



Semantix

## Elastic Essential I

Aula 3

Quem sou eu?

Eu sou Rodrigo Augusto Rebouças.

Engenheiro de dados da Semantix  
Instrutor do Semantix Academy

Você pode me encontrar em:  
[rodrigo.augusto@semantix.com.br](mailto:rodrigo.augusto@semantix.com.br)





# API de pesquisa

Comandos básicos do Search APIs

# Pesquisa Funcionamento

- Buscar todos os documentos
  - GET cliente/\_search
- Pesquisar algo em todos os documento
  - GET cliente/\_search?q=hadoop
    - Faz uso do campo “\_all”
- Pesquisar em um atributo específico
  - GET cliente/search?q=nome:João
  - GET cliente/\_search?q=nome:João&q=idade:20



# Query DSL

- Definição de consultas
  - Query DSL (Domain Specific Language)
  - Baseada em JSON

- Exemplo:

GET cliente/\_search?q=Hadoop

GET cliente/\_search?q=Hadoop

```
{  
  "query": {  
    "match_all": {}  
  }  
}
```

- Estrutura do json de busca
  - Took: Tempo em milissegundo
  - Timed\_out: Tempo de limite excedido
  - \_shards: Quantidade usadas (sucesso e falha)
  - Hits: Informação do resultado
    - Total: Quantidade de documentos encontrados
    - Max\_score: Valor de semelhança da consulta (0 à 1)
      - Score é calculado com uso do algoritmo BM25

$$\text{score}(D, Q) = \sum_{i=1}^n \text{IDF}(q_i) \cdot \frac{f(q_i, D) \cdot (k_1 + 1)}{f(q_i, D) + k_1 \cdot \left(1 - b + b \cdot \frac{|D|}{\text{avgdl}}\right)},$$

# Pesquisa Múltiplos Índices

- Pesquisar em todos os índices
  - GET \_all/\_search?q=Windows
  - Cuidado para não fazer consultas lentas
- Pesquisar em índices específicos
  - GET produto,cliente/\_search?q=Windows
  - GET produto,cliente/\_count?q=Windows
  - Caso o índice não exista
    - index\_not\_found\_exception



## Limitação e Paginação na Pesquisa



# Pesquisa Limitação e Paginação

- Para pesquisas com muitos documentos
  - Difícil visualização
  - Limitar a quantidade de documentos
    - Size = número de documentos
  - Paginação
    - From = Documento que irá visualizar
  - Resposta máxima
    - $\text{from} + \text{size} \leq \text{index.max\_result\_window} (10.000)$
    - Scroll
- Limitar o número de documentos
  - `GET cliente/_search?q=hadoop&size=100`
- Paginação, visualizar x documentos por paginação
  - `GET cliente/_search?q=hadoop&size=100&from=500`

- Paginação na estrutura na busca

`GET cliente/_search`

```
{  
  "from" : 0, "size" : 10,  
  "query" : {  
    ...  
  }  
}
```

# Exemplo Limitação e Paginação

- Mostrar os 10 primeiros documentos (1ª Página)

GET \_search?&size=10

```
{
  "query": {
    "match_all": {}
  }
}
```

- Mostrar os documentos de 31 à 40 (4ª Página)

GET \_search?&size=10&from=30

```
{
  "query": {
    "match_all": {}
  }
}
```

- Paginação

- Ex com paginação de 10 documentos.
  - 1ª Página - size =10 & from = 0 (Default)
  - 2ª Página - size =10 & from = 10
  - 10ª Página - size =10 & from = 90
- Fórmula
  - Primeiro documento da busca = From + 1
  - Último documento da busca = From + Size
  - Página = From / Size + 1

# Exercícios Pesquisa e paginação

1. Pesquisar no índice produto os documentos com os seguintes atributos:
  - a) Nome = mouse
  - b) Quantidade = 30
  - c) Descrição = USB
  - d) Nome = hd e descrição = windows
  - e) Nome = memória e descrição = GB
2. Pesquisar todos os índices, limitando a pesquisa em 5 documentos em cada página e visualizar a 4 página (Documentos entre 16 á 20 )



## Gerenciamento de índices



# Índice Gerenciamento

- Create Index
- Get Index
- Indices Exists
- Delete Index
- Open / Close Index API



# Índice Criação

- Configuração básica

PUT teste

```
{  
  "settings" : {  
    "index" : {  
      "number_of_shards" : 1,  
      "number_of_replicas" : 1  
    }  
  },  
  "mappings" : { ... },  
  "aliases": { ... }  
}
```

- Boa prática, cada shard ter entre 20 GB à 50 GB (Não é uma regra)
  - Precisa testar o índice para saber o valor ideal

- Consultar índice
  - GET teste
  - GET teste/\_search
  - GET teste/\_settings
  - GET teste/\_mapping
  - GET teste/\_alias
  - GET teste/\_stats
- Verificar se o índice existe
  - HEAD teste

# Índice Remoção

- Deletar um índice

PUT teste1

DELETE teste1

- Deletar múltiplos índices

PUT indice1

PUT indice2

PUT indice3

PUT indice4

GET ind\*

DELETE ind\*

HEAD ind\*

# Índice Fechamento e Abertura

- Fechar índice
  - Diminuir sobrecarga no cluster
    - Manter metadados
  - Bloquear
    - Leitura
    - Gravação
  - Não manter por muito tempo
    - Quando um nó deixar o cluster
      - As cópias serão perdidas
    - Solução: Frozen Index
  - Exemplo
    - POST teste/\_close
    - POST test\*/\_close
- Abrir índice
  - POST teste/\_open
  - POST test\*/\_open



# Maapeamento

Comandos básicos para Mappings



# Índice Mapeamento

- Elasticsearch define automaticamente no índice os tipos dos campos
- Visualizar o mapeamento
  - SQL -> schema da tabela
- Exemplo
  - GET cliente/\_mapping
  - Não é possível alterar o tipo do dado
    - Reindex
  - É possível criar novos atributos

# Mapeamento Tipo de dados

- Core
  - Texto
    - text and keyword
  - Numérico
    - long, integer, short, byte, double, float
  - Date
  - Binary
  - Boolean
  - Ip
- Complexos
  - Object
  - Nested

# Mapeamento Pesquisa

- Mapeamento pelo índice
  - GET cliente/\_mapping
  - GET client\*/\_mapping
- Mapeamento pelo atributo
  - GET cliente/\_mapping/field/conhecimento
  - GET cliente/\_mapping/field/conhe\*
  - GET cliente/\_mapping/field/nome,conhecimento
- Mapeamento de todos os índices
  - GET \_mapping

# Mapeamento Exemplo

- Exemplo de Criação

PUT cliente1

```
{}
```

PUT cliente1/\_mapping/

```
{
```

```
  "properties": {
```

```
    "nome": { "type": "text" },
```

```
    "idade": { "type": "long" },
```

```
    "conhecimento": { "type" : "keyword"}
```

```
  }
```

```
}
```



# Reindex





# Índice Reindex

- Alterar o mapeamento
- Forma básica para Reindexar
  - Configura o novo índice
  - Indexa o índice de entrada (source) para o destino (dest)

- Exemplo

POST \_reindex

```
{  
  "source": {  
    "index": "teste1"  
  },  
  "dest": {  
    "index": "new_teste"  
  }  
}
```

# Exercícios Índices

1. Visualizar as configurações do índice produto
2. Visualizar o mapeamento do índice produto
3. Visualizar o mapeamento do atributo nome do índice produto
4. Inserir o campo **data** do tipo date no índice produto
5. Adicionar o documento:
  - `_id: 6, "nome": "teclado", "qtd": 100, "descricao": "USB", "data": "2020-09-18"`
6. Reindexar o índice produto para produto2, com o campo quantidade para o tipo short
7. Visualizar o mapeamento do índice produto2
8. Fechar o índice produto
9. Pesquisar todos os documentos no índice produto
10. Abrir o índice produto



# Semantix

## Obrigado!

Alguma pergunta?



Você pode me encontrar em:  
[rodrigo.augusto@semantix.com.br](mailto:rodrigo.augusto@semantix.com.br)

**GET SMARTER**