



Sesión 1: Big Data



Objetivos

1. **Conocer el clúster:** descubrir la manera de acceder al cluster y entender cómo darle el mejor uso durante las clases.
2. **Explorar el acceso a tecnologías Big Data:** acceder a las tecnologías y lenguajes como Spark, Hive, HBase, Sqoop y Python.

Especificaciones

El material está compuesto por imágenes y códigos necesarios para lograr los objetivos planteados. Adicionalmente pueden existir ejercicios para profundizar el aprendizaje los cuales quedan a criterio del alumno desarrollarlos para afianzar lo visto en clase.

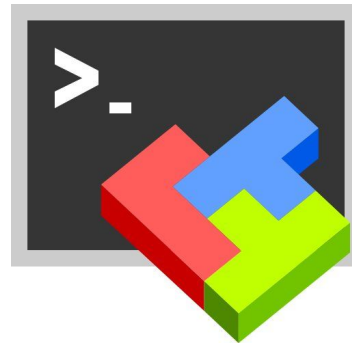
1. Accediendo al Cluster

Para acceder al cluster necesitas de un nombre de servidor o IP, además de un usuario y contraseña proporcionadas oportunamente.

- a. Primero necesitamos instalar un cliente en nuestra pc-windows para acceder remotamente, entre las opciones disponibles están Putty y MobaXterm.



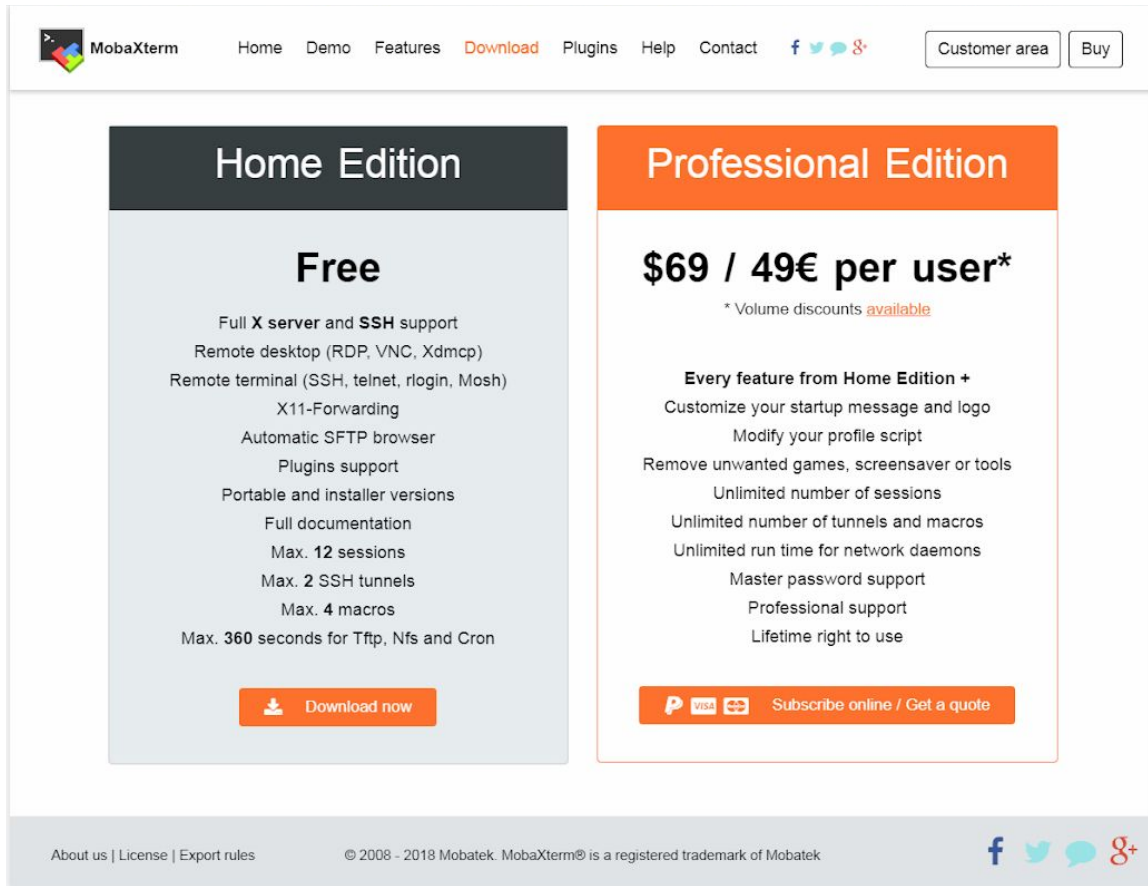
Putty



MobaXterm

El material también está disponible en
<https://github.com/BigDataAcademyPeru/sesion1-BigData>

- b. Trabajaremos con MobaXterm así que la descargamos desde <https://mobaxterm.mobatek.net/download.html> eligiendo la versión portable



The screenshot shows the MobaXterm website with a navigation bar at the top containing links: Home, Demo, Features, Download, Plugins, Help, Contact, and social media icons. There are also buttons for 'Customer area' and 'Buy'.

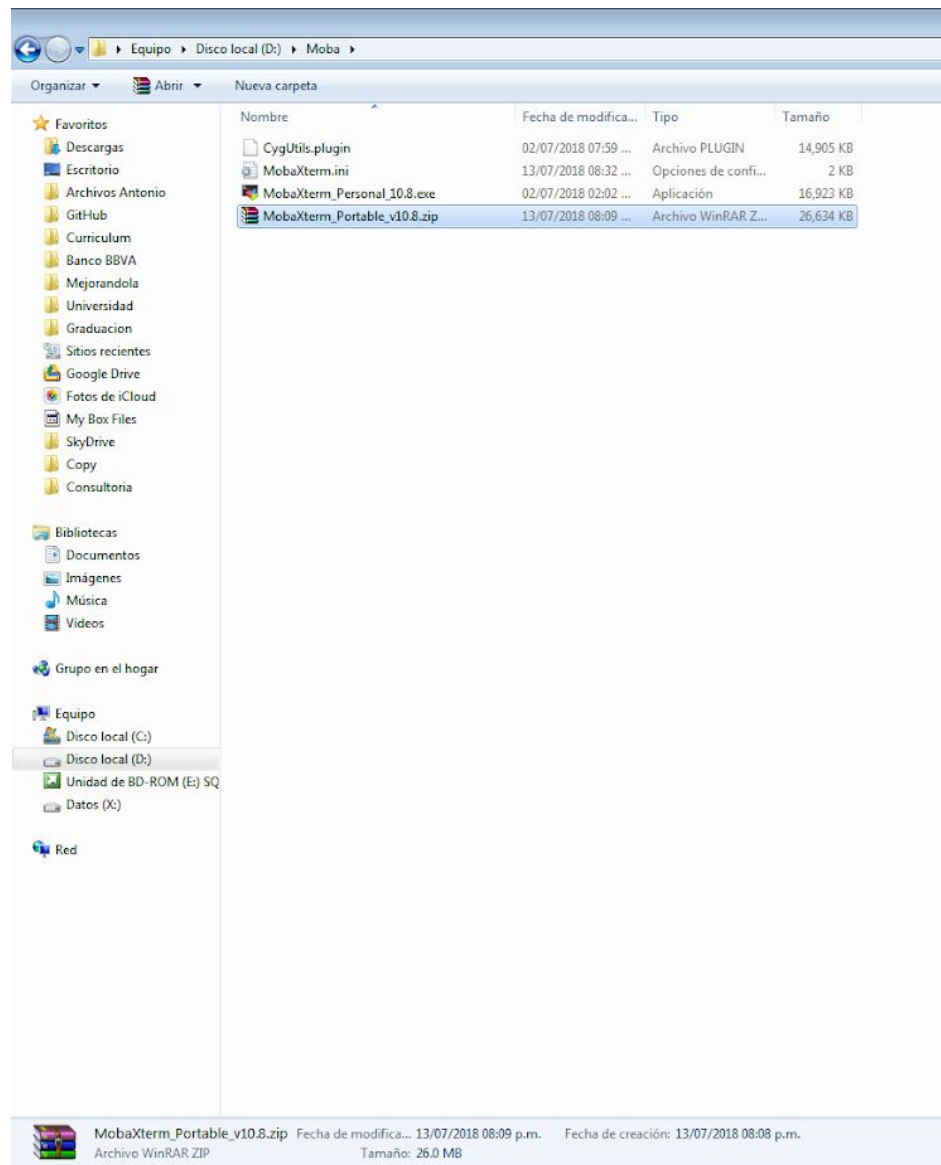
The main content area features two columns:

- Home Edition** (dark grey header):
 - Free**
 - Full **X server** and **SSH** support
 - Remote desktop (RDP, VNC, Xdmcp)
 - Remote terminal (SSH, telnet, rlogin, Mosh)
 - X11-Forwarding
 - Automatic SFTP browser
 - Plugins support
 - Portable and installer versions
 - Full documentation
 - Max. **12** sessions
 - Max. **2** SSH tunnels
 - Max. **4** macros
 - Max. **360** seconds for Tftp, Nfs and Cron
 - [Download now](#) button
- Professional Edition** (orange header):
 - \$69 / 49€ per user***
 - * Volume discounts [available](#)
 - Every feature from Home Edition +**
 - Customize your startup message and logo
 - Modify your profile script
 - Remove unwanted games, screensaver or tools
 - Unlimited number of sessions
 - Unlimited number of tunnels and macros
 - Unlimited run time for network daemons
 - Master password support
 - Professional support
 - Lifetime right to use
 - [Subscribe online / Get a quote](#) button with payment icons (P, VISA, MC)

The footer contains links for 'About us | License | Export rules', copyright information '© 2008 - 2018 Mobatek. MobaXterm® is a registered trademark of Mobatek', and social media icons for Facebook, Twitter, LinkedIn, and Google+.

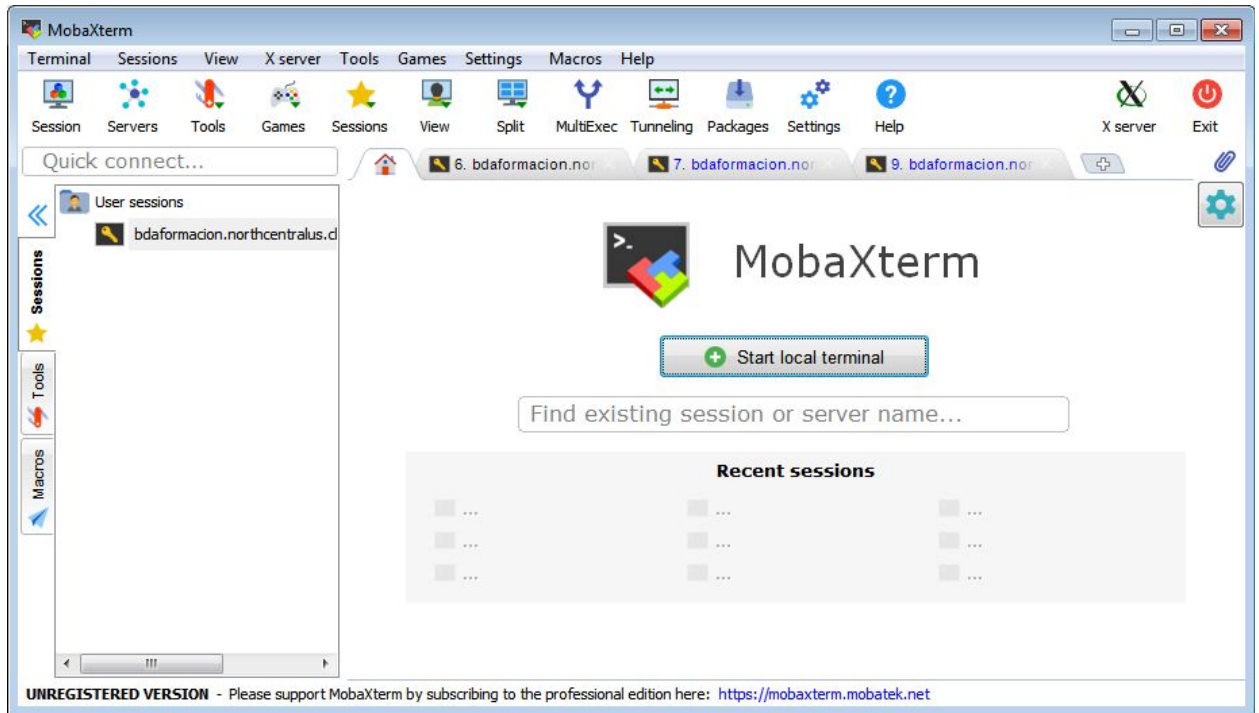
El material también está disponible en
<https://github.com/BigDataAcademyPeru/sesion1-BigData>

c. Guardar el .zip en una carpeta fácil de ubicar y descomprimirlo



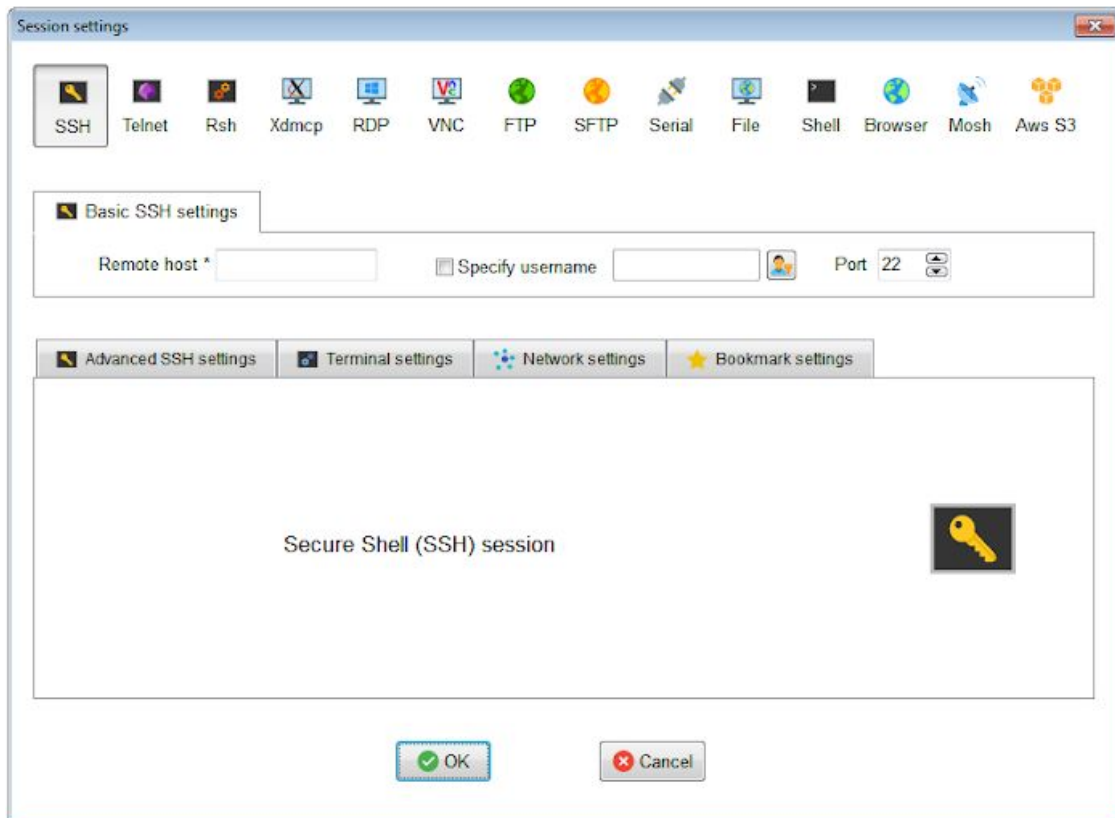
El material también está disponible en
<https://github.com/BigDataAcademyPeru/sesion1-BigData>

- d. Ejecutamos “MobaXterm_Personal_10.8.exe” y aparecerá la pantalla de bienvenida

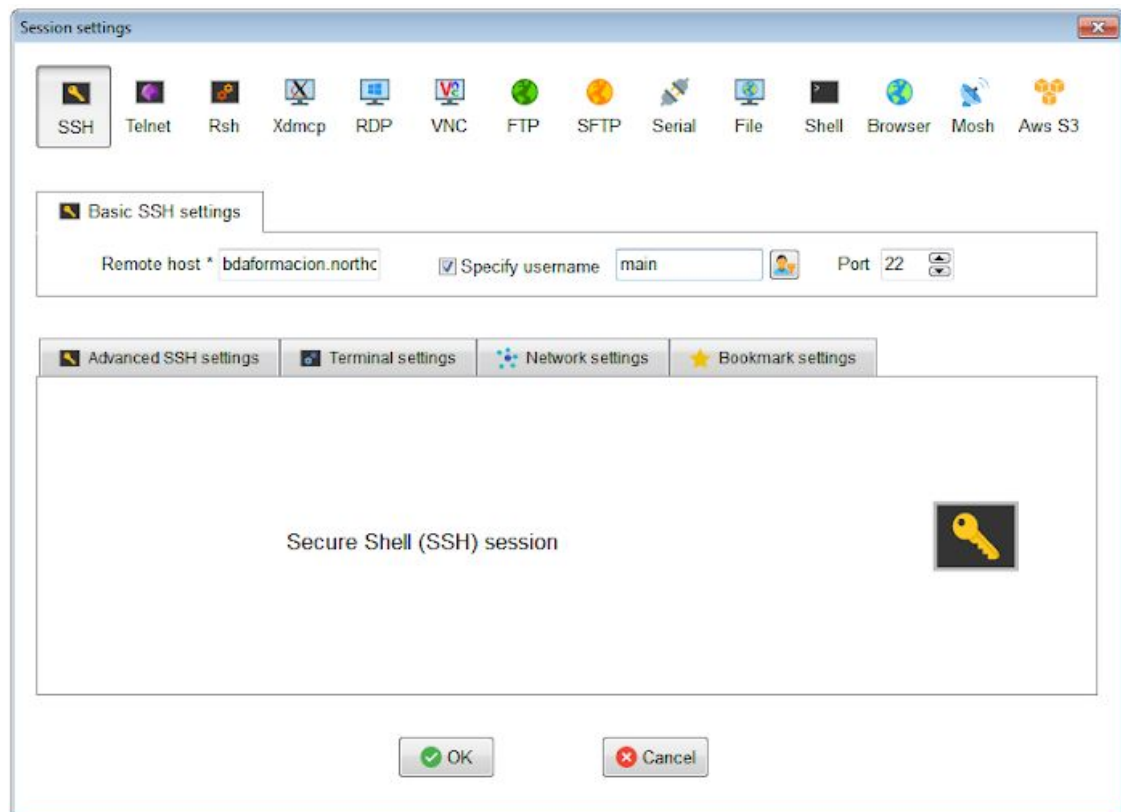


El material también está disponible en
<https://github.com/BigDataAcademyPeru/sesion1-BigData>

- e. Buscamos el botón **Session** para que nos habrá una ventana en la cuál damos clic en **SSH**

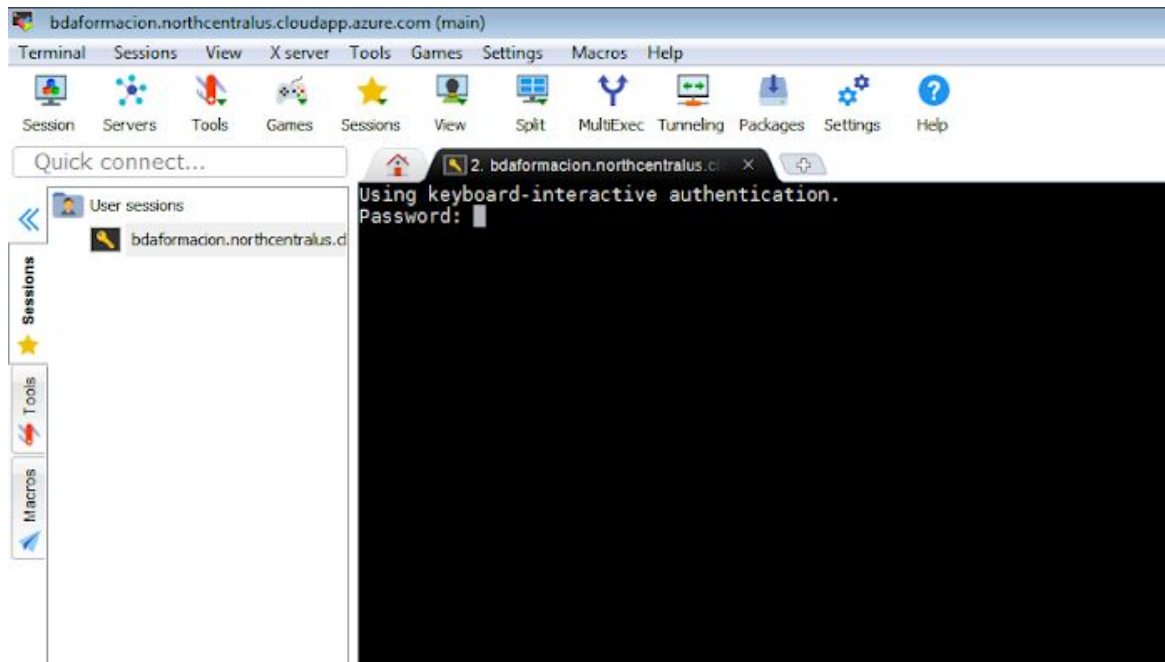


- f. En la misma ventana ingresamos nuestras credenciales Host (ip/nombre del servidor) y especificamos nuestro username



El material también está disponible en
<https://github.com/BigDataAcademyPeru/sesion1-BigData>

- g. Se abrirá una pantalla pidiéndonos ingresar la contraseña

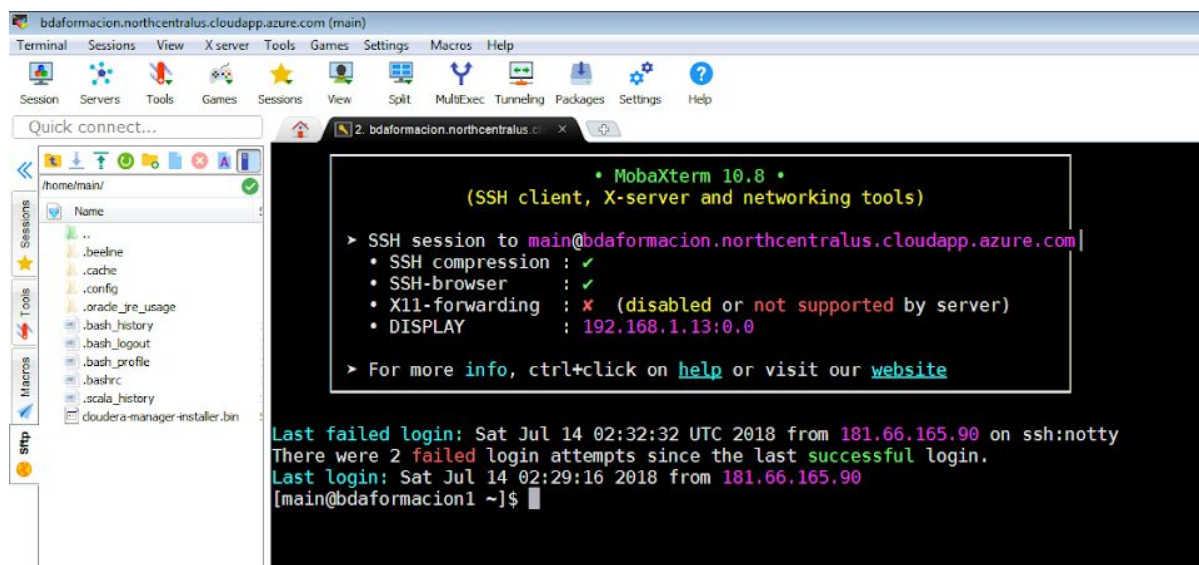


El material también está disponible en
<https://github.com/BigDataAcademyPeru/sesion1-BigData>

- h. Aparecerá una ventana pidiendote guardar la contraseña, le damos **Yes**



- i. En la ventana se mostrará en la parte izquierda se pueden observar los archivos y directorios de la carpeta actual. La pantalla negra es la consola o terminal por donde interactuamos con nuestro clúster.

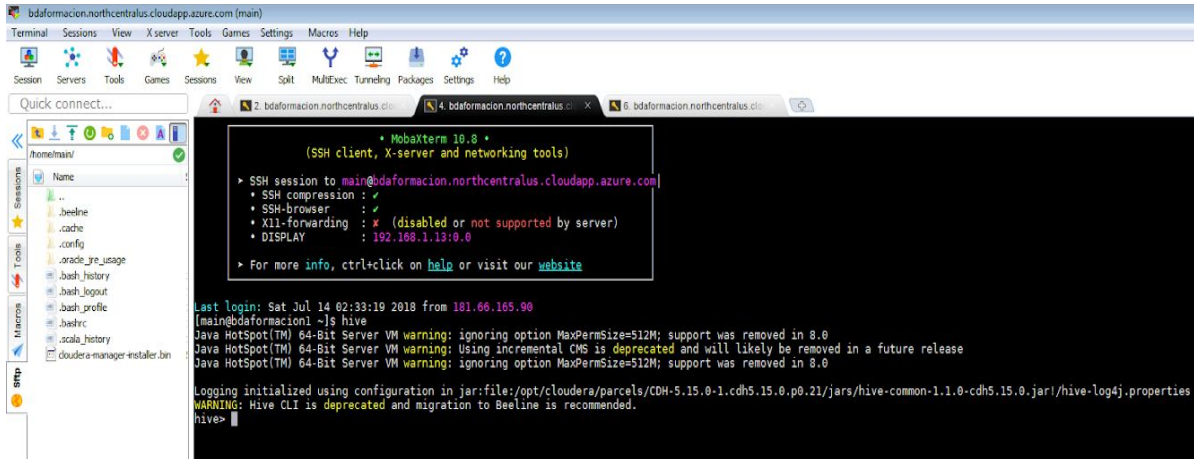


El material también está disponible en
<https://github.com/BigDataAcademyPeru/sesion1-BigData>

b. Apache Hive

Para acceder a Hive después de logearnos escribimos

```
> hive
```



```
bdaformacion.northcentralus.cloudapp.azure.com (main)
Terminal Sessions View X server Tools Games Settings Macros Help
Session Servers Tools Games Sessions View Split MultiExec Tunneling Packages Settings Help
Quick connect...
/home/main/
Sessions
  Name
  ...
  .beeline
  .cache
  .config
  .coracle_gre_usage
  .bash_history
  .bash_logout
  .bash_profile
  .bashrc
  .local_history
  cloudera-manager-installer.bin
Tools
Macros
SSH

* MobaXterm 10.8 *
(SSH client, X-server and networking tools)
> SSH session to main@bdaformacion.northcentralus.cloudapp.azure.com
  * SSH compression : ✓
  * SSH-browser : ✓
  * X11-forwarding : x (disabled or not supported by server)
  * DISPLAY : 192.168.1.13:0.0
  * For more info, ctrl+click on help or visit our website

Last login: Sat Jul 14 02:33:19 2018 from 101.66.165.90
[main@bdaformacion1 ~]$ hive
Java HotSpot(TM) 64-Bit Server VM warning: ignoring option MaxPermSize=512M; support was removed in 8.0
Java HotSpot(TM) 64-Bit Server VM warning: Using incremental CMS is deprecated and will likely be removed in a future release
Java HotSpot(TM) 64-Bit Server VM warning: ignoring option MaxPermSize=512M; support was removed in 8.0

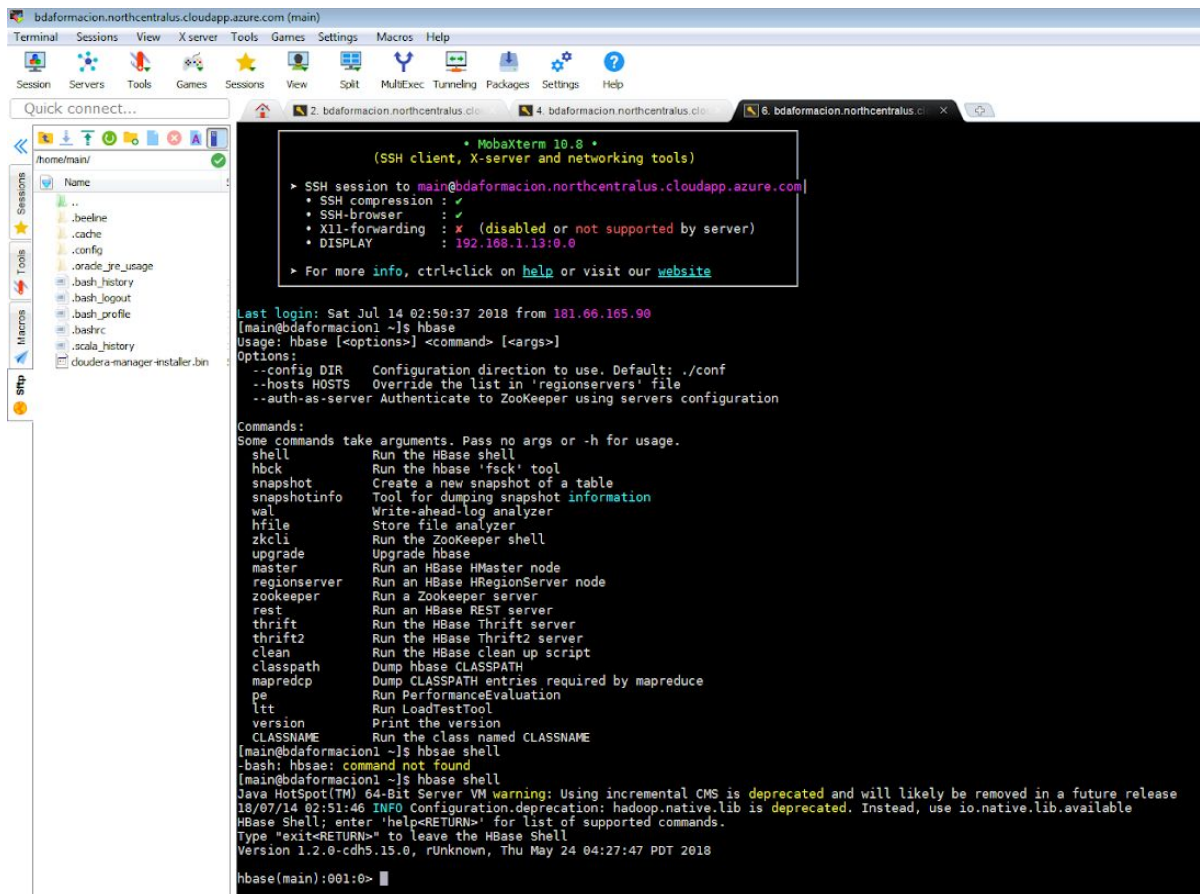
Logging initialized using configuration in jar:file:/opt/cloudera/parcels/CDH-5.15.0-1.cdh5.15.0.p0.21/jars/hive-common-1.1.0-cdh5.15.0.jar!/hive-log4j.properties
WARNING: Hive CLI is deprecated and migration to Beeline is recommended.
hive>
```

El material también está disponible en
<https://github.com/BigDataAcademyPeru/sesion1-BigData>

c. Hbase

Para acceder a Hbase después de logearnos escribimos

`hbase shell`



```

bdaformacion.northcentralus.cloudapp.azure.com (main)
Terminal Sessions View X-server Games Settings Macros Help
Session Servers Tools Games Sessions View Split MultiExec Tunneling Packages Settings Help

Quick connect...

/home/main/
Sessions
  Name
  ..
  .beeline
  .cache
  .config
  .oracle_jre_usage
  .bash_history
  .bash_logout
  .bash_profile
  .bashrc
  .scala_history
  cloudera-manager-installer.bin
Tools
Macros
SFTP

+ MobaXterm 10.8 +
(SSH client, X-server and networking tools)

> SSH session to main@bdaformacion.northcentralus.cloudapp.azure.com
+ SSH compression : ✓
+ SSH-browser : ✓
+ X11-forwarding : ✗ (disabled or not supported by server)
+ DISPLAY : 192.168.1.13:0.0
> For more info, ctrl+click on help or visit our website

Last login: Sat Jul 14 02:50:37 2018 from 181.66.165.90
[main@bdaformacion1 ~]$ hbase
Usage: hbase [<options>] [<command>] [<args>]
Options:
  --config DIR      Configuration direction to use. Default: ./conf
  --hosts HOSTS     Override the list in 'regionservers' file
  --auth-as-server  Authenticate to ZooKeeper using servers configuration

Commands:
Some commands take arguments. Pass no args or -h for usage.
shell              Run the HBase shell
hbck               Run the hbase 'fsck' tool
snapshot           Create a new snapshot of a table
snapshotinfo       Tool for dumping snapshot information
wal                Write-ahead-log analyzer
hfile              Store file analyzer
zkcli              Run the ZooKeeper shell
upgrade            Upgrade hbase
master             Run an HBase HMaster node
regionserver       Run an HBase HRegionServer node
zookeeper          Run a Zookeeper server
rest              Run an HBase REST server
thrift             Run the HBase Thrift server
thrift2            Run the HBase Thrift2 server
clean              Run the HBase clean up script
classpath          Dump hbase CLASSPATH
mapredcp           Dump CLASSPATH entries required by mapreduce
pe                Run PerformanceEvaluation
ltt               Run LoadTestTool
version            Print the version
CLASSNAME          Run the class named CLASSNAME
[main@bdaformacion1 ~]$ hbase shell
.bash: hbase: command not found
[main@bdaformacion1 ~]$ hbase shell
Java HotSpot(TM) 64-Bit Server VM warning: Using incremental CMS is deprecated and will likely be removed in a future release
18/07/14 02:51:46 INFO Configuration.deprecation: hadoop.native.lib is deprecated. Instead, use io.native.lib.available
HBase Shell; enter 'help<RETURN>' for list of supported commands.
Type "exit<RETURN>" to leave the HBase Shell
Version 1.2.0-cdh5.15.0, rUnknoWn, Thu May 24 04:27:47 PDT 2018

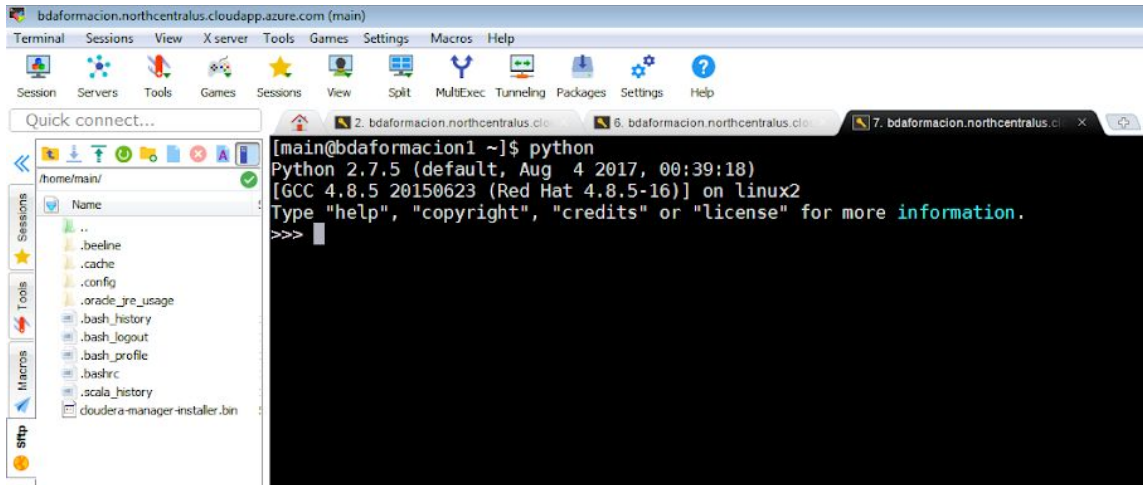
hbase(main):001:0>
  
```

El material también está disponible en
<https://github.com/BigDataAcademyPeru/sesion1-BigData>

d. Python

Para acceder a Python ingresamos en la terminal

```
> python
```



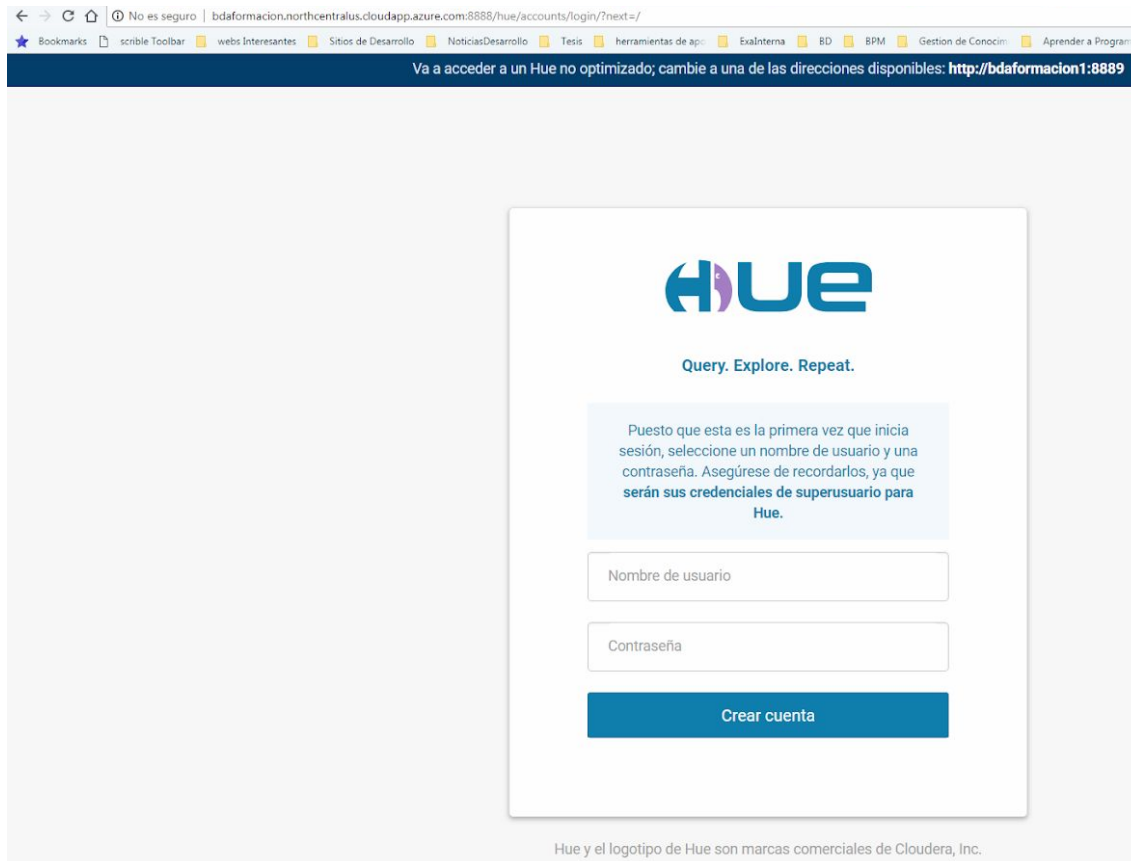
```
[main@bdaformacion1 ~]$ python
Python 2.7.5 (default, Aug 4 2017, 00:39:18)
[GCC 4.8.5 20150623 (Red Hat 4.8.5-16)] on linux2
Type "help", "copyright", "credits" or "license" for more information.
python>
```

El material también está disponible en
<https://github.com/BigDataAcademyPeru/sesion1-BigData>

e. Hue

Para acceder a Hue ingresamos al navegador y entramos a

<http://bdaformacion.northcentralus.cloudapp.azure.com:8888>



El material también está disponible en
<https://github.com/BigDataAcademyPeru/sesion1-BigData>