ACD\_BDD2.3\_Session\_4\_Assignment\_3

Pig Wordcount script

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

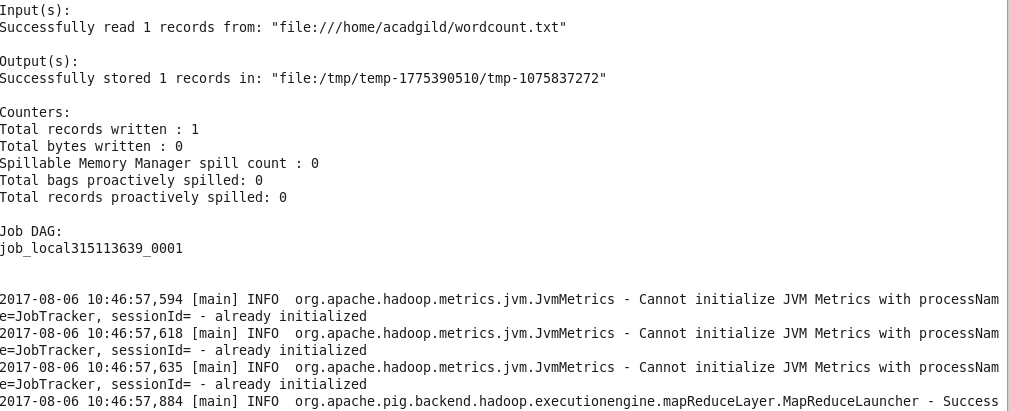
words = FOREACH lines GENERATE FLATTEN(TOKENIZE(line)) as word;

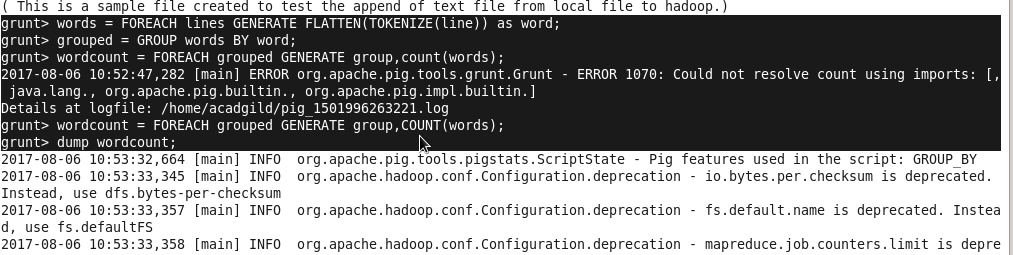
grouped = GROUP words BY word;

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

DUMP wordcount;

Commands used





Output:-

