ASSIGNMENT 7.1 - EXPLORING APACHE PIG

TASK 1: Write a program to implement word count using Pig.

Input Data:

souund sar sam sar sound sar sound sam sam sam sar sound sound is the an big data bif is the big sound sar sam data sound sar
sam sound sar sam sam sar sound data big

PROGRAM:

Step 1 : load_lines = load 'wordcount' as (lines:chrararray);

Step 2: load_tokens= FOREACH load_lines generate TOKENIZE(lines) as tokens;

Step 3: load_words= FOREACH load_tokens generate FLATTEN(tokens) as words;

Step 4: load_grp =GROUP load_words by words;

Step 5: load_count=FOREACH load_grp generate group,COUNT(load_words);

Step 6: load_finalwordcount= ORDER load_count by $1;

After every step using DUMP command we can check the execution flow.

DUMP load_finalwordcount;

Output:

```
cess : 1
(souund,1)
(bif,1)
(an,1)
(the,2)
(is,2)
(data,3)
(big,3)
(sound,8)
(sar,8)
(sam,8)
```