

# Mapping

Mapping é uma definição de schemas

Elasticsearch tem tipos defaults, mas as vezes eles precisam ser customizados

```
curl -XPUT 127.0.0.1:9200/movies -d '{
  "mappings": {
    "properties": {
      "year": {"type": "date"}
    }
  }
}'
```

# Mappings mais comuns

## field types

string, byte, short, integer, long,  
float, double, boolean, date

```
"properties": {  
  "user_id" : {  
    "type": "long"  
  }  
}
```

## field index

you want this field to be  
indexed for full-text search?

```
"properties": {  
  "genre" : {  
    "index": "not_analyzed"  
  }  
}
```

## field analyzer

Define filters, tokenizer, token  
filter

```
"properties": {  
  "description" : {  
    "analyzer": "english"  
  }  
}
```

# Mais sobre analyzers

- character filters
  - remover codificação HTML, converter & em e etc.
- tokenizer
  - dividir palavras em espaços, pontuação ou caracteres especiais
- token filter
  - letras minúsculas, stemming, sinónimos, stopwords

# Opções de Analyzers

- **Standard**

- divide em palavras, remove pontuação, transforma em letras minúsculas. Opção se o idioma é desconhecido.

- **Simple**

- divide em qualquer coisa que não seja uma letra e transforma em minúsculas

- **Whitespace**

- divide em espaços mas não transforma em minúsculas

- **Language** (ex. portuguese)

- trata de questões específicas do idioma, como stopwords e stemming