

# Pig Installation

Prepared by: Dev Jethva (23MDS003)

Branch: MTech (Data Science)

## ❖ Prerequisites

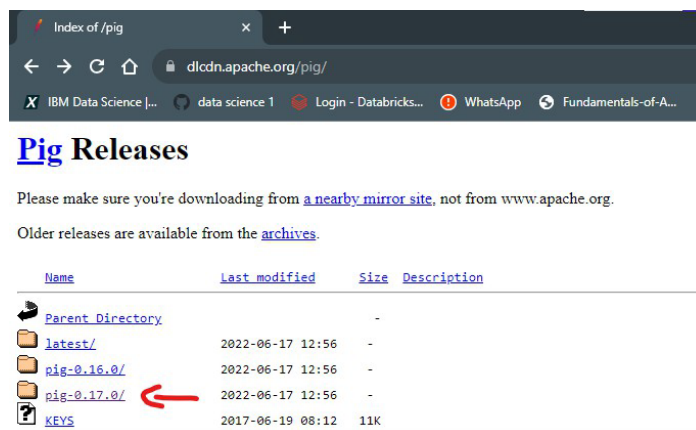
1. Java 8 development Kit (JDK)
2. Hadoop 3.2.4 must be installed.
3. 7zip

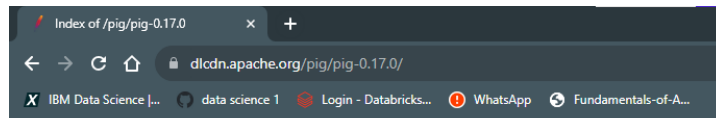
## ❖ Download PIG file

The first step is to download pig from the apache official server , it can be downloaded at following link :

<https://dlcdn.apache.org/pig/>

under this link go to the pig-0.17.0/ and then download [pig-0.17.0.tar.gz](#) this file. The package size is about 220MB.





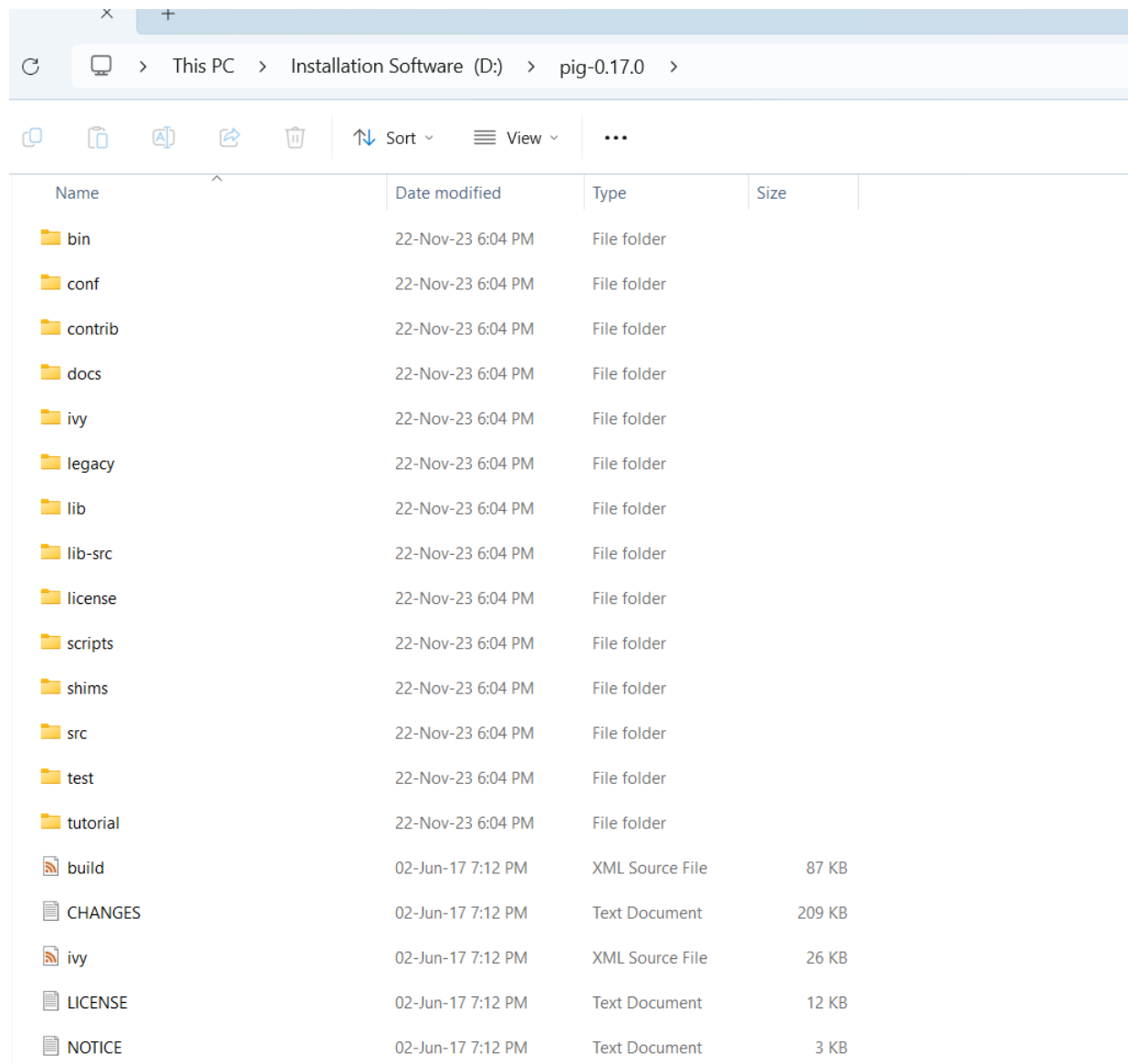
## Index of /pig/pig-0.17.0

Name	Last modified	Size	Description
Parent Directory		-	
README.txt	2017-06-16 18:10	1.4K	
RELEASE_NOTES.txt	2017-06-16 18:10	1.9K	
pig-0.17.0-src.tar.gz	2017-06-16 18:11	15M	
pig-0.17.0-src.tar.gz.asc	2017-06-16 18:11	488	
pig-0.17.0-src.tar.gz.md5	2017-06-16 18:11	56	
pig-0.17.0.tar.gz	2017-06-16 18:11	220M	
pig-0.17.0.tar.gz.asc	2017-06-16 18:11	488	
pig-0.17.0.tar.gz.md5	2017-06-16 18:11	52	

### ❖ Unzip and Install PIG

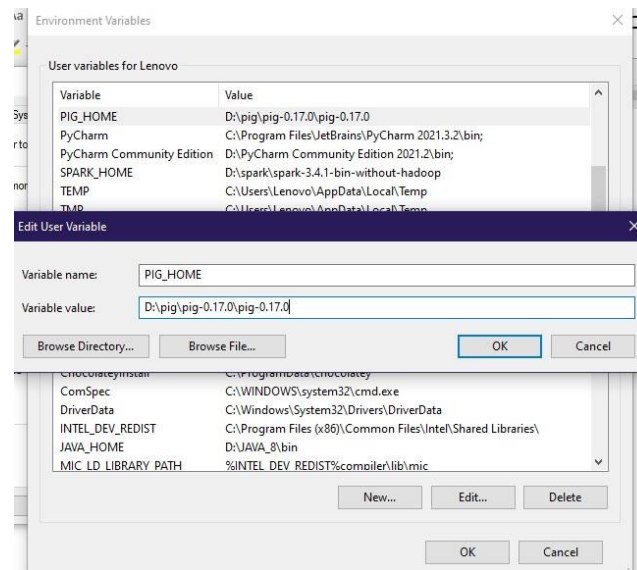
After Downloading the PIG, we need to Unzip the pig-0.17.0.tar.gz file.

After unzipping the folder it will be look like this:

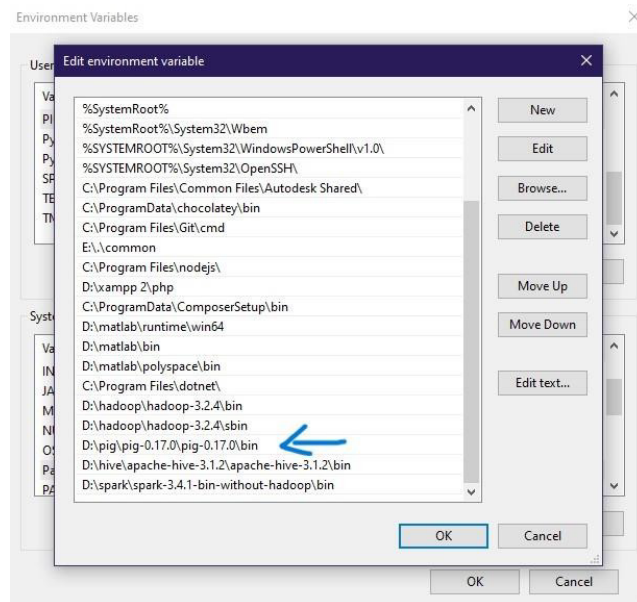


### ❖ **Setting Up Environment Variables**

- Another important step in setting up a work environment is to set your Systems environment variable.
- Go to Control Panel > System > click on the “Advanced system settings” link to edit environment variables.
- Open environment Variable and click on “New” in “User Variable”.
- On clicking “New”, we get the below screen.



- The last step in setting the Environment variable is setting Path in System Variable.

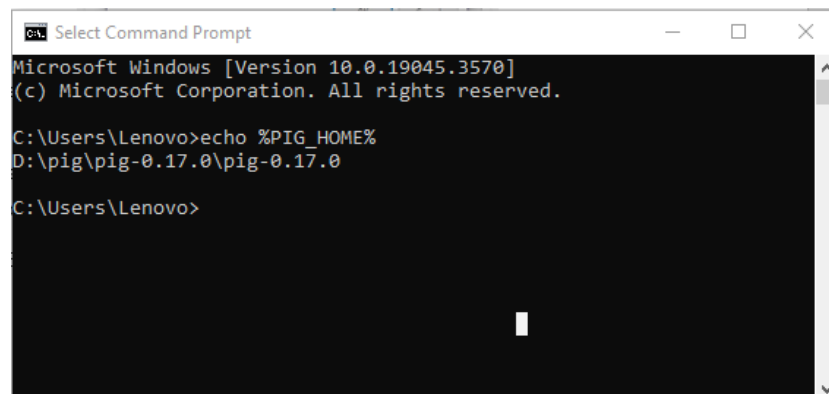


## ❖ Verify the Paths

Now we need to verify that what we have done is correct and reflecting.

- Open a NEW Command Window
- Run following commands

```
echo %PIG_HOME%
```



```
Microsoft Windows [Version 10.0.19045.3570]
(c) Microsoft Corporation. All rights reserved.

C:\Users\Lenovo>echo %PIG_HOME%
D:\pig\pig-0.17.0\pig-0.17.0

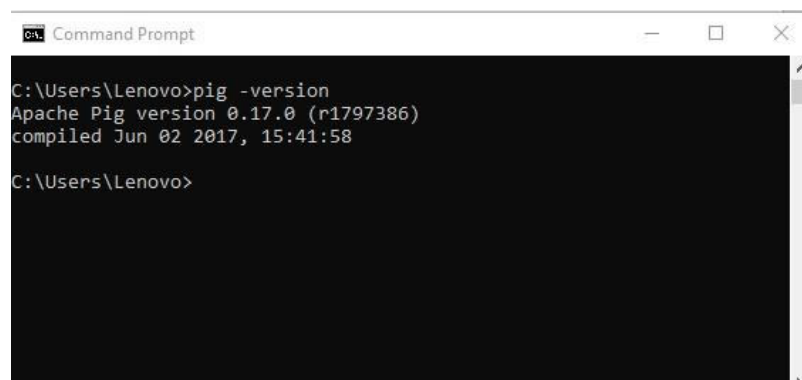
C:\Users\Lenovo>
```

## ❖ Verifying Setup

- We are done with setting up the PIG on our System.
- Now we need to check if everything works smoothly...
- Open a cmd window, run the below command to test the connection and PIG.

`pig -version`

Upon running the command we should get the version of PIG. i.e 0.17.0 in our case.



```
C:\Users\Lenovo>pig -version
Apache Pig version 0.17.0 (r1797386)
compiled Jun 02 2017, 15:41:58

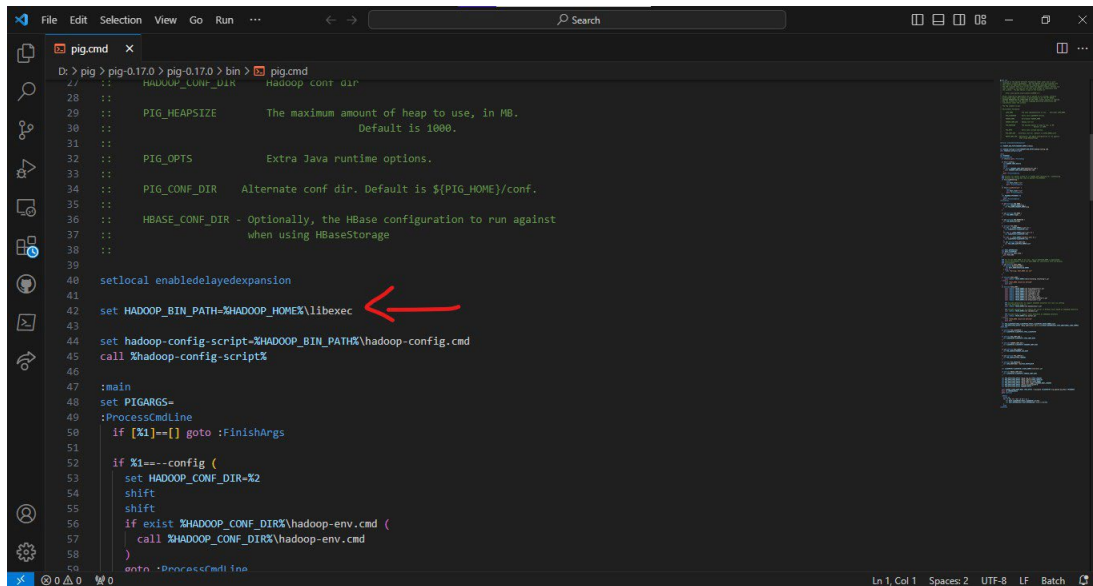
C:\Users\Lenovo>
```

Don't worry some of us can get the below error after running `pig -version`

'-Xmx1000M' is not recognized as an internal or external command, operable program or batch file.

To resolve this we will need to perform the following steps:-

1. Open the pig. cmd file in edit mode. We can find the file in the bin folder.
2. Now we need to change the value of the HADOOP\_BIN\_PATH



```
File Edit Selection View Go Run ... Search
D:\> pig > pig-0.17.0 > pig-0.17.0 > bin > pig.cmd
28 :: HADOOP_CONF_DIR Hadoop conf dir
29 ::
30 :: PIG_HEAPSIZE The maximum amount of heap to use, in MB.
31 :: Default is 1000.
32 ::
33 :: PIG_OPTS Extra Java runtime options.
34 ::
35 :: PIG_CONF_DIR Alternate conf dir. Default is ${PIG_HOME}/conf.
36 ::
37 :: HBASE_CONF_DIR - Optionally, the HBase configuration to run against
38 :: when using HBaseStorage
39 ::
40 setlocal enabledelayedexpansion
41
42 set HADOOP_BIN_PATH=%HADOOP_HOME%\libexec
43
44 set hadoop-config-script=%HADOOP_BIN_PATH%\hadoop-config.cmd
45 call %hadoop-config-script%
46
47 :main
48 set PIGARGS=
49 :ProcessCmdLine
50 if [%1]==[] goto :FinishArgs
51
52 if %1==--config (
53 set HADOOP_CONF_DIR=%2
54 shift
55 shift
56 if exist %HADOOP_CONF_DIR%\hadoop-env.cmd (
57 call %HADOOP_CONF_DIR%\hadoop-env.cmd
58 )
59 goto :ProcessCmdLine
60
Ln 1, Col 1 Spaces: 2 UTF-8 LF Batch
```

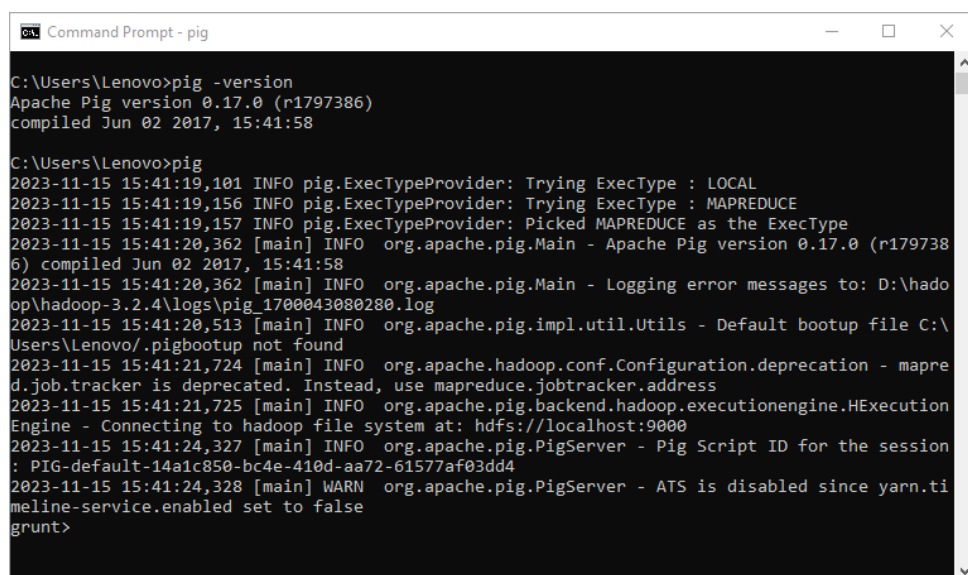
3. Save the file.

The next step is to verify the setup once again. So, we need to execute the pig-version command once again.

## ❖ Starting PIG

Now we need to start a new Command Prompt remember to run it as administrator to avoid permission issues and execute the below commands:

pig



```
Command Prompt - pig
C:\Users\Lenovo> pig -version
Apache Pig version 0.17.0 (r1797386)
compiled Jun 02 2017, 15:41:58

C:\Users\Lenovo> pig
2023-11-15 15:41:19,101 INFO pig.ExecTypeProvider: Trying ExecType : LOCAL
2023-11-15 15:41:19,156 INFO pig.ExecTypeProvider: Trying ExecType : MAPREDUCE
2023-11-15 15:41:19,157 INFO pig.ExecTypeProvider: Picked MAPREDUCE as the ExecType
2023-11-15 15:41:20,362 [main] INFO org.apache.pig.Main - Apache Pig version 0.17.0 (r1797386) compiled Jun 02 2017, 15:41:58
2023-11-15 15:41:20,362 [main] INFO org.apache.pig.Main - Logging error messages to: D:\hadoop\hadoop-3.2.4\logs\pig_1700043080280.log
2023-11-15 15:41:20,513 [main] INFO org.apache.pig.impl.util.Utils - Default bootup file C:\Users\Lenovo/.pigbootup not found
2023-11-15 15:41:21,724 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address
2023-11-15 15:41:21,725 [main] INFO org.apache.pig.backend.hadoop.executionengine.HExecutionEngine - Connecting to hadoop file system at: hdfs://localhost:9000
2023-11-15 15:41:24,327 [main] INFO org.apache.pig.PigServer - Pig Script ID for the session : PIG-default-14a1c850-bc4e-410d-aa72-61577af03dd4
2023-11-15 15:41:24,328 [main] WARN org.apache.pig.PigServer - ATS is disabled since yarn.timeline-service.enabled set to false
grunt>
```

We can see grunt> once the pig starts.