**gitt**

**INFRAESTRUCTURA BIG DATA CURSO BDBI2017**

La universidad cuenta con dos “infraestructuras Big Data” a las cuales se puede acceder únicamente desde dentro de la universidad:

1. **Cluster de Hadoop basado en Hortonworks**

El cluster actual son 2 nodos de 4 cpu y 6 gb de RAM cada uno

bigd-hadoop1 172.16.250.73

bigd-hadoop2 172.16.250.74

Los nombres de usuario son estudiante1, estudiante2, .....

La password de todos es password

**¿Cómo conectarse? Carece de interfaz visual, hacerlo vía SSH o putty**

**Via SSH:**

Descargar uno de estos programas:

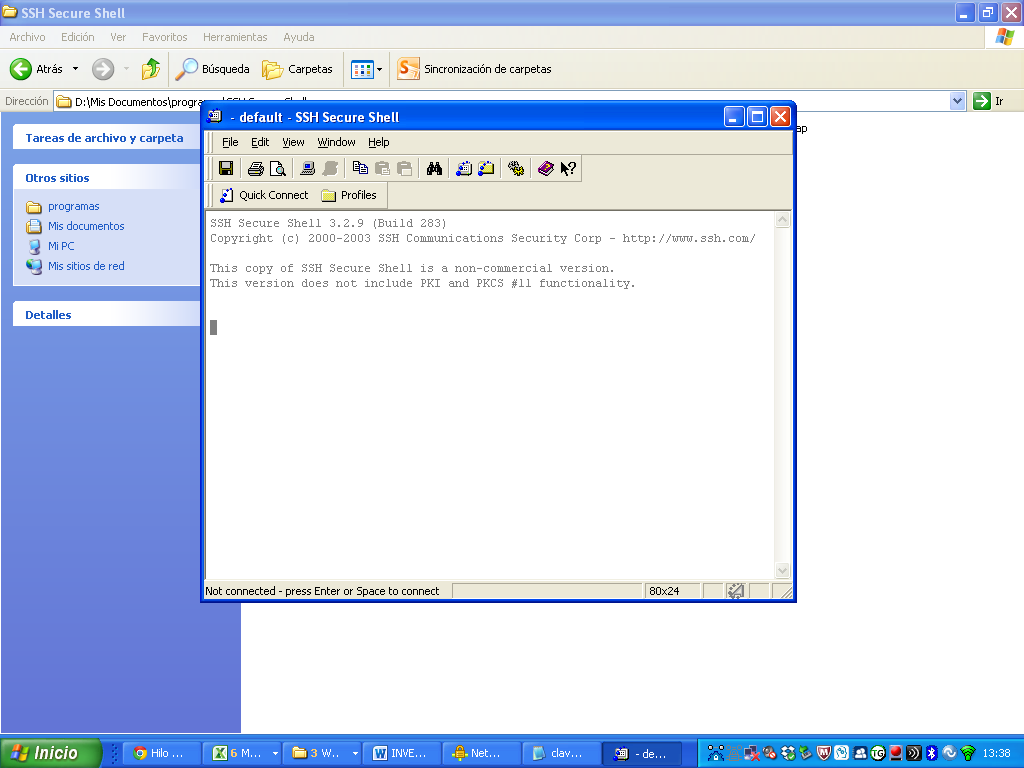
Putty:

https://www.google.es/url?sa=t&rct=j&q=&esrc=s&source=web&cd=1&cad=rja&uact=8&ved=0CCEQFjAA&url=http%3A%2F%2Fwww.chiark.greenend.org.uk%2F~sgtatham%2Fputty%2Fdownload.html&ei=keg0VbSfForsO4SfgLAG&usg=AFQjCNEawi7s0aRUeJP3qKnncgvPiSqZYA

SSH:

<https://www.dropbox.com/s/s93aqiu6ln77xuw/SshClient.exe?dl=0>

Ejemplo de acceso desde SSH:

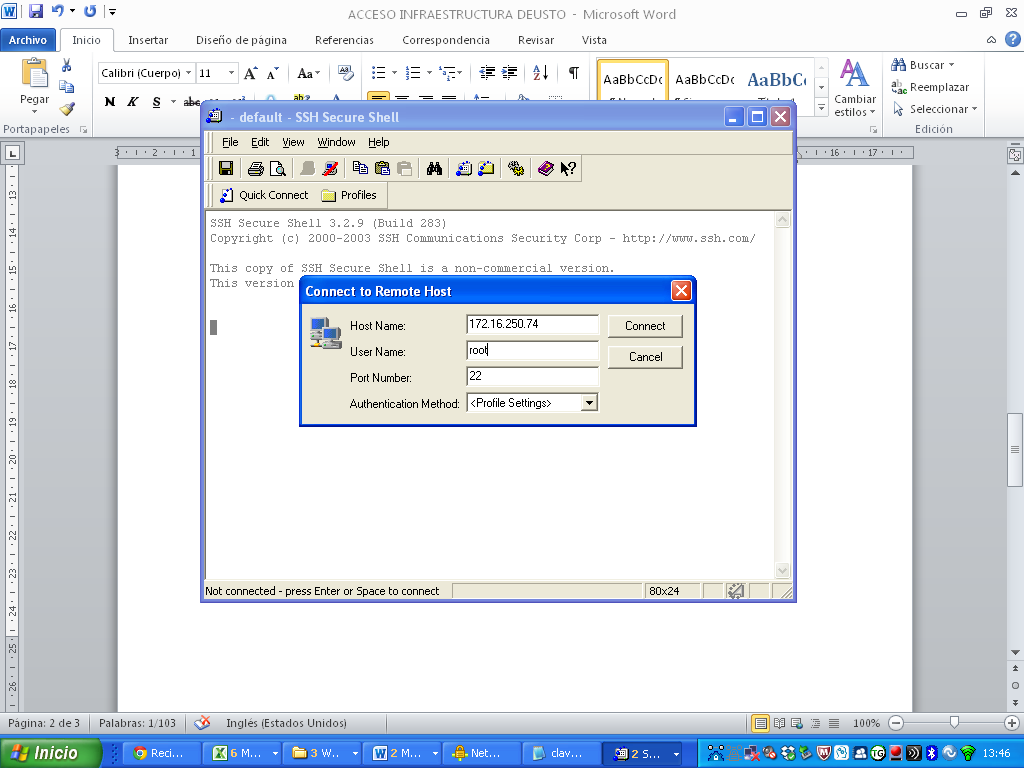


Pinchamos en Quick Connect y rellenamos con estos datos:

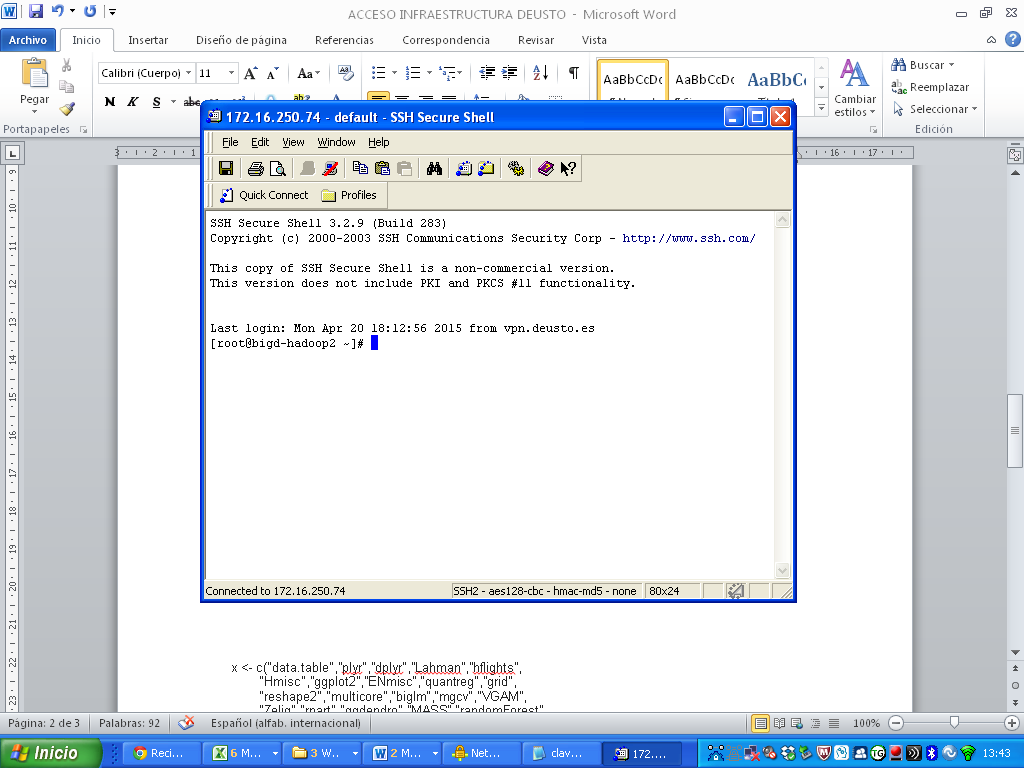
Hostname: 172.16.250.74

Username:

Password:



Pantalla una vez accedido:



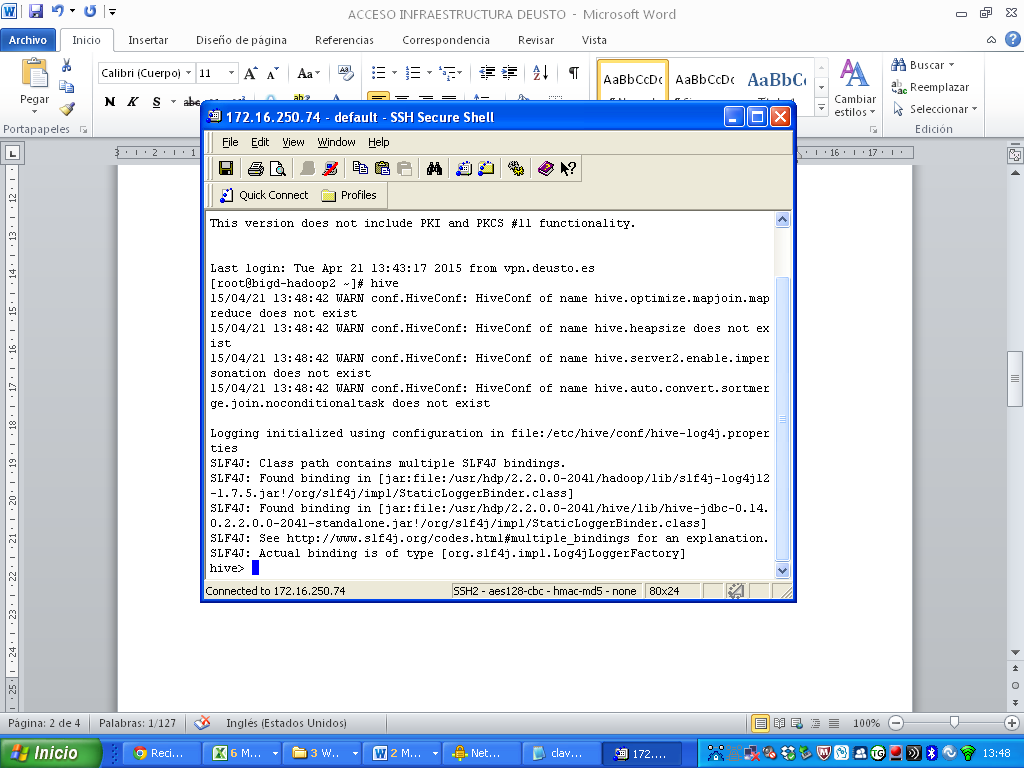
Una vez ahí, se pueden lanzar los servicios necesarios, por ejemplo, HIVE:

|  |
| --- |
| > hive |

Y se puede navegar por los schema y las tablas y hacer queries.

Nota: hay que poner ; al final.

|  |
| --- |
| > show schemas; |



1. **Cluster de Hadoop basado en Cloudera**

[bigd-cloudera.deusto.es](http://bigd-cloudera.deusto.es/)

IP 172.16.250.77

**¿Cómo acceder? Vía SSH-Putty o vía web**

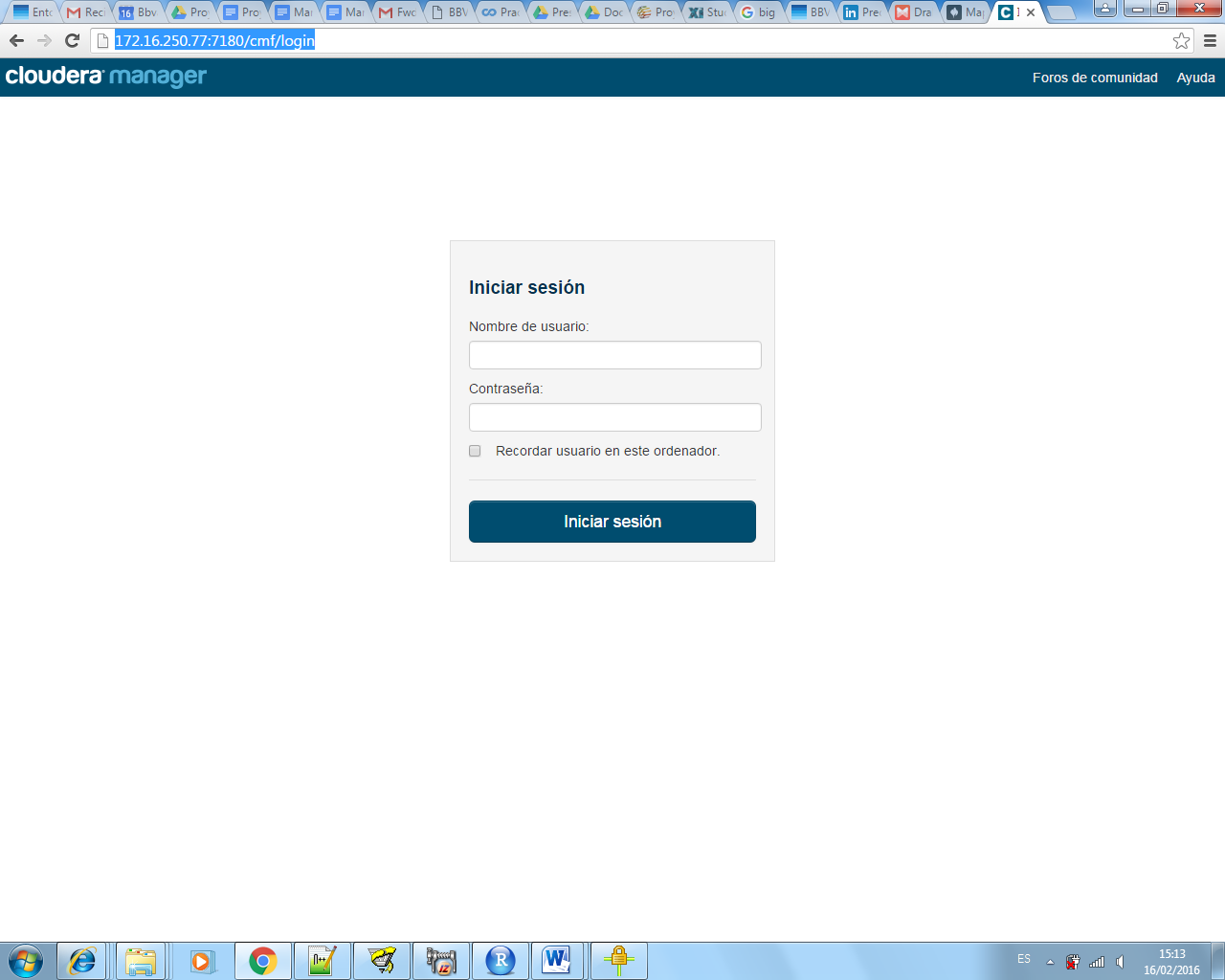
Usuarios SHELL

Password de todos los usuarios : password

estudiante1, estudiante2, .....

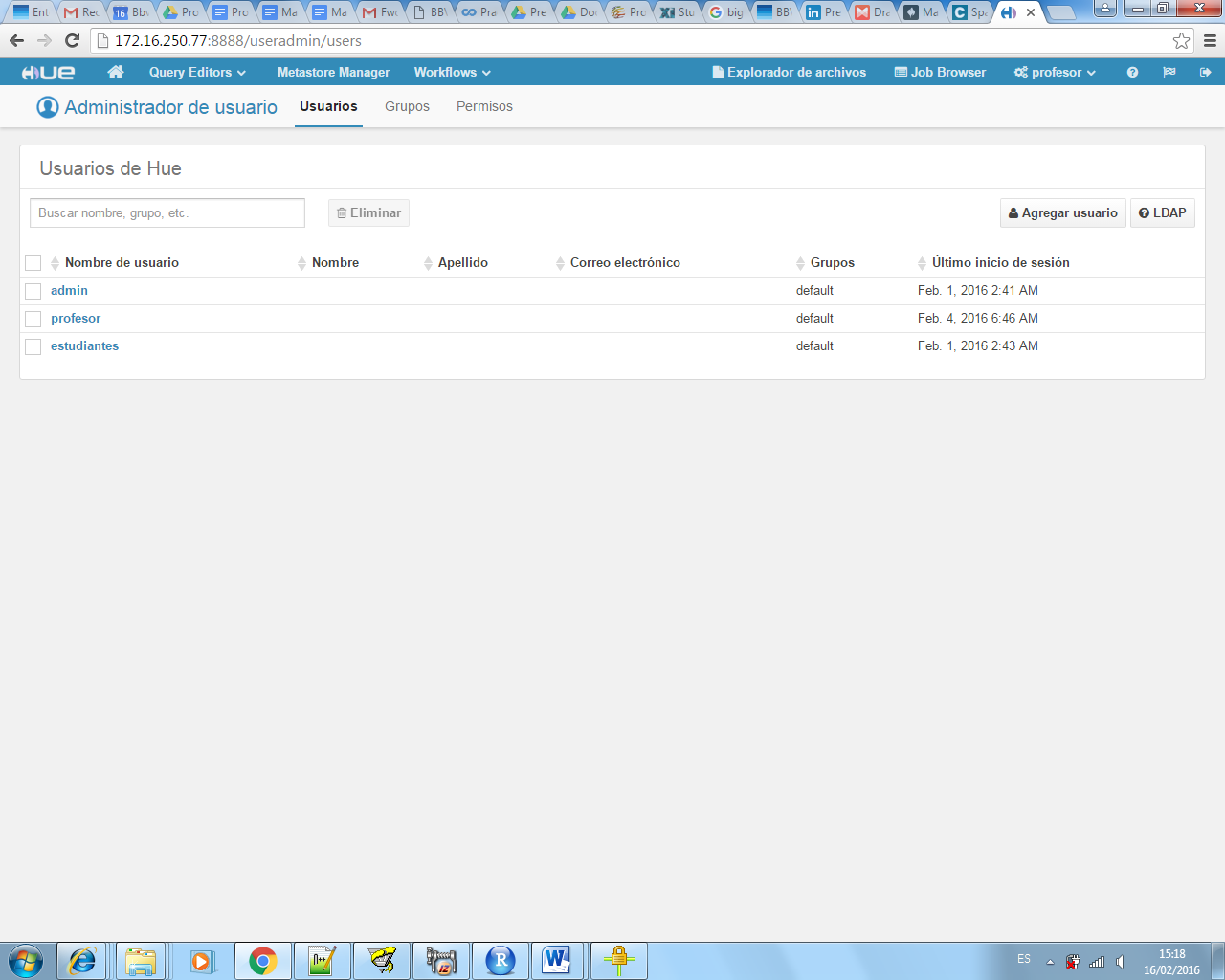
Vía WEB

**Cludera manager http://172.16.250.77:7180/cmf/login**



**Cloudera HUE** [**http://172.16.250.77:8888/**](http://172.16.250.77:8888/)

Usuarios:

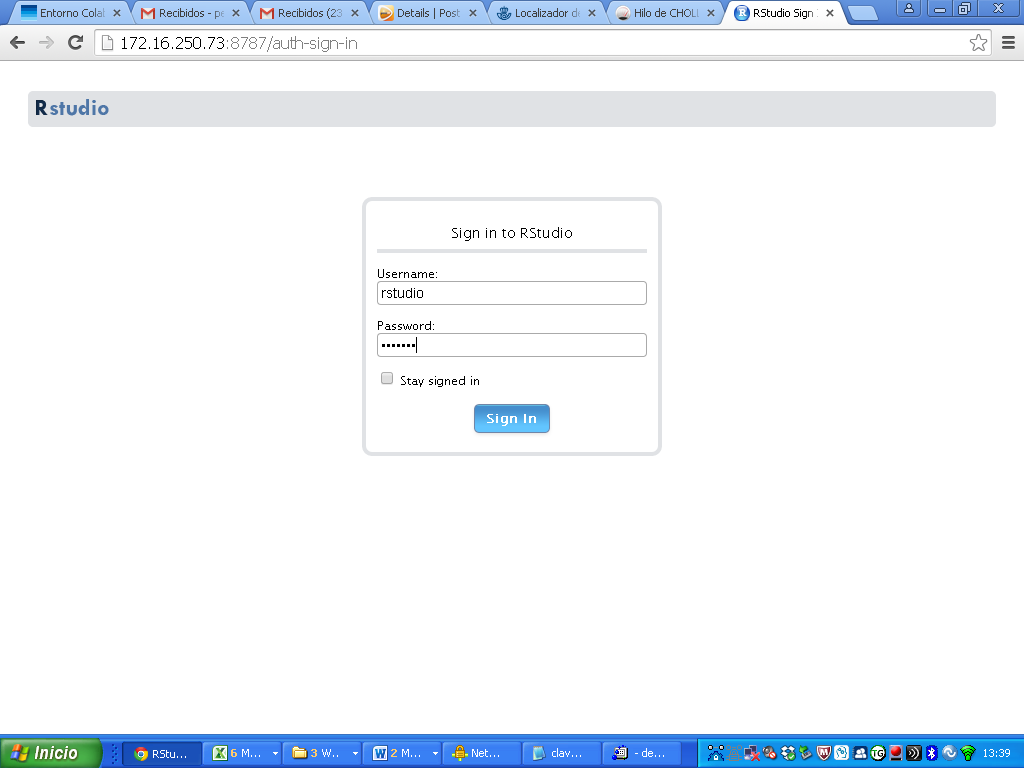


**ACCESO A RSTUDIO CLUSTER 1**

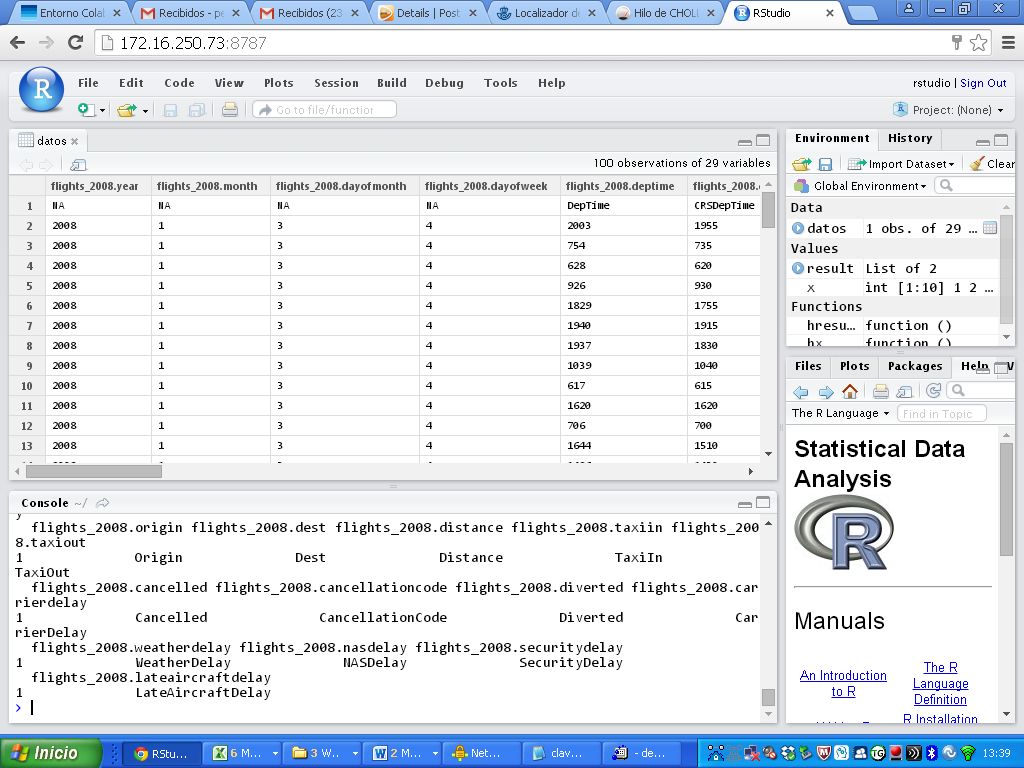
**Vía WEB:** http://172.16.250.73:8787/

Los nombres de usuario son estudiante1, estudiante2….

La password de todos es password



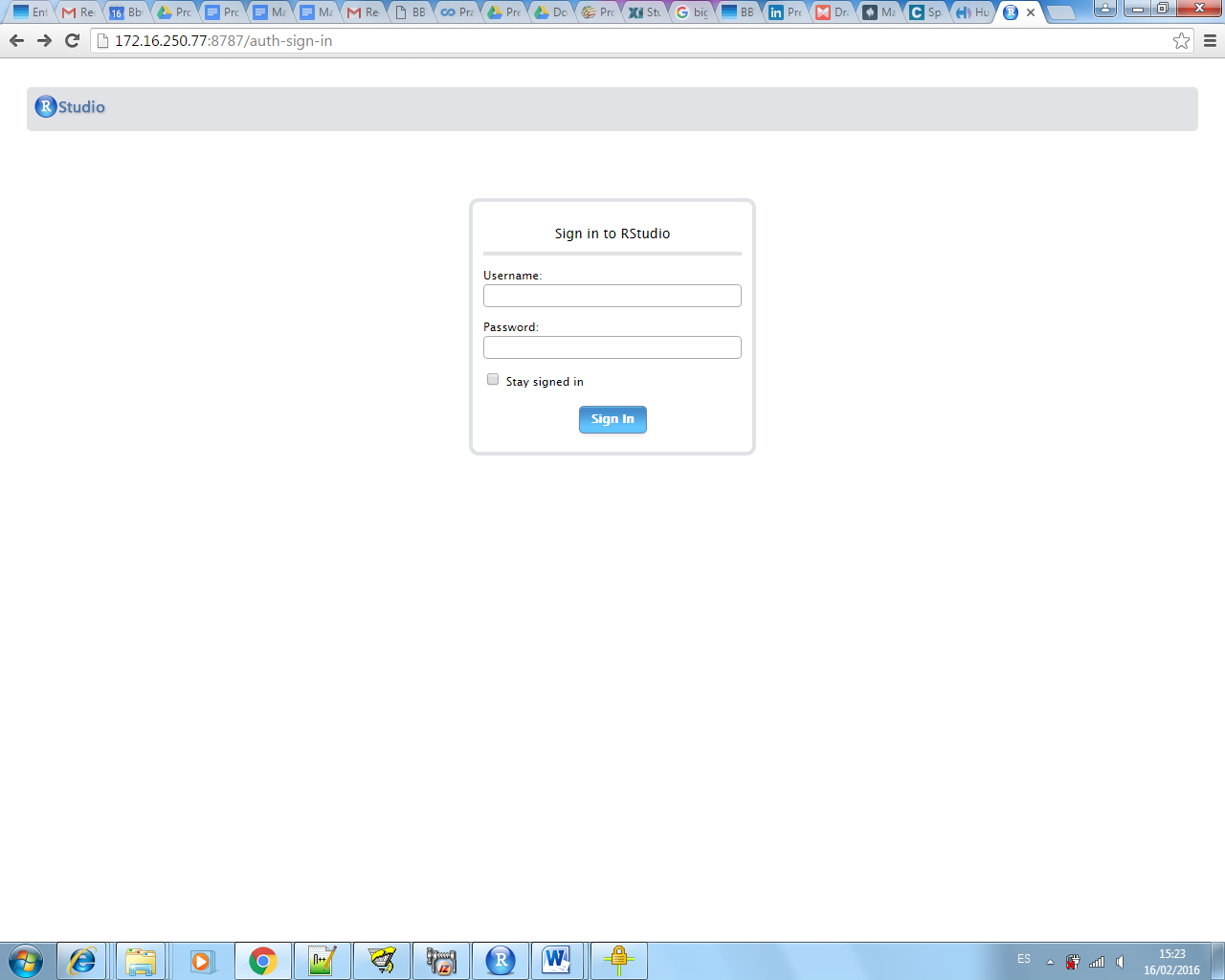
Pantalla una vez accedido:

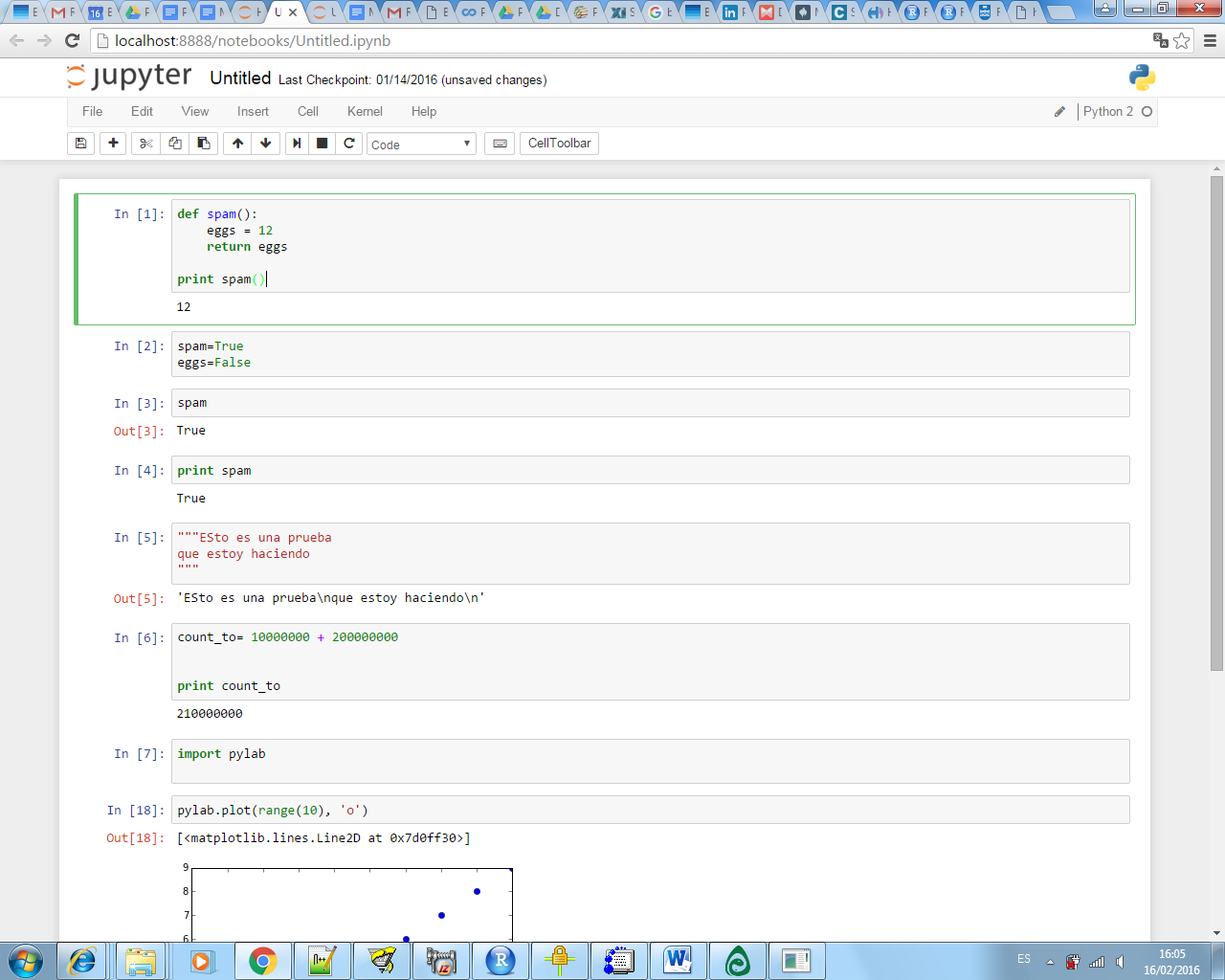


Durante el curso veremos cómo poder conectar R con los clusters de Hadoop.

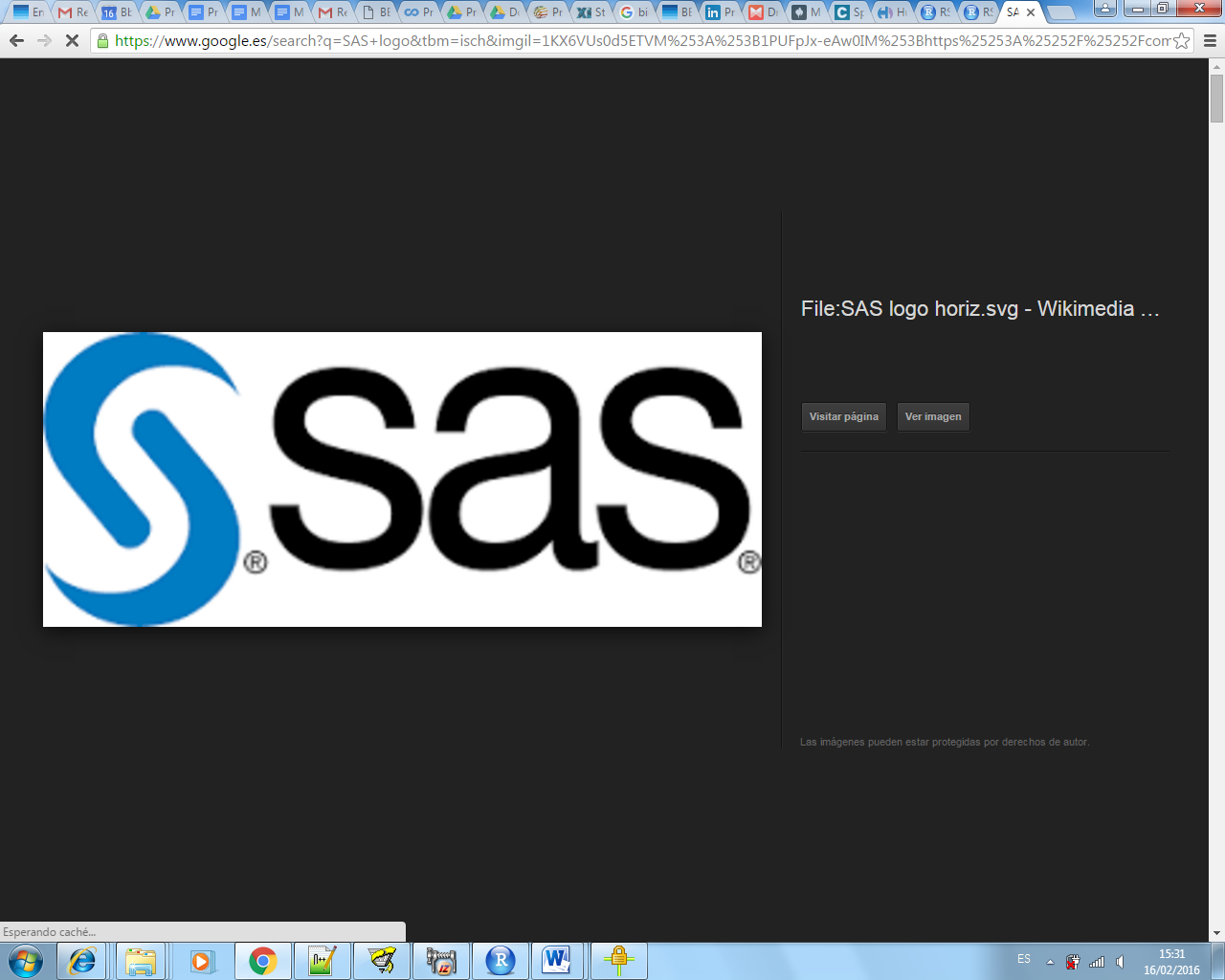
**ACCESO A RSTUDIO CLUSTER 2**

**http://172.16.250.77:8787/**





Aunque no sea objeto de materia del curso, los dos servidores 172.16.250.73 y **172.16.250.77** tienen instalado Python como alternativa a R.



Lo tenemos instalado en los ordenadores de la universidad.

Para poder disponer de él hay que entrar con un usuario y password específico al arrancar la sesión en el ordenador.