**Assignment 4.3**

**Problem Statement**

Write a program to implement wordcount using Pig.

Share the screenshots of the commands used with its associated output.

**INPUT**

**wordCount**

hi this is aditya from acadgild

hi aditya we are from acadgild

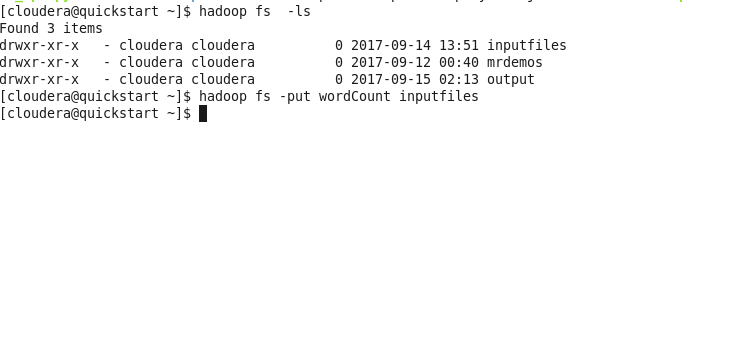
how can we help you

hi acadgild i am writing word count in pig

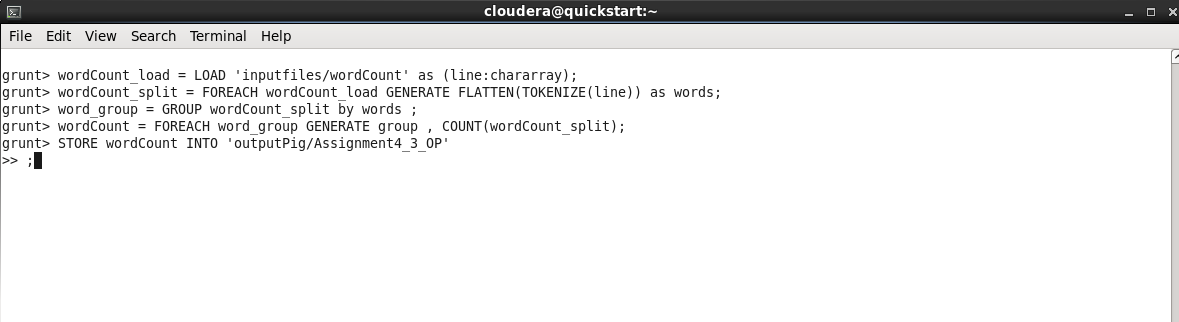
so you are using pig

good luck

**Moving This file to HDFS**



**Pig CODE :**



**WORDCOUNT.pig**

wordCount\_load = LOAD 'inputfiles/wordCount' as (line:chararray);

wordCount\_split = FOREACH wordCount\_load GENERATE FLATTEN(TOKENIZE(line)) as words;

word\_group = GROUP wordCount\_split by words ;

wordCount = FOREACH word\_group GENERATE group , COUNT(wordCount\_split);

STORE wordCount INTO 'outputPig/Assignment4\_3\_OP' ;

**wordCount\_load** : we are loading our input file into the Pig

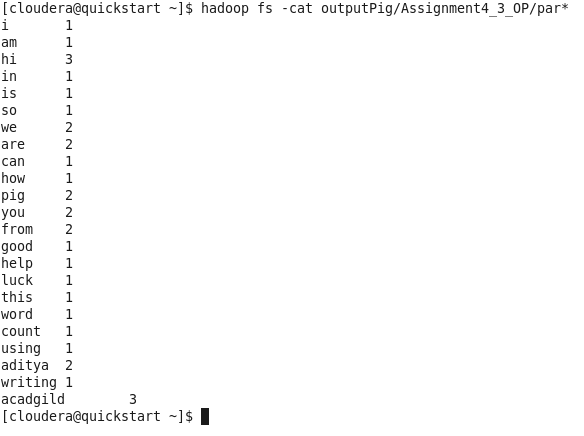
**wordCount\_split :** we are splitting the words on the default delimeter , the tokenize will return bag of words , we will flatten or we will create a tuple of words by using flatten.

**word\_group :** grouping the words on the basis of words which we split

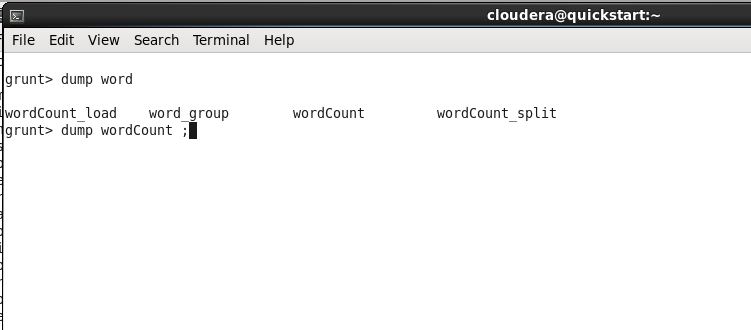
**wordCount :**  Transforming and doing aggregation or count on the wordCount\_split

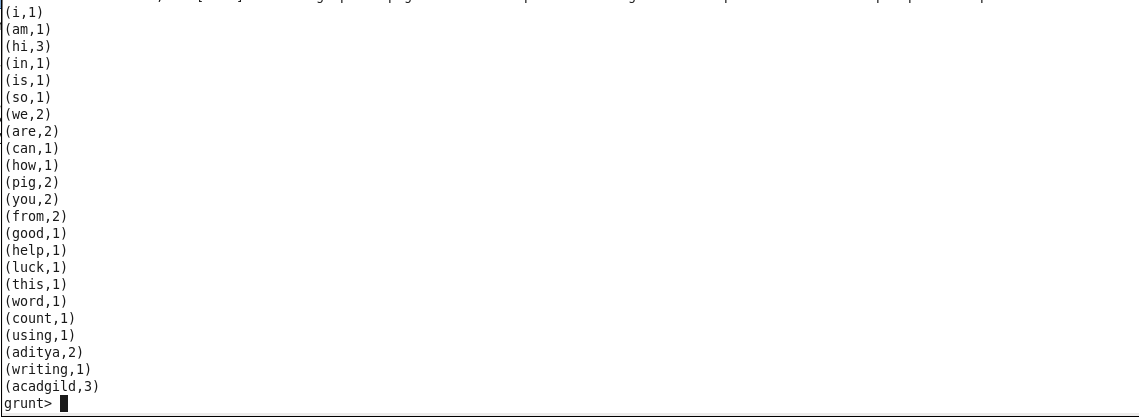
After that storing the result in the 'outputPig/Assignment4\_3\_OP’

**OUTPUT :**



**IF WE DUMP** wordCount **we will get result in grunt shell**





**NOTE: GitHub Repository will Contain**

1)pig Scipt : Assignment\_4\_3.pig

2)Document for the flow

3)Input file