

## Challenge 1

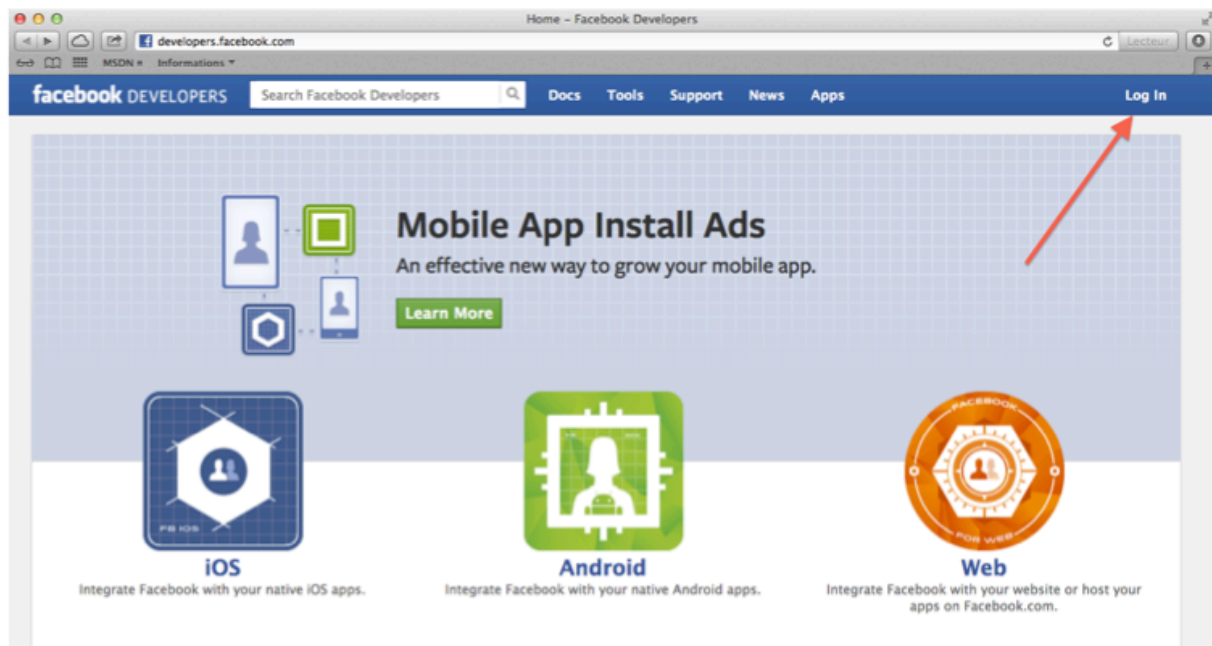
### Hadoop vs. Pig

#### Pre-requisites

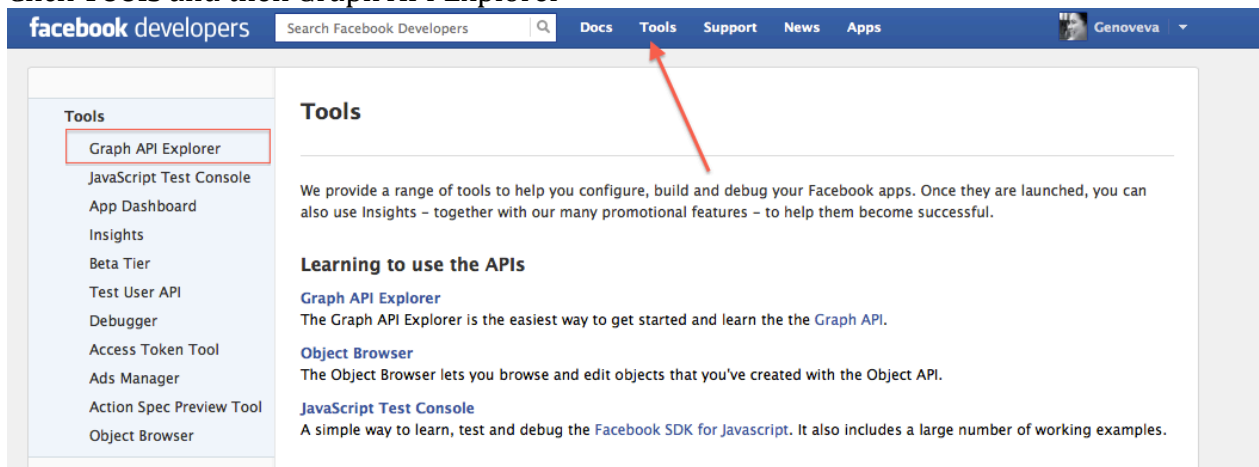
- Get Pig at <http://pig.apache.org/releases.html>
- <http://pig.apache.org/docs/r0.9.1/basic.html>
- <http://pig.apache.org/docs/r0.9.1/func.html>

#### Create a simple collection

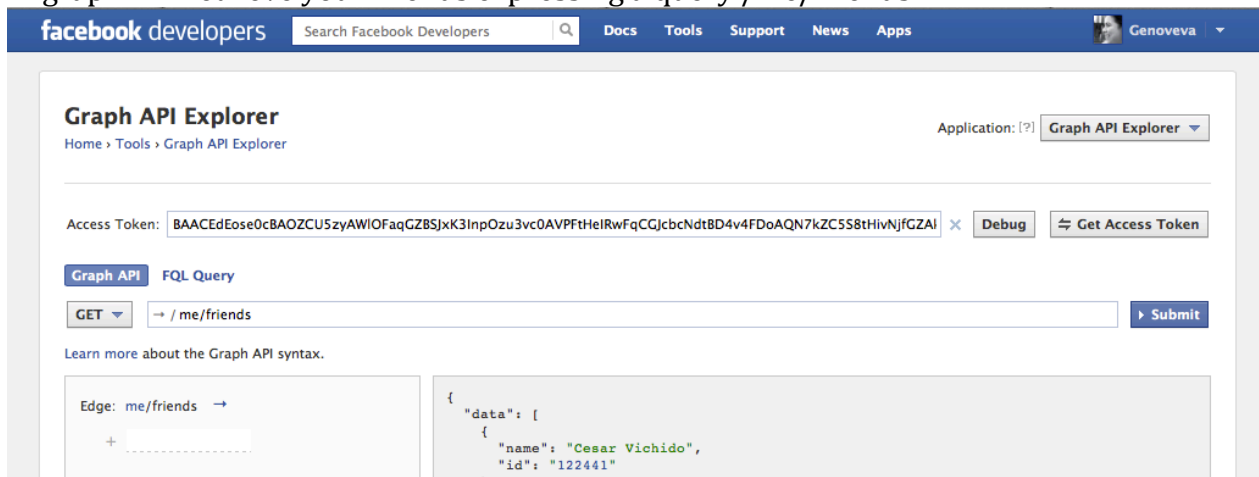
1. Go to <https://developers.facebook.com> and **sign in** into the portal with your Facebook account



2. Click **Tools** and then Graph API Explorer



3. In graph API retrieve your friends expressing a query `/me/friends`



4. Copy paste the list on a text file name it "Friends.txt"

5. Run Pig

- a. Mac users
  - `export HADOOP_OPTS="-Djava.security.krb5.realm=OX.AC.UK -Djava.security.krb5.kdc=kdc0.ox.ac.uk:kdc1.ox.ac.uk"`
  - `export JAVA_HOME=/Library/Java/Home`
  - `java -cp pig-0.11.1.jar org.apache.pig.Main -x local`

6. Define a first program for counting the words of your text:

```
A = LOAD 'Friends.txt' USING TextLoader() AS (words:chararray);
B = FOREACH A GENERATE FLATTEN(TOKENIZE(*));
C = GROUP B BY $0;
D = FOREACH C GENERATE group, COUNT(B);
STORE D INTO 'Friends3';
```