

## Chapter 4 – Pig Latin

### Exercise

1. Write a Pig Latin program to read u.user and find the users who are female *and* scientists. Store the results in a file.

```
users = LOAD 'u.user' USING PigStorage('|') AS (userid:int,  
age:int, gender:chararray, occupation:chararray, zipcode:int);  
fem_sci= FILTER users BY gender == 'F' AND occupation ==  
'scientist';  
STORE fem_sci INTO 'female_scientists';
```

2. What is the average age of the users in u.user? Hint: use the AVG function. Dump the result to the screen.

```
users = LOAD 'u.user' USING PigStorage('|') AS (userid:int,  
age:int, gender:chararray, occupation:chararray, zipcode:int);  
grp = GROUP users ALL;  
avg_age = FOREACH grp GENERATE AVG(users.age);  
DUMP avg_age;
```

3. Create a script using a text editor. It should load the u.user file, find the 3 most common occupations and store the results in HDFS. Execute the script using "pig -f script"

```
users = LOAD 'u.user' USING PigStorage('|') AS (userid:int,  
age:int, gender:chararray, occupation:chararray, zipcode:int);  
grp = GROUP users BY occupation;  
counted = FOREACH grp GENERATE group, COUNT(users) AS num:long;  
ordered = ORDER counted BY num DESC;  
result = LIMIT ordered 3;  
STORE result INTO 'common_occupations';
```