# PIG PROGRAMS

**Bash 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

Command to run it: **source bash.sh**

**Programs:**

1. Create a program to **filter** out student data based on City Chennai

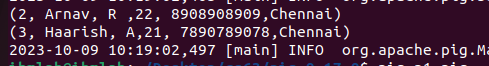
**student\_details.txt:** 001,Rajiv,Reddy,21,9848022337,Hyderabad 002,Arnav,R,22,8908908909,Chennai 003,Haarish,A,21,7890789078,Chennai 004,Preethi,Agarwal,21,9848022330,Pune 005,Trupthi,Mohanthy,23,9848022336,Bhuwaneshwar 006,Archana,Mishra,23,9848022335,Chennai 007,Komal,Nayak,24,9848022334,trivendram 008,Bharathi,Nambiayar,24,9848022333,Chennai

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);

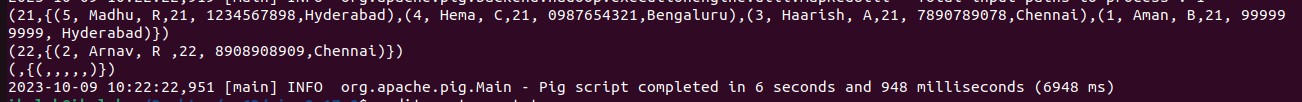
filter\_data = FILTER student\_details BY city == 'Chennai'; Dump filter\_data;



1. Create a program to **group** student data based on age

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;

1. Create a program to **Join** two separate data files of custome.txt and order.txt based id and order by id.

## 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

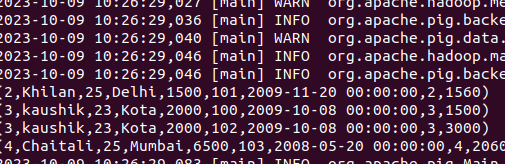
customers = LOAD 'customer.txt' USING PigStorage(',') as (id:int, name:chararray, age:int,

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;



1. Create a program to obtain **union** of customer1 and customer2 dataset

## 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.

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;

