Storing Data

David Arroyo Menéndez

October 9, 2019

Basics

```
$ python3 for.py # arrays
$ python3 lists.py
$ python3 diccionarios.py
$ python3 conjuntos.py
```

Ficheros

```
$ python3 write.py
$ python3 list2file.py
$ python3 json-example.py
$ python3 items.py
$ python3 csv-example.py
$ python3 rss-example.py
```

Databases

```
$ python3 sqlite.py
$ python3 sqlite2.py
$ python3 mongo-tutorial.py
$ python3 mysql-example.py
```

JSON

JSON is a format file based in pairs key value, very used in the Javascript world.

JSON source (I)

```
python3 json2dict.py
python3 json2file.py
python3 json2html-example.py
python3 json2list.py
python3 json2pandas.py
python3 json-example.py
python3 jsonuk2.py
python3 jsonuk.py
python3 list2json.py
python3 mergejson.py
python3 object2json.py
python3 pandas2json.py
python3 tuple2json.py
```

JSON source (II)

```
python3 data3.py
python3 dict2json.py
python3 exer1-intf-table.py
python3 file2json.py
python3 genderapi.py
python3 lulu.py
python3 perceval2pandas.py
python3 printfield.py
python3 todos.py
python3 user2photo.py
```

Pandas

Pandas is a Python library for manage dataframes such as statistical software as GNU R.

Pandas (Source)

```
$ python pandas-example.py
$ python pandas-10min.py
$ python pandas-creating-dataframe-from-arrays.py
$ python pandas-creating-dataframe.py
$ python3 pandas-plot.py
$ python3 jsonpandas.py
```

Pytables

PyTables is a package for managing hierarchical datasets and designed to efficiently and easily cope with extremely large amounts of data.

```
$ python3 objecttre.py
```

NetCDF

NetCDF is a binary format commonly used in climatology, meteorology and oceanography applications (e.g., weather forecasting, climate change) and GIS applications.

NetCDF

```
$ python3 netcdf-example.py
$ python3 netcdf-example2.py
$ python3 netcdf-example3.py
$ python3 netcdf-example4.py
```

HDF5

HDF5 lets you store huge amounts of numerical data, and easily manipulate that data from NumPy.

HDF5

```
$ python3 h5_cmprss.py
$ python3 h5_crtgrpar.py
$ python3 h5_crtgrpd.py
$ python3 h5_crtgrp.py
$ python3 h5_rdwt.py
```

RDF

RDFLib

```
python3 literal.py
python3 merging.py
python3 n3.py
python3 namespace.py
python3 navigating.py
python3 persistence.py
python3 rdflib-example2.py
python3 rdflib-example.py
python3 rdflib-statements.py
python3 serialize-to-n3.py
python3 sparql_query_example.py
python3 urirefs.py
```

Elasticsearch: Introduction

This document is only my personal notes about this software.
Elasticsearch is a highly scalable open-source full-text search and analytics engine. It allows you to store, search, and analyze big volumes of data quickly and in near real time. It is generally used as the underlying engine/technology that powers applications that have complex search features and requirements.

Elasticsearch: Installation

https://www.elastic.co/guide/en/elasticsearch/reference/current/_installation.html

Elasticsearch: Checking

```
http://localhost:9200/
```

Elasticsearch: List indexes

```
http://localhost:9200/_cat/indices
```

Elasticsearch: List indexes in json

```
$ curl 'http://localhost:9200/_cat/indices?pretty' -H "Accept:
```

Elasticsearch: Delete index

```
$ curl -XDELETE localhost:9200/commits
```

Elasticsearch: Show an index

```
$ wget http://localhost:9200/dam-index/
```

Elasticsearch: Searching in an index

```
$ wget -c http://localhost:9200/dam-index/_search
```

Elasticsearch: The cluster state API allows to get a comprehensive state information of the whole cluster.

```
$ wget http://localhost:9200/_cluster/state
```

Elasticsearch: Master

```
http://localhost:9200/_cat/master?v
```

Elasticsearch: Show nodes

```
$ wget -c http://localhost:9200/_cat/nodes?h=ip,port,heapPercer
$ curl http://localhost:9200/_cat/nodes?v
```

Elasticsearch: Show cluster health

```
http://localhost:9200/_cluster/health
```

Elasticsearch: Elasticdump

json to the index

```
$ elasticdump --input=git_openstack_data.json --output=http://l
$ elasticdump --input=git_openstack_mapping.json --output=http:
```

index data to json

```
$ elasticdump  --input=http://production.es.com:9200/my_index -
$ elasticdump  --input=http://production.es.com:9200/my_index -
```

License