**Department of Computer Science & Engineering**

# CSL76: BIG DATA LABORATORY TERM: Sep2022- Jan2022

**PIG REPORT**

***By***

# Manoj Kumar M 1MS19CS086

**Submitted to**

# Dr. S. Rajarajeswari Assistant Professor

## M S RAMAIAH INSTITUTE OF TECHNOLOGY

**(Autonomous Institute, Affiliated to VTU)**

**BANGALORE-560054**

[**www.msrit.edu**](http://www.msrit.edu/)

2022

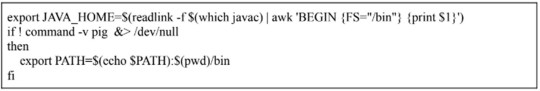
**PIG**

## Installation Steps:

1. 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. 1.Create a new folder inside Desktop , name the Folder as your USN <1ms19cs068>
2. 2.Move the Downloaded Pig File to USN <1ms19cs068> Folder.
3. 3.Right Click on that File and Extract inside the USN <1ms19cs068> Folder.
4. Open Terminal
5. Cd into pig file inside your USN folder cd /Desktop/1ms19cs068/pig-0.17.0
6. Create a new Bash.sh file and paste this code to set the PATH



1. To launch pig Hadoop should also be configured.
2. So check in the $PATH if Hadoop is present along with pig and if not present go to Hadoop folder and
3. source the Bash.sh File once again.
4. Then check if pig is working fine by using the command pig -- version.
5. Execution: pig <filename.pig>

**PROGRAMS**

### Program to filter those entries where city is “Chennai”

student\_details = LOAD

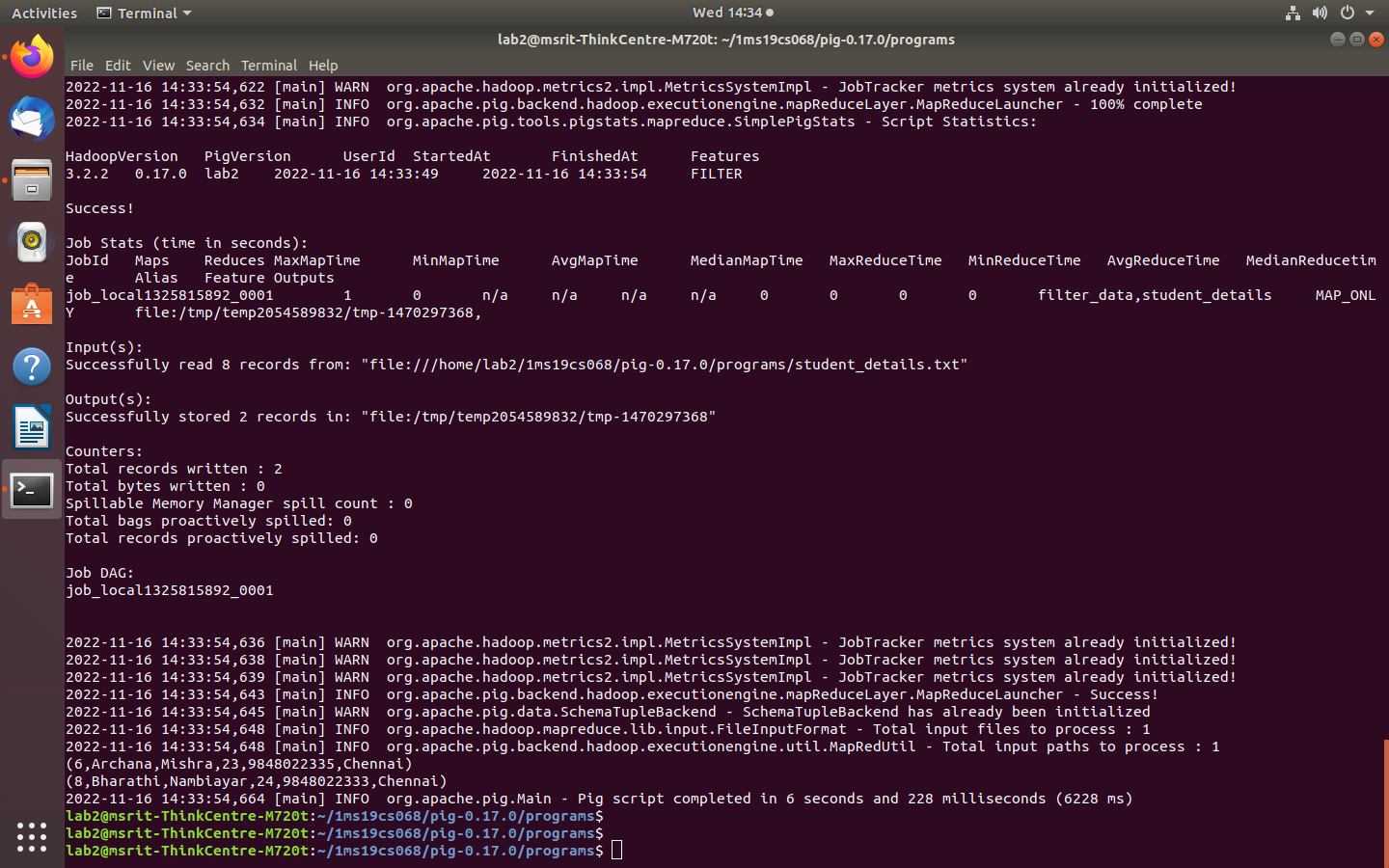
'/home/lab2/1MS19CS068/pig-0.17.0/programs/student\_details.txt' USING PigStorage(',') as (id:int, firstname:chararray, lastname:chararray, age:int, phone:chararray,city:chararray);

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

**Input file**: 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

**Output:**



* 1. **Program to group entries by age.**

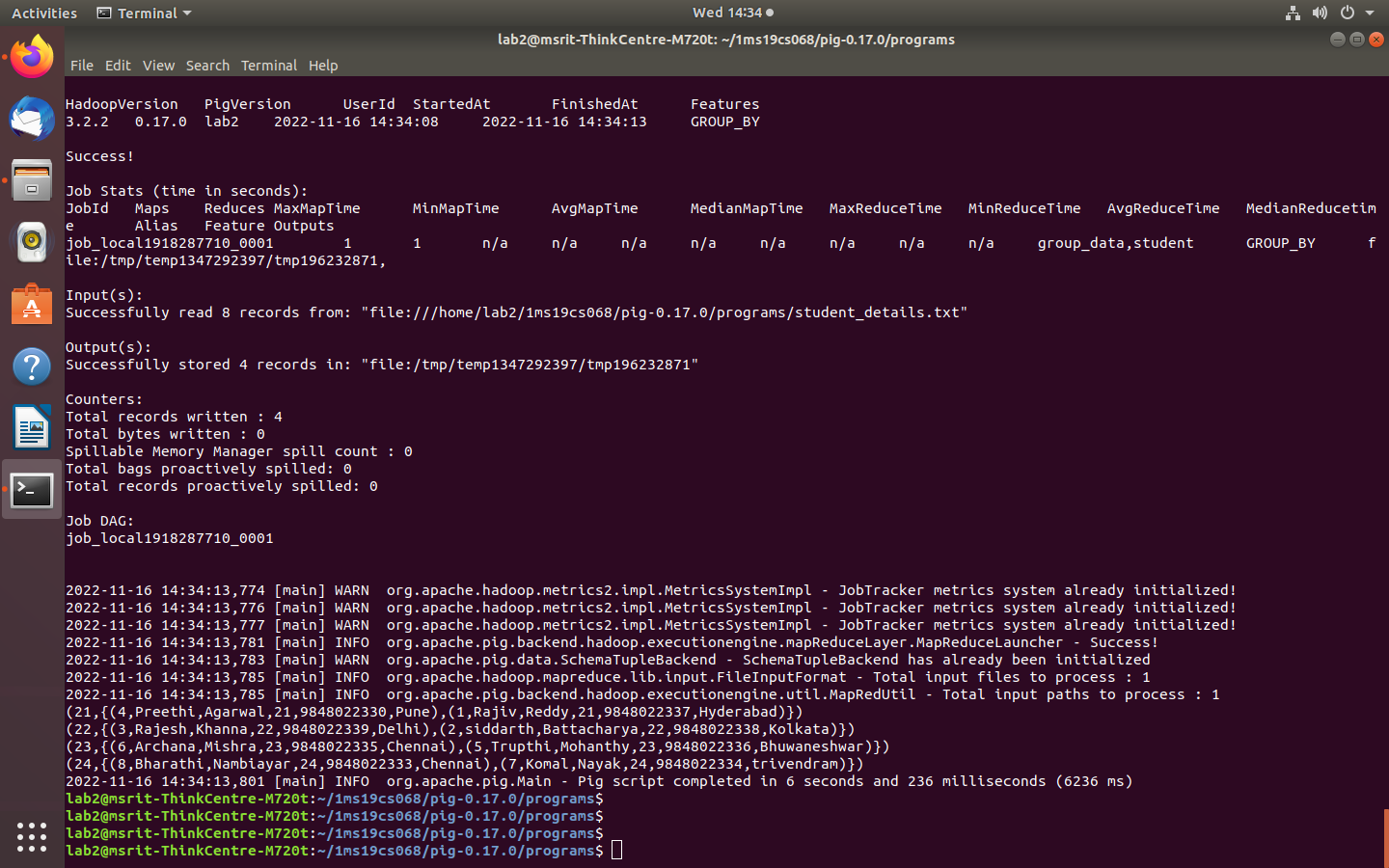
student = LOAD

'/home/lab2/1MS19CS068/pig-0.17.0/programs/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;

**Input file**: student\_details.txt

**Output:**



* 1. **Program to sort the entries by age in the descending order.**

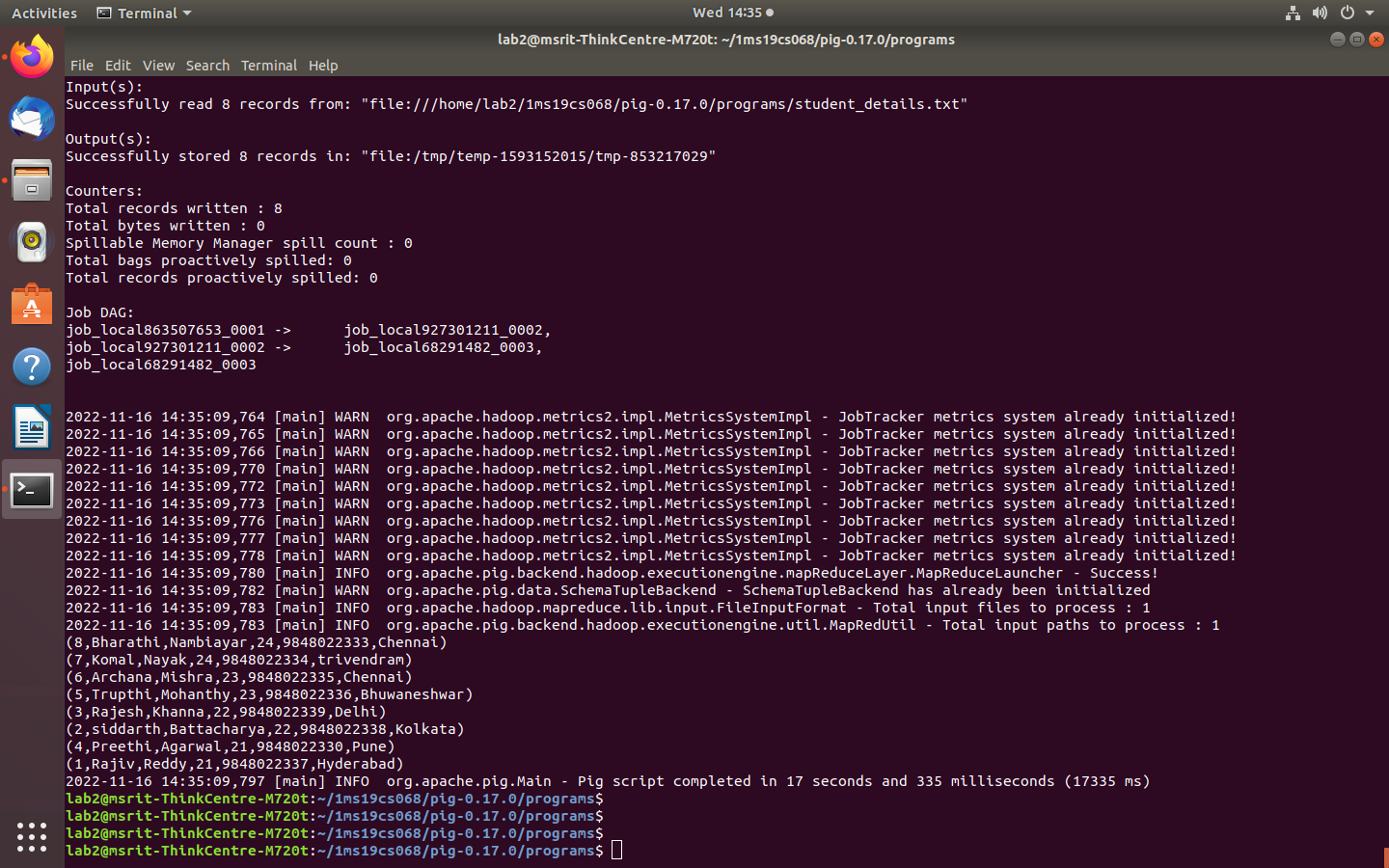
student = LOAD

'/home/lab2/1MS19CS086/pig-0.17.0/programs/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; Dump student\_order;

**Input file**: student\_details.txt

**Output:**



* 1. **Program to join two separate data files into one data file.**

customers = LOAD

'/home/lab2/1MS19CS068/pig-0.17.0/programs/customer.txt' USING PigStorage(',') as (id:int, name:chararray,age:int,address:chararray, salary:int); orders = LOAD '/home/lab2/1MS19CS068/pig-0.17.0/programs/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

### Input File: customer.txt and order.txt 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

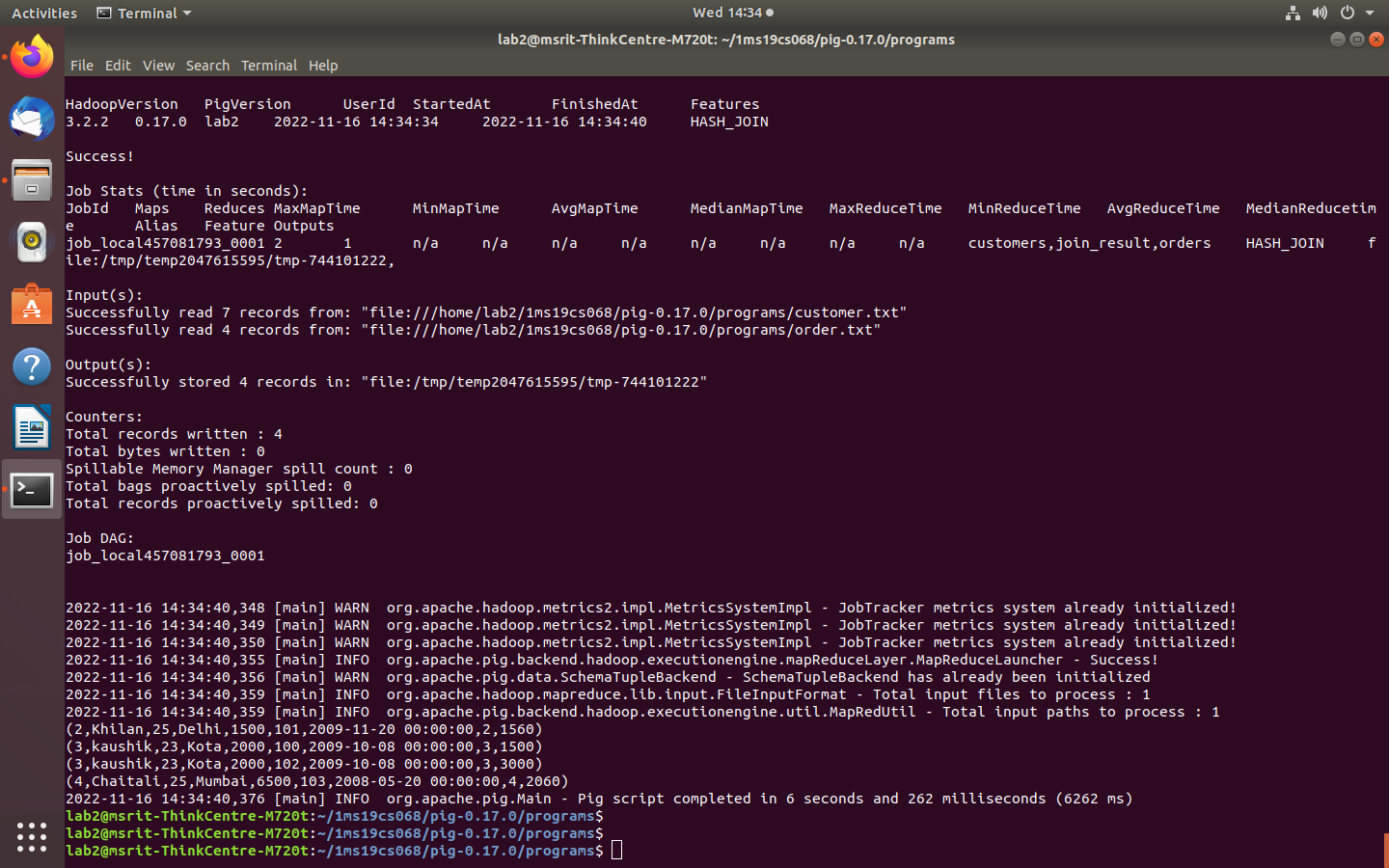
102,2009-10-08 00:00:00,3,3000

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

**Output:**



* 1. **Program to perform union operation on two data sets.**

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

### Input Files: Customer1.txt

001,Rajiv,Reddy,98480223 7,Hyderabad 002,siddarth,Battacharya,9 848022338,Kolkata 003,Rajesh,Khanna,984802 2339,Delhi 004,Preethi,Agarwal,98480 22330,Pune 005,Trupthi,Mohanthy,984 8022336,Bhuwaneshwar 006,Archana,Mishra,98480 22335,Chennai.

**Customer2.txt** 7,Komal,Nayak,984802233 4,trivendram.

8,Bharathi,Nambiayar,9848 022333,Chennai.

**Output:**

