2016



Ing. Juan Zaldumbide

Integrantes:

* Bastidas Geovanny
* Chicaiza Jessica

**Proyecto Bases de Datos Multidimensionales**

Tabla de contenido

[Introducción 2](#_Toc453155012)

[Objetivo General 2](#_Toc453155013)

[Objetivo Específico 2](#_Toc453155014)

[Desarrollo del proyecto 2](#_Toc453155015)

[Consulta de datos 10](#_Toc453155016)

[Análisis de gráficos 13](#_Toc453155017)

[**Conclusiones** 14](#_Toc453155018)

[Recomendaciones 15](#_Toc453155019)

[Conclusiones 15](#_Toc453155020)

# Introducción

En este documento se especificará la arquitectura propuesta en el proyecto en couchDB .

En el cual se recopilaran datos de diferentes tipos para crear una base de datos unificada de tipo NoSQL y poder analizar los datos obtenidos.

# Objetivo General

* Realizar una recopilación de datos (tweets) de las regiones del Ecuador y analizar los datos obtenidos.

# Objetivo Específico

* Recopilar tweets de las diferentes provincias del Ecuador para aguparlas por regiones y obtener un resultado total por cada region.
* Analizar y comparar los datos obtenidos.
* Utilizar gráficos estadísticos para representar los datos de manera visual y sean mas entendibles.
* Preparar una interfaz web para que nos muestren los datos recopilados.

# 

# Desarrollo del proyecto

La arquitectura propuesta para el diseño de la base de datos NoSQL está basada en la recopilación de tweets en couchDB.

De diferentes lugares que pertenecen a ciertas provincias del Ecuador

**Costa**

Prov\_Manabi

* Pedernales
* Perdernales1

Prov\_Esmeraldas

* Esmeraldas
* Esmeraldas1
* Atacames

Prov\_SantaElena

**Sierra**

Prov\_Pichincha

* Quito
* Amaguaña

Prov\_Guaranda

**Oriente**

Prov\_Pastaza

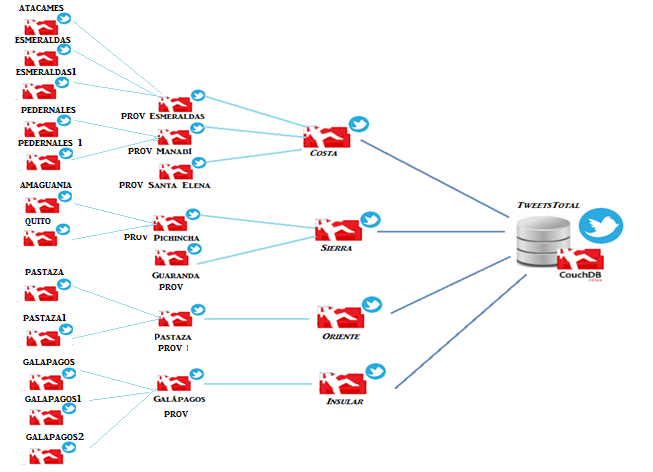
* Pastaza
* Pastaza1

**Insular**

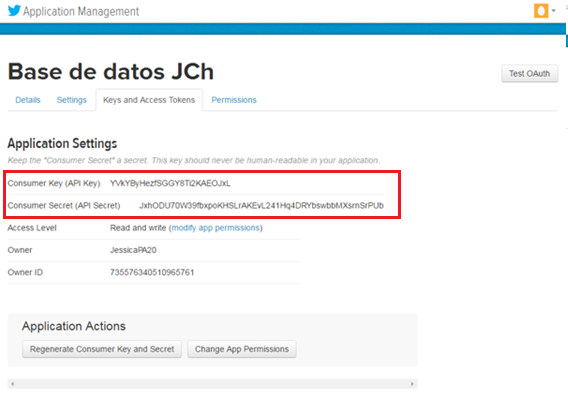
Prov\_Galápagos

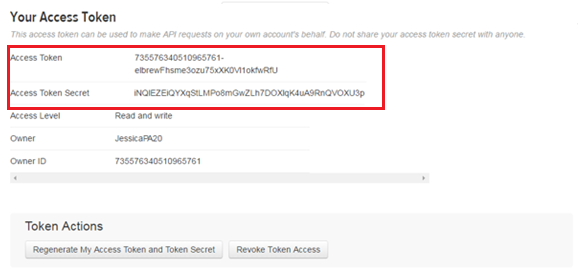
* Galápagos
* Galápagos2
* Galápgos3

Cabe recalcar que algunos os datos fueron obtenidos desde diferentes host los cuales se unificaron por provincias y a su vez por regiones



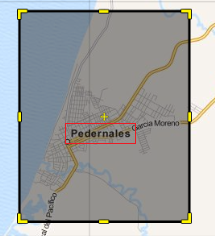
Para poder realizar la recoleccion de datos necesitamos crearnos una cuenta de twitter para poder obtener las credenciasl las cuales son indispensables para la recoleccion de los datos.

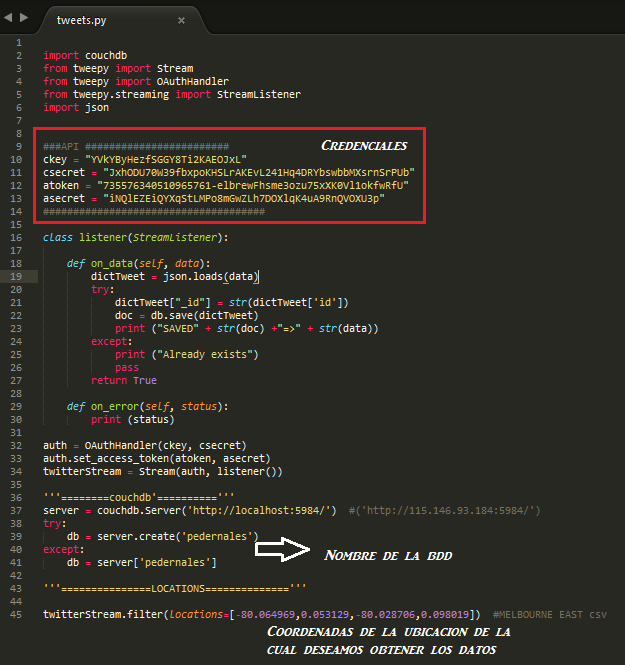




Estas **credenciales** deberan ser ser introduccidas en los archivos .py encargados de la recoleccion de tweets, ademas en este documento debemos colocar **el nombre de la base de datos** en la cual se para a recolectar los datos y tambien **la ubicación en coordenadas en formato csv** del lugar del cual vamos a obtener los datos.

Por ejemplo podemos utilizar : <http://boundingbox.klokantech.com/>





Otro paso importante para que poder recocger los datos es entrar al cmd y colocar los comandos

**pip install tweepy**

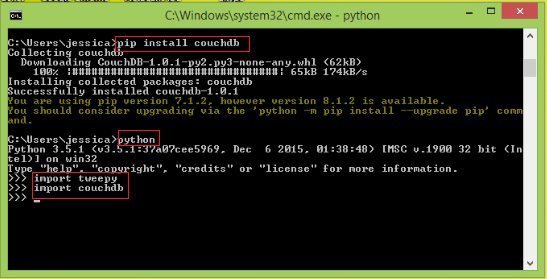
**pip install couchdb**

**pyhon**

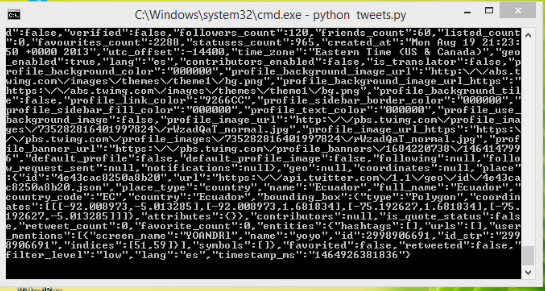
**>>>import tweepy**

**>>>import couchdb**

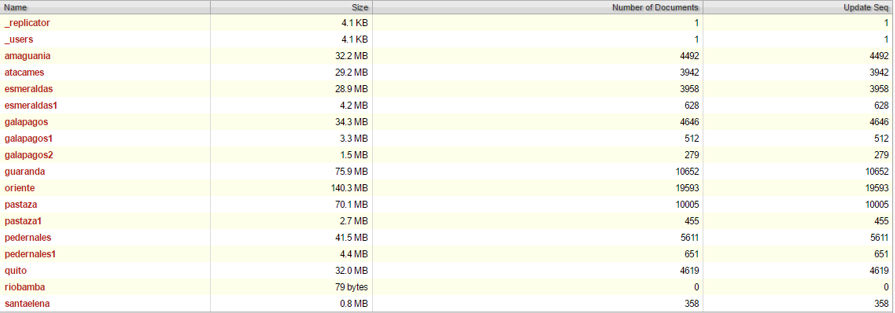




Par continuar con la recoleción de tweets debemos ejecutar estos archivos.py



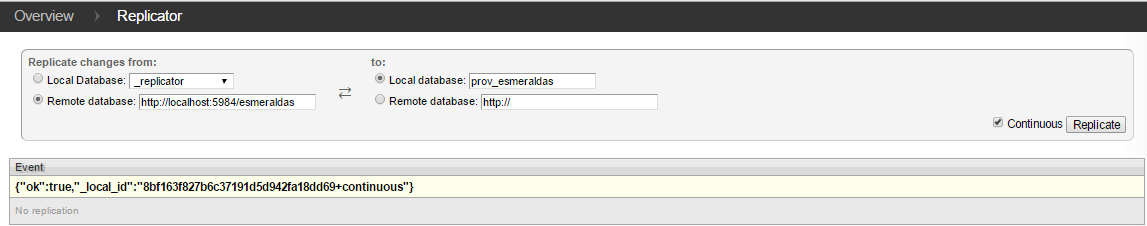
Los datos recolectados fueron:

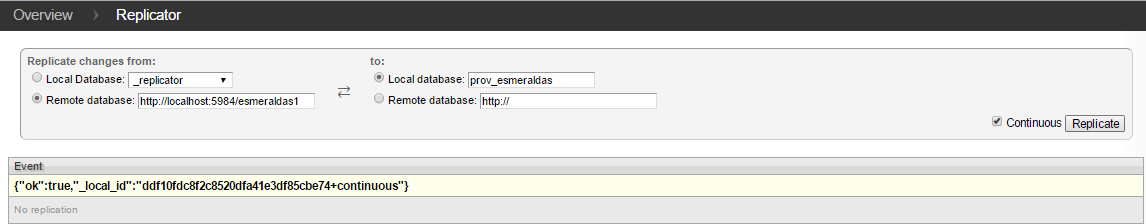


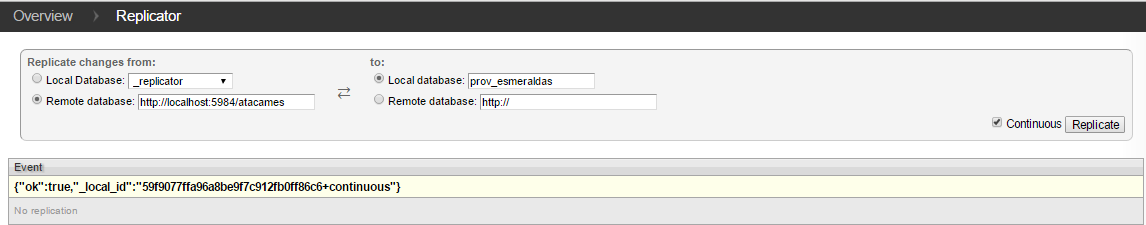
Ahora procederemos a agrupar estas bases de datos por provincias a las que pertenece para posteriormente agruparlos por regiones.

**Costa**

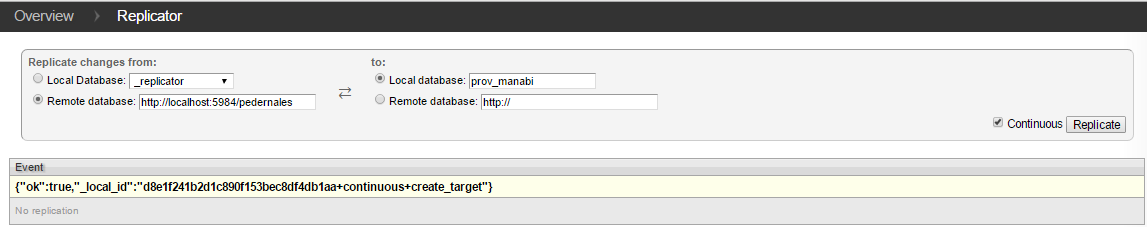
**Prov\_Esmeraldas (esmeraldas, esmeraldas1, atacames)**

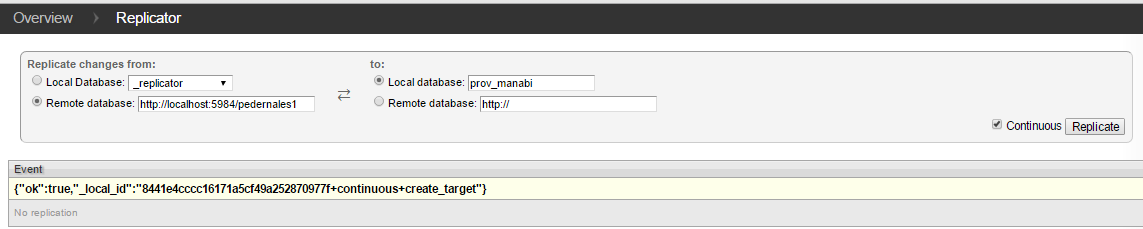




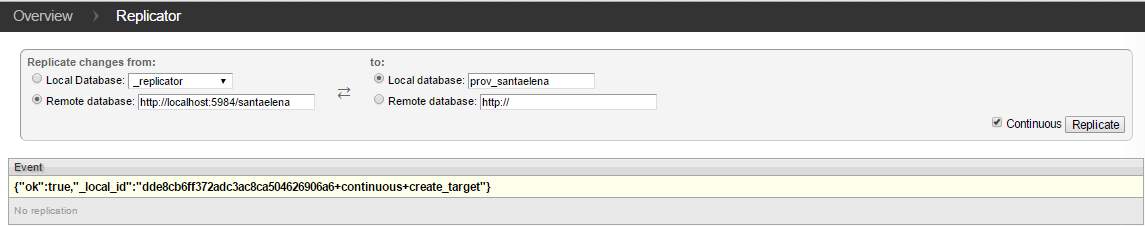


**Prov\_Manabi (pedernales, pedernales1)**



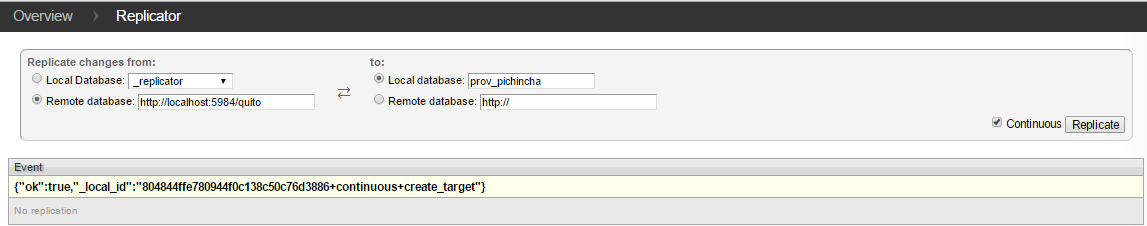


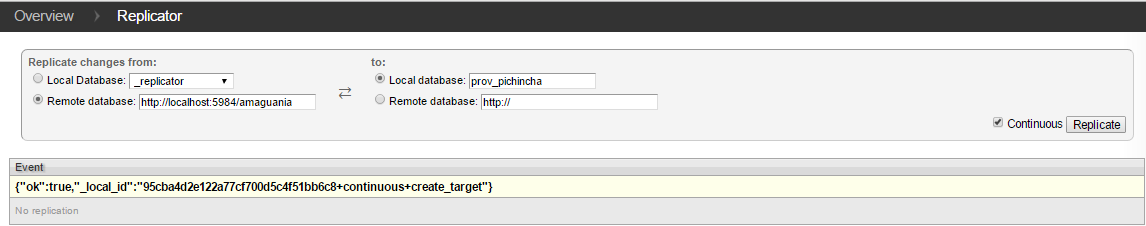
**Prov\_SantaElena (santaelena)**



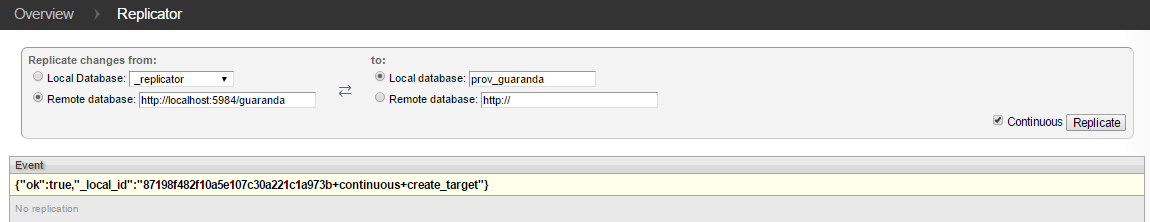
**Sierra**

**Prov\_Pichincha (quito,amaguaña)**



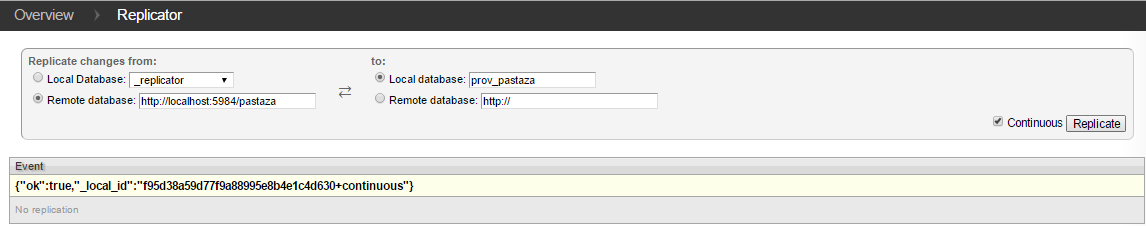


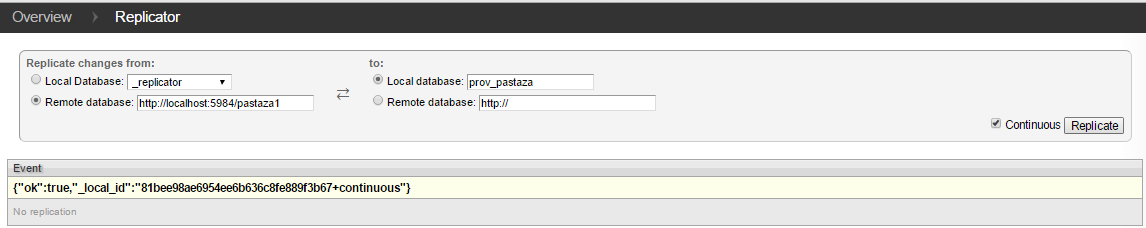
**Prov\_guaranda (guaranda)**



**Oriente**

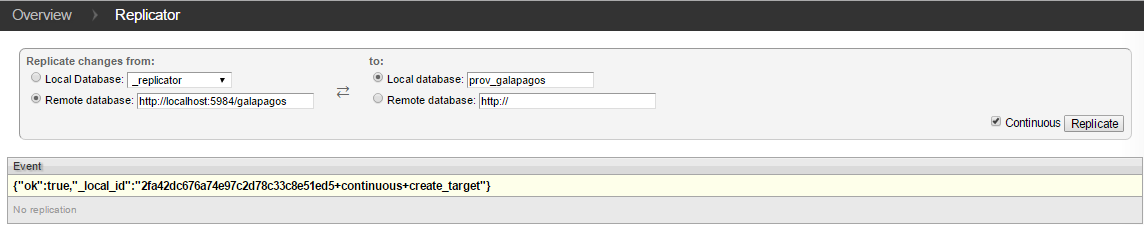
**Prov\_Pastaza (pastaza, pastaza1)**

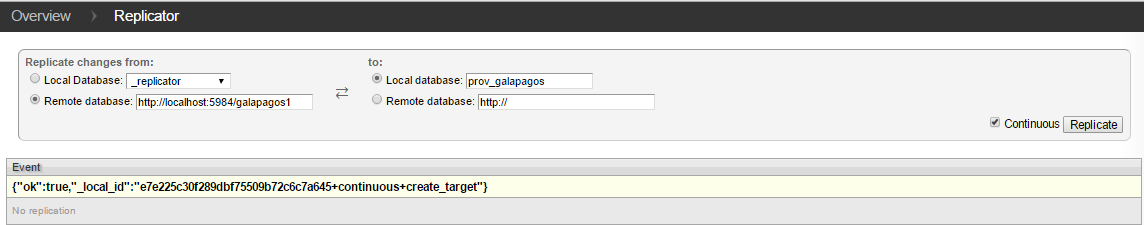


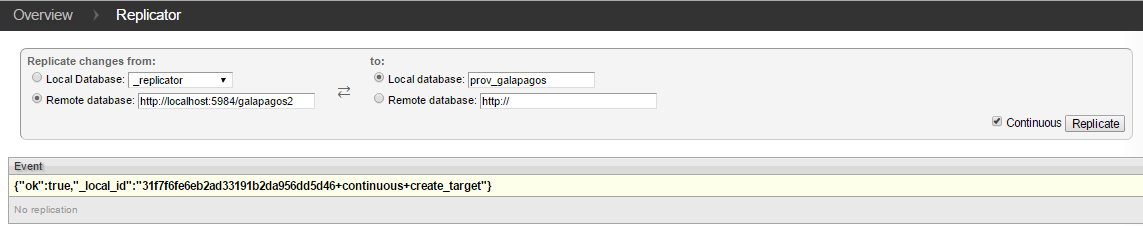


**Insular**

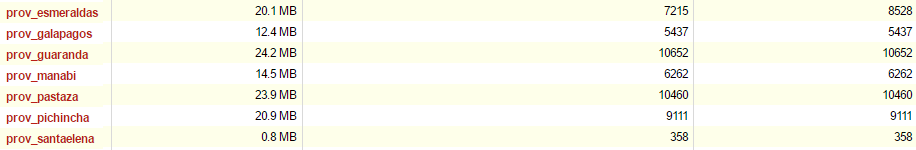
**Prov\_Galapagos (galapagos, galapagos 1, galapagos2)**





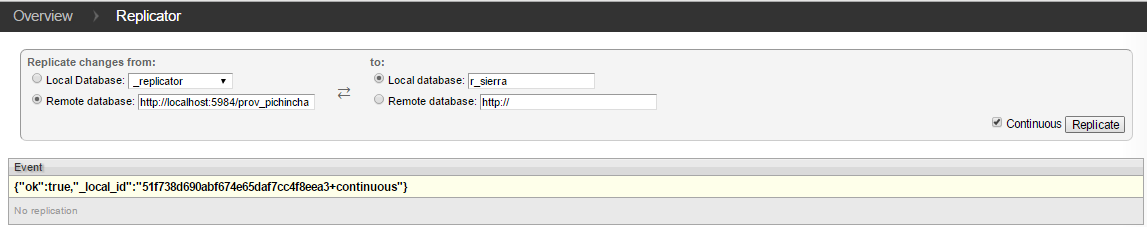


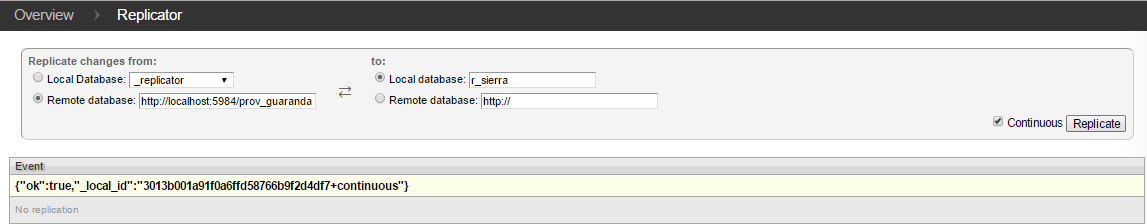
**Total obtenido por provincias**



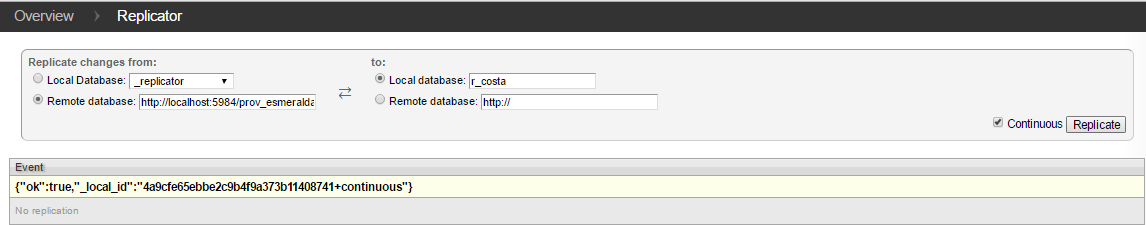
Ahora las agruparemos por regiones

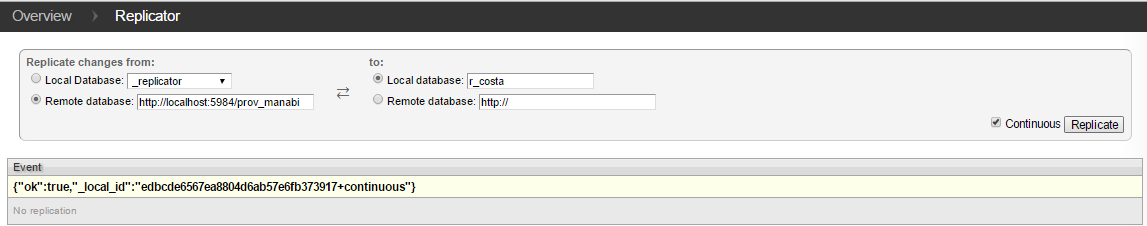
**Sierra (prov\_pichincha,prov\_guaranda)**

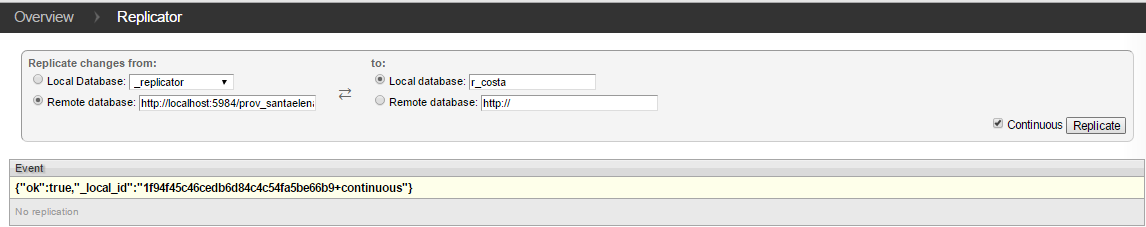




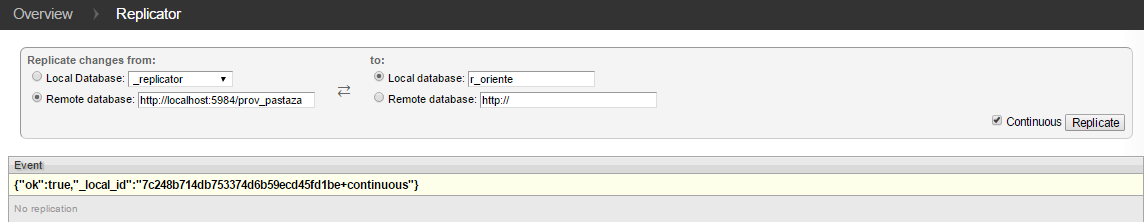
**Costa (prov\_esmeraldas, prov\_manabi, prov\_santaelena)**



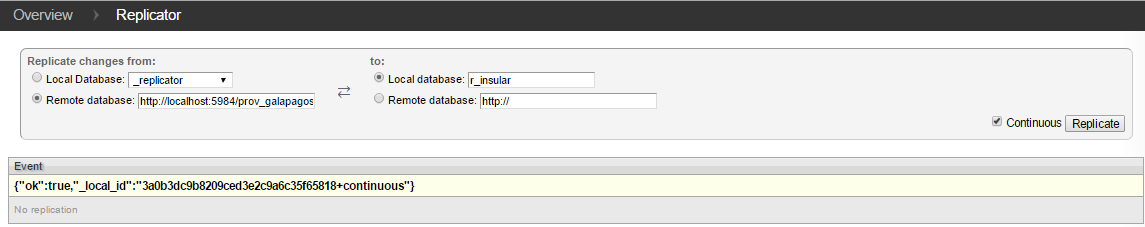




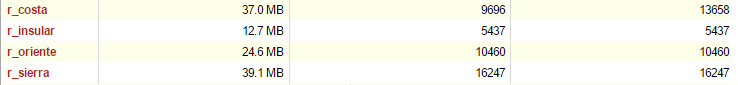
**Oriente (prov\_pastaza)**



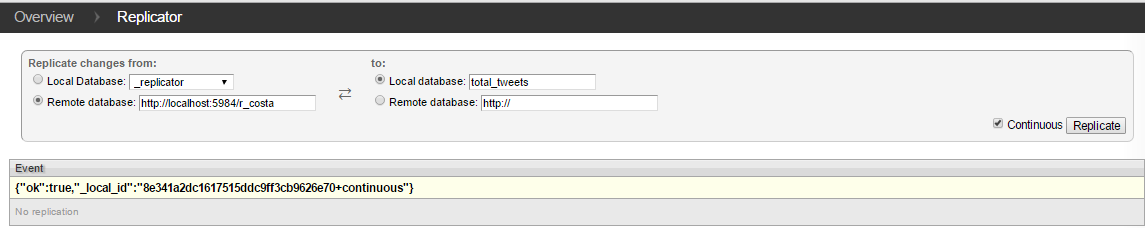
**Insular (prov\_galapagos)**

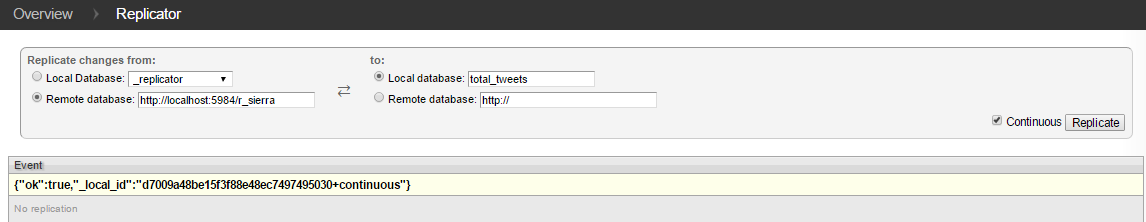


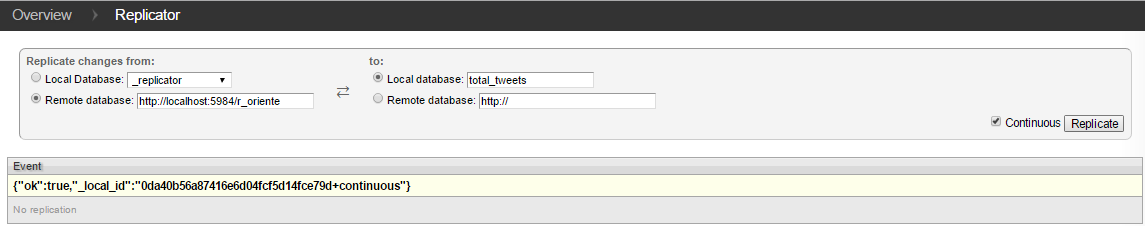
**Totales obtenidos por región**

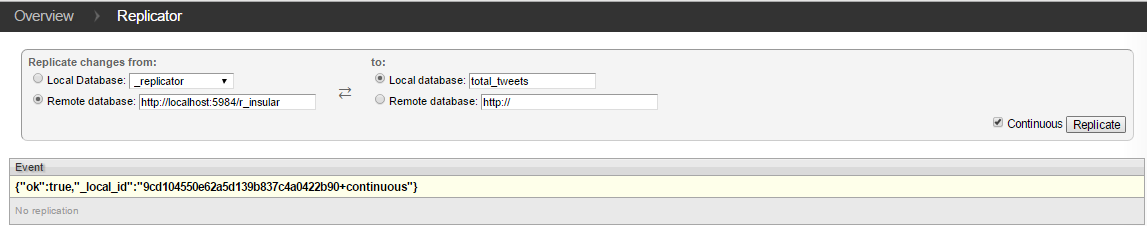
****

Crearemos una sola base de datos de contendra los tweets de las cuatro provincias del Ecuador **total\_tweets** (r\_costa, r\_sierra, r\_oriente, r\_galapagos).









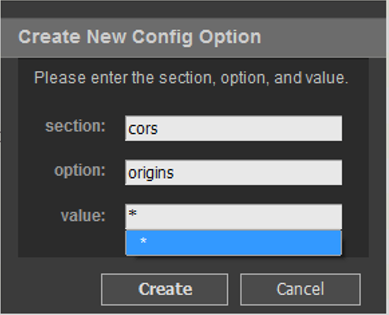
Nuestra BDD **tweets\_total** cuenta con **29407** elementos



Ahora debemos cambiar algunas configuraciones del couchDB



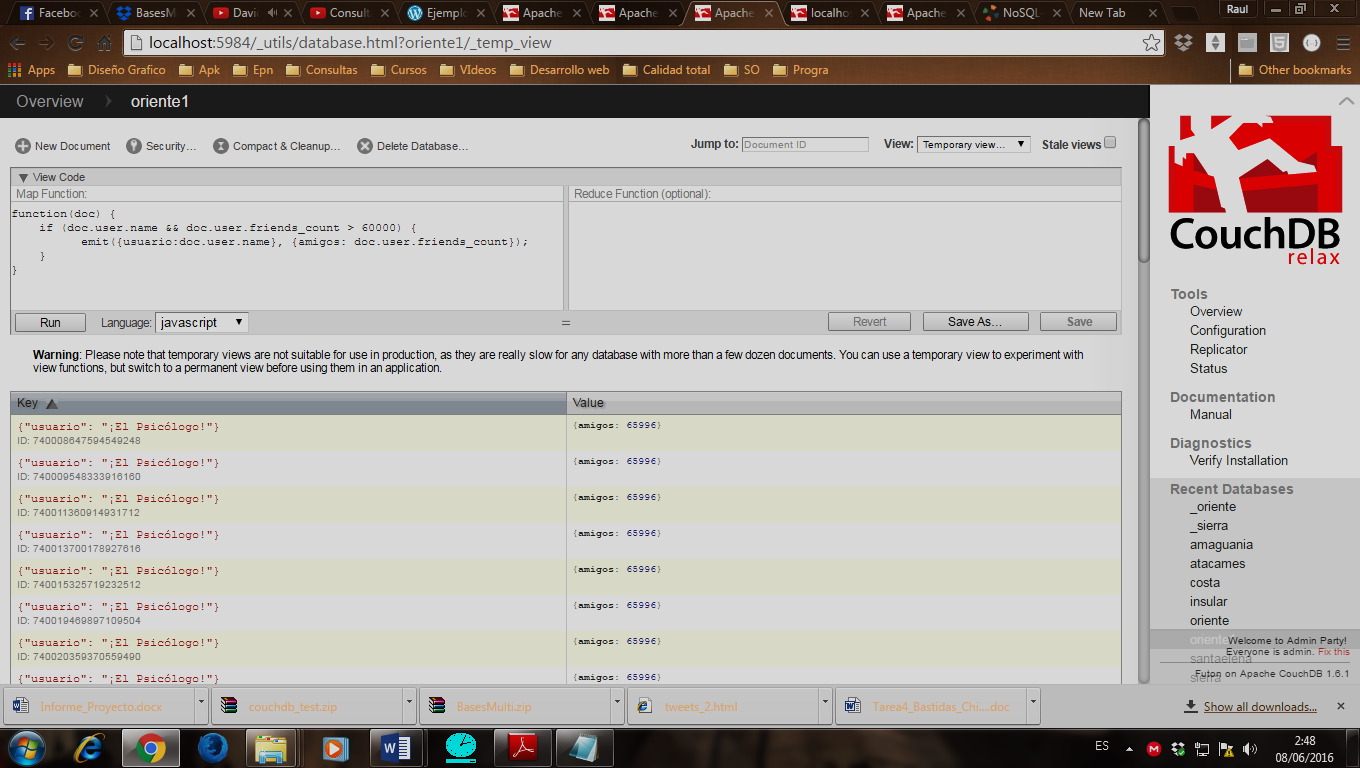


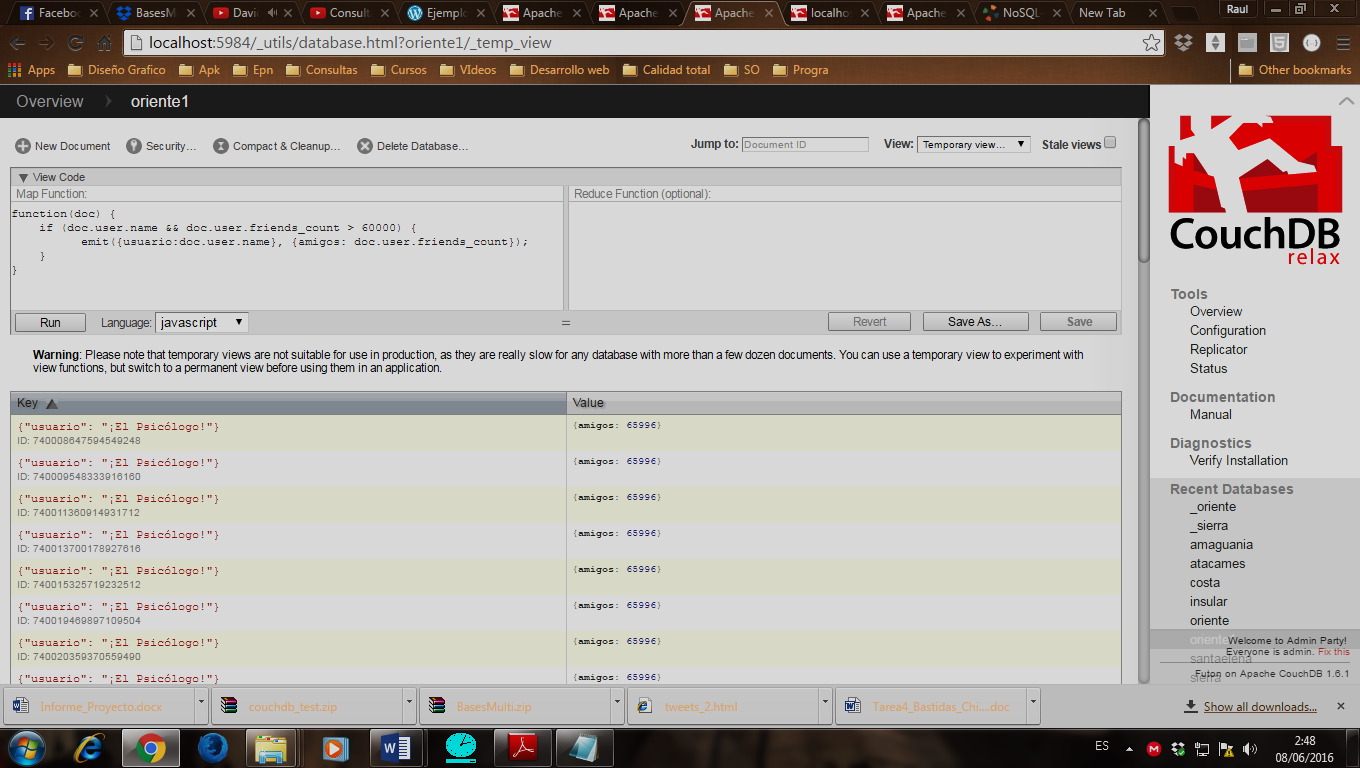
Y creamos una nueva

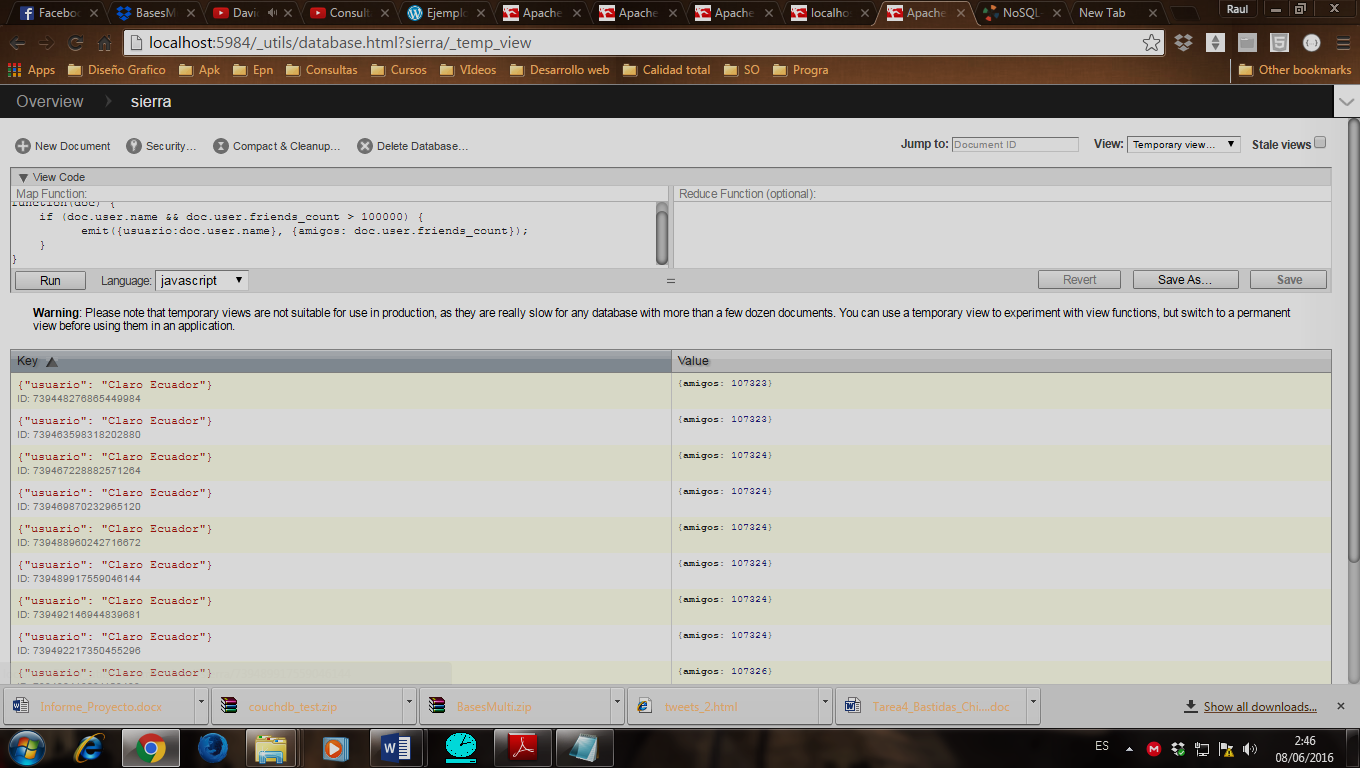
**Consulta de datos**

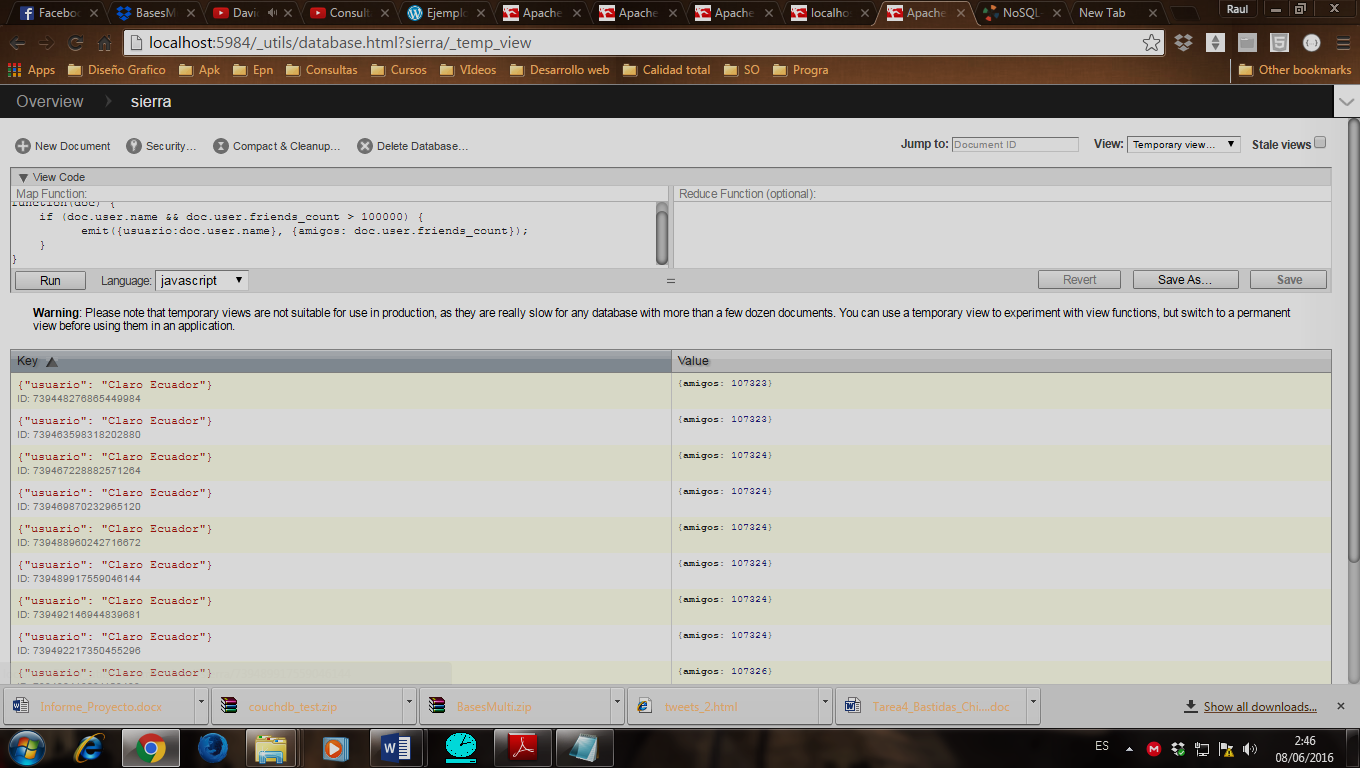
* Una vez que hemos recolectado datos procedentes de los tweets procedemos a obtener informacion.
* Para eso usaremos la herramienta que proporciona CouchDB para las consultas en la BD llamado Map Function.

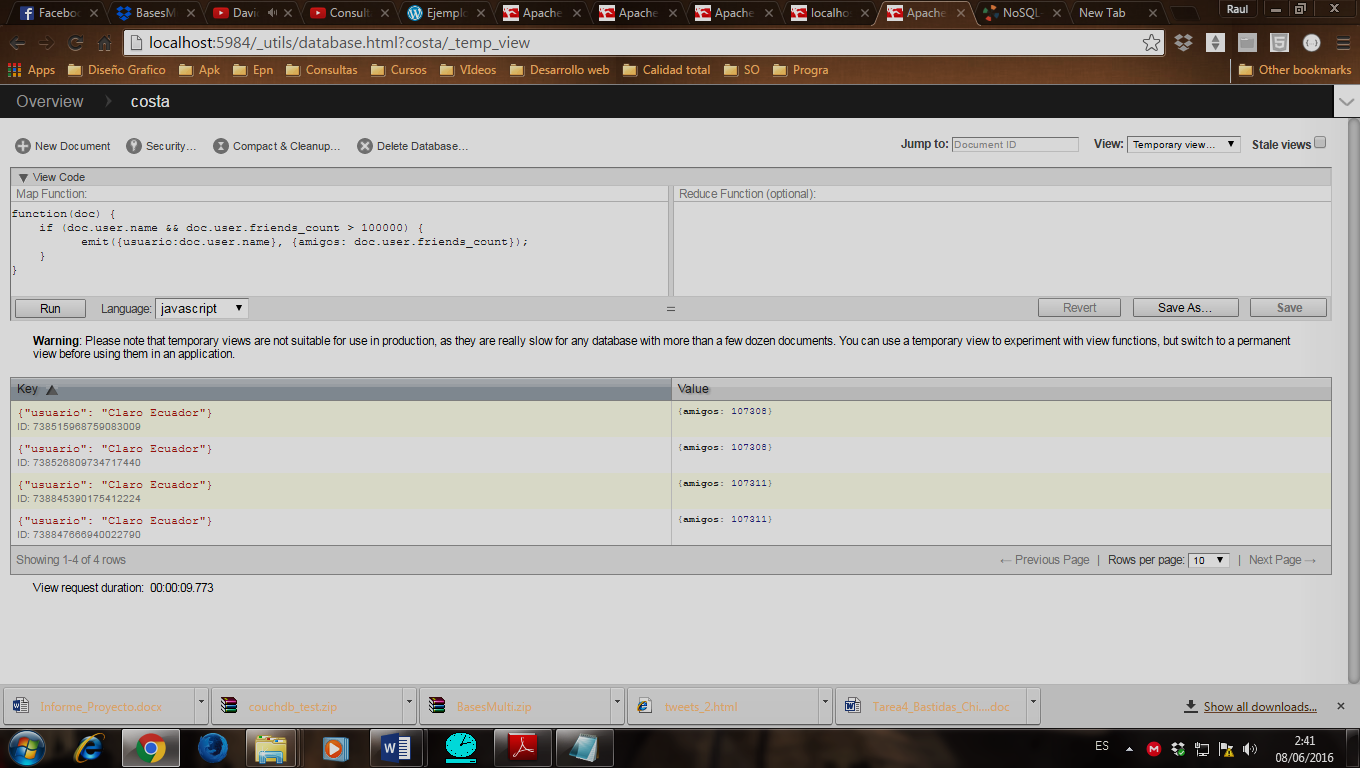
***Consulta: Averiguar el mayor # de amigos que tengan los usuarios en las diferentes regiones del Ecuador.***

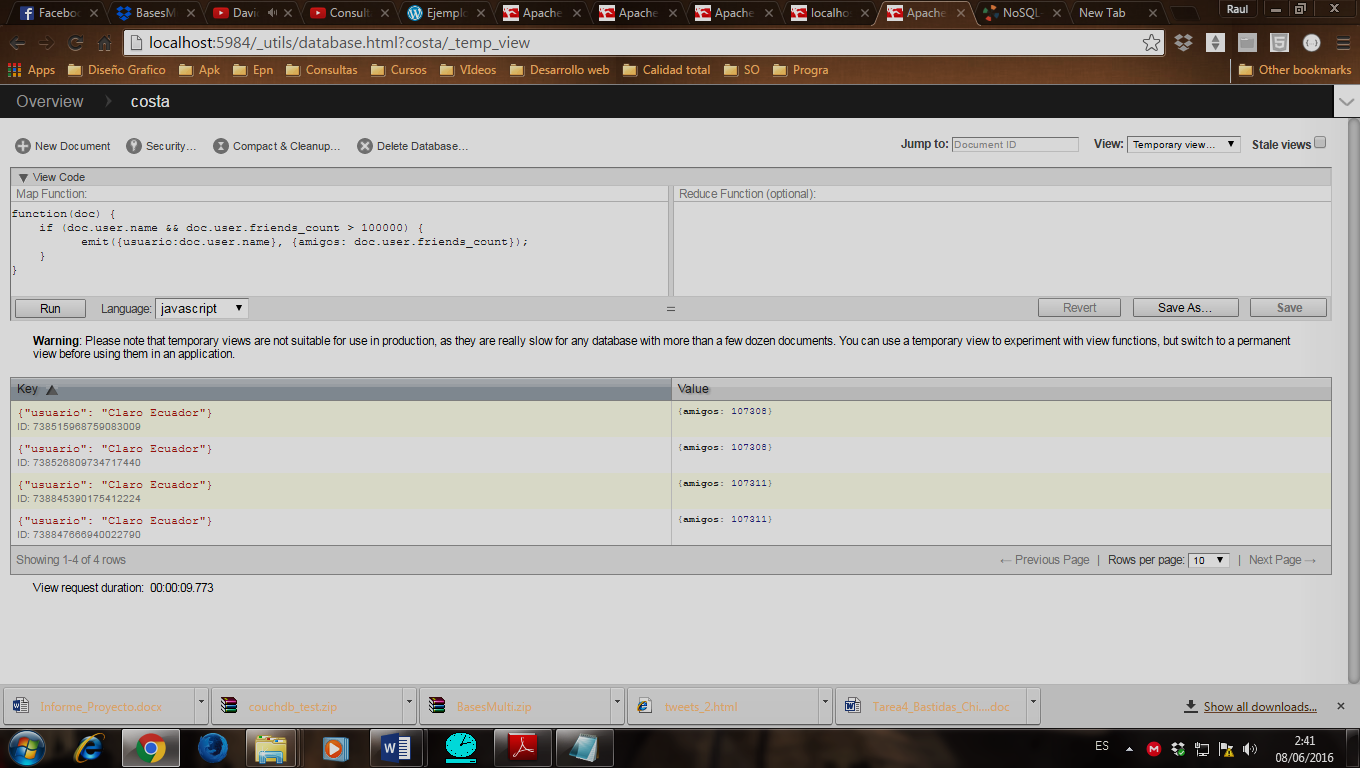


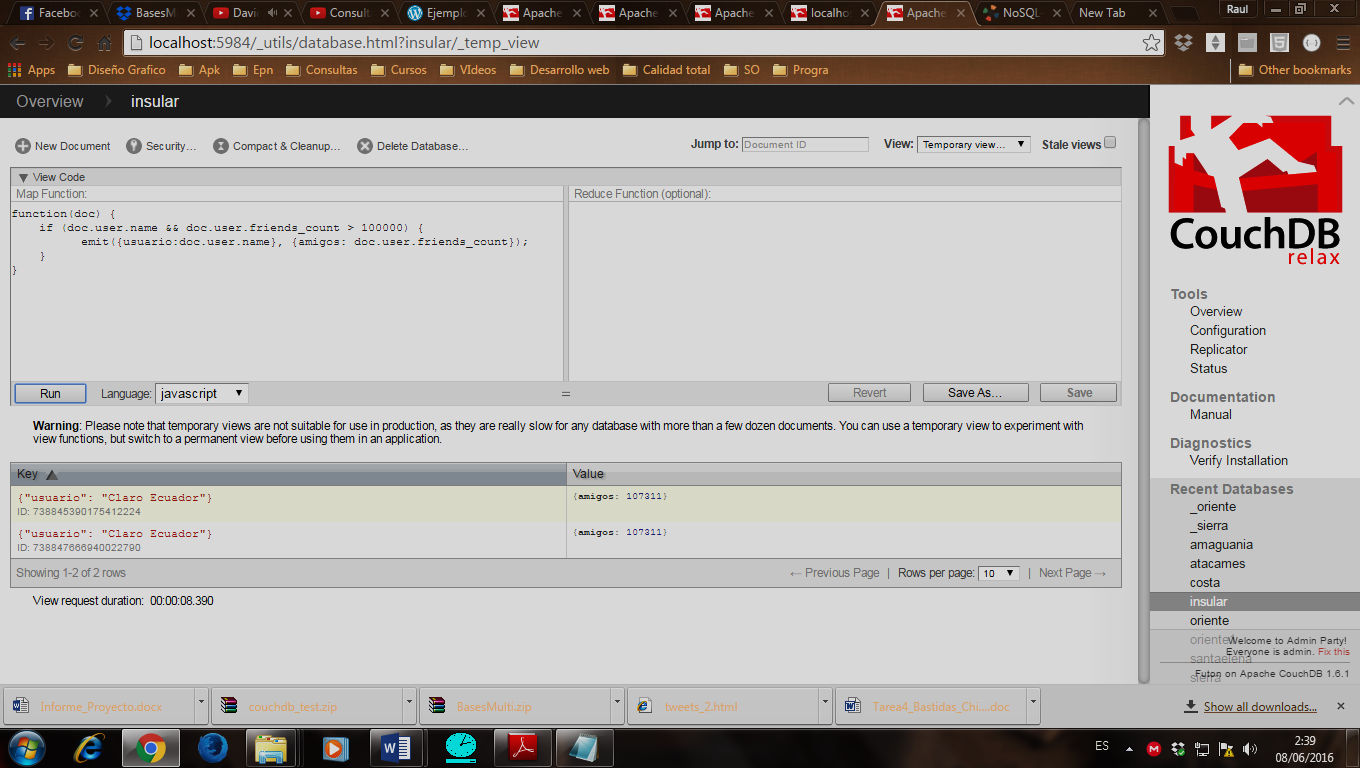


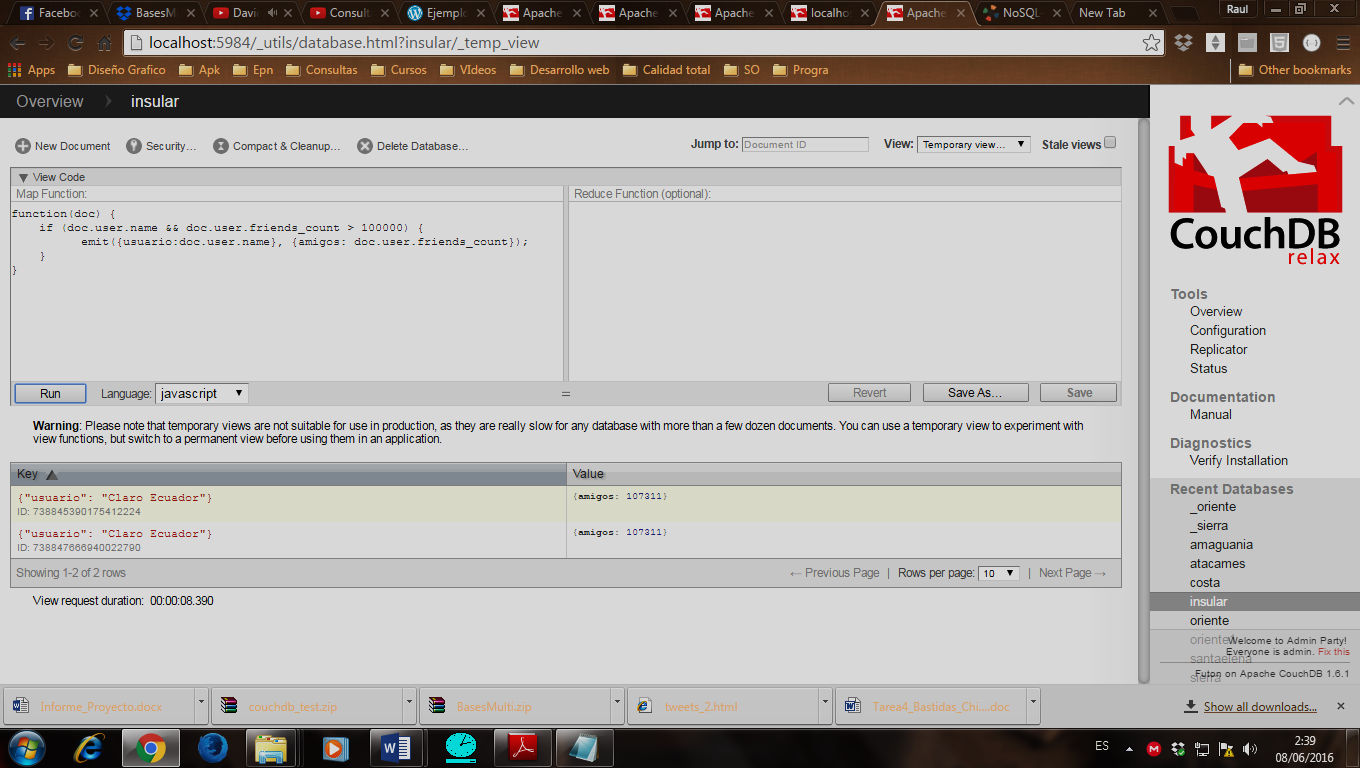


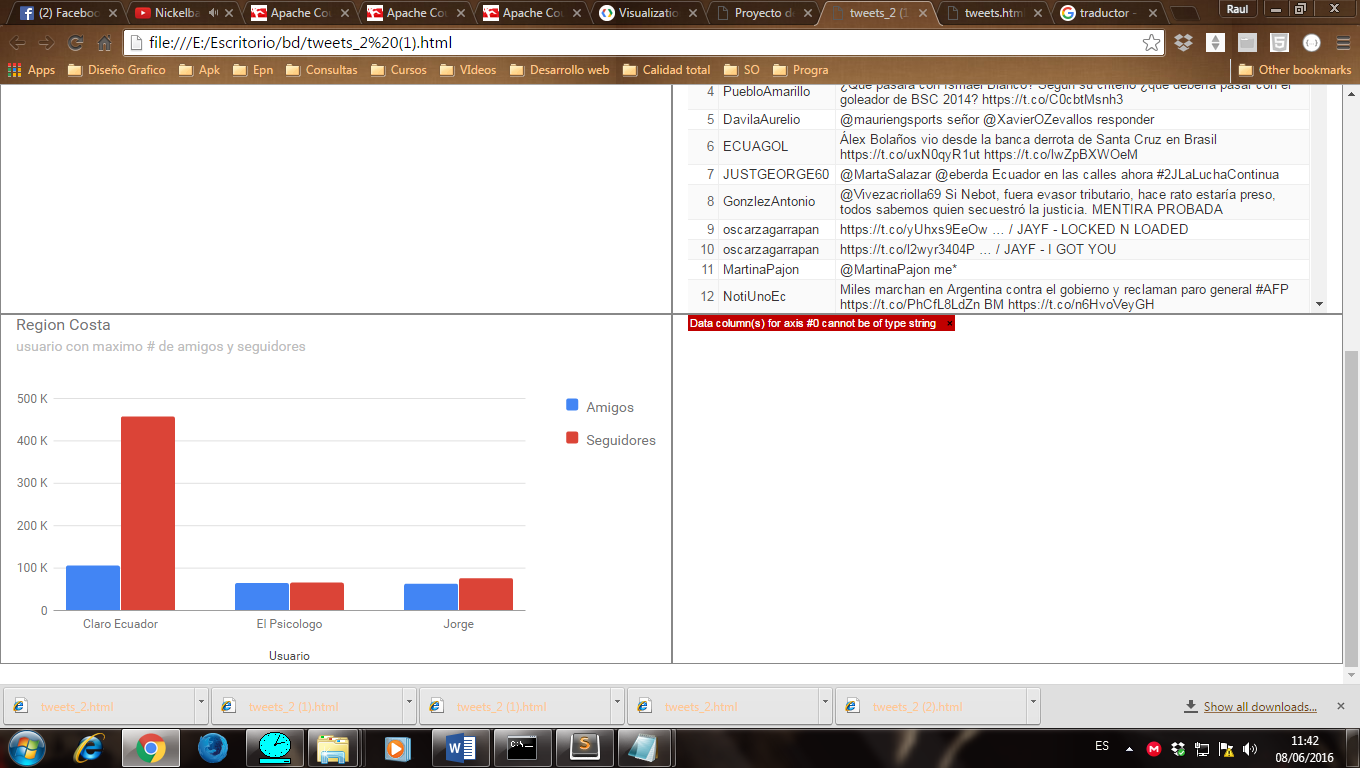




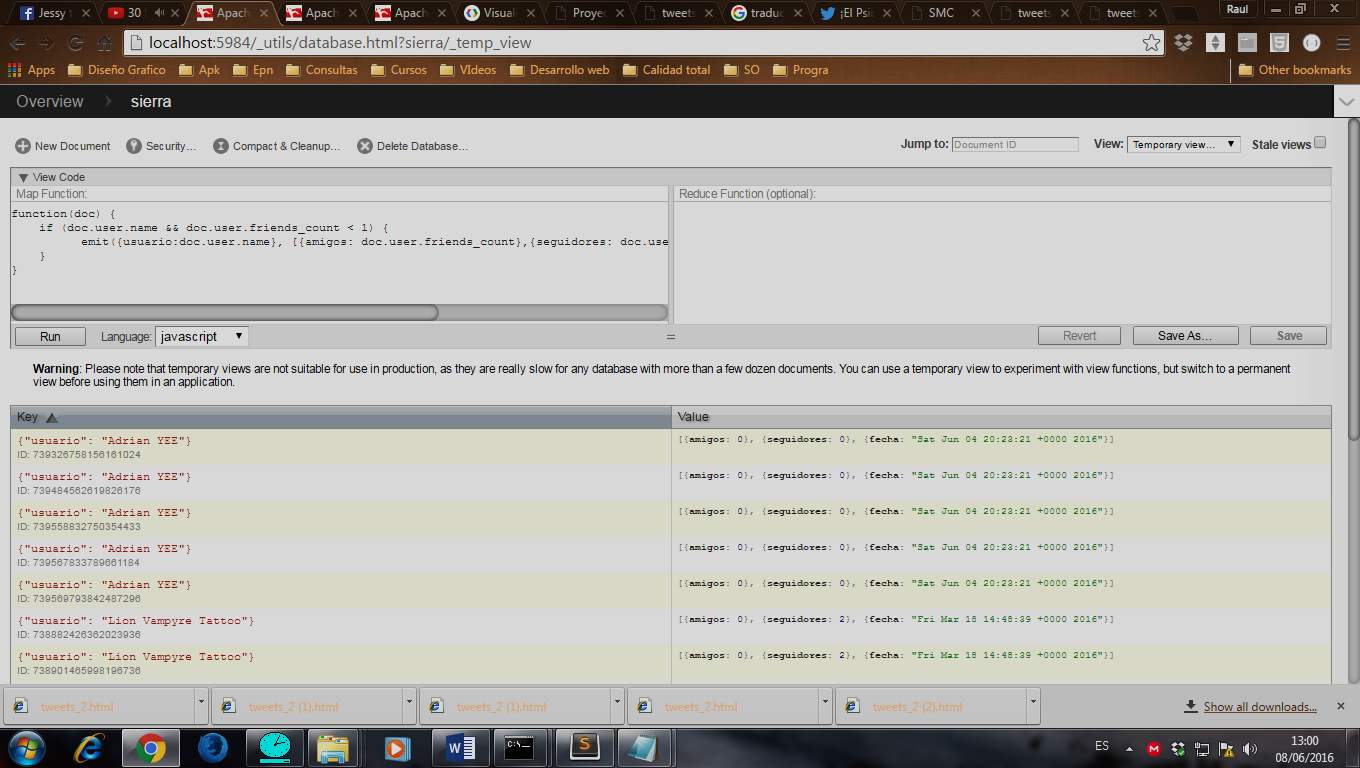


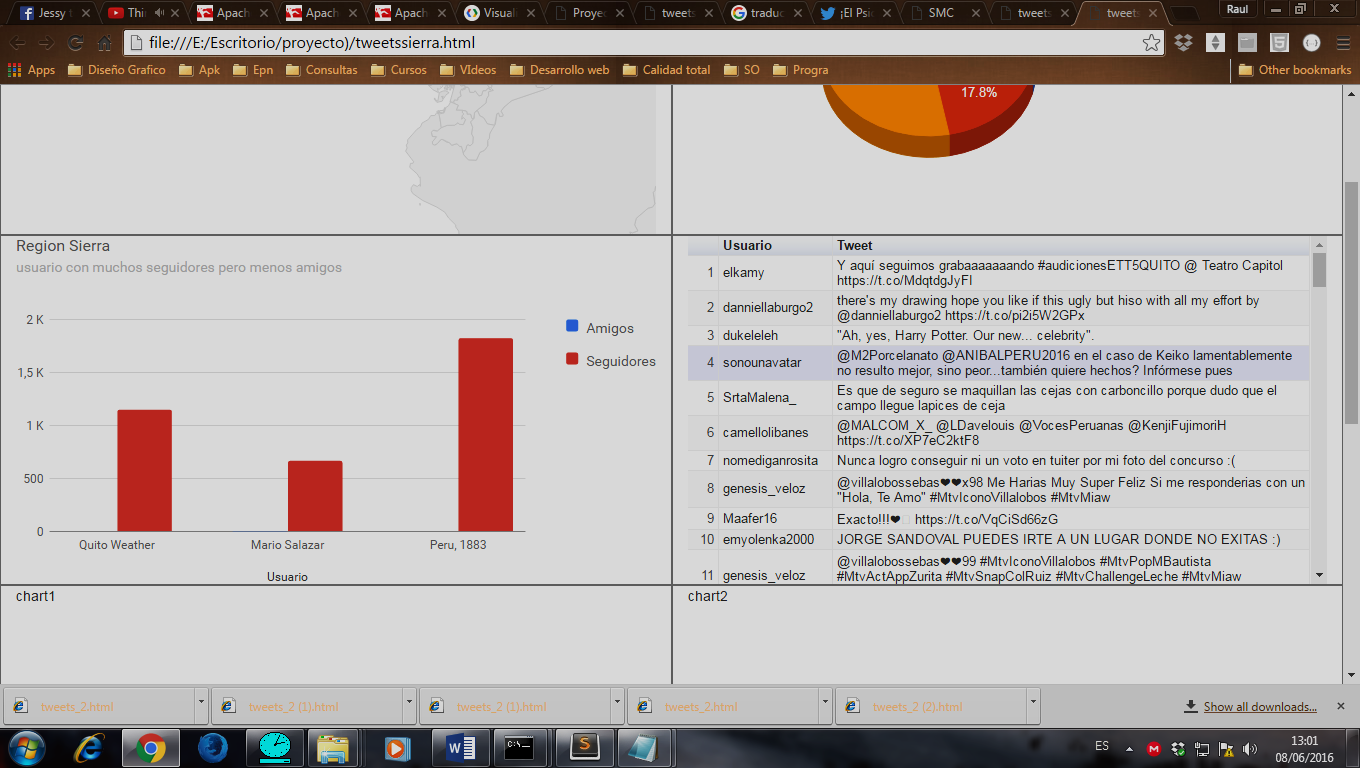






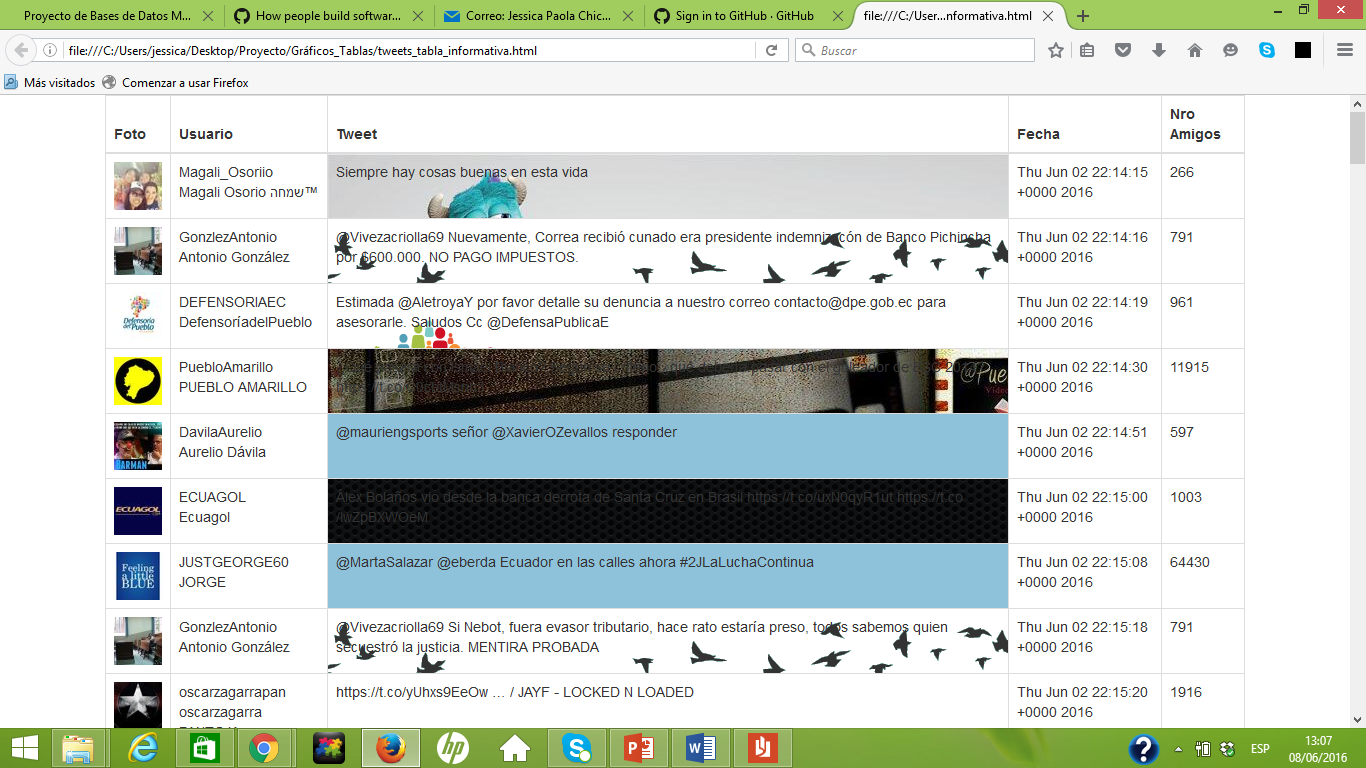
***Consulta: Encontrar usuarios que tengan muchos seguidores pero casi nada de amigos en la region sierra***

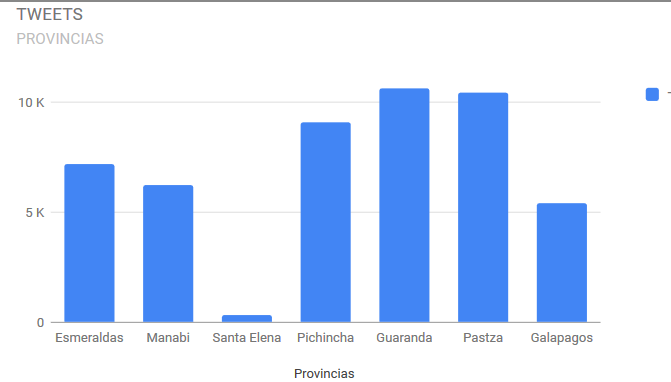


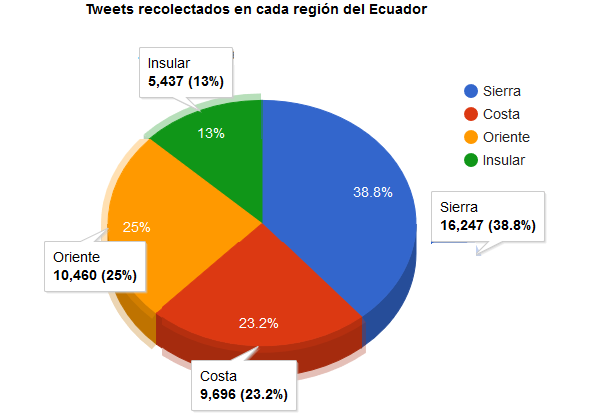


# Análisis de gráficos

Mostramos información recolectada a manera de una tabla.







# Conclusiones

* La provincia que mas tweets recolectó fue la provincia de Guaranda y la provincia que recogio menos tweets fue Santa Elena.
* En conlusion podemos observar que en la region Sierra es la que mas alto porcentaje tiene va con el 38,8 seguido del Oriente con 25% , la Costa con el 23,2% y por ultimo la region Insular 13%.
* Los tweets recolectados durante la semana en diferentes regiones del ecuador demuestra que los ecuatorianos utilizan esta plataforma grandemente.
* Los tweets reolectados fueron segundo a segundo, incluso pasada la media noche.
* Cabe destacar que en regiones como pedernales no hubo mucha recolección de tweets y esto se debe a que esa provincia esta pasando por una situación critica.
* Entre las regiones, el que mayór recaudo de tweets tuvo fue la Sierra, dejando en cuarto lugar a la región Insular.