



Semantix

Elastic Essential I

Aula 4

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





Query e Filtros

Conceitos

Índice invertido

Tipos de queries

Queries e Filtros

- Como fazer as buscas?
 - Queries
 - Quão bem a busca corresponde com o documento
 - Características
 - Tem score
 - Não são armazenados na cache
 - Filtros
 - A busca corresponde com o documento
 - Sim ou não
 - Características
 - Não possui score (order by)
 - Armazenados em cache automaticamente
 - Acelerar o desempenho

Inverted Index

- Índice Invertido
- Índice Remissivo
 - As páginas que estão localizados os principais termos

Índice Remissivo

A	K	Q
Alef..... 3	Kepler..... 2	Quociente 3
Análise 1		
B	L	R
Bola	Limite..... 2	Razão 3
aberta 3	infinito..... 4	Riemman..... 4
fechada 4	M	
	Matemática 2	
F	N	S
Função 1	Napier..... 2	Somatório 3
H		
História	P	T
da Matemática 1	Polinômios 2	Topologia 3

Como salvar todo o dado e como buscar?

Queries Tipos

- Query DSL
 - <https://www.elastic.co/guide/en/elasticsearch/reference/current/query-dsl.html>
- Todas as queries calculam o “_score”
 - Se não
 - Utilizar o “constant_score”
- Filtrar os dados antes da busca textual
 - Ganhar desempenho
- Term
- Terms
- Range
- Match
- Exists
- Missing
- Prefix
- Wildcard
- Regexp
- Fuzzy
- Ids

Exemplo Query - Term

GET cliente/_search

```
{  
  "query": {  
    "term": {  
      "nome": "joão"  
    }  
  }  
}
```

- Qual o “took” e o “_score”?

```
{  
  "took" : 5,  
  ...  
  "hits" : [  
    {  
      "_index" : "cliente",  
      "_type" : "_doc",  
      "_id" : "1",  
      "_score" : 0.9808291,  
      "_source" : {  
        "nome" : "João",  
        "idade" : 20,  
        "conhecimento" : "Windows, Office, Hadoop, Elastic"  
      }  
    }  
  ],  
  ...  
}
```

Exemplo Query – Term com constant_score

GET cliente/_search

```
{
  "query": {
    "constant_score": {
      "filter": {
        "term": {
          "nome": "joão"
        }
      }
    }
  }
}
```

- Qual o “took” e o “_score”?

```
{
  "took" : 1,
  ...
  "hits" : [
    {
      "_index" : "cliente",
      "_type" : "_doc",
      "_id" : "1",
      "_score" : 1.0,
      "_source" : {
        "nome" : "João",
        "idade" : 20,
        "conhecimento" : "Windows, Office, Hadoop, Elastic"
      }
    }
  ]
}
```


Exemplo Query - Terms

GET cliente/_search

```
{  
  "query": {  
    "terms": {  
      "idade": [30,20]  
    }  
  }  
}
```



Bool Query

Consultas booleanas



Bool Query

- Filtrar um dataset grande
- Estrutura flexível
- Atributos
 - Must - And
 - Should – OR
 - Must_not - Not and
 - Filter – Filtrar mais dados antes de atender as outras cláusulas

Estrutura Bool Query

GET cliente/_search

```
{  
  "query": {  
    "bool": {  
      "must": [ { ... } ],  
      "must_not": [ { ... } ],  
      "should": [ { ... } ],  
      "filter": [ { ... } ]  
    }  
  }  
}
```

Exemplo Bool Query

- 1 Consulta booleana com 1 atributo

GET cliente/_search

```
{
  "query": {
    "bool": {
      "should": {
        "term": {
          "idade": "30"
        }
      }
    }
  }
}
```

Exemplo Bool Query

- 1 consulta booleana com N atributos

GET cliente/_search

```
{
  "query" : {
    "bool" : {
      "must" : [
        {"match" : {"estado": "sp" }},
        {"match" : {"ativo": "sim" }}
      ]
    }
  }
}
```


Exemplo Bool Query

- N consultas booleanas com N atributos

GET cliente/_search

```
{
  "query": {
    "bool": {
      "must": {"match": {"setor": "vendas"}},
      "should": [
        {"match": {"tags": "imutabilidade"}},
        {"match": {"tags": "larga escala"}}
      ],
      "must_not": {"match": {"nome": "inativo"}}
    }
  }
}
```

Exercícios Query e Filtros

- Realizar todas as buscas a seguir no índice produto
- 1. Buscar no termo nome o valor mouse
- 2. Buscar no termo nome os valores mouse e teclado
- 3. Realizar a mesma busca do item 1 e 2, desconsiderando o score
- 4. Buscar os documentos que contenham a palavra “USB” no atributo descrição
- 5. Buscar os documentos que contenham a palavra “USB” e não contenham a palavra “Linux” no atributo descrição
- 6. Buscar os documentos que podem ter a palavra “memória” no atributo nome ou contenham a palavra “USB” e não contenham a palavra “Linux” no atributo descrição



Ordem de busca



Ordem Busca

- Quantas vezes o termo aparece no atributo
- Tamanho do atributo
- Tamanho do termo
- Quantas vezes o termo aparece em todos os documentos
- Exemplo:

GET cliente/_search

```
{  
  "query": {  
    "match": {  
      "conhecimento": "sqoop hive impala elk"  
    }  
  }  
}
```

Operador Busca

- Padrão é o “or”
- "operator": "and"

GET cliente/_search

```
{  
  "query": {  
    "match": {  
      "conhecimento": {  
        "query": "sqoop hive",  
        "operator": "and"  
      }  
    }  
  }  
}
```

Operador Busca

- Definir o mínimo de % que estejam na consulta
 - "minimum_should_match": " valor em %"
- Exemplo

GET cliente/_search

```
{  
  "query": {  
    "match": {  
      "Hobby": {  
        "query": "sqoop hive impala",  
        "minimum_should_match": "50%"  
      }  
    }  
  }  
}
```


Operador Busca

- Definir o mínimo de % que estejam na consulta
 - "minimum_should_match": Número
- Exemplo

GET cliente/_search

```
{  
  "query": {  
    "match": {  
      "Hobby": {  
        "query": "sqoop hive impala",  
        "minimum_should_match": 2  
      }  
    }  
  }  
}
```

Múltiplos Atributos

```
{  
  "query": {  
    "bool": {  
      "should": [  
        { "match": { "endereco": "pinheiros" } },  
        { "match": { "cidade": "pinheiros" } },  
        { "match": { "estado": "pinheiros" } }  
      ]  
    }  
  }  
}
```

Múltiplos Atributos

```
{  
  "query": {  
    "multi_match": {  
      "query": "pinheiros",  
      "type": "most_fields",  
      "fields":["endereco", "cidade", "estado"]  
    }  
  }  
}
```

- OBS: Não pode usar junto com operator e minimum_should_match

Exercícios Ordem de Busca

- Realizar todas as buscas a seguir no índice produto
- 1. Buscar os documentos que contenham as palavras “Windows” e “Linux” no atributo descrição
- 2. Buscar os documentos que contenham as palavras “Windows”, “Linux” ou “USB” no atributo descrição
- 3. Buscar os documentos que contenham pelo menos 2 palavras da seguinte lista de palavras: “Windows”; “Linux” e “USB” no atributo descrição
- 4. Buscar os documentos que contenham pelo menos 50 % da seguinte lista de palavras: “Windows”; “Linux” e “USB” no atributo descrição

Consultas por intervalo

Range Query

- Atributos para controlar o intervalo
 - gte – Maior que ou igual a
 - gt – Maior que
 - lte – Menor que ou igual a
 - lt – Menor que

Exemplo Range Query

- Consultar o campo idade maior ou igual a 10

GET cliente/_search

```
{
  "query": {
    "range" : {
      "idade" : {
        "gte" : 10
      }
    }
  }
}
```

Exemplo Range Query

- Consultar o campo idade menor ou igual a 25

GET cliente/_search

```
{
  "query": {
    "range" : {
      "idade" : {
        "lte" : 25
      }
    }
  }
}
```

Exemplo Range Query

- Consultar o campo idade entre 10 e 25

GET cliente/_search

```
{  
  "query": {  
    "range" : {  
      "idade" : {  
        "gte" : 10,  
        "lte" : 25  
      }  
    }  
  }  
}
```



Consultas por intervalo de tempo



Opções Tempo

○ Propriedades com data

- "format": "dd/MM/yyyy | yyyy"
- "time_zone": "+03:00"

○ Exemplos

- Now: Agora
- +1d: Adiciona 1 dia
- -1M: Subtrai 1 mês

y	Anos
M	Meses
w	Semanas
d	Dias
H	Horas
h	Horas
m	Minutos
s	Segundos

Exemplo Range de data

- Intervalo com diferentes formatos

GET cliente/_search

```
{  
  "query": {  
    "range" : {  
      "data" : {  
        "gte": "01/01/2012",  
        "lte": "2013",  
        "format": "dd/MM/yyyy | yyyy"  
      }  
    }  
  }  
}
```


Exemplo Range de data

- Intervalo com diferentes formatos

GET cliente/_search

```
{  
  "query": {  
    "range" : {  
      "data" : {  
        "gte" : "now-1d",  
        "lt" : "now"  
      }  
    }  
  }  
}
```

Exemplo Range de data

- Intervalo com diferentes formatos

GET cliente/_search

```
{  
  "query": {  
    "range" : {  
      "timestamp" : {  
        "gte": "2015-01-01 00:00:00",  
        "lte": "now",  
        "time_zone": "+03:00"  
      }  
    }  
  }  
}
```

Exemplo Range de data

- Campo padrão de tempo - Timestamp

GET cliente/_search

```
{  
  "query":{  
    "range":{  
      "@timestamp":{  
        "gte":"2015-08-04T11:00:00",  
        "lt":"2015-08-04T12:00:00"  
      }  
    }  
  }  
}
```

Exercícios Consultas por Intervalo

1. Verificar se existe o índice populacao
2. Executar as consultas no índice populacao
 - a) Mostrar os documentos com o atributo "Total Population" menor que 100
 - b) Mostrar os documentos com o atributo "Median Age" maior que 70
 - c) Mostrar os documentos 50 (Zip Code: 90056) à 60 (Zip Code: 90067) do índice de populacao
3. Importar através do Kibana o arquivo weekly_MSFT.csv Pré-visualizar o documento (Guia Arquivos/dataset/weekly_MSFT.csv) com o índice bolsa
4. Executar as consultas no índice bolsa
 - a) Visualizar os documentos do dia 2019-01-01 à 2019-03-01. (hits = 9)
 - b) Visualizar os documentos do dia 2019-04-01 até agora. (hits = 3)



Semantix

Obrigado!

Alguma pergunta?



Você pode me encontrar em:
rodrigo.augusto@semantix.com.br

GET SMARTER