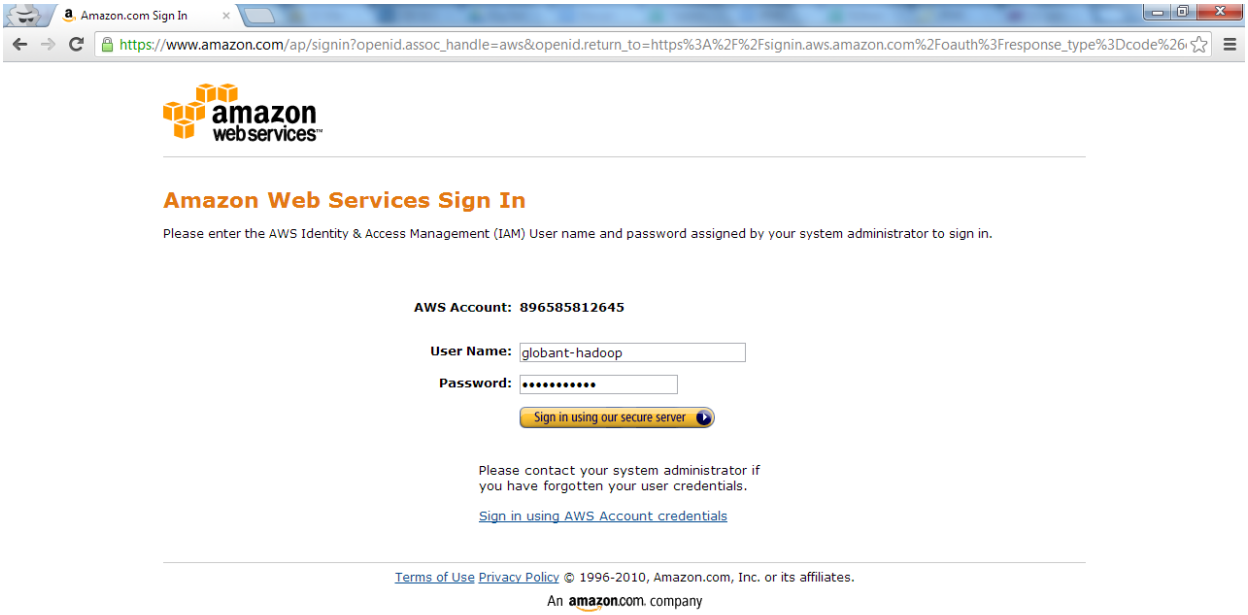
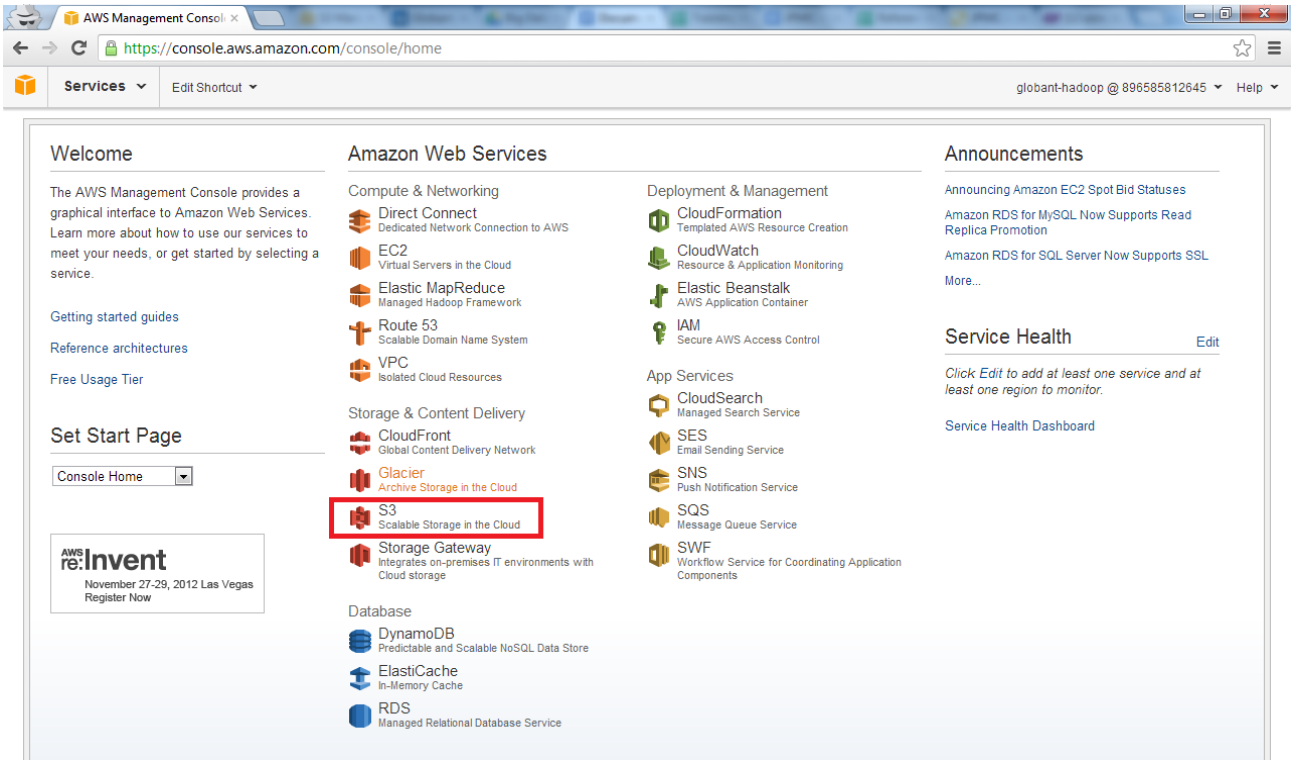
Tweet Wordcount MapReduce

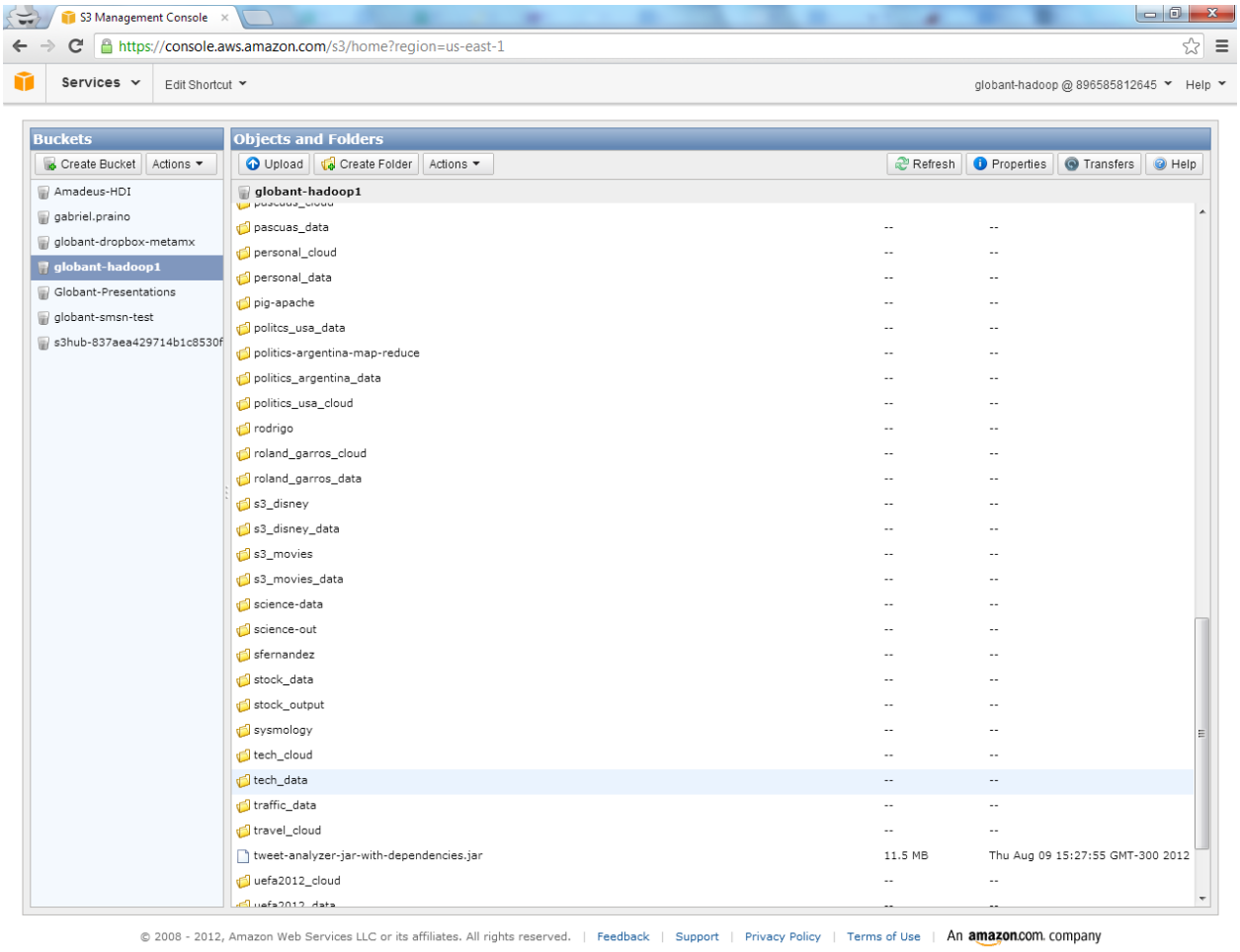
- Log into this Amazon WebServices console with these credentials
    - Username:‘globant-hadoop’
    - Password ‘Gl0b4nt2011’



- Navigate to the S3 section



- Open the ‘globant-hadoop1’ bucket
- Scroll down to the ‘tech-data’ folder



- Download any (or all) of the .json files.
- Load it into the HDFS at a destination folder of your choosing
- Create a User Java class which implements Writable and contains a ‘lang’ field and a ‘screen_name’ field.
- Create a Tweet Java class which implements Writable and contains a ‘text’ field, and a ‘user’ field (of type User).
- Write an InputFormat capable of parsing the .json files and for each tweet generating a Tweet object.
- Create a MapReduce job which processes Tweets from the .json files and outputs the english speaking users who mentioned one of the following terms:
    - apple
    - ios
    - itunes

- google
- android
- gmail
- microsoft
- windows
- hotmail