

# Module 8: Project - Hadoop Implementation

---

PIG Installation on Ubuntu

edureka!

**edureka!**

© 2014 Brain4ce Education Solutions Pvt. Ltd.

# PIG Installation on Ubuntu

## Table of Contents

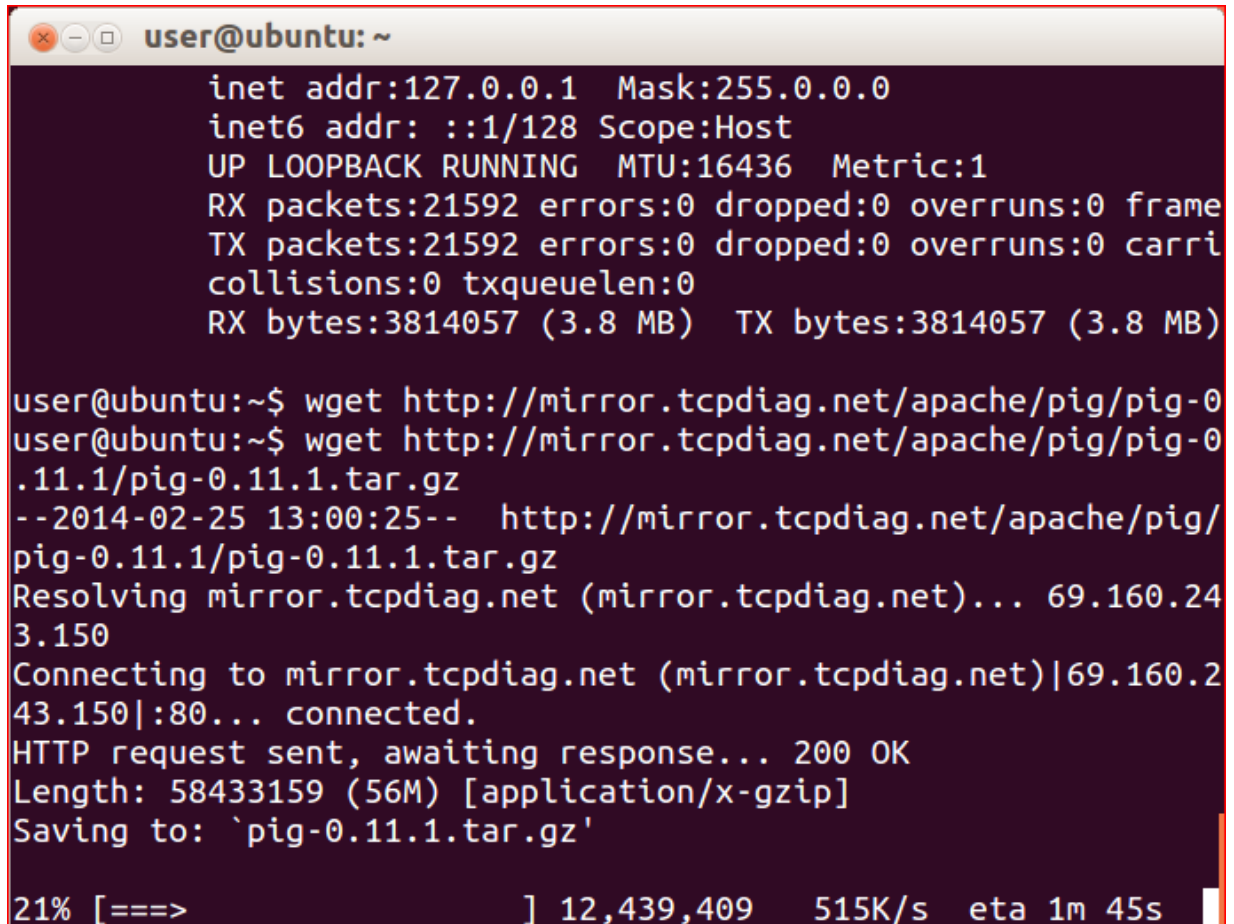
<b>Step 1:</b> Download the tar file of Pig and extracting it.....	2
<b>Step 2:</b> Extract the file.....	3
<b>Step 3:</b> Setting path .....	4
<b>Step 4:</b> Logout from your Ubuntu and log in again. ....	4
<b>Step 5:</b> Confirming if pig is installed correctly or not. ....	5

edureka!

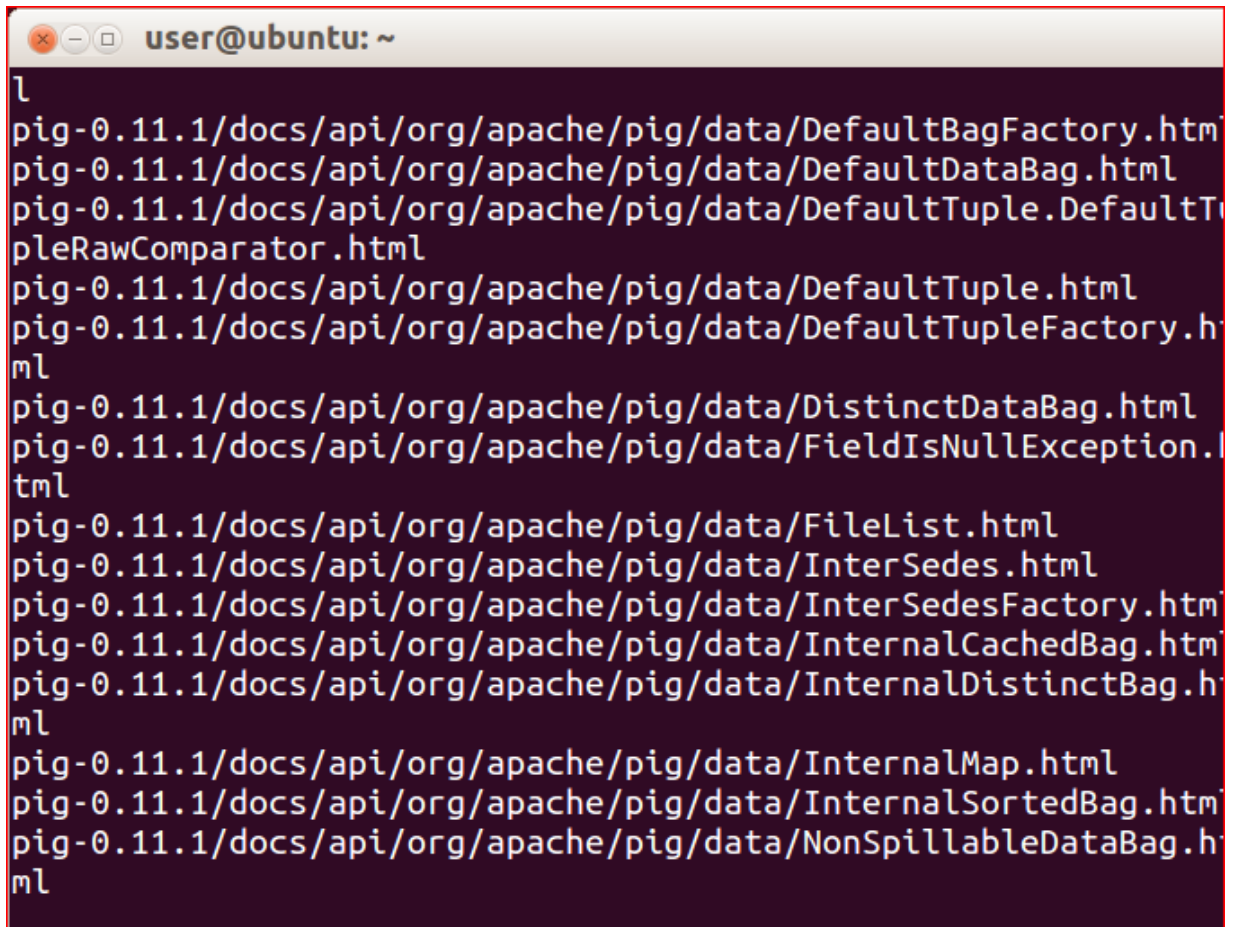
# PIG installation guide

**Step 1:** Download the tar file of Pig and extracting it

Command: `wget http://mirror.tcpdiag.net/apache/pig/pig-0.11.1/pig-0.11.1.tar.gz`



```
user@ubuntu: ~  
    inet addr:127.0.0.1  Mask:255.0.0.0  
    inet6 addr: ::1/128 Scope:Host  
    UP LOOPBACK RUNNING  MTU:16436  Metric:1  
    RX packets:21592 errors:0 dropped:0 overruns:0 frame  
    TX packets:21592 errors:0 dropped:0 overruns:0 carri  
    collisions:0 txqueuelen:0  
    RX bytes:3814057 (3.8 MB)  TX bytes:3814057 (3.8 MB)  
  
user@ubuntu:~$ wget http://mirror.tcpdiag.net/apache/pig/pig-0  
user@ubuntu:~$ wget http://mirror.tcpdiag.net/apache/pig/pig-0  
.11.1/pig-0.11.1.tar.gz  
--2014-02-25 13:00:25--  http://mirror.tcpdiag.net/apache/pig/  
pig-0.11.1/pig-0.11.1.tar.gz  
Resolving mirror.tcpdiag.net (mirror.tcpdiag.net)... 69.160.24  
3.150  
Connecting to mirror.tcpdiag.net (mirror.tcpdiag.net)|69.160.2  
43.150|:80... connected.  
HTTP request sent, awaiting response... 200 OK  
Length: 58433159 (56M) [application/x-gzip]  
Saving to: `pig-0.11.1.tar.gz'  
  
21% [==>                ] 12,439,409   515K/s   eta 1m 45s
```

**Step 2:** Extract the file.**Command:** `tar -xvf pig-0.11.1.tar.gz`A terminal window titled 'user@ubuntu: ~' with a dark background and light text. It shows the output of the command 'tar -xvf pig-0.11.1.tar.gz'. The output lists various files from the pig-0.11.1 package, including API documentation for DefaultBagFactory, DefaultDataBag, DefaultTuple, TupleRawComparator, DistinctDataBag, FieldIsNullException, FileList, InterSedes, InterSedesFactory, InternalCachedBag, InternalDistinctBag, InternalMap, InternalSortedBag, and NonSpillableDataBag.

```
l
pig-0.11.1/docs/api/org/apache/pig/data/DefaultBagFactory.html
pig-0.11.1/docs/api/org/apache/pig/data/DefaultDataBag.html
pig-0.11.1/docs/api/org/apache/pig/data/DefaultTuple.DefaultT
pleRawComparator.html
pig-0.11.1/docs/api/org/apache/pig/data/DefaultTuple.html
pig-0.11.1/docs/api/org/apache/pig/data/DefaultTupleFactory.h
ml
pig-0.11.1/docs/api/org/apache/pig/data/DistinctDataBag.html
pig-0.11.1/docs/api/org/apache/pig/data/FieldIsNullException.l
tml
pig-0.11.1/docs/api/org/apache/pig/data/FileList.html
pig-0.11.1/docs/api/org/apache/pig/data/InterSedes.html
pig-0.11.1/docs/api/org/apache/pig/data/InterSedesFactory.htm
pig-0.11.1/docs/api/org/apache/pig/data/InternalCachedBag.htm
pig-0.11.1/docs/api/org/apache/pig/data/InternalDistinctBag.h
ml
pig-0.11.1/docs/api/org/apache/pig/data/InternalMap.html
pig-0.11.1/docs/api/org/apache/pig/data/InternalSortedBag.htm
pig-0.11.1/docs/api/org/apache/pig/data/NonSpillableDataBag.h
ml
```

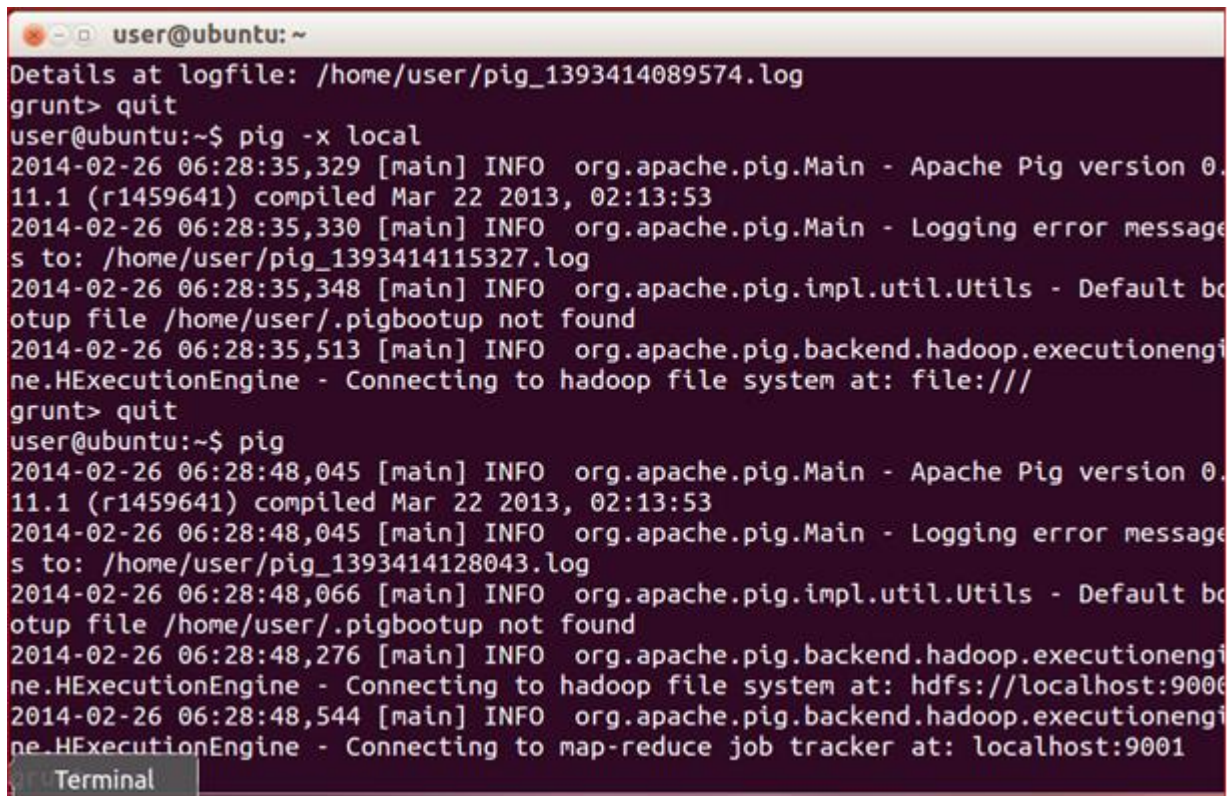
### Step 3: Setting path

Command: `sudo gedit /etc/profile pig`

A screenshot of a gedit window titled 'profile (/etc) - gedit'. The window has a menu bar with 'File', 'Edit', 'View', 'Search', 'Tools', 'Documents', and 'Help'. Below the menu bar is a toolbar with icons for 'Open', 'Save', 'Print', 'Undo', 'Redo', 'Cut', 'Copy', 'Paste', 'Find', and 'Replace'. The main text area shows the contents of the /etc/profile file, which includes comments about pam\_umask, a shell script to source files in /etc/profile.d, and four export statements for PIG\_INSTALL, PATH, JAVA\_HOME, and PIG\_CLASSPATH. The status bar at the bottom indicates 'Plain Text', 'Tab Width: 8', 'Ln 1, Col 1', and 'INS'.

Command: `source /etc/profile`

### Step 4: Logout from your Ubuntu and log in again.

**Step 5:** Confirming if pig is installed correctly or not.**Command:** pigA terminal window titled 'user@ubuntu: ~' showing the execution of the 'pig' command. The terminal output displays log messages from the Apache Pig main class and its components. It shows the version (0.11.1), compilation date (Mar 22 2013), and logging paths. It also indicates that the default bootstrap file is not found and that it is connecting to the Hadoop file system and map-reduce job tracker at localhost:9000 and localhost:9001 respectively. The terminal text is as follows:

```
Details at logfile: /home/user/pig_1393414089574.log
grunt> quit
user@ubuntu:~$ pig -x local
2014-02-26 06:28:35,329 [main] INFO org.apache.pig.Main - Apache Pig version 0.11.1 (r1459641) compiled Mar 22 2013, 02:13:53
2014-02-26 06:28:35,330 [main] INFO org.apache.pig.Main - Logging error messages to: /home/user/pig_1393414115327.log
2014-02-26 06:28:35,348 [main] INFO org.apache.pig.impl.util.Utils - Default bootstrap file /home/user/.pigbootstrap not found
2014-02-26 06:28:35,513 [main] INFO org.apache.pig.backend.hadoop.executionengine.HExecutionEngine - Connecting to hadoop file system at: file:///
grunt> quit
user@ubuntu:~$ pig
2014-02-26 06:28:48,045 [main] INFO org.apache.pig.Main - Apache Pig version 0.11.1 (r1459641) compiled Mar 22 2013, 02:13:53
2014-02-26 06:28:48,045 [main] INFO org.apache.pig.Main - Logging error messages to: /home/user/pig_1393414128043.log
2014-02-26 06:28:48,066 [main] INFO org.apache.pig.impl.util.Utils - Default bootstrap file /home/user/.pigbootstrap not found
2014-02-26 06:28:48,276 [main] INFO org.apache.pig.backend.hadoop.executionengine.HExecutionEngine - Connecting to hadoop file system at: hdfs://localhost:9000
2014-02-26 06:28:48,544 [main] INFO org.apache.pig.backend.hadoop.executionengine.HExecutionEngine - Connecting to map-reduce job tracker at: localhost:9001
```

```
export PIG_INSTALL=/home/user/pig-0.11.1 export PATH=$PATH:$PIG_INSTALL/bin
export JAVA_HOME=/usr/lib/jvm/java-1.6.0-openjdk-i386 export
PIG_CLASSPATH=/home/user/hadoop-1.2.0/conf/
```