

**Poornima Institute of Engineering & Technology**  
**Department of Computer Engineering**  
**Session 2020-21**  
**IV year---VIII Semester**  
**Subject: Big Data Analytics (8CS4-21)**  
**Experiment-7**

Topic: -Pig Latin scripts to sort, group, join, project, and filter your data.

**Apache Pig - Group Operator**

```
student_details = LOAD '/user/cloudera/Sec-A/Pig/student_details.txt' USING PigStorage(',') as
(id:int, firstname:chararray, lastname:chararray, age:int, phone:chararray, city:chararray);
group_data = GROUP student_details by age;
Dump group_data;
```

**Output:**

```
(21,{(4,Preethi,Agarwal,21,9848022330,Pune),(1,Rajiv,Reddy,21,9848022337,Hyderabad)})
(22,{(3,Rajesh,Khanna,22,9848022339,Delhi
),(2,siddarth,Battacharya,22,9848022338,Kolkata)})
(23,{(6,Archana,Mishra,23,9848022335,Chennai
),(5,Trupthi,Mohanthi,23,9848022336,Bhuwaneshwar)})
(24,{(8,Bharathi,Nambiayar,24,9848022333,Chennai),(7,Komal,Nayak,24,9848022334,trivendram)})
```

**Grouping by Multiple Columns**

```
group_multiple = GROUP student_details by (age, city);
Dump group_multiple;
```

**Output:**

```
((21,Pune),(4,Preethi,Agarwal,21,9848022330,Pune))
((21,Hyderabad),(1,Rajiv,Reddy,21,9848022337,Hyderabad))
((22,Delhi),(3,Rajesh,Khanna,22,9848022339,Delhi))
((22,Kolkata),(2,siddarth,Battacharya,22,9848022338,Kolkata))
((23,Chennai),(6,Archana,Mishra,23,9848022335,Chennai))
((23,Bhuwaneshwar),(5,Trupthi,Mohanthi,23,9848022336,Bhuwaneshwar))
((24,Chennai),(8,Bharathi,Nambiayar,24,9848022333,Chennai))
((24,trivendram),(7,Komal,Nayak,24,9848022334,trivendram))
```

**Join**

```
customers = LOAD '/user/cloudera/Sec-A/Pig/customers.txt' USING PigStorage(',') as (id:int,
name:chararray, age:int, address:chararray, salary:int);
orders = LOAD '/user/cloudera/Sec-A/Pig/orders.txt' USING PigStorage(',') as (oid:int,
date:chararray, customer_id:int, amount:int);
```

**Self Join:**

```
customers1 = LOAD '/user/cloudera/Sec-A/Pig/customers.txt' USING PigStorage(',') as (id:int,
name:chararray, age:int, address:chararray, salary:int);
customers2 = LOAD '/user/cloudera/Sec-A/Pig/customers.txt' USING PigStorage(',') as (id:int,
name:chararray, age:int, address:chararray, salary:int);
customers3 = JOIN customers1 BY id, customers2 BY id;
Dump customers3;
```

**Output:**

(1,Ramesh,32,Ahmedabad,2000,1,Ramesh,32,Ahmedabad,2000)  
(2,Khilan,25,Delhi,1500,2,Khilan,25,Delhi,1500)  
(3,kaushik,23,Kota,2000,3,kaushik,23,Kota,2000)  
(4,Chaitali,25,Mumbai,6500,4,Chaitali,25,Mumbai,6500)  
(5,Hardik,27,Bhopal,8500,5,Hardik,27,Bhopal,8500)  
(6,Komal,22,MP,4500,6,Komal,22,MP,4500)  
(7,Muffy,24,Indore,10000,7,Muffy,24,Indore,10000)

**Inner Join:**

```
coustomer_orders = JOIN customers BY id, orders BY customer_id;  
Dump coustomer_orders;
```

**Output:**

(2,Khilan,25,Delhi,1500,101,2009-11-20 00:00:00,2,1560)  
(3,kaushik,23,Kota,2000,100,2009-10-08 00:00:00,3,1500)  
(3,kaushik,23,Kota,2000,102,2009-10-08 00:00:00,3,3000)  
(4,Chaitali,25,Mumbai,6500,103,2008-05-20 00:00:00,4,2060)

**Outer Join:****Left Outer Join:**

```
outer_left = JOIN customers BY id LEFT OUTER, orders BY customer_id;  
Dump outer_left;
```

**Output:**

(1,Ramesh,32,Ahmedabad,2000,,,)   
(2,Khilan,25,Delhi,1500,101,2009-11-20 00:00:00,2,1560)  
(3,kaushik,23,Kota,2000,100,2009-10-08 00:00:00,3,1500)  
(3,kaushik,23,Kota,2000,102,2009-10-08 00:00:00,3,3000)  
(4,Chaitali,25,Mumbai,6500,103,2008-05-20 00:00:00,4,2060)  
(5,Hardik,27,Bhopal,8500,,,)   
(6,Komal,22,MP,4500,,,)   
(7,Muffy,24,Indore,10000,,,)

**Right Outer Join:**

```
outer_right = JOIN customers BY id RIGHT, orders BY customer_id;  
Dump outer_right
```

**Output:**

(2,Khilan,25,Delhi,1500,101,2009-11-20 00:00:00,2,1560)  
(3,kaushik,23,Kota,2000,100,2009-10-08 00:00:00,3,1500)  
(3,kaushik,23,Kota,2000,102,2009-10-08 00:00:00,3,3000)  
(4,Chaitali,25,Mumbai,6500,103,2008-05-20 00:00:00,4,2060)

**Full Outer Join**

```
outer_full = JOIN customers BY id FULL OUTER, orders BY customer_id;  
Dump outer_full;
```

**Output**

```
(1,Ramesh,32,Ahmedabad,2000,,,)
(2,Khilan,25,Delhi,1500,101,2009-11-20 00:00:00,2,1560)
(3,kaushik,23,Kota,2000,100,2009-10-08 00:00:00,3,1500)
(3,kaushik,23,Kota,2000,102,2009-10-08 00:00:00,3,3000)
(4,Chaitali,25,Mumbai,6500,103,2008-05-20 00:00:00,4,2060)
(5,Hardik,27,Bhopal,8500,,,)
(6,Komal,22,MP,4500,,,)
(7,Muffy,24,Indore,10000,,,)

```

**Apache Pig - Filter Operator**

```
student_details = LOAD '/user/cloudera/Sec-A/Pig/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;
```

**Output:**

```
(8,Bharathi,Nambiayar,24,9848022333,Chennai)
```