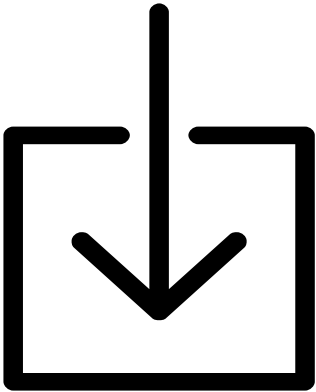


Você pode importar praticamente qualquer coisa



- Scripts podem enviar documentos em lote através de REST API
- [Logstash](#) e [Beats](#) podem ingerir dados de logs, S3, bancos de dados e outros.
- Sistemas AWS pode ingerir dados em streaming com [lambda](#) ou [kinesis firehose](#)
- [kafka](#), [spark](#), e outros tem plug-ins de integração com Elasticsearch

Script

- Lê dados de um Sistema Distribuído
- Transforma em Inserções JSON em lote (bulk)
- Envia via HTTP/REST para seu cluster Elasticsearch

```
1 import csv
2 import re
3
4 csvfile = open('ml-latest-small/movies.csv', 'r')
5
6 reader = csv.DictReader( csvfile )
7 for movie in reader:
8     print ("{ \"create\" : { \"_index\": \"movies\", \"_id\" : \"\" , movie['movieId'], \"\" } }", sep='')
9     title = re.sub(" \\.\\$)", "", re.sub("'", "", movie['title']))
10    year = movie['title'][-5:-1]
11    if (not year.isdigit()):
12        year = "2016"
13    genres = movie['genres'].split('|')
14    print ("{ \"id\": \"\", movie['movieId'], "\", \"title\": \"\", title, "\", \"year\":", year, ", \"genre\":[", end='', sep='
15    for genre in genres[:-1]:
16        print("\"", genre, "\",", end='', sep='')
17    print("\"", genres[-1], "\",", end = '', sep='')
18    print ("] }")
```

Script

```
import csv
import re

csvfile = open('ml-latest-small/movies.csv', 'r')

reader = csv.DictReader( csvfile )
for movie in reader:
    print ("{ \"create\" : { \"_index\": \"movies\", \"_id\" : \"\", movie['movieId'], \"\" } }, sep=")
    title = re.sub(" \\.*\$\"", "", re.sub("''", "", movie['title']))
    year = movie['title'][-5:-1]
    if (not year.isdigit()):
        year = "2016"
    genres = movie['genres'].split('|')
    print ("{ \"id\": \"\", movie['movieId'], "\", \"title\": \"\", title, "\", \"year\":", year, "\", \"genre\":[", end=", sep=")
    for genre in genres[:-1]:
        print("\\\"", genre, "\",", end=", sep=")
    print("\\\"", genres[-1], "\"\", end = ", sep=")
    print (" ] }")
```

