**SESSION 8: PIG – DEMO I**

**ASSIGNMENT 1**

**WORD COUNT EXAMPLE**

lines = LOAD 'word\_count.txt' AS (line: chararray); -- load input file, taking whole row as a column

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

-- TOKENIZE creates a bag of words, and FLATTEN convers the bag into a tuple.

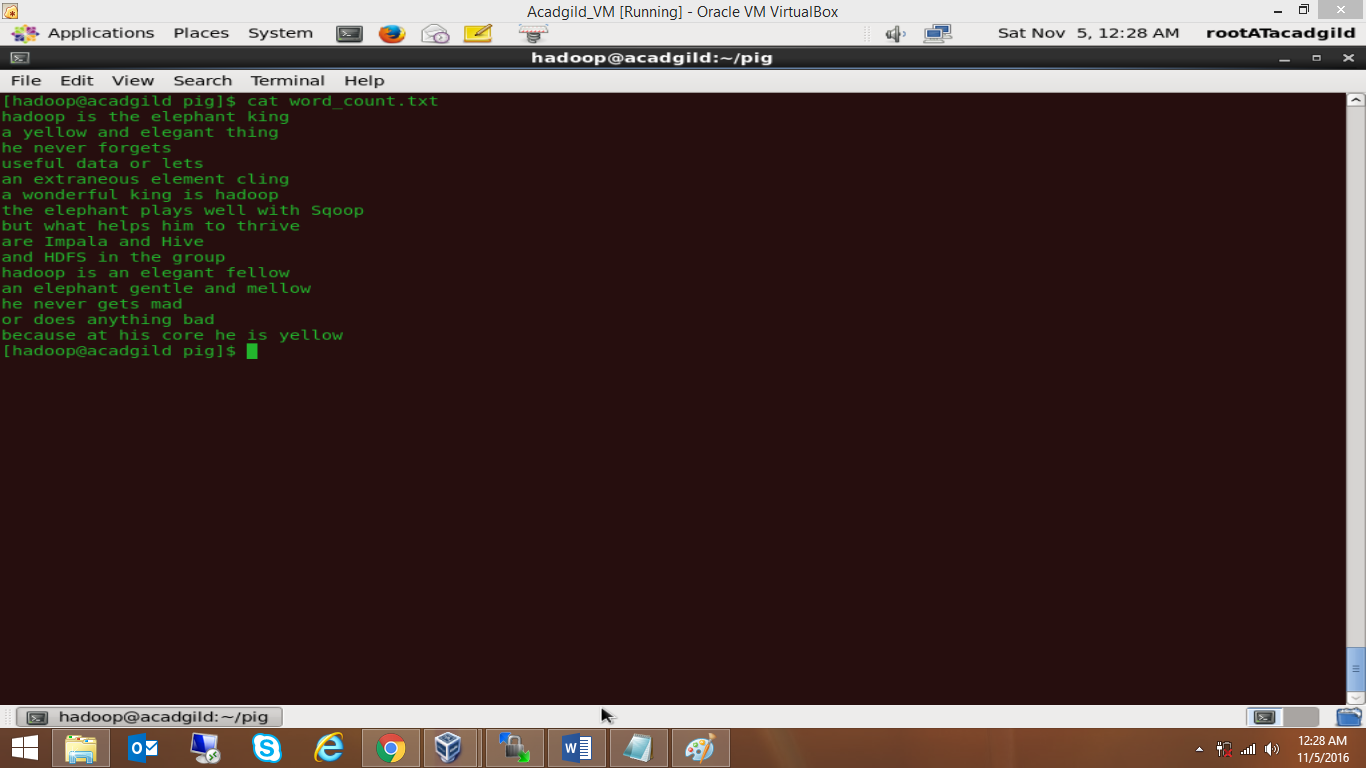
grouped\_words = GROUP words by word;

-- words are grouped together

count = FOREACH grouped\_words GENERATE group, COUNT(words); -- count is calculated

STORE count into 'wcount\_ouput'; -- output is stored in a directory

**INPUT FILE**



**OUTPUT**

