Input file

[cloudera@quickstart pig]$ cat wc.txt

Hadoop is good.

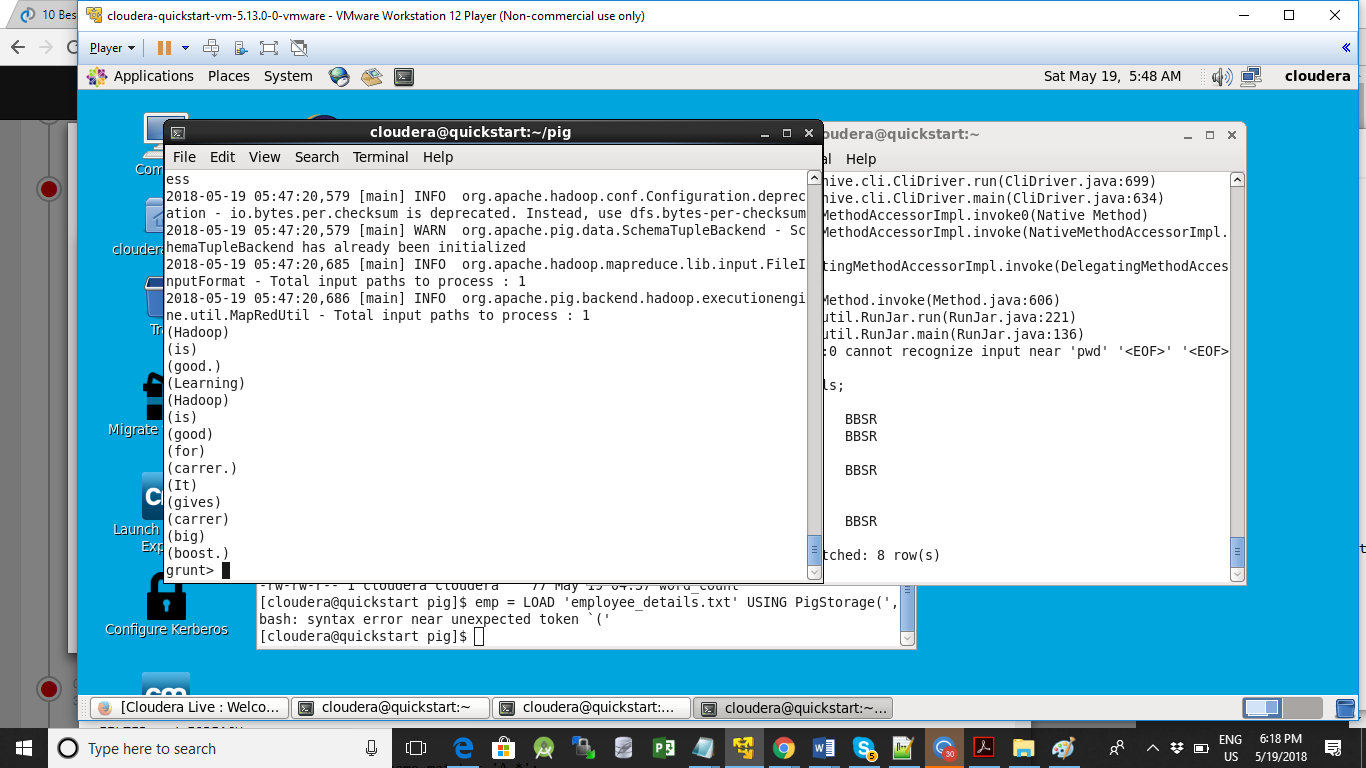
Learning Hadoop is good for carrer.

It gives carrer big boost.[cloudera@quickstart pig]$

Loading file in alias table

wc = LOAD 'wc.txt' AS (line:chararray);

Words = FOREACH wc GENERATE FLATTEN(TOKENIZE(line,' ')) AS word;



(Hadoop)

(is)

(good.)

(Learning)

(Hadoop)

(is)

(good)

(for)

(carrer.)

(It)

(gives)

(carrer)

(big)

(boost.)

Output of :

grp= GROUP Words by word;

(It,{(It)})

(is,{(is),(is)})

(big,{(big)})

(for,{(for)})

(good,{(good)})

(gives,{(gives)})

(good.,{(good.)})

(Hadoop,{(Hadoop),(Hadoop)})

(boost.,{(boost.)})

(carrer,{(carrer)})

(carrer.,{(carrer.)})

(Learning,{(Learning)})

Output of :

wordcount = FOREACH grp GENERATE group, COUNT(Words);

ne.util.MapRedUtil - Total input paths to process : 1

(It,1)

(is,2)

(big,1)

(for,1)

(good,1)

(gives,1)

(good.,1)

(Hadoop,2)

(boost.,1)

(carrer,1)

(carrer.,1)

(Learning,1)