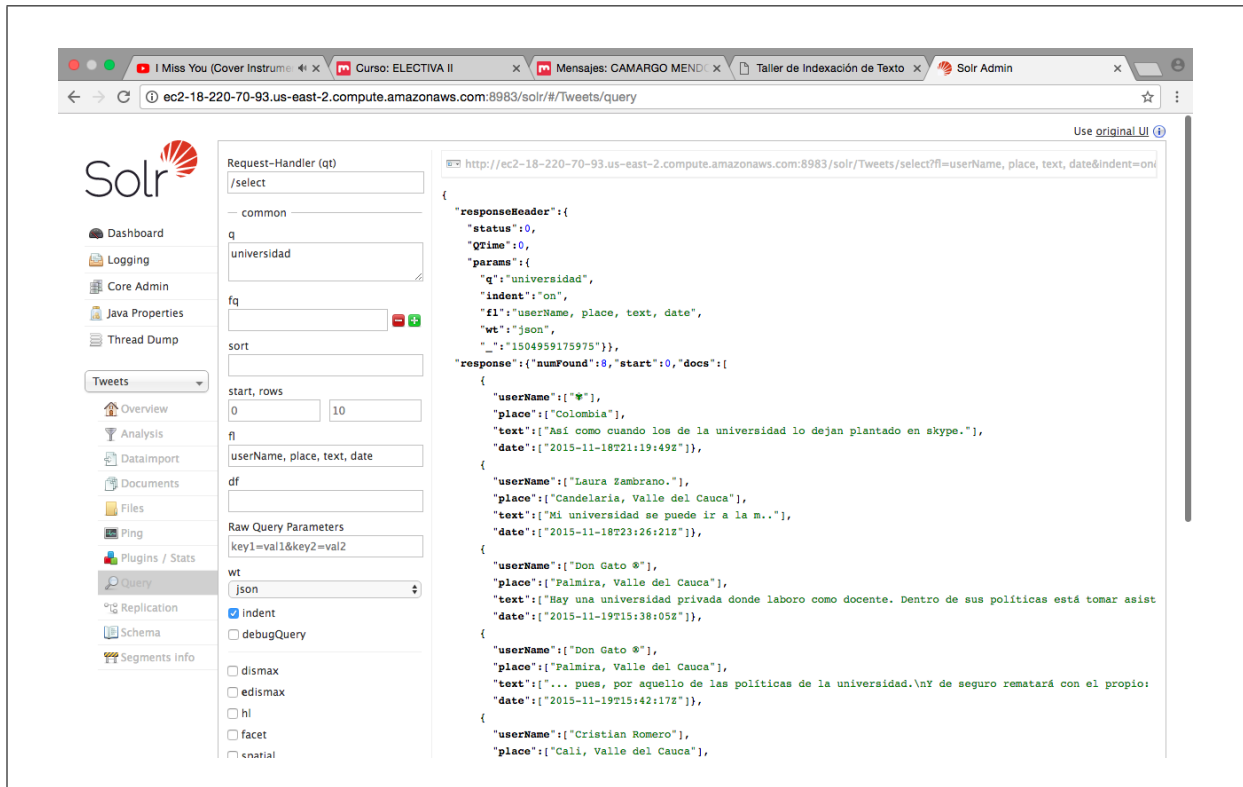


Viviana Márquez

Big Data - Prof. Camargo

Taller 5 - Septiembre 9, 2017

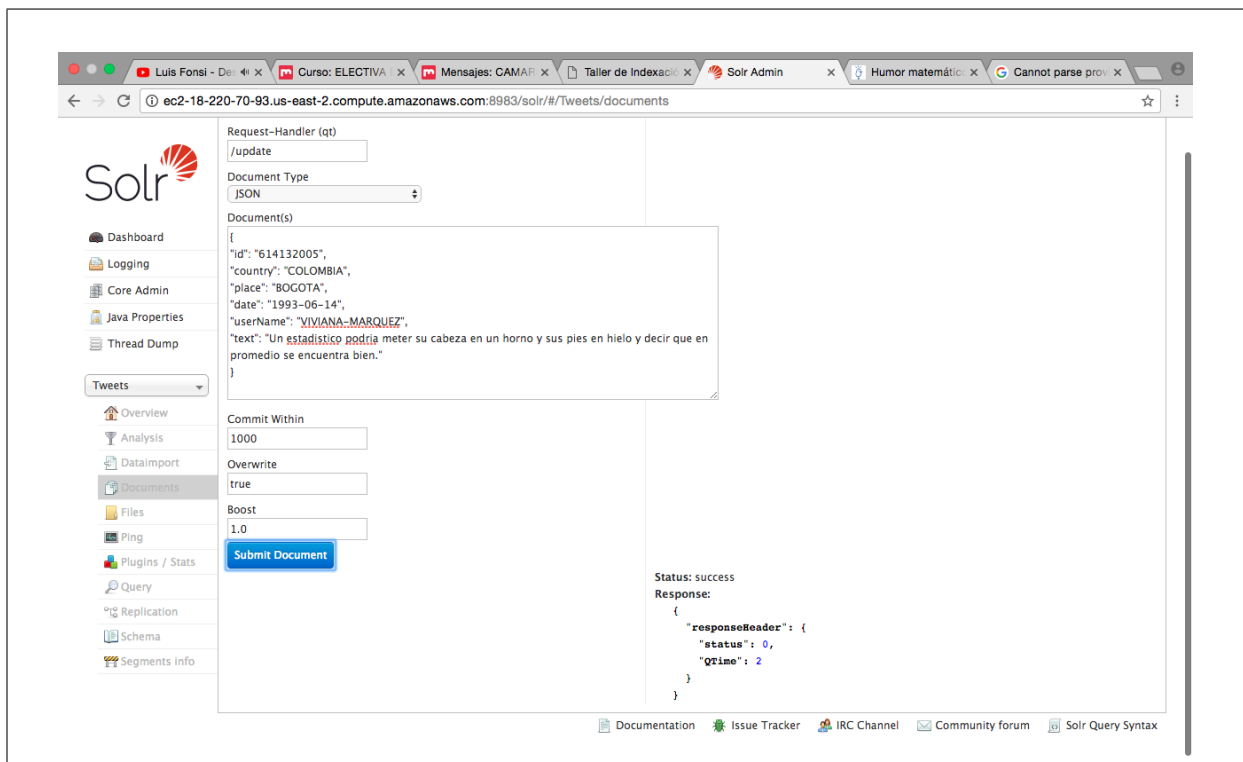
## 1. Primer paso: Realizar consulta en la colección Tweets.



The screenshot shows the Solr Admin interface with the 'Tweets' collection selected. The query is: `http://ec2-18-220-70-93.us-east-2.compute.amazonaws.com:8983/solr/#/Tweets/select?q=universidad&indent=on`. The response is a JSON array of documents, each representing a tweet. The first document is:

```
{
  "username": "Laura Zambrano",
  "place": "Candelaria, Valle del Cauca",
  "text": "Mi universidad se puede ir a la m..",
  "date": "2015-11-18T23:26:21Z"
}
```

## 2. Segundo paso: Añadir documento a la colección Tweets y verificar que está.

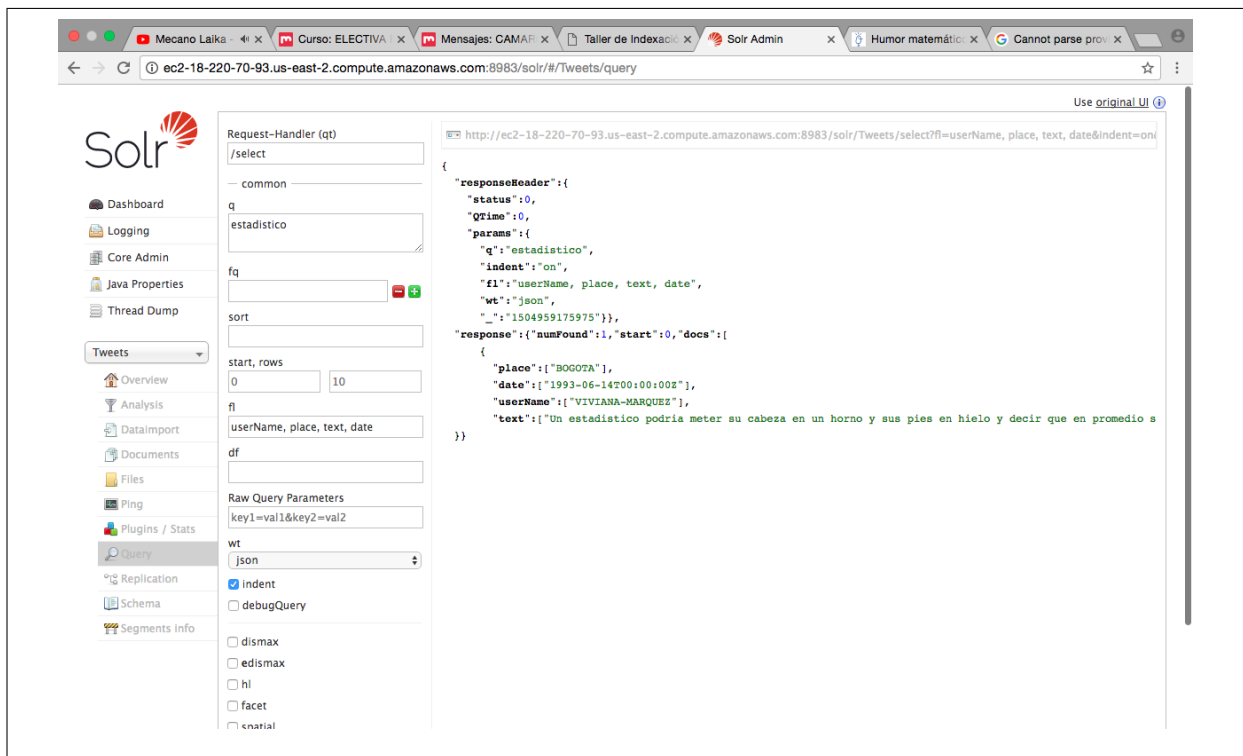


The screenshot shows the Solr Admin interface with the 'Documents' tab selected for the 'Tweets' collection. A new document is being added with the following fields:

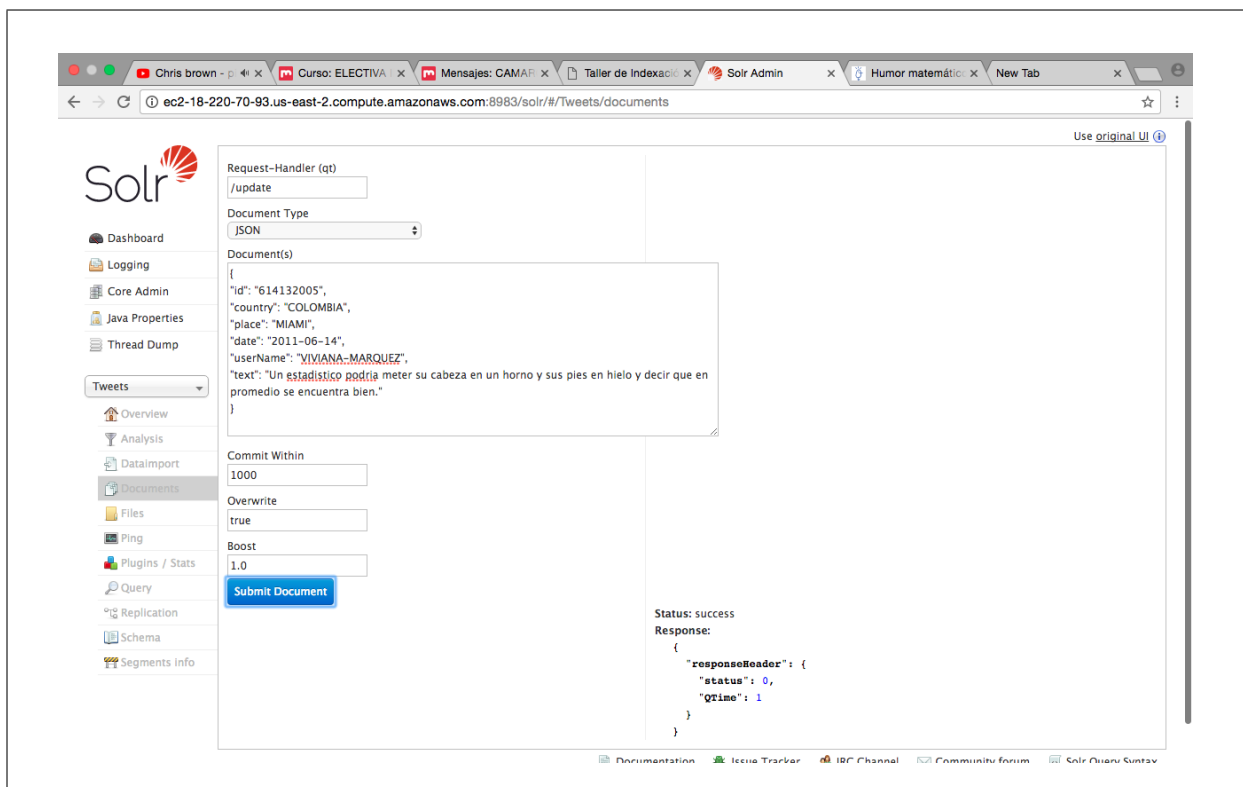
```
{
  "id": "614132005",
  "country": "COLOMBIA",
  "place": "BOGOTA",
  "date": "1993-06-14",
  "username": "VIVIANA-MARQUEZ",
  "text": "Un estadístico podría meter su cabeza en un horno y sus pies en hielo y decir que en promedio se encuentra bien."
}
```

The document is submitted successfully, and the response is:

```
{
  "responseHeader": {
    "status": 0,
    "QTime": 2
  }
}
```



3. Tercer paso: Actualizar documento en la colección Tweets y verificar que está.



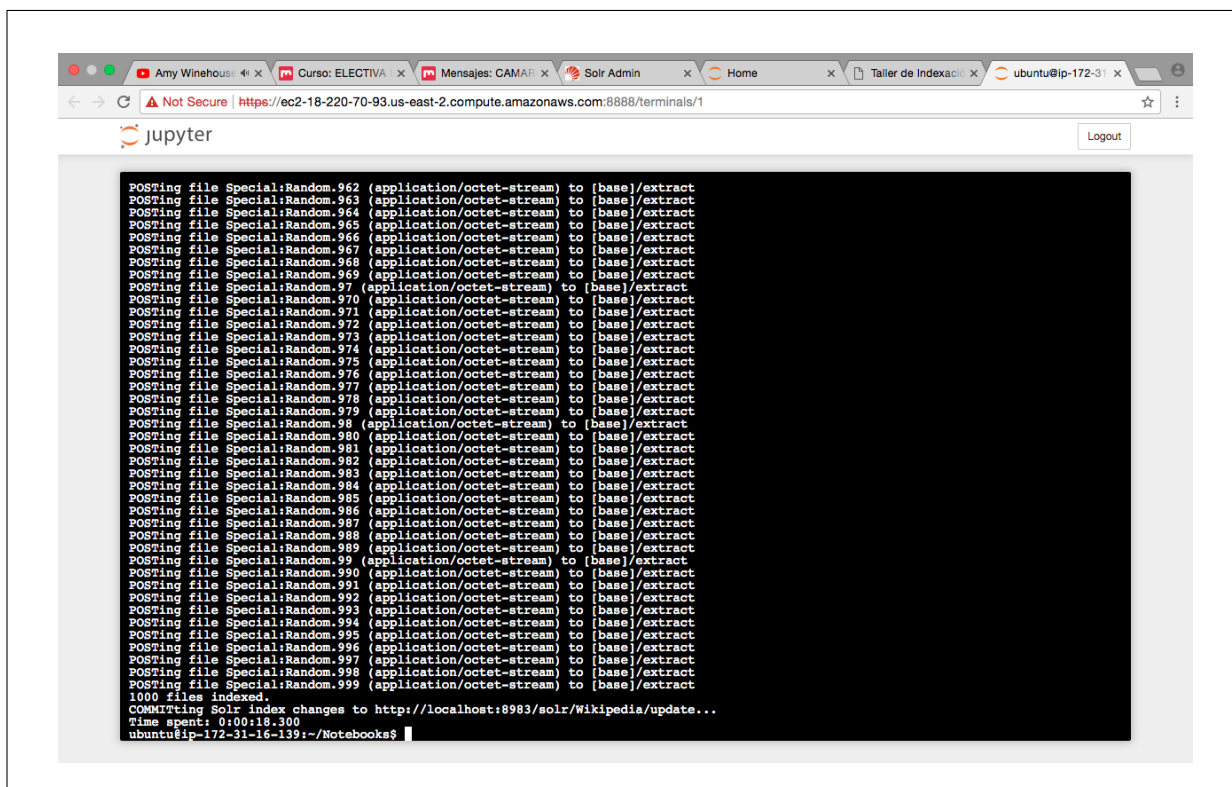
The screenshot shows the Solr Admin interface. On the left is a sidebar with navigation links: Dashboard, Logging, Core Admin, Java Properties, Thread Dump, Tweets (selected), Overview, Analysis, DataImport, Documents, Files, Ping, Plugins / Stats, Query, Replication, Schema, and Segments Info. The main area is divided into two panels. The left panel contains query configuration fields: Request-Handler (qt) set to /select, common, q set to 614132005, fq is empty, sort is empty, start/rows set to 0/10, fl set to userName, place, text, date, df is empty, Raw Query Parameters set to key1=val1&key2=val2, wt set to json, and checkboxes for indent (checked), debugQuery, dismax, edismax, hl, facet, and cnatral. The right panel shows the JSON response from the query, which includes a status of 0, qtime of 0, and a single document with fields for place (MIAMI), date (2011-06-14T00:00:00Z), username (VIVIANA-MARQUEZ), and text (Un estadístico podría meter su cabeza en un horno y sus pies en hielo y decir que en promedio s).

4. Cuarto paso: Eliminar documento en la colección Tweets desde la terminal y verificar que ya no está.

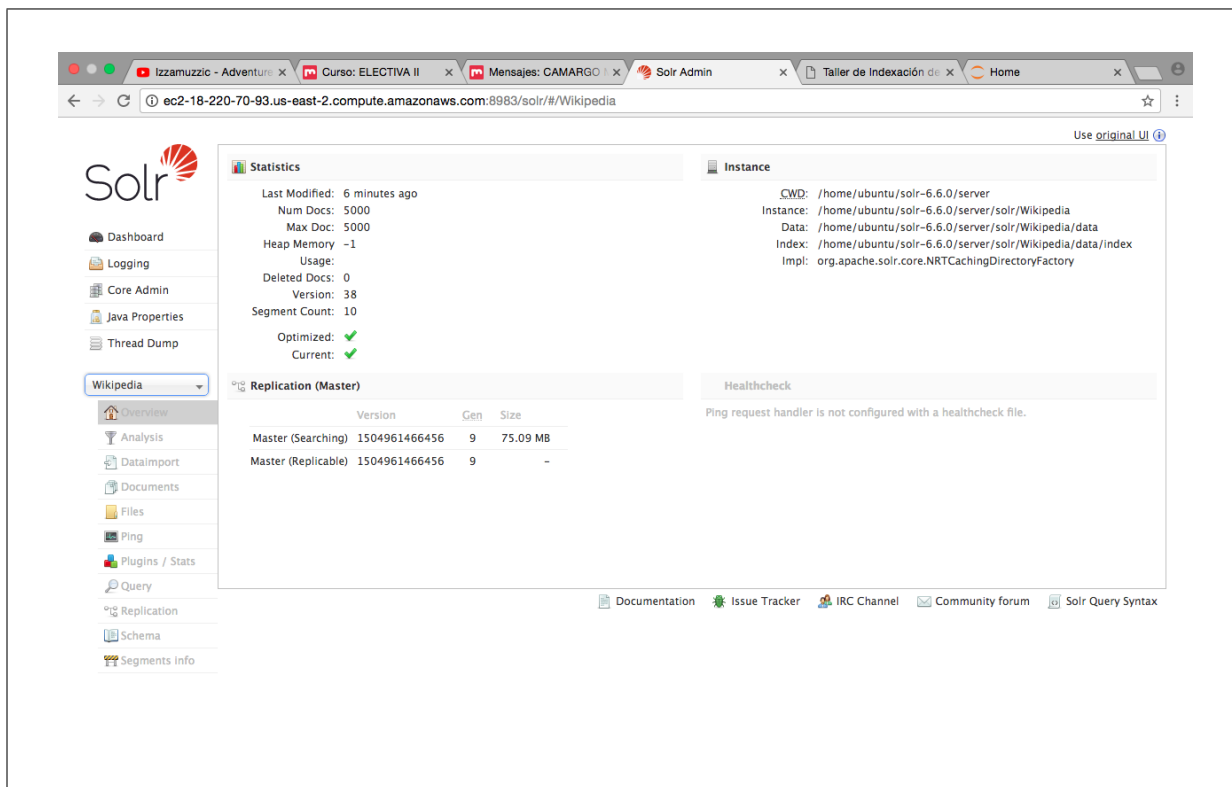
The screenshot shows a Jupyter Notebook interface with a terminal window. The terminal output shows the execution of a curl command to delete a document from the Solr index. The command is: `curl -XDELETE http://localhost:8983/solr/Tweets/update?qt=/select&indent=on&wt=json&key1=614132005`. The output shows the command being executed and the response being received.

The screenshot shows the Solr Admin interface after the deletion of the tweet. The query configuration fields are the same as in the previous screenshot. The JSON response in the right panel now shows "numFound": 0, indicating that no documents were found for the given query.

5. Insertar páginas descargadas en el taller de crawling en la colección Wikipedia a través del comando (ubicación) -c Wikipedia (ubicación taller)




6. Cantidad de documentos en la colección Wikipedia y hacer búsquedas.



La Reine De Saba x Curso: ELECTIVA II x Mensajes: CAMARGO x Solr Admin x Taller de Indexación de x Home x

ec2-18-220-70-93.us-east-2.compute.amazonaws.com:8983/solr/#/Wikipedia/query

Use original UI



- Dashboard
- Logging
- Core Admin
- Java Properties
- Thread Dump
- Wikipedia
  - Overview
  - Analysis
  - Dataimport
  - Documents
  - Files
  - Ping
  - Plugins / Stats
  - Query
  - Replication
  - Schema
  - Segments Info

Request-Handler (qt)

/select

common

q

1993

fq

sort

start, rows

0 10

fl

id, title

df

Raw Query Parameters

key1=val1&key2=val2

wt

json

☒ indent

☐ debugQuery

☐ dismax

☐ edismax

☐ hl

☐ facet

☐ enrtial

```
{
  "responseHeader": {
    "status": 0,
    "qtime": 9,
    "params": {
      "q": "1993",
      "indent": "on",
      "fl": "id, title",
      "wt": "json",
      "_": "1504959175975"
    }
  },
  "response": {
    "numFound": 331, "start": 0, "docs": [
      {
        "id": "/home/ubuntu/Notebooks/614132005/Taller1-02/Special:Random.48",
        "title": ["Esmer Yarim - Wikipedia"]
      },
      {
        "id": "/home/ubuntu/Notebooks/506161014/Taller1-02/Especial:Aleatoria.640",
        "title": ["Anexo:Asteroides (29001)-(30000) - Wikipedia, la enciclopedia libre"]
      },
      {
        "id": "/home/ubuntu/Notebooks/614132005/Taller1-02/Special:Random.54",
        "title": ["1993 Big Sky Conference Men's Basketball Tournament - Wikipedia"]
      },
      {
        "id": "/home/ubuntu/Notebooks/506142036/Taller1-02/Especial:Aleatoria.245",
        "title": ["Francisco Rojas Villegas - Wikipedia, la enciclopedia libre"]
      },
      {
        "id": "/home/ubuntu/Notebooks/506142036/Taller1-02/Especial:Aleatoria.390",
        "title": ["Nunca tuve tanto Blues - Wikipedia, la enciclopedia libre"]
      },
      {
        "id": "/home/ubuntu/Notebooks/614132005/Taller1-02/Special:Random.913",
        "title": ["Pakistani cricket team in the West Indies in 1992-93 - Wikipedia"]
      },
      {
        "id": "/home/ubuntu/Notebooks/614132005/Taller1-02/Special:Random.396",
        "title": ["Lassie (1954 TV series) - Wikipedia"]
      },
      {
        "id": "/home/ubuntu/Notebooks/614132005/Taller1-02/Special:Random.934",

```