



Proyecto 1mer Bimestre Animales



Objetivos generales

Diseñar un Data warehouse en el cual confluyan varias fuentes de información estructurada, no estructurada y semi estructurada. Esta parte del proyecto pertenece a las etapa de “preparar y explorar datos”.

Objetivos Específicos

- Recopilar tweets acerca del animales con los subtemas de perros, gatos y cerdos.
- Importar datos de mysql a couchdb.
- Exportar datos de un archivo cvs a couchdb a través de python.
- Preparar y exportar datos a couchdb.
- Visualizar los datos a través de una interfaz en HTML.



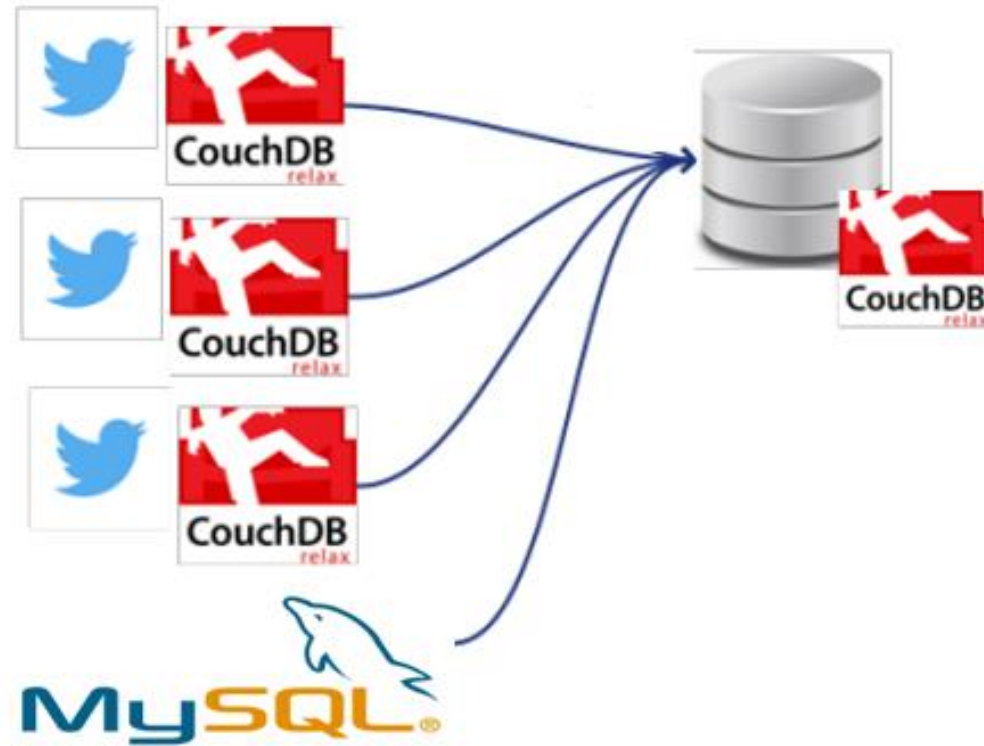
Recursos y Herramientas

- XAMPP
- Python 3.5
- Notepad++
- CouchDb
- MySql



Arquitectura

- El tema principal es Animales de los cuales existe 3 bases de datos que se accederá remotamente a couchdb principal. Las 3 bases son los subtemas que son perros, gatos, cerdos. Mientras que la base MySQL se exporta a un archivo csv y se la ingresa al couchdb principal.



Extracción de datos de base de datos MySQL

The screenshot shows the phpMyAdmin web interface. The browser address bar displays the URL: `localhost/phpmyadmin/sql.php?server=1&db=basededatos&table=catsvdogs&pos=0`. The interface includes a sidebar on the left with a database tree structure showing 'basededatos' and its sub-databases, including 'catsvdogs'. The main panel shows the 'Examinar' (Browse) tab selected. A message at the top indicates that the current selection does not contain a unique column, affecting grid editing and linking. Below this, a green status bar confirms that 25 rows (0-24 of 50 total) are displayed. The SQL query `SELECT * FROM `catsvdogs`` is shown in the query editor. Below the query, there are controls for pagination (page 1 of 2), a 'Mostrar todo' (Show all) checkbox, a 'Número de filas' (Number of rows) dropdown set to 25, and a 'Filtrar filas' (Filter rows) input field with the placeholder 'Buscar en esta tabla'. The query results are displayed in a table with 12 columns (COL 1 to COL 12) and 4 rows of data (Alabama, Arizona, Arkansas, California).

La selección actual no contiene una columna única. La edición de la grilla y los enlaces de copiado, eliminación y edición no es posible.

Mostrando filas 0 - 24 (total de 50, La consulta tardó 0,0000 segundos.)

```
SELECT * FROM `catsvdogs`
```

☐ Perfilando [Editar en línea] [Ejecutar]

1 > >> | ☐ Mostrar todo | Número de filas: 25 | Filtrar filas:

+ Opciones

COL 1	COL 2	COL 3	COL 4	COL 5	COL 6	COL 7	COL 8	COL 9	COL 10	COL 11	COL 12
Alabama	1.828	59,5	1.088	44,1	807	1,7	1.410	27,4	501	2,5	1.252
Arizona	2.515	59,5	1.497	40,1	1.008	1,8	1.798	29,6	743	1,9	1.438
Arkansas	1148	62,4	716	47,9	550	2,0	1.097	30,6	351	2,3	810
California	12.974	52,9	6.865	32,8	4.260	1,6	6.687	28,3	3.687	1,9	7.118

Exportar la base a un archivo CSV

localhost/phpmyadmin/tbl_export.php?db=basededatos&table=catsvdogs&single_table=true

phpMyAdmin

Reciente Favoritas

Nueva
basededatos
Nueva
catsvdogs
information_schema
mysql
performance_schema
phpmyadmin
test

Servidor: 127.0.0.1 » Base de datos: basededatos » Tabla: catsvdogs

Examinar Estructura SQL Buscar Insertar Exportar Importar Privilegios

Exportando filas de la tabla "catsvdogs"

Exportar plantillas:

Nueva plantilla:

Nombre de plantilla

Plantillas existentes:

Plantilla: -- Seleccionar plantilla --

Método de exportación:

☒ Rápido - mostrar sólo el mínimo de opciones de configuración

☐ Personalizado - mostrar todas las opciones de configuración posibles

Formato:

CSV

Exportar la base a un archivo CSV

localhost/phpmyadmin/tbl_export.php?db=basededatos&table=catsvdogs&single_table=true

phpMyAdmin

Reciente Favoritas

Nueva
basededatos
Nueva
catsvdogs
information_schema
mysql
performance_schema
phpmyadmin
test

Servidor: 127.0.0.1 » Base de datos: basededatos » Tabla: catsvdogs

Examinar Estructura SQL Buscar Insertar Exportar Importar Privilegios

Exportando filas de la tabla "catsvdogs"

Exportar plantillas:

Nueva plantilla:

Nombre de plantilla

Plantillas existentes:

Plantilla: -- Seleccionar plantilla --

Método de exportación:

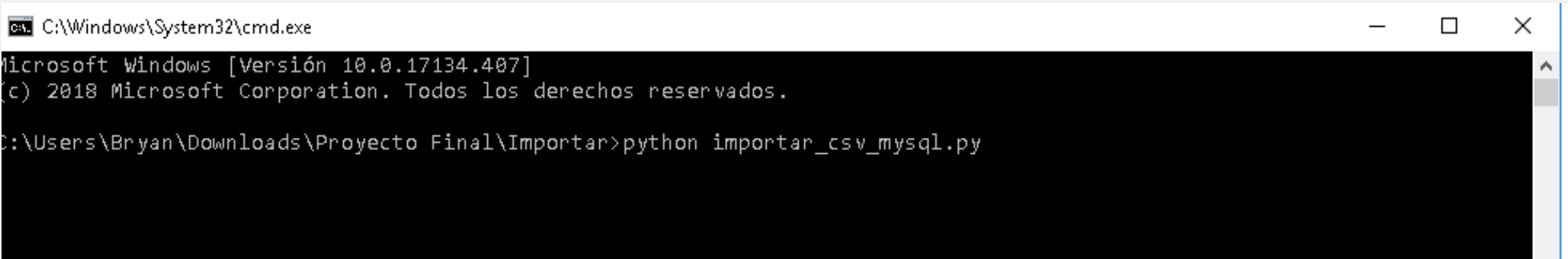
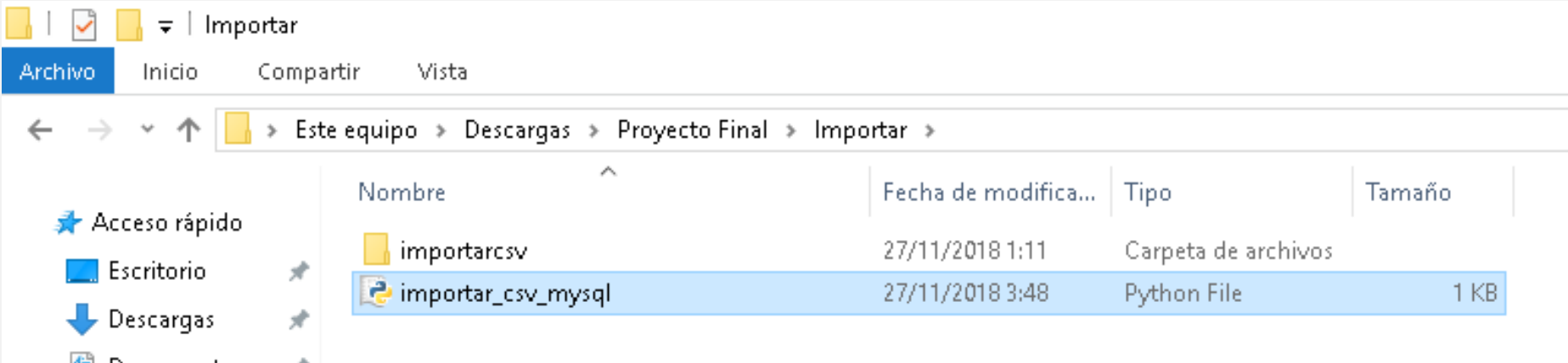
☒ Rápido - mostrar sólo el mínimo de opciones de configuración

☐ Personalizado - mostrar todas las opciones de configuración posibles

Formato:

CSV

Ejecutar el script Python para ingresar datos a couchdb



importarcsv

25.3 KB

52

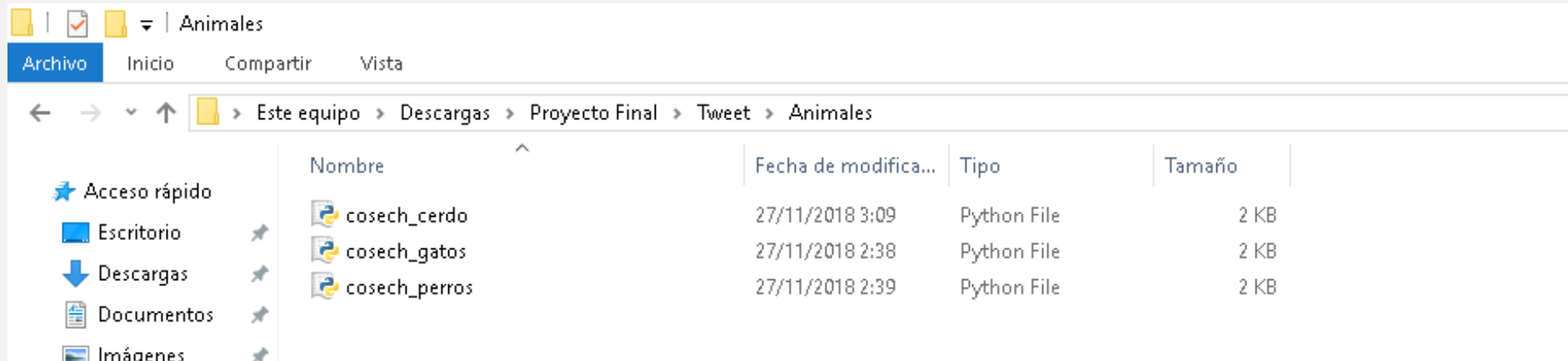


Código Python

```
C:\Users\Bryan\Downloads\Proyecto Final\Importar\importar_csv_mysql.py - Notepad++
Archivo  Editar  Buscar  Vista  Codificación  Lenguaje  Configuración  Herramientas  Macro  Ejecutar  Plugins  Ventana  ?
importar_csv_mysql.py x  gatos.html x  estadisticagatos.html x

1  import couchdb
2  import csv
3  server = couchdb.Server('http://user:user@localhost:5984/')
4  try:
5      #Si no existe la Base de datos la crea
6      db = server.create('importarcsv')
7  except:
8      #Caso contrario solo conectarse a la base existente
9      db = server['importarcsv']
10 doc=open('importarcsv\catsvdogs.csv','r')
11 reader = csv.reader(doc)
12 for row in reader:
13     ingreso = db.save({'Localization' : row[0], 'NumberofHouseholds(in1000)' : row[1], 'Percentageofhouseholdswithpets' : row[2],
14     'NumberofPetHouseholds(in1000)' : row[3], 'PercentageofDogOwners' : row[4], 'DogOwningHouseholds(1000s)' : row[5], 'MeanNumberofDogsperhousehold' : row[6],
15     'PercentageofCatOwners' : row[8], 'CatOwningHouseholds' : row[9], 'MeanNumberofCats' : row[10], 'CatPopulation' : row[11]})
16 doc.close
```

Exportar datos Tweeter de perros, gatos y cerdos



```
C:\Windows\System32\cmd.exe
Microsoft Windows [Versión 10.0.17134.407]
(c) 2018 Microsoft Corporation. Todos los derechos reservados.

C:\Users\Bryan\Downloads\Proyecto Final\Tweet\Animales>python cosech_cerdo.py
```

```
C:\Windows\System32\cmd.exe
Microsoft Windows [Versión 10.0.17134.407]
(c) 2018 Microsoft Corporation. Todos los derechos reservados.

C:\Users\Bryan\Downloads\Proyecto Final\Tweet\Animales>python cosech_gatos.py
```

```
C:\Windows\System32\cmd.exe
Microsoft Windows [Versión 10.0.17134.407]
(c) 2018 Microsoft Corporation. Todos los derechos reservados.

C:\Users\Bryan\Downloads\Proyecto Final\Tweet\Animales>python cosech_perros.py_
```

Código Python

```
import couchdb #Libreria de CouchDB (requiere ser instalada primero)
from tweepy import Stream #tweepy es la libreria que trae tweets desde la API de Twitter (requiere ser instalada primero)
from tweepy import OAuthHandler
from tweepy.streaming import StreamListener
import json #Libreria para manejar archivos JSON

###Credenciales de la cuenta de Twitter#####
#Poner aqui las credenciales de su cuenta privada, caso contrario la API bloqueara esta cuenta de ejemplo
ckey = "m4Fq2Pr4yHn1YLLg6nmPYxXYZ"
csecret = "OWC0Z1D9sT4aMiBY5xDrRjFIQqT3KbU8oSaNEsFEkKPHZCASe4"
atoken = "999027411613356032-NvGF9YveYjVjQq4sf61x5IbFDeOKBej"
asecret = "ZHTEn2rxBoKLkFa57Ksm3Hs4jYtwimhgYcsq8TtAXVXv"
#####

class listener(StreamListener):

    def on_data(self, data):
        dictTweet = json.loads(data)
        try:
            dictTweet["_id"] = str(dictTweet['id'])
            #Antes de guardar el documento puedes realizar parseo, limpieza y cierto analisis o filtrado de datos previo
            #a guardar en documento en la base de datos
            doc = db.save(dictTweet) #Aqui se guarda el tweet en la base de couchDB
            print ("Guardado " + "> " + dictTweet["_id"])
        except:
            print ("Documento ya existe")
            pass
        return True

    def on_error(self, status):
        print (status)

auth = OAuthHandler(ckey, csecret)
auth.set_access_token(atoken, asecret)
twitterStream = Stream(auth, listener())

#Setear la URL del servidor de couchDB
server = couchdb.Server('http://user:user@localhost:5984/')
try:
    #Si no existe la Base de datos la crea
    db = server.create('harvestcerdo')
except:
    #Caso contrario solo conectarse a la base existente
    db = server['harvestcerdo']

#Aqui se define el bounding box con los limites geograficos donde recolectar los tweets
twitterStream.filter(track=['serpiente','serpientes','animales'])
#twitterStream.filter(locations=[-78.586922,-0.395161,-78.274155,0.021973])
```

Realizar una plantilla para mostrar todas las tablas de la base de datos.

[Home](#) [Perros](#) [Gatos](#) [Cerdos](#) [Estadística](#)

Arquitectura



Proyecto Base de Datos Multidimencional



Bryan España y Lizeth Toasa

Diseñar un Data warehouse en el cual confluyan varias fuentes de información estructurada, no estructurada y semi estructurada. Esta parte del proyecto pertenece a las etapa de "preparar y explorar datos" El repositorio unificado de datos será de tipo NoSQL, las fuentes de información pueden ser diversas, por ejemplo csv's, RDBMS, NoSQL, etc.



Gracias

Realizado por: Bryan España
Lizeth Toasa