**Problem Statement**

**Write a program to implement wordcount using pig**

**Share the screenshot of the command used with its output**

Consider a file named “animal” having entries of few animal names, these names are repeated as shown in the following image:



The problem statement is to find out how many times each word( animal names) have appeared in the file.

This can be done by writing a pig script which in turn converts it into a MapReduce program

The first step is to load the file into a relation. This can be done using the Load command.

Syntax: Animal\_Name = LOAD ‘/home/acadgild/pig/animal’ AS (animal\_name : chararaay);



The above command loads the “animal” into a relation named “Animal\_Name” .The relation will be formed as following:

**grunt> dump Animal\_Name;**



Flatten command is applied to each line in the relation. The bag is converted to a tuple by this command, The syntax is as follows:

Words = FOREACH Animal\_Name GENERATE FLATTEN (TOKENIZE(animal\_name)) as word;



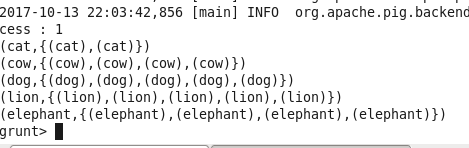
In the next step the words are grouped together so the count logic can be applied on them.

Syntax: grouped = GROUP Words by word;

C:\Users\p.muthurajaiah\AppData\Local\Microsoft\Windows\INetCache\Content.Word\group.png

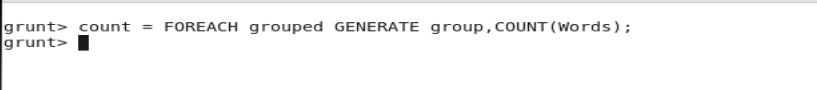
Group command output is as follows:

**grunt> dump grouped;**



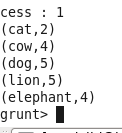
Now we can apply count logic on the grouped words.

Syntax: count = FOREACH grouped GENERATE group, COUNT(Words);



The final output shows the count of each words in the file.

**grunt> dump count;**



**Complete Word Count program in PIG**

lines = LOAD '/user/hadoop/HDFS\_File.txt' AS (line:chararray);

words = FOREACH lines GENERATE FLATTEN(TOKENIZE(line)) as word;

grouped = GROUP words BY word;

wordcount = FOREACH grouped GENERATE group, COUNT(words);

DUMP wordcount;