Pig Installation Guide

# 1. Download Apache

|  |
| --- |
| wget https://dlcdn.apache.org/pig/pig-0.17.0/pig-0.17.0.tar.gz |

A screenshot of a computer

Description automatically generated



# 2. Setup Pig

|  |
| --- |
| tar -xzf pig-0.17.0.tar.gz |



|  |
| --- |
| mv pig-0.17.0 pig |



# 3. Declare environment variables

|  |
| --- |
| nano ~/.bashrc |
| export PIG\_HOME=/home/hadoopngocphung/pig  export PIG\_CLASSPATH=$HADOOP\_CONF\_DIR  export PATH=$PATH:$PIG\_HOME/bin |

A screen shot of a computer program

Description automatically generated

# 4. Check Pig

|  |
| --- |
| pig -help |

A screenshot of a computer screen

Description automatically generated

# 5. Thực thi pig

Pig has 6 execution modes

**- Local Mode:** Runs on a single machine. All files are installed and run on the local file system and server.

|  |
| --- |
| pig -x local |

**- Tez Local Mode:** Same as local mode, except Pig internally calls the Tez execution engine.

|  |
| --- |
| pig -x tez\_local |

**- Spark Local Mode:** Same as local mode, except Pig internally calls the Spark execution engine.

|  |
| --- |
| pig -x spark local |

**- Map Reduce Mode:** Runs in mapreduce mode, requires access to Hadoop cluster and HDFS. This is the default running mode, may not need to provide the -x parameter

|  |
| --- |
| pig |

Or

|  |
| --- |
| pig -x mapreduce |

**- Tez Mode:** Needs access to Hadoop cluster and HDFS system

|  |
| --- |
| pig -x tez |

**- Spark Mode:** Needs access to Spark, Yarn or Mesos cluster and HDFS file system. Pig Scripts running on Spark can take advantage of dynamic allocation. This feature can be enabled by enabling spark.dynamicAllocation.enabled.

|  |
| --- |
| pig -x spark |