**Pig in class exercise**

Two parts to this exercise please make sure you screen print and label

**Part 1**

Firstly load the following files to your local filesystem to be accessed by PIG preferably from Zeppelin.

Make a directory called “/user/pigexercise” and then copy the following files

1. Emp.txt
2. Emp2.txt
3. Orders.txt
4. Customers.txt
5. Display the data from each file and do a screen shot labelling which file it belongs too.
6. Do a union on the Emp and Emp2 txt files and call the result joinedEmp
7. Split joinedEmp into Emp11 where salary < 30000 and into Emp12 where salary<20000 and salary <25000 display Emp11 and Emp12
8. Find from Emp where employee is equal to 103
9. Create a subset of Emp of just id and name called subemp and display it
10. Create a join on Emp and Customer called Empcust and display it
11. Create a left outer join on Emp and Customer called Empcust\_left and display it
12. Create a right outer join on Emp and Customer called Empcust\_right and display it
13. Create a full outer join on Emp and Customer called Empcust\_full and display it

**Part 2**

Load daily\_show\_guests to /user/pigexercise

A = load '/user/pigexercise/daily\_show\_guests.txt' using PigStorage(',') AS (year:chararray,occupation:chararray,date:chararray,group:chararray,gusetlist:chararray);

1. Generate B with just the occupation and date
2. Generate C with the date formatted to “MM/dd/yy”
3. Generate D as a filter between 1/11/99 and 6/11/99
4. Generate E group by occupation
5. Generate F as the count of all groups in E showing the group and count
6. Generate G order by descending count