# **PIG - INSTALLATION**

# **Downloading Pig**

Download the latest version of Pig by visiting the following link https://dlcdn.apache.org/pig/pig-0.17.0/pig-0.17.0.tar.gz

- 1.Create a new Folder inside Desktop, name the Folder as your USN <1ms18cs017>.
- 2. Move the Downloaded Pig File to USN <1ms18cs017> Folder.
- 3.Right Click on that File and Extract inside the USN <1ms18cs017> Folder.

### **Open Terminal**

Navigate to Extracted Hadoop Folder cd ~/Desktop/1MS18CS017/pig-0.17.0

### Create a New File named Bash.sh

Copy the Below code and Paste inside Bash.sh and save that File.

```
export JAVA_HOME=$(readlink -f $(which javac) | awk 'BEGIN {FS="/bin"} {print $1}') if ! command -v pig &> /dev/null then export PATH=$(echo $PATH):$(pwd)/bin fi
```

To launch pig hadoop should be configured.

```
rootekali:~/Desktop/IMS18CS017/pig-0.17.0# echo SPATH
/root/.nvm/versions/node/v14.17.0/bin:/root/anaconda3/condabin:/usr/local/bin:/usr/local/bin:/usr/bin:/sbin:/sbin:/sbin:/shin:/snap/bin:/root/Desktop/IMS18CS017/pig-0.17.0# pin
rootekali:~/Desktop/IMS18CS017/pig-0.17.0# pig
2022-07-03 20:31:32,276 IMF0 [main] pig.ExecTypeProvider (ExecTypeProvider.java:selectExecType(41)) - Trying ExecType: LOCAL
2022-07-03 20:31:32,277 IMF0 [main] pig.ExecTypeProvider (ExecTypeProvider.java:selectExecType(41)) - Trying ExecType: MARREDUCE
2022-07-03 20:31:32,277 IMF0 [main] pig.ExecTypeProvider (ExecTypeProvider.java:selectExecType(41)) - Picked MARREDUCE as the ExecType
2022-07-03 20:31:32,300 [main] IMF0 org.apache.pig.Main - Apache Pig veron ol.17.0 (17)97386) compiled Jun 02 2017, 15:41:58
2022-07-03 20:31:32,300 [main] IMF0 org.apache.pig.Main - Logging error messages to: /root/Desktop/IMS18CS017/pig-0.17.0/pig_1656860492296.log
2022-07-03 20:31:32,310 [main] IMF0 org.apache.pig.mpl.util.utils - Default bootup file /root/.pigbootup not found
2022-07-09 20:31:32,331 [main] ERROR org.apache.pig.Main - ERROR 4010: Cannot find hadoop configurations in classpath (neither hadoop-site.xml nor core-site
xml was found in the classpath). If you plan to use local mode. please put - x local option in command line
Details at logfile: /root/Desktop/IMS18CS017/pig-0.17.0# pig script completed in 103 milliseconds (103 ms)
rootekali:~/Desktop/IMS18CS017/pig-0.17.0# |
```

Error: No Hadoop in \$PATH variable

### set up Hadoop

1. Navigate to Hadoop Folder to set up Hadoop.

```
cd ~/Desktop/<1ms18cs017>/hadoop-3.3.2
```

2.Execute Bash.sh inside hadoop Folder

```
source Bash.sh
```

# After this u can see the Hadoop Folder inside PATH Variable

```
root@kali:-/Desktop/IMS18CS017/hadoop-3.3.2# source Bash.sh
17/pig-0.17.0/bin:/root/Desktop/IMS18CS017/hadoop-3.3.2/bin
root@kali:-/Desktop/IMS18CS017/hadoop-3.3.2# pig
2072-07-08.20.37:31.486 IMFO pig.ExecTypeProvider: Trying ExecType : LOCAL
2022-07-08.20.37:31.487 IMFO pig.ExecTypeProvider: Trying ExecType : MAPREDUCE
2022-07-08.20.37:31.487 IMFO pig.ExecTypeProvider: Picked MAPREDUCE
2022-07-08.20.37:31.518 [main] IMFO org.apache.pig.Main - Apache Pig version 0.17.0 (r1797386) compiled Jun 02 2017, 15:41:58
2022-07-08.20.37:31.518 [main] IMFO org.apache.pig.Main - Logging error messages to: /root/Desktop/IMS18CS017/hadoop-3.3.2/pig_1656860851512.log
2022-07-08.20.37:31.518 [main] IMFO org.apache.pig.main - Logging error messages to: /root/Desktop/IMS18CS017/hadoop-3.3.2/pig_1656860851512.log
2022-07-08.20.37:31.510 [main] IMFO org.apache.pig.impl.util.Utils - Default bootup file /root/.pigbootup not found
2022-07-08.20.37:31.710 [main] IMFO org.apache.pig.org.comf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.
2022-07-08.20.37:31.710 [main] IMFO org.apache.pig.backend.hadoop.executionengine.HExecutionEngine - Connecting to hadoop file system at: file://
2022-07-08.20.37:31.777 [main] IMFO org.apache.pig.PigServer - Pig Script ID for the session: PIG-default-2044730a-33bc-4319-08a7-db65222c63a1
2022-07-08.20.37:31.777 [main] IMFO org.apache.pig.PigServer - Pig Script ID for the session: PIG-default-2044730a-33bc-4319-08a7-db6522c263a1
2022-07-08.20.37:31.777 [main] IMFO org.apache.pig.PigServer - ATS is disabled since yarn.timeline-service.enabled set to false
2021-07-08.20.37:31.777 [main] IMFO org.apache.pig.PigServer - ATS is disabled since yarn.timeline-service.enabled set to false
```

Verify JAVA\_HOME variable to be set to Java Path and PATH variable has your USN Pig Folder and Hadoop. If any previous PATH set to Hadoop or pig Folder remove that inside .bashrc file.

Verify Pig is Installed or not by executing **pig or pig –version** command.if command gives pig prompt then pig is Successfully Installed and Configured.

```
Iroot@kali:~/Desktop/1MS18CS017/hadoop-3.3.2# pig --version
Apache Pig version 0.17.0 (r1797386)
compiled Jun 02 2017, 15:41:58
root@kali:~/Desktop/1MS18CS017/hadoop-3.3.2#
```

# **PIG PROGRAMS**

### 1.Filter

# Create a new file with .pig extension

student\_details = LOAD '/home/msrit/Downloads/pig/test/Filter/student\_details.txt' USING PigStorage(',') as (id:int, firstname:chararray, lastname:chararray, age:int, phone:chararray, city:chararray);

filter data = FILTER student details BY city == 'Chennai';

Dump filter data;

### Execution: pig <name.pig>

```
root@kali:-/Desktop/IMS18CS917/pig-8.17.8/programs/Filter# ls
filter.pig student_details.txt
root@kali:-/Desktop/IMS18CS917/pig-8.17.8/programs/Filter# ls
filter.pig student_details.txt
root@kali:-/Desktop/IMS18CS917/pig-8.17.8/programs/Filter# cat filter.pig
student_details.txd
root@kali:-/Desktop/IMS18CS917/pig-8.17.8/programs/Filter# cat filter.pig
student_details = LOAD 'student_details.txt' USING PigStorage(',') as (id:int, firstname:chararray, lastname:chararray, age:int, phone:chararray, city:chararray);
filter_data = FILTER student_details BY city == 'Chennai';

Dump filter_data;
root@kali:-/Desktop/IMS18CS917/pig-8.17.8/programs/Filter# cat student_details.txt
001.Rajiv,Reddy.21.9048022337, Hyderabad
002.siddarray.22.9048022337, Hyderabad
003.sidarray.22.9048022339, Delni
904.Freeth.Agarwal.21.948022338, Delni
904.Freeth.Agarwal.21.948022338, Delni
905.Trupthi.Mohanthy.23.9848022335, Chennai
007.Komal.Mayak.24.9848022335, Chennai
008.Bharathi.Mambiayar.24.9848022335, Chennai
008.Bharathi.Mambiayar.24.9848022335, Chennai
008.Bharathi.Mambiayar.24.9848022335, Chennai
0006&Rali:-/Desktop/IMS18CS017/pig-0.17.8/programs/Filter# pig filter.pig
```

```
(6,Archana,Mishra,23,9848022335,Chennai)
(8,Bharathi,Nambiayar,24,9848022333,Chennai)
2022-07-03 20:55:40,414 [main] INFO org.apache.pig.Main - Pig script completed in 6 seconds and 408 milliseconds (6408 ms)
root@kali:~/Desktop/1M518CS017/pig-0.17.0/programs/Filter#
```

# Student details.txt

001, Rajiv, Reddy, 21, 9848022337, Hyderabad

002, siddarth, Battacharya, 22, 9848022338, Kolkata

003, Rajesh, Khanna, 22, 9848022339, Delhi

004, Preethi, Agarwal, 21, 9848022330, Pune

005, Trupthi, Mohanthy, 23, 9848022336, Bhuwaneshwar

006, Archana, Mishra, 23, 9848022335, Chennai

007, Komal, Nayak, 24, 9848022334, trivendram

008, Bharathi, Nambiayar, 24, 9848022333, Chennai

### 2. GROUPING

student = LOAD 'student\_details.txt' USING PigStorage(',') as (id:int, firstname:chararray, lastname:chararray, age:int, phone:chararray, city:chararray);

group data = GROUP student by age;

Dump group\_data;

### 3. JOIN

customers = LOAD 'customer.txt' USING PigStorage(',') as (id:int, name:chararray, age:int, address:chararray, salary:int);

orders = LOAD 'order.txt' USING PigStorage(',') as (oid:int, date:chararray, customer\_id:int, amount:int);

join\_result = JOIN customers BY id, orders BY customer\_id;

Dump join result

### //Customer.txt

- 1,Ramesh,32,Ahmedabad,2000.00
- 2,Khilan,25,Delhi,1500.00
- 3,kaushik,23,Kota,2000.00
- 4, Chaitali, 25, Mumbai, 6500.00
- 5, Hardik, 27, Bhopal, 8500.00
- 6,Komal,22,MP,4500.00
- 7, Muffy, 24, Indore, 10000.00

### //Order.txt

102,2009-10-08 00:00:00,3,3000

100,2009-10-08 00:00:00,3,1500

101,2009-11-20 00:00:00,2,1560

103,2008-05-20 00:00:00,4,2060

# 4. Sorting

student = LOAD 'student\_details.txt' USING PigStorage(',') as (id:int, firstname:chararray, lastname:chararray, age:int, phone:chararray, city:chararray);

```
student_order = ORDER student BY age DESC;
student_limit = LIMIT student_order 4;
Dump student limit;
```

### 5.Union

cust1 = LOAD 'customer1.txt' USING PigStorage(',') as (id:int, firstname:chararray, lastname:chararray, phone:chararray, city:chararray);

cust2 = LOAD 'customer2.txt' USING PigStorage(',') as (id:int, firstname:chararray, lastname:chararray, phone:chararray, city:chararray);

```
cust = UNION cust1, cust2;
Dump cust
```

#### //Customer1.txt

001,Rajiv,Reddy,9848022337,Hyderabad 002,siddarth,Battacharya,9848022338,Kolkata 003,Rajesh,Khanna,9848022339,Delhi 004,Preethi,Agarwal,9848022330,Pune 005,Trupthi,Mohanthy,9848022336,Bhuwaneshwar 006,Archana,Mishra,9848022335,Chennai.

#### //Customer2.txt

7, Komal, Nayak, 9848022334, triven dram.

8, Bharathi, Nambiayar, 9848022333, Chennai.