

EXTRACCIÓN Y OBTENCIÓN DE DATOS

IVONNE VIRGINIA VEGA PEDRAZA
Ivonne.vega01@epn.edu.ec

JOSE GEOVANNY DIAS AMAGUAÑA
Jose.dias@epn.edu.ec

ESCUELA POLITÉCNICA NACIONAL ESCUELA DE FORMACIÓN DE TECNÓLOGOS

RESUMEN— En el presente documento se mostrará el diseño de un Data Warehouse en el cual convergen diferentes fuentes de información como Tableau, Twitter y Data.gov

I. INTRODUCCIÓN

Un data warehouse es un repositorio unificado para todos los datos que recogen los diversos sistemas de una empresa. El repositorio puede ser físico o lógico y hace hincapié en la captura de datos de diversas fuentes sobre todo para fines analíticos y de acceso. [1]

Tableau es la plataforma de análisis integral más eficaz, segura y flexible para los datos. Además es la única plataforma de inteligencia de negocios que convierte los datos en información útil. [2]

Twitter es un servicio de comunicación bidireccional con el que puedes compartir información de diverso tipo de una forma rápida, sencilla y gratuita. [3]

Data.gov es un sitio web del gobierno de los EE.UU, apunta a mejorar el acceso público a conjuntos de datos de alto valor, legibles por máquina, generados por el Poder Ejecutivo del Gobierno Federal. El sitio es un depósito de información del gobierno federal, estatal, local y tribal, disponible para el público. [3]

II. OBJETIVOS

- Obtener información de tres fuentes distintas.
- Unificar la información de las diferentes fuentes.
- Visualizar en una interfaz web la información obtenida.

III. DESARROLLO

Se han seleccionado 3 temas para la recolección de datos:

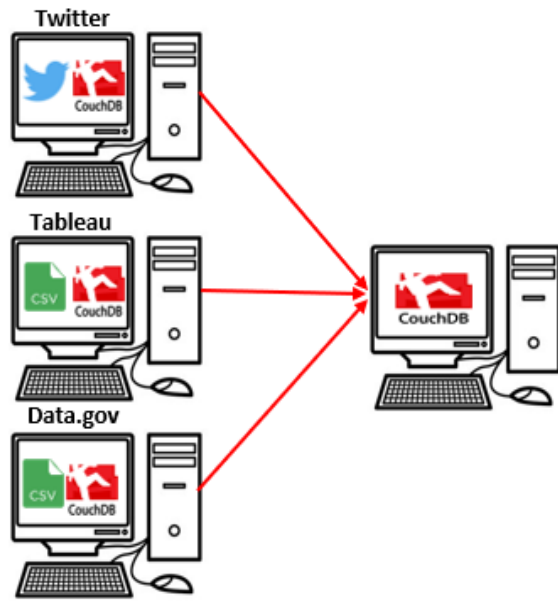
1. Mascotas.
2. Entretenimiento.
3. Tecnología.

Para realizar este proyecto se utilizó los siguientes recursos.

1. CouchDB 2.2.
2. Python 5.0
3. USBWebserver v8.6 (Portable).
4. Editores de texto (Brackets).
5. Navegador Web



A) ARQUITECTURA



B) Recolección de información de Twitter.

Para la recolección de datos de Twitter se utilizó el siguiente script:

```

1.cosecha - mascotas.py - C:\Users\PERSONAL\Desktop\eto\BD_MULT\proyecto bl\1.cose...
File Edit Format Run Options Window Help

import couchdb #Libreria de CouchDB (requiere ser instalada primero)
from tweepy import Stream #tweepy es la libreria que trae tweets desde la API de
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 bloquea
ckey = "unksC6a2wXM5JoZ5cFwqx8D8F6"
csecret = "EJ1qOoTUSQIFNLQpTgJnDFk6IJsjpt7EzeD7sgBGOG143HCWm0"
atoken = "999027538243588098-gAoMQKaN0Y5vHB3g5Q9558jXCM5gQ"
asecret = "UoQU3N0S1bSMRHBekjYjGleBjuwaqPUpnKYMfspiQ26jN"
#####
i = 1

class listener(StreamListener):
    def on_data(self, data):
        global i
        dictTweet = json.loads(data)
        try:
            dictTweet["_id"] = str(dictTweet['id'])
            #Antes de guardar el documento puedes realizar parseo, limpieza y ci
            #a guardar en documento en la base de datos
            doc = db.save(dictTweet) #Aqui se guarda el tweet en la base de couc
            print (str(i) + " Guardado " + "=" + dictTweet["_id"])
            i += 1
            if i > 500:
                print("Datos completos satisfactoriamente...")
                #break
        except:
            print (str(i) + "Documento ya existe")
            pass
        return True

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

auth = OAuthHandler(ckey, csecret)

```

En el cual tenemos que setear la url del servidor de couchDB:

```
server = couchdb.Server('http://ivonne:ivonne@localhost:5984/')
```

Crear la base de datos:

```
db = server.create('mascotas_1')
```

Definir el filtro por el cual se quiere buscar los tweets:

```
twitterStream.filter(track=["mimascota","perros","gatos","pets","dogs","cats"])
```

Al momento de ejecutar los scripts se pudo observar como se iba guardando los datos en couchDB.

```

Python 3.5.0 Shell
File Edit Shell Debug Options Window Help
Python 3.5.0 (v3.5.0:374f501f4567, Sep 13 2015, 02:27:37) [MSC v.1900 64 bit (AMD64)] on win32
Type "copyright", "credits" or "license()" for more information.
>>>
RESTART: C:\Users\PERSONAL\Desktop\eto\BD_MULT\proyecto bl\3.cosecha - tecnolo
gia.py
1 Guardado => 1067448120735395846
2 Guardado => 1067448120420782080
3 Guardado => 1067448120819236864
4 Guardado => 1067448120945070080
5 Guardado => 1067448121414819840
6 Guardado => 1067448121104494592
7 Guardado => 1067448121502851072
8 Guardado => 1067448121553289217
9 Guardado => 1067448121154768996
10 Guardado => 1067448122048036864
11 Guardado => 1067448122270351361
12 Guardado => 1067448121691701254
13 Guardado => 1067448122236907520
14 Guardado => 106744812232750080
15 Guardado => 1067448122723508224
16 Guardado => 1067448122744455168
17 Guardado => 1067448122807396112
18 Guardado => 1067448122702553088
19 Guardado => 1067448123285536769
20 Guardado => 1067448123176423427
21 Guardado => 1067448123557965824
22 Guardado => 1067448122698280960
23 Guardado => 1067448124262764544
24 Guardado => 1067448124413812736
25 Guardado => 1067448125923692546
26 Guardado => 1067448126141882368
27 Guardado => 1067448126213042176
28 Guardado => 1067448126477357057
29 Guardado => 1067448126456455168
30 Guardado => 1067448123104972800
31 Guardado => 1067448127052021760
32 Guardado => 1067448126758404096
33 Guardado => 1067448127513333760

```

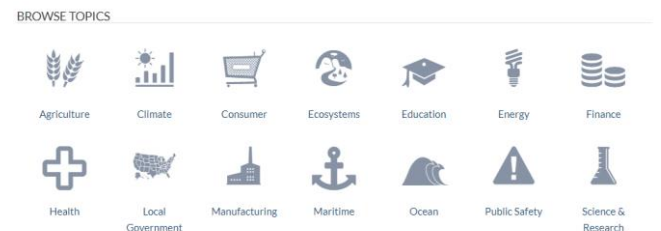
entretenimiento_1	1.5 MB	569
mascotas_1	1.5 MB	547
tecnologia_1	1.6 MB	571

C) Recolección de datos de Tableau.

Para obtener los datos de Tableau se ingreso a la siguiente página: <https://www.tableau.com/es-es>, en la cual se procedió a descargar los archivos Excel.

D) Recolección de datos en Data.gov

Para obtener los datos de Data.gov se ingresó a la siguiente página: <https://www.data.gov/> en la cual se puede escoger datos entre diversos temas, dichos documentos pueden ser archivos CSV o Excel.



- E) Transformar archivos Excel a CSV y guardar en CouchDB con formato JSON.

Primero tenemos que transformar los archivos de Excel en CSV. Simplemente al archivo *.xlsx lo guardamos como CSV (delimitados por comas).

Nombre de archivo: catsvdogs
Tipo: CSV (delimitado por comas)

Se nos guarda el archivo en *.csv con el cual podremos ingresar a nuestra base de datos NoSQL.

catsvdogs	27/11/2018 12:29	Archivo de valores...
catsvdogs	25/11/2018 19:55	Hoja de cálculo d...

- F) Ingreso de datos de archivo CSV a CouchDB.

Para realizar este proceso hemos escrito un pequeño Script en lenguaje Python.

```

cvstojason.py - D:\Universidad\06. Sexto Semestre\BDD Multidimensionales\Proyectos\Proye...
File Edit Format Run Options Window Help

import csv, json
import couchdb

csvFilePath = "catsvdogs.csv"
jsonFilePath = "pruebal.json"

db = None
#=====
#Conexión CouchDB
#=====
url = "http://geo:geo@localhost:5984/"
db = "mascotas_2"
couch = couchdb.Server(url)

if db in couch:
    db = couch['mascotas_2']
    print("Existing Data Base")
else:
    db = couch.create('mascotas_2')
    print("Creating Data Base")

#=====
#=====
data = {}
#Leer archivo CSV y agregar datos al Diccionario
with open(csvFilePath) as csvFile:
    csvReader = csv.DictReader(csvFile)
    for csvRow in csvReader:
        print("llega")
        print(json.dumps(csvRow)) #Solo agregar a CouchDB // db.save(csvRow)
        db.save(csvRow)
        #nombre = csvRow['nombre']
        #data[nombre] = csvRow

#print("")
#print (data)

#Escribir datos del diccionario a un archivo JSON
#with open(jsonFilePath, "w") as jsonFile:

```

En el cual importamos las librerías necesarias para poder ejecutar el código y no nos marque error.

```
import csv, json
import couchdb
```

Se ingresa el nombre del archivo con la extensión *.csv, el cual contiene el dataset que se guardará en la base de datos.

```
csvFilePath = "catsvdogs.csv"
```

Realizamos la conexión a CouchDB y creamos la base de datos con el nombre seleccionado.

```
url = "http://geo:geo@localhost:5984/"
db = "mascotas_2"
couch = couchdb.Server(url)
```

```
if db in couch:
    db = couch['mascotas_2']
    print("Existing Data Base")
else:
    db = couch.create('mascotas_2')
    print("Creating Data Base")
```

El Script leerá cada línea del documento CSV en formato JSON e irá guardando automáticamente en CouchDB mediante un ciclo "For".

```

#Leer archivo CSV y agregar datos al Diccionario
with open(csvFilePath) as csvFile:
    csvReader = csv.DictReader(csvFile)
    for csvRow in csvReader:
        print("llega")
        print(json.dumps(csvRow)) #Solo agregar a CouchDB // db.save(csvRow)
        db.save(csvRow)

```

Ahora podemos observar cómo se crea la base de datos y como se van guardando los datos:

entretenimiento_1	1.5 MB	569
entretenimiento_2	175.6 KB	235
entretenimiento_3	68.5 KB	194
mascotas_1	1.5 MB	547
mascotas_2	24.4 KB	50
mascotas_3	304.2 KB	281
tecnologia_1	1.6 MB	571
tecnologia_2	491.9 KB	1898
tecnologia_3	2.2 MB	5899

Fuente: Tableau

Tema: Mascotas

```
Python 3.5.0 Shell
File Edit Shell Debug Options Window Help
, "Number of Households (in 1000)": "2.515", "Percentage of Dog Owners": "40.1",
"Location": "Arizona", "Percentage of households with pets": "59.5", "Mean Num
ber of Cats": "1.9", "Cat Owning Households": "743",
{"": "", "Dog Population (in 1000)": "1.097", "Percentage of Cat Owners": "30.6",
, "Number of Pet Households (in 1000)": "716", "Cat Population": "810", "Mean Nu
mber of Dogs per household": "2.0", "Dog Owning Households (1000s)": "550", "Num
ber of Households (in 1000)": "1148", "Percentage of Dog Owners": "47.9", "Locat
ion": "Arkansas", "Percentage of households with pets": "62.4", "Mean Number of
Cats": "2.3", "Cat Owning Households": "351",
{"": "", "Dog Population (in 1000)": "6.687", "Percentage of Cat Owners": "28.3",
, "Number of Pet Households (in 1000)": "6.865", "Cat Population": "7.118", "Mea
n Number of Dogs per household": "1.6", "Dog Owning Households (1000s)": "4.260",
, "Number of Households (in 1000)": "12.974", "Percentage of Dog Owners": "32.8",
, "Location": "California", "Percentage of households with pets": "52.9", "Mean
Number of Cats": "1.9", "Cat Owning Households": "3.687",
{"": "", "Dog Population (in 1000)": "1.349", "Percentage of Cat Owners": "32.3",
, "Number of Pet Households (in 1000)": "1.217", "Cat Population": "1191", "Mean
Number of Dogs per household": "1.6", "Dog Owning Households (1000s)": "845", "
Number of Households (in 1000)": "1.986", "Percentage of Dog Owners": "42.5", "L
ocation": "Colorado", "Percentage of households with pets": "61.3", "Mean Number
of Cats": "1.9", "Cat Owning Households": "642",
{"": "", "Dog Population (in 1000)": "507", "Percentage of Cat Owners": "31.9",
, "Number of Pet Households (in 1000)": "728", "Cat Population": "796", "Mean Num
ber of Dogs per household": "1.3", "Dog Owning Households (1000s)": "379", "Numbe
r of Households (in 1000)": "1.337", "Percentage of Dog Owners": "28.3", "Locati
on": "Connecticut", "Percentage of households with pets": "54.4", "Mean Number o
f Cats": "1.9", "Cat Owning Households": "427",
{"": "", "Dog Population (in 1000)": "1.63", "Percentage of Cat Owners": "33.7",
, "Number of Pet Households (in 1000)": "188", "Cat Population": "187", "Mean Num
ber of Dogs per household": "1.4", "Dog Owning Households (1000s)": "113", "Numbe
r of Households (in 1000)": "334", "Percentage of Dog Owners": "33.7", "Location
": "Delaware", "Percentage of households with pets": "56.6", "Mean Number of Cat
s": "1.7", "Cat Owning Households": "113",
{"": "", "Dog Population (in 1000)": "42", "Percentage of Cat Owners": "31.6", "
Number of Pet Households (in 1000)": "63", "Cat Population": "38", "Mean Number
of Dogs per household": "1.1", "Dog Owning Households (1000s)": "38", "Number o
f Households (in 1000)": "287", "Percentage of Dog Owners": "13.1", "Location": "
District of Columbia", "Percentage of households with pets": "21.9", "Mean Numbe
r of Cats": "1.9", "Cat Owning Households": "33",

```

Tema: Entretenimiento

```
Python 3.5.0 Shell
File Edit Shell Debug Options Window Help
, "Country": "Poland", "key": "7", "Points": "90", "Song.Quality": "3.860.999.9
01", "Normalized.Points": "0.059681698", "Home.Away.Country": "Away", "Semi.Fina
l.Number": "NA", "Happiness": "4.999.230.765", "duration": "18.278.354", "Artist
": "Ich Troje", "Is.Final": "1", "valence": "0.13689471", "timeasignature": "4",
, "danceability": "0.491622961", "Song": "Keine Grenzenadnych granic", "Home.Awa
y.Region": "Home", "liveness": "0.110196832", "u00efu00bbu00bfCod": "168", "Y
ear": "2003", "energy": "0.126677225", "acousticness": "0.690863124", "mode": "1
", "Song.In.English": "0",
llega
{"Group.Solo": "NA", "speechiness": "0.029959432", "tempo": "132.062", "Region":
"Former Socialist Bloc", "loudness": "4.963", "Artist.gender": "NA", "Place": "
24", "Country": "Latvia", "key": "0", "Points": "5", "Song.Quality": "0.15704788
6", "Normalized.Points": "0.00331565", "Home.Away.Country": "Home", "Semi.Final.
Number": "NA", "Happiness": "5.556.521.738", "duration": "17.846.667", "Artist":
"F.L.Y.", "Is.Final": "1", "valence": "0.415208033", "timeasignature": "4", "da
nceability": "0.549474469", "Song": "Hello From Mars", "Home.Away.Region": "Home
", "liveness": "0.161618276", "u00efu00bbu00bfCod": "167", "Year": "2003", "e
nergy": "0.712074212", "acousticness": "0.02490548", "mode": "1", "Song.In.Engli
sh": "1",
llega
{"Group.Solo": "NA", "speechiness": "0.030357394", "tempo": "85.073", "Region":
"Western Europe", "loudness": "6.722", "Artist.gender": "NA", "Place": "2", "Cou
ntry": "Belgium", "key": "7", "Points": "165", "Song.Quality": "6.253.894.443",
, "Normalized.Points": "0.105416446", "Home.Away.Country": "Away", "Semi.Final.Num
ber": "NA", "Happiness": "4.954", "duration": "24.789.907", "Artist": "Urban Tra
d", "Is.Final": "1", "valence": "0.700236428", "timeasignature": "4", "danceabi
lity": "0.557138512", "Song": "Sanomi", "Home.Away.Region": "Away", "liveness":
"0.158820038", "u00efu00bbu00bfCod": "168", "Year": "2003", "energy": "0.80991
3211", "acousticness": "0.098444229", "mode": "0", "Song.In.English": "0",
llega
{"Group.Solo": "NA", "speechiness": "NA", "tempo": "NA", "Region": "Former Socia
list Bloc", "loudness": "NA", "Artist.gender": "NA", "Place": "21", "Country": "
Estonia", "key": "NA", "Points": "14", "Song.Quality": "0.502635996", "Normaliz
e d.Points": "0.00928382", "Home.Away.Country": "Away", "Semi.Final.Number": "NA",
, "Happiness": "549.114.094", "duration": "NA", "Artist": "Ruffus", "Is.Final":
"1", "valence": "NA", "timeasignature": "NA", "danceability": "NA", "Song": "Eigh
ties Coming Back", "Home.Away.Region": "Home", "liveness": "NA", "u00efu00bbu
00bfCod": "169", "Year": "2003", "energy": "NA", "acousticness": "NA", "mode": "
NA", "Song.In.English": "1",
>>> |
Ln: 54 Col: 4

```

Tema: Tecnología

```
Python 3.5.0 Shell
File Edit Shell Debug Options Window Help
{"Date": "", "Percent of Usage": "0.14", "Mobile Operating System": "Sony Ericss
on"}
llega
{"Date": "", "Percent of Usage": "0.27", "Mobile Operating System": "Linux"}
llega
{"Date": "", "Percent of Usage": "0.1", "Mobile Operating System": "bada"}
llega
{"Date": "", "Percent of Usage": "0.74", "Mobile Operating System": "Nokia Unkno
wn"}
llega
{"Date": "", "Percent of Usage": "0.12", "Mobile Operating System": "LG"}
llega
{"Date": "", "Percent of Usage": "0.04", "Mobile Operating System": "Playstation
"}
llega
{"Date": "", "Percent of Usage": "0", "Mobile Operating System": "WinCE"}
llega
{"Date": "", "Percent of Usage": "0.02", "Mobile Operating System": "Nintendo 3D
S"}
llega
{"Date": "", "Percent of Usage": "0", "Mobile Operating System": "Nintendo"}
llega
{"Date": "", "Percent of Usage": "0", "Mobile Operating System": "webOS"}
llega
{"Date": "", "Percent of Usage": "0", "Mobile Operating System": "WinXP"}
llega
{"Date": "", "Percent of Usage": "0.01", "Mobile Operating System": "Brew"}
llega
{"Date": "", "Percent of Usage": "0", "Mobile Operating System": "Motorola"}
llega
{"Date": "", "Percent of Usage": "0", "Mobile Operating System": "Win7"}
llega
{"Date": "", "Percent of Usage": "0.01", "Mobile Operating System": "MeeGo"}
llega
{"Date": "", "Percent of Usage": "0", "Mobile Operating System": "Maemo 5"}
llega
{"Date": "", "Percent of Usage": "0.02", "Mobile Operating System": "WinVista"}
llega
{"Date": "", "Percent of Usage": "0", "Mobile Operating System": "Other"}
>>> |

```

Fuente: Data.gov

Tema: Mascotas

```
Python 3.5.0 Shell
File Edit Shell Debug Options Window Help
{"imound.no": "K18-113887", "image_alt_text": "Image Copyright HLP Inc. 2018",
"City": "", "animal_type": "Cat", "Zip": "", "Animal Color": "Unknown", "jurisdi
ction": "JURISDICTION", "obfuscated_longitude": "", "Record Type": "LOST", "Anim
al Breed": "Domestic Shorthair", "Date Type": "Date Lost", "Date": "11/18/2018 1
2:00:00 AM", "Animal Name": "Howard", "Age": "", "Obfuscated_Address": "S 107TH
E DUTHIE HILL RD", "Animal Gender": "Unaltered", "Current Location": "King Count
y Pet Adoption Center 21615 64TH AVE S KENT, WA 98032", "Memo": "Received on: 20
18-11-21<p/> Found Near: 26400 SE DUTHIE HILL RD, SAMMAMISH WA 98075<p/> Descrip
tion: Tan Unaltered Lizard Lizard<p/> Age: NO AGE<p/> Current Location: King Cou
nty Pet Adoption Center 21615 64TH AVE S KENT, WA 98032<p/>", "Animal ID": "A575
123", "Image": "https://petharbor.com/get_image.asp?RES=Detail&LOCATION=KING&ID=
A575123", "State": "WA", "Link": "https://petharbor.com/pet.asp?uid=KING.A57512
3", "obfuscated_latitude": "", "Temperament": "", "Data_Source": "Regional Anima
l Services of King County"}
{"imound.no": "K18-113887", "image_alt_text": "Image Copyright HLP Inc. 2018",
"City": "", "animal_type": "Cat", "Zip": "", "Animal Color": "Unknown", "jurisdi
ction": "JURISDICTION", "obfuscated_longitude": "", "Record Type": "LOST", "Anim
al Breed": "Domestic Shorthair", "Date Type": "Date Lost", "Date": "11/18/2018 1
2:00:00 AM", "Animal Name": "Howard", "Age": "", "Obfuscated_Address": "S 107TH
WA 98092<p/> Description: Gray / White Spayed Female Domestic Shorthair Cat<p/>
Age: 4 MONTHS<p/> Current Location: King County Pet Adoption Center 21615 64TH
AVE S KENT, WA 98032<p/>", "Animal ID": "A570611", "Image": "https://petharbor.c
om/get_image.asp?RES=Detail&LOCATION=KING&ID=A570611", "State": "WA", "Link": "h
ttps://petharbor.com/pet.asp?uid=KING.A570611", "obfuscated_latitude": "", "Tem
perament": "BLUE", "Data_Source": "Regional Animal Services of King County"}
{"imound.no": "K18-112646", "image_alt_text": "Image Copyright HLP Inc. 2018",
"City": "AUBURN", "animal_type": "Cat", "Zip": "98092", "Animal Color": "Gray /
White", "jurisdiction": "KING COUNTY", "obfuscated_longitude": "", "Record Type"
: "FOUND", "Animal Breed": "Domestic Shorthair", "Date Type": "Received on", "Da
te": "10/03/2018 12:00:00 AM", "Animal Name": "Buffy", "Age": "4 MONTHS", "Obfus
cated_Address": "14800 SE LAKE HOLM RD", "Animal Gender": "Spayed Female", "Curr
ent Location": "King County Pet Adoption Center 21615 64TH AVE S KENT, WA 98032",
, "Memo": "Received on: 2018-10-03<p/> Found Near: 14800 SE LAKE HOLM RD, AUBURN
WA 98092<p/> Description: Gray / White Spayed Female Domestic Shorthair Cat<p/>
Age: 4 MONTHS<p/> Current Location: King County Pet Adoption Center 21615 64TH
AVE S KENT, WA 98032<p/>", "Animal ID": "A570611", "Image": "https://petharbor.c
om/get_image.asp?RES=Detail&LOCATION=KING&ID=A570611", "State": "WA", "Link": "h
ttps://petharbor.com/pet.asp?uid=KING.A570611", "obfuscated_latitude": "", "Tem
perament": "BLUE", "Data_Source": "Regional Animal Services of King County"}

```

í para buscar

Tema: Entretenimiento


```
Python 3.5.0 Shell
File Edit Shell Debug Options Window Help

{"Account Link": "https://www.instagram.com/atx311/"}
{"Account Platform": "Twitter", "Department": "3-1-1", "Account Name": "atx_311", "Account Link": "https://twitter.com/ATX_311"}
{"Account Platform": "Facebook", "Department": "Animal Services", "Account Name": "Austin Animal Center", "Account Link": "https://www.facebook.com/austinanimal services"}
{"Account Platform": "Instagram", "Department": "Animal Services", "Account Name": "austinanimalcenter", "Account Link": "https://Instagram.com/austinanimalcenter/"}
{"Account Platform": "Twitter", "Department": "Animal Services", "Account Name": "@austinanimals", "Account Link": "https://Twitter.com/austinanimals"}
{"Account Platform": "YouTube", "Department": "Animal Services", "Account Name": "Austin Animal Center", "Account Link": "https://www.YouTube.com/user/austinet s2011"}
{"Account Platform": "Facebook", "Department": "Austin Energy", "Account Name": "Austin Energy", "Account Link": "https://www.facebook.com/austinenenergy?ref=ts&f ref=ts"}
{"Account Platform": "Facebook", "Department": "Austin Energy", "Account Name": "Austin Energy Regional Science Festival", "Account Link": "https://www.facebook .com/austinenenergysciencefest"}
{"Account Platform": "Facebook", "Department": "Austin Energy", "Account Name": "Austin Energy Green Building", "Account Link": "https://www.facebook.com/aegree nbuiding?fref=ts"}
{"Account Platform": "Facebook", "Department": "Austin Energy", "Account Name": "Austin Energy Electric Vehicles", "Account Link": "https://www.facebook.com/Aus tinEnergyElectricVehicles?fref=ts"}
{"Account Platform": "Google Plus", "Department": "Austin Energy", "Account Name": "Austin Energy", "Account Link": "https://plus.google.com/+austinenenergy/video s"}
{"Account Platform": "LinkedIn", "Department": "Austin Energy", "Account Name": "Austin Energy", "Account Link": "https://www.linkedin.com/company/austin-energy /"}
{"Account Platform": "Picasa", "Department": "Austin Energy", "Account Name": "A ustin Energy", "Account Link": "https://picasaweb.google.com/1093755180416213065 55"}
{"Account Platform": "Twitter", "Department": "Austin Energy", "Account Name": " @austinenenergy", "Account Link": "https://Twitter.com/austinenenergy"}
{"Account Platform": "Vimeo", "Department": "Austin Energy", "Account Name": "Au stin Energy", "Account Link": "https://vimeo.com/austinenenergy"}

Escribe aquí para buscar
```

Tema: Tecnología

```
Python 3.5.0 Shell
File Edit Shell Debug Options Window Help

ges/New-York-NY/NYC-DADS/11150458886342", "Likes/Followers/Visits/Downloads": " 183", "Agency": "HRA", "Platform": "Facebook"}
{"Date Sampled": "08/08/2011 12:00:00 AM", "Url": "http://www.facebook.com/profi le.php?id=1570569347", "Likes/Followers/Visits/Downloads": "190", "Agency": "MOP D", "Platform": "Facebook"}
{"Date Sampled": "08/08/2011 12:00:00 AM", "Url": "http://www.facebook.com/timeb ankany", "Likes/Followers/Visits/Downloads": "227", "Agency": "DFTA", "Platform ": "Facebook"}
{"Date Sampled": "08/08/2011 12:00:00 AM", "Url": "http://www.facebook.com/coolr oofs?sk=wall", "Likes/Followers/Visits/Downloads": "194", "Agency": "DOB: Cool R oofs", "Platform": "Facebook"}
{"Date Sampled": "08/08/2011 12:00:00 AM", "Url": "http://www.facebook.com/nycgo .nl", "Likes/Followers/Visits/Downloads": "243", "Agency": "NYC & Co", "Platform ": "Facebook"}
{"Date Sampled": "08/08/2011 12:00:00 AM", "Url": "http://www.facebook.com/pages /NYC-Mayors-Office-of-Immigrant-Affairs/118622031512497", "Likes/Followers/Visits /Downloads": "237", "Agency": "MOIA", "Platform": "Facebook"}
{"Date Sampled": "08/08/2011 12:00:00 AM", "Url": "http://www.facebook.com/NYCHA yorsCAU", "Likes/Followers/Visits/Downloads": "234", "Agency": "CAU", "Platform ": "Facebook"}
{"Date Sampled": "08/08/2011 12:00:00 AM", "Url": "http://www.facebook.com/pages /New-York-NY/NYC-Information-Technology-Telecommunications/104786059565184", "Li kes/Followers/Visits/Downloads": "238", "Agency": "DOITT", "Platform": "Facebook " }
{"Date Sampled": "08/08/2011 12:00:00 AM", "Url": "http://www.facebook.com/pages /New-York-NY/NYC-Charter-Revision-Commission/110528715643388", "Likes/Followers /Visits/Downloads": "287", "Agency": "City Charter", "Platform": "Facebook"}
{"Date Sampled": "08/08/2011 12:00:00 AM", "Url": "http://www.facebook.com/pages /NYC-Mayors-Office-of-Veterans-Affairs/128003537214726", "Likes/Followers/Visits /Downloads": "257", "Agency": "Vets", "Platform": "Facebook"}
{"Date Sampled": "08/08/2011 12:00:00 AM", "Url": "http://www.facebook.com/pages /New-York-NY/HOPE-2011/157690657606772", "Likes/Followers/Visits/Downloads": "28 2", "Agency": "DHS", "Platform": "Facebook"}
{"Date Sampled": "08/08/2011 12:00:00 AM", "Url": "http://www.facebook.com/nycgo .de", "Likes/Followers/Visits/Downloads": "272", "Agency": "NYC & Co", "Platform ": "Facebook"}
{"Date Sampled": "08/08/2011 12:00:00 AM", "Url": "http://www.facebook.com/nycgo .fr", "Likes/Followers/Visits/Downloads": "277", "Agency": "NYC & Co", "Platform ": "Facebook"}

Escribe aquí para buscar
```

Cada vez que ejecutábamos un Script se visualizaba como los datos se leían en formato JSON y se guardaban en la base de datos NoSQL de CouchDB.

G) APLICACIÓN WEB.

Se creó una interface interactiva para poder visualizar la información recopilada de las diferentes fuentes.

La interface cuenta con un menú de navegación que permitirá al usuario navegar de forma fácil y segura.

Se puede seleccionar el tema de interés y mediante un selector de etiquetas elegir la fuente de donde fue obtenida la información.



Seleccionamos la fuente y la interfaz nos direcciona a una nueva página donde mostrara la información de guardada en nuestra base de datos NoSQL de CouchDB.

Mascotas - Fuente Twitter (1)					
Foto	Usuario	Tweet	Fecha	Conteo de seguidores	DESCRIPCIONES
	m0gumogu vitamin c	RT @espreux: nothing will ever be as pure as minceop treating and taking to the pushies as if they were real cats https://t.co/1T3p2gC07	Mon Dec 03 01:35:57 +0000 2018	171	Support exo & ry 🍷 • InayaAllah tahun depan lulu 🍷 • Pejung TA 🍷
	Neutral_Beings Matthew Shadwell (More in Pic)	👉 SHE WANTS THE DICK!	Mon Dec 03 01:35:57 +0000 2018	973	"What? Not 'm straight as an arrow"? Male Writer 🍷 Literate & Descriptive 🍷 18+ 🍷 No art is mine!
	Valeriana13 Valeriana	RT @Sexy_Girl_Hot_X: Así puse mi árbol de navidad el año pasado, quieren ver cómo estoy poniendo el de este año? Solo apoyen con su RT 🍷 mis...	Mon Dec 03 01:35:57 +0000 2018	196	Ser alegre y divertido
	JulissaJ_ Julie	RT @Prismeeus: my phone made a slideshow of yesterday when i literally only saved pictures of crying cats https://t.co/K21AR03q2n	Mon Dec 03 01:35:58 +0000 2018	201	null
	thvrlg	Love Cats https://t.co/vuafACDKvT	Mon Dec 03 01:35:58 +0000 2018	5005	All About Fashion
	RTn_Perrros RTn_Perrros	Cómo ayudan los perros a los niños con TDAH https://t.co/yKFPsc3NC3 https://t.co/k2Ay24CNE	Mon Dec 03 01:36:03 +0000 2018	949	Noticias sobre el cuidado de tu perro 🍷 #Mascotas #Perros
	aroci172 aroci	RT @Prismeeus: my phone made a slideshow of yesterday when i literally only saved pictures of crying cats https://t.co/K21AR03q2n	Mon Dec 03 01:36:03 +0000 2018	494	be happy 🍷 716 🍷 518 #BLM
	L_adrian13 jorge green		Mon Dec 03 01:36:04 +0000 2018	983	
	EduardSerranoCa Eduard Serrano Cala	RT @LaFallas: Venga, y ahora, como siempre, la pregunta del idiota. ¿Cómo no lo vimos venir? Me pregunto qué estarán cenando todos esos.	Mon Dec 03 01:36:04 +0000 2018	2651	null
	LexliMoore2 Lexli	Ugh i love cats 🍷	Mon Dec 03 01:36:04 +0000 2018	389	ecu women's soccer #2
	bill_bazaro	RT @nowthatbully: Baby Leopard 🍷	Mon Dec 03 01:36:04 +0000 2018	392	DJ/Producer/Artist

IV. CONCLUSIONES

Para diseñar una buena arquitectura de un Data Warehouse es necesario conocer bien los requerimientos del negocio y hacer un estudio profundo de las fuentes externas que nos van a suministrar los datos.

Tener un Data Warehouse dentro de una organización es muy útil debido a que ayuda y facilita la toma de decisiones ya que permite tener un buen control sobre toda la información que necesita la organización para poder tomar decisiones, basándose en datos confiables y que pueden tener a la mano en cualquier momento.

V. REFERENCIAS

- [1] «PowerData,» PowerData, [En línea]. Available:
] <https://www.powerdata.es/data-warehouse>. [Último acceso: 29 11 2018].
- [2] «Tableau,» Tableau, [En línea]. Available:
] [//www.tableau.com/es-es/products/what-is-tableau#UWd33zZDTkMuUIYY.99](https://www.tableau.com/es-es/products/what-is-tableau#UWd33zZDTkMuUIYY.99). [Último acceso: 29 11 2018].
- [3] «WebEmpresa,» [En línea]. Available:
] <https://www.webempresa.com/blog/que-es-twitter-como-funciona.html>. [Último acceso: 29 11 2018].
- [4] «Wikipedia,» [En línea]. Available:
] <https://en.wikipedia.org/wiki/Data.gov>. [Último acceso: 29 11 2018].
- [5] M. Rouse, «SearchDataCenter,» [En línea]. Available:
] <https://searchdatacenter.techtarget.com/es/definicion/CouchDB>. [Último acceso: 03 11 2018].

VI. ANEXO

<https://github.com/GeovannyDias/Datawarehouse-with-CouchDB>