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.

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

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.

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>

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,trivendram.
8,Bharathi,Nambiayar,9848022333,Chennai.