

ELASTICSEARCH

Nikola Milikić

nikola.milikic@fon.bg.ac.rs

Elasticsearch (ES)

Vebsajt: <https://www.elastic.co/downloads/elasticsearch>

Besplatan i open-source

Zasnovan na Apache Lucene

Verzije:

- Prva zvanična verzija v0.4 objavljena u februaru 2010
- Poslednja verzija v6.6.1 od 19.02.2019.

Pisan u Javi

Kako proširuje Lucene?

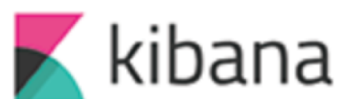
Dodaje RESTful servise

Visoke performanse i dostupnost

- jednostavno klasterovanje

Elastic Stack

User Interface



Store, Index
& Analyze



Ingest



Karakteristike

- Skalira se horizontalno
- Visoka pouzdanost, ugradjeno repliciranje

Instalacija (Linux)

```
$ wget https://artifacts.elastic.co/downloads/elasticsearch/elasticsearch-6.6.1.zip
```

```
$ unzip elasticsearch-6.6.1.zip
```

```
$ cd elasticsearch-6.6.1/
```

```
$ ./bin/elasticsearch
```

Instalacija (Windows)

Skinuti poslednju verziju sa URL-a

<https://artifacts.elastic.co/downloads/elasticsearch/elasticsearch-6.6.1.zip>

Raspakovati u željeni folder (npr. C:/elasticsearch/)

Pokrenuti fajl /bin/elasticsearch.bat, npr.

```
C:/elasticsearch/elasticsearch-6.6.1/bin/elasticsearch.bat
```

Provera da li je server pokrenut

```
← → ↻ 🏠 ⓘ localhost:9200

{
  name: "kHfqrfg",
  cluster_name: "elasticsearch",
  cluster_uuid: "WP77RxSKSuGx7mUkOBFrXA",
- version: {
    number: "6.6.1",
    build_flavor: "default",
    build_type: "tar",
    build_hash: "1fd8f69",
    build_date: "2019-02-13T17:10:04.160291Z",
    build_snapshot: false,
    lucene_version: "7.6.0",
    minimum_wire_compatibility_version: "5.6.0",
    minimum_index_compatibility_version: "5.0.0"
  },
  tagline: "You Know, for Search"
}
```


Poređenje koncepata sa SQL-om

MySQL	Elastic Search
Baza	Indeks
Tabela	Tip Biće izbačeno od verzije 7.0.0
Red	Dokument
Kolona	Polje
Šema	Mapiranje
Indeksiranje	Sve je indeksirano
SQL	Query DSL

Distribuirana i visoko dostupna

Više servera (node) mogu raditi u **klasteru**

- Ponašaju se kao jedan servis

Šardovi (shardes)

- Indeksi se nalaze na različitim šardovima
- Svaki šard može imati nula ili više replika
 - Replike šardova čuvati na različitim serverima

Master instanca

- Automatsko detektovanje Master-a u slučaju otkaza
- Odgovoran za raspodelu i upravljanje opterećenjem šardova (load ballancing)

Nema šemu, koristi “dokumente”

- Ne treba definisati šemu unapred, vrši se dinamičko mapiranje u Elasticsearch tipove podataka
- Nema potrebe ažurirati šemu tokom korišćenja (ALTER u sql-u)
- Može se definisati mapiranje (šema) kako bi se prilagodilo indeksiranje

Linkovi za današnji čas

- Instalirati elasticsearch-head (Chrome plugin GUI za Elasticsearch):
<https://chrome.google.com/webstore/detail/elasticsearch-head/ffmkiejjmecolpfloofpjologoblkegm>
- Instalirati Postman (Chrome plugin za izvršenje HTTP poziva):
<https://chrome.google.com/webstore/detail/postman/fhbjgbiflinjbdgggehcdcbncdddomop?hl=en>
- Skinuti Eclipse projekat (grana (class-start)):
<https://github.com/milikicn/elasticsearch-tutorial>

Primer za vežbanje

U našem primeru ćemo imati:

- klaster of jednog servera (node)
- dva indeksa:
 - **book** – sa jednim tipom **book**
 - **author** – sa jednim tipom **author**

Struktura dokumenta tipa *author*

```
{  
  "name": "Larry",  
  "lastname": "Niven",  
  "gender": "male"  
}
```

Struktura dokumenta tipa *book*

```
{
  "title": "The Mote in God's Eye",
  "isbn": "0-671-21833-6",
  "year": 1974,
  "score": 4.1,
  "authors" : [
    {
      "id" : 1
    },
    {
      "id" : 2
    }
  ]
}
```

Zadatak 1: Kreiranje dokumenta

U indeksu author, kreirati dokument tipa author pod id-jem 1 sa datim podacima.

Parametri	Vrednost
HTTP method	PUT
URL	http://localhost:9200/author/author/1
content-type	application/json
body	<pre>{ "name": "Larry", "lastname": "Niven", "gender": "male" }</pre>

Zadatak 2: Brisanje dokumenta

Iz indeksa author, obrisati dokument pod id-jem 1.

Parametri	Vrednost
HTTP method	DELETE
URL	<code>http://localhost:9200/author/author/1</code>

Zadatak 3: Brisanje indeksa

Obrisati indeks author.

Parametri	Vrednost
HTTP method	DELETE
URL	<code>http://localhost:9200/author</code>

Zadatak 4: Kreiranje indeksa

Kreirati prazan indeks author.

Parametri	Vrednost
HTTP method	PUT
URL	<code>http://localhost:9200/author</code>

Eksplicitno mapiranje

- Definisanje strukture indeksa, tipova podataka i njihovih karakteristika, npr.
 - Koji stringovi treba da se posmatraju kao tekst
 - Koja polja sadrže brojeve, datume i geolokacije
 - Format datuma
- Poboljšava performanse jer Elasticsearch ne mora da pogađa tip podatka (npr. datum se često predstavlja kao long u programskim jezicima)
- Koristi drugačije analizatore i podešavanja za indeksiranje i za pretragu

Primer šeme (mapiranja) indeksa

```
{
  "mappings": {
    "author": {
      "properties": {
        "name": {
          "type": "text"
        },
        "lastname": {
          "type": "text"
        },
        "gender": {
          "type": "text"
        }
      }
    }
  }
}
```

Zadatak 5: Postavljanje mapiranja za indeks

Postaviti mapiranje za indeks author.

Parametri	Vrednost
HTTP method	PUT
URL	<code>http://localhost:9200/author</code>
content-type	<code>application/json</code>
body	(na sledećem slajdu)

Zadatak 5: Postavljanje mapiranja za indeks

Parametri

Vrednost

body

```
{
  "properties": {
    "name": {
      "type": "text"
    },
    "lastname": {
      "type": "text"
    },
    "gender": {
      "type": "text"
    }
  }
}
```

Zadatak 6: Indeksiranje više dokumenata odjednom

U jednom upitu, u indeks author indeksirati podatke o četiri autora (tip author).

Parametri	Vrednost
HTTP method	PUT
URL	http://localhost:9200/author/author/_bulk
content-type	application/json
body	<pre>{ "index":{ "_id": 1} } { "name": "Larry", "lastname": "Niven", "gender": "male" } { "index":{ "_id": 2} } { "name": "Jerry", "lastname": "Pournelle", "gender": "male" } { "index":{ "_id": 3} } { "name": "Terry", "lastname": "Pratchett", "gender": "male" } { "index":{ "_id": 4} } { "name": "Alice", "lastname": "Hoffman", "gender": "female" }</pre>

Zadatak 7: Kreiranje dokumenta sa kolekcijom

U indeksu book, kreirati dokument tipa book sa datim atributima i kolekcijom id-jeva autora.

Parametri	Vrednost
HTTP method	PUT
URL	http://localhost:9200/book/book/1
content-type	application/json
body	<pre>{ "title": "The Mote in God's Eye", "isbn": "0-671-21833-6", "year": 1974, "score": 4.1, "authors" : [{ "id": 1 }, { "id": 2 }] }</pre>

Zadatak 8: Indeksiranje više dokumenata odjednom

U jednom upitu, kreirati dokumente u indeksu book, tipa book, sa datim podacima o knjigama i id-jevima autora.

Parametri	Vrednost
HTTP method	PUT
URL	http://localhost:9200/book/book/_bulk
content-type	application/json
body	<pre>{ "index":{ "_id": 2} } { "title": "The Color of Magic", "isbn": "0-86140-324-X", "year": 1983, "score": 4, "authors" : [{ "id" : 3 }]} { "index":{ "_id": 3} } { "title": "Good Omens: The Nice and Accurate Prophecies of Agnes Nutter, Witch", "isbn": "0-575-04800-X", "year": 1980, "score": 4.2, "authors" : [{ "id" : 3 }]} { "index":{ "_id": 4} } { "title": "The Rules of Magic", "isbn": "978-1501137488", "year": 2017, "score": 4, "authors" : [{ "id" : 4 }]}</pre>

Upit 1: Vraćanje svih dokumenata iz indeksa

Vratiti sve knjige iz indeksa book.

Parametri	Vrednost
HTTP method	PUT
URL	http://localhost:9200/book/_search
content-type	application/json
body	<pre>{ "query" : { "match_all" : { } } }</pre>

Upit 2: Pretraga po vrednosti atributa (URL)

Vratiti sve knjige iz indeksa book koje su izdate 1980. godine.

Parametri	Vrednost
HTTP method	GET
URL	<code>http://localhost:9200/book/_search?q=year%3A1980</code>
content-type	
body	

Query DSL

- Domenski specifičan jezik kreiran u JavaScript-u
- Služi za definisanje uslova pretrage

Upit 3: Pretraga po vrednosti atributa (DSL)

Vratiti sve knjige iz indeksa book koje su izdate 1980. godine.

Parametri	Vrednost
HTTP method	POST
URL	http://localhost:9200/book/_search
content-type	application/json
body	<pre>{ "query" : { "term" : { "year" : "1980" } } }</pre>

Upit 4: Pretraga po logičkom uslovu

Vratiti sve knjige iz indeksa book koje su izdate 1980. godine ili pre.

Parametri	Vrednost
HTTP method	POST
URL	http://localhost:9200/book/_search
content-type	application/json
body	<pre>{ "query" : { "range" : { "year" : { "lte" : "1980" } } } }</pre>

Upit 5: Sortiranje

Vratiti sve knjige iz indeksa book koje su izdate 1980. godine ili pre, sortirane po oceni u opadajućem redosledu.

Parametri	Vrednost
HTTP method	POST
URL	http://localhost:9200/book/_search
content-type	application/json
body	<pre>{ "sort" : [{ "score" : { "order" : "desc" } }], "query" : { "range" : { "year" : { "lte" : "1980" } } } }</pre>

Upit 6: Paginacija

Vratiti jednu knjigu, počevši od prve knjige, iz indeksa book koje su izdate 1980. godine ili pre.

Parametri	Vrednost
HTTP method	POST
URL	http://localhost:9200/book/_search
content-type	application/json
body	<pre>{ "from" : 0, "size" : 1, "query" : { "range" : { "year" : { "lte" : "1980" } } } }</pre>

Analizatori teksta

Analizatori konvertuju tekst u tokene iliti termine na osnovu kojih se vrši pretraga.

Default analizator rečenicu podeli na reči (tokene), ukloni znake interpunkcije i spusti slova (lowercase).

"The QUICK brown foxes jumped over the lazy dog!"

će biti pretvorena u skup termina

[quick, brown, fox, jump, over, lazy, dog]

Analiza se vrši:

- Prilikom indeksiranja
- Prilikom izvršenja upita

term vs. match

- term upit traži tačno pojavljivanje termina
- match upit prvo analizira query, pa tim rezultatom vrši pretragu

Upit 7: Vredost tekstualnog tipa (term)

Vratiti sve autore muškog pola.

Parametri	Vrednost
HTTP method	POST
URL	http://localhost:9200/author/_search
content-type	application/json
body	<pre>{ "query" : { "term" : { "gender" : "MALE" } } }</pre>

NEMA REZULTATA

Upit 8: Vredost tekstualnog tipa (match)

Vratiti sve autore muškog pola.

Parametri	Vrednost
HTTP method	POST
URL	http://localhost:9200/author/_search
content-type	application/json
body	<pre>{ "query" : { "match" : { "gender" : "MALE" } } }</pre>

Upit 9: Pretraga po nekim rečima

Vratiti sve knjige koje u naslovu imaju neku od reči iz fraze "color of magic".

Parametri	Vrednost
HTTP method	POST
URL	http://localhost:9200/book/_search
content-type	application/json
body	<pre>{ "query" : { "query_string" : { "default_field" : "title", "query" : "color of magic" } } }</pre>

Default operator je OR

Upit 10: Pretraga po svim rečima

Vratiti sve knjige koje u naslovu imaju sve reči iz fraze "color of magic".

Parametri	Vrednost
HTTP method	POST
URL	http://localhost:9200/book/_search
content-type	application/json
body	<pre>{ "query" : { "query_string" : { "default_field" : "title", "default_operator": "AND", "query" : "color of magic" } } }</pre>

Upit 11: Pretraga po početku reči

Vratiti sve knjige čiji naslov počinje sa "good o".

Parametri	Vrednost
------------------	-----------------

HTTP method	POST
-------------	------

URL	http://localhost:9200/book/_search
-----	------------------------------------

content-type	application/json
--------------	------------------

body	<pre>{ "query" : { "query_string" : { "default_field" : "title", "default_operator": "AND", "query" : "good o*" } } }</pre>
------	---

Upit 12: fuzzy pretraga

Vratiti sve knjige koje u naslovu imaju reč "magig" tolerišući grešku u jednom slovu.

Parametri	Vrednost
HTTP method	POST
URL	http://localhost:9200/book/_search
content-type	application/json
body	<pre>{ "query" : { "fuzzy" : { "title" : "magig" } } }</pre>

Upit 13: Pretraga po više kriterijuma

Vratiti sve knjige koje u naslovu imaju reč "magig" tolerišući grešku u jednom slovu.

Parametri	Vrednost
HTTP method	POST
URL	<code>http://localhost:9200/book/_search</code>
content-type	<code>application/json</code>
body	(na sledećem slajdu)

Upit 13: Pretraga po više kriterijuma

Parametri

Vrednost

body

```
{
  "query" : {
    "bool" : {
      "must" : [
        {
          "term" : { "authors.id" : "3" }
        },
        {
          "query_string" : {
            "default_field" : "title",
            "query" : "magic"
          }
        }
      ]
    }
  }
}
```

Korisni linkovi

- Postman upiti sa današnjeg časa:
<https://documenter.getpostman.com/view/611615/nst2-elasticsearch-bookstore/RVnWiK6p>
- Još primera Elasticsearch upita:
<https://dzone.com/articles/23-useful-elasticsearch-example-queries>