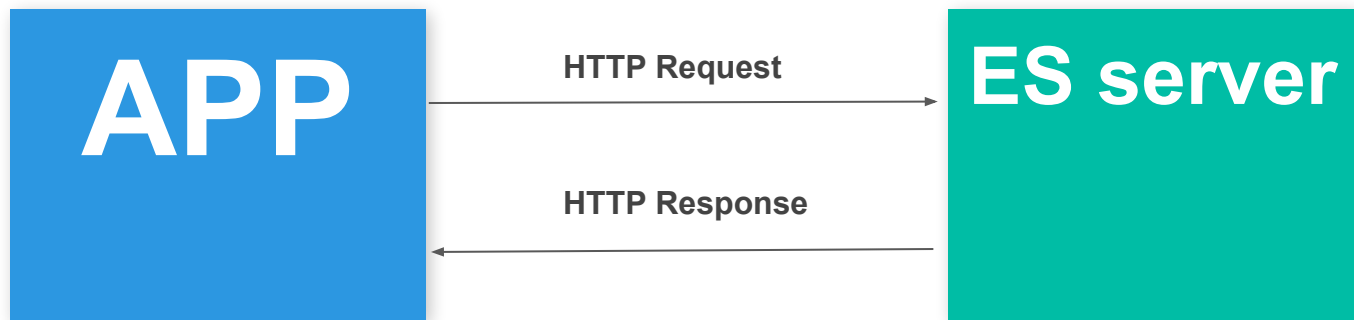
Como o ES funciona - RESTful Architecture



Como o ES funciona - RESTful Architecture



Como o ES funciona - RESTful Architecture



Index & documents - NoSQL

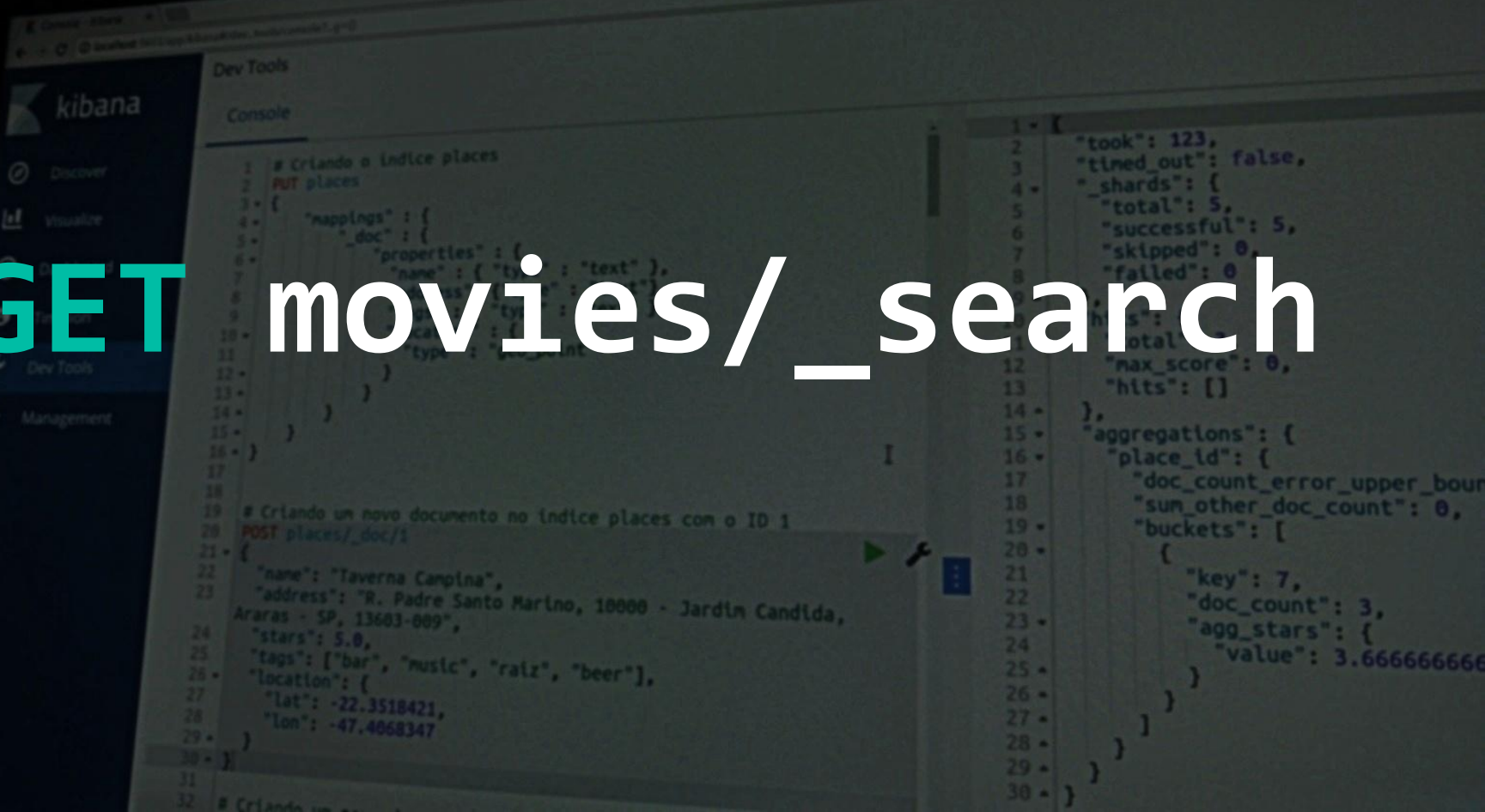


Documents - NoSQL

Os documents são estruturados no formato JSON.

```
{  
  "title": "The Godfather",  
  "year": 1972,  
  "director": "Francis Ford Coppola",  
  "stars": [  
    "Marlon Brando",  
    "Al Pacino",  
    "James Caan"  
  ],  
  "genres": ["Crime", "Drama"]  
}
```

GET movies/_search





Dev Tools

Console Search Profiler Grok Debugger

```
1 GET /shakespeare/_search
2 {
3   "query": {
4     "bool": {
5       "must": [
6         {
7           "match": {
8             "play_name": {
9               "fuzziness": "AUTO",
10              "query": "Kariolanus"
11            }
12          }
13        },
14        {
15          "match_phrase": {
16            "text_entry": "Methinks thou"
17          }
18        }
19      ]
20    }
21  }
22 }
```

```
1 {
2   "took": 6,
3   "timed_out": false,
4   "_shards": {
5     "total": 5,
6     "successful": 5,
7     "skipped": 0,
8     "failed": 0
9   },
10  "hits": {
11    "total": 1,
12    "max_score": 13.890961,
13    "hits": [
14      {
15        "_index": "shakespeare",
16        "_type": "doc",
17        "_id": "25107",
18        "_score": 13.890961,
19        "_source": {
20          "type": "line",
21          "line_id": 25108,
22          "play_name": "Coriolanus",
```

LET'S CODE!



Query Match

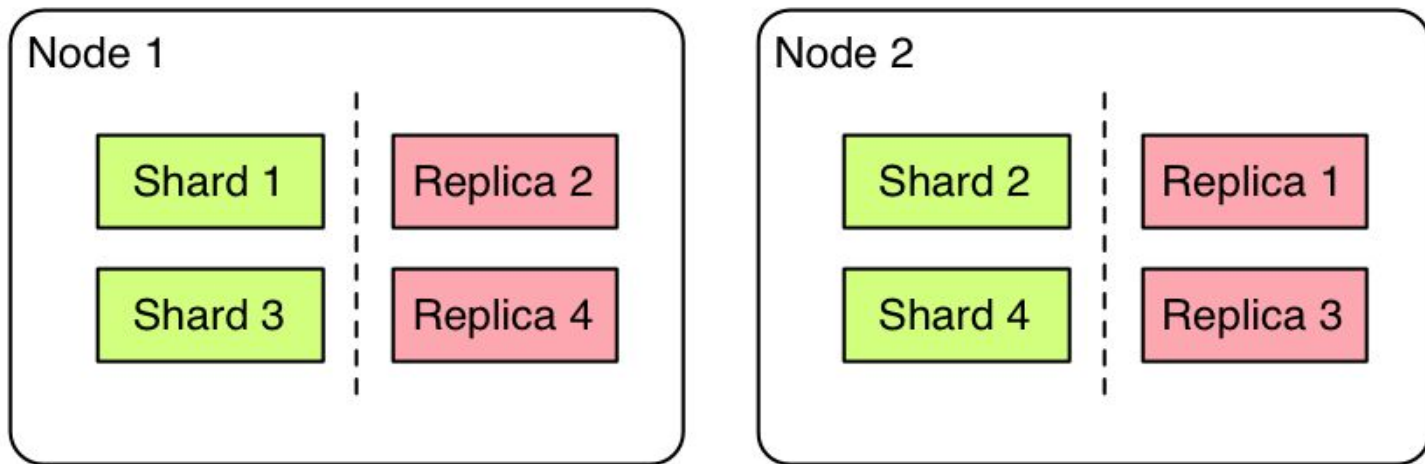
Bool Query

Filter Context

Score

Cluster, Node, Shard & Replica

ElasticSearch Cluster



Elasticsearch é distribuído naturalmente.

Sua aplicação não precisa se preocupar com isso.

Elasticsearch clients

Official

.NET

Groovy

Java

Javascript

Perl

PHP

Python

Ruby

Community

.NET

OCaml

Perl

PHP

Python

R

Ruby

Rust

Scala

Smalltalk

Vert.x

B4J

C++

Clojure

ColdFusion (CFML)

Erlang

Go

Groovy

Haskell

Java

JavaScript

kotlin

Lua

Github case



2 billion documents

8 million repositórios

4 million usuários ativos

128 shards, cerca 120 gigabytes cada

15 terabytes de informação

Tinder case



+280.000.000 queries por dia

+100.000.000.000 por mes

+60 elasticsearch nodes

LET'S CODE AGAIN!



Geolocation API

Geopoint

THAT'S ALL



Leonardo Antunes

Github: antunesleo

Email: antunesleo4@gmail.com

Twitter: antunesleo_