

Anexo 9: Manual de instalación Apache Pig

Descarga e instalación:

```
bigtexts@bigtexts-1:~$ wget http://apache.mirrors.pair.com/pig/latest/pig-0.13.0.tar.gz
bigtexts@bigtexts-1:~/Downloads$ sudo tar -xvzf pig-0.13.0.tar.gz -C /usr/local/
bigtexts@bigtexts-1:~/Downloads$ cd /usr/local/
bigtexts@bigtexts-1:~/Downloads$ sudo ln -s pig-0.13.0 pig
hduser@bigtexts-1:/usr/local$ sudo chown -R hduser:hadoop pig
hduser@bigtexts-1:/usr/local$ sudo chown -R hduser:hadoop pig-0.13.0
```

```
bigtexts@bigtexts-1:~$ cd ~
bigtexts@bigtexts-1:~$ nano .bashrc
```

```
#Pig
export PIG_INSTALL=/usr/local/pig
export PATH=$PATH:$PIG_INSTALL/bin
```

Cerrar la consola y volverla a abrir

```
bigtexts@bigtexts-1:~$ su -l hduser
bigtexts@bigtexts-1:~$ pig version
bigtexts@bigtexts-1:~$ pig
```

Pig en maven:

Para que funcionara pig en maven se adiciona el repositorio

```
<repositories>
  <repository>
    <id>central</id>
    <url>http://repo1.maven.org/maven2/</url>
  </repository>
</repositories>
```

luego se adiciona la dependencia

```
<dependency>
  <groupId>org.apache.pig</groupId>
  <artifactId>pig</artifactId>
```

```
<version>0.13.0</version>
</dependency>
```

Pero dicho jar está compilado para hadoop 1.x y no para hadoop 2.x, para lo cual lo que hacemos es que tomamos del directorio de instalación de pig \$PIG_INSTALL el jar pig-0.13.0-withouthadoop-h2.jar que es el que si está compilado para hadoop 2.x y reemplazamos el que está en /home/hduser/.m2/repository/org/apache/pig/pig/0.13.0/, con esto ya no se generan errores de incompatibilidad entre las librerías de hadoop-client y pig

```
hduser@smaug:~/Downloads$ cd $PIG_INSTALL
hduser@smaug:/usr/local/pig$ cp pig-0.13.0-withouthadoop-h2.jar
/home/hduser/.m2/repository/org/apache/pig/pig/0.13.0/
hduser@smaug:/usr/local/pig$ cd
/home/hduser/.m2/repository/org/apache/pig/pig/0.13.0/
hduser@smaug:~/.m2/repository/org/apache/pig/pig/0.13.0$ mv
pig-0.13.0-withouthadoop-h2.jar pig-0.13.0.jar
```

Instalar Hadoop local (Necesario para poder contar con el protocolo hdfs en el cliente)

```
bigtexts@bigtexts-1:~$ wget
http://apache.spinellicreations.com/hadoop/common/stable/hadoop-2.5.0.tar.gz
```

```
bigtexts@bigtexts-1:/usr/local$ sudo tar -xvzf hadoop-2.5.0.tar.gz -C /usr/local/
bigtexts@bigtexts-1:~$ cd /usr/local/
bigtexts@bigtexts-1:/usr/local$ sudo mv hadoop-2.5.0/ hadoop
bigtexts@bigtexts-1:~$ nano .bashrc
bigtexts@bigtexts-1:~$ nano .profile
```

```
export HADOOP_INSTALL=/usr/local/hadoop
export PATH=$HADOOP_INSTALL/bin:$PATH
```

Configuración Pig a un Hadoop remoto

<http://netezzaadmin.wordpress.com/2013/09/24/configuring-pig-to-work-with-a-remote-hadoop-cluster/>

En el servidor:

```
bigtexts@bigtexts-1:/usr/local/hadoop/etc/hadoop$ cd /usr/local/hadoop/etc/hadoop/  
bigtexts@bigtexts-1:/usr/local/hadoop/etc/hadoop$ sudo tar -czvf client.tar.gz core-site.xml  
hadoop-env.sh hdfs-site.xml log4j.properties mapred-site.xml ssl-client.xml.example  
bigtexts@bigtexts-1:/usr/local/hadoop/etc/hadoop$ mv client.tar.gz ~/Desktop/
```

En el cliente:

```
wilson@wilson-VirtualBox:~$ mkdir hadoop.conf  
wilson@wilson-VirtualBox:~$ cd ~/Downloads/  
wilson@wilson-VirtualBox:~/Downloads$ cd ~/hadoop.conf/  
wilson@wilson-VirtualBox:~/hadoop.conf$ tar -xzf client.tar.gz  
wilson@wilson-VirtualBox:~$ nano .bashrc  
wilson@wilson-VirtualBox:~$ nano .profile  
export HADOOP_CONF_DIR=$HOME/hadoop.conf
```

Editar la ip del archivo core-site.xml

```
<configuration>  
  <property>  
    <name>fs.default.name</name>  
    <value>hdfs://192.168.0.7:9000</value>  
  </property>  
</configuration>
```

```
wilson@wilson-VirtualBox:~$ pig
```