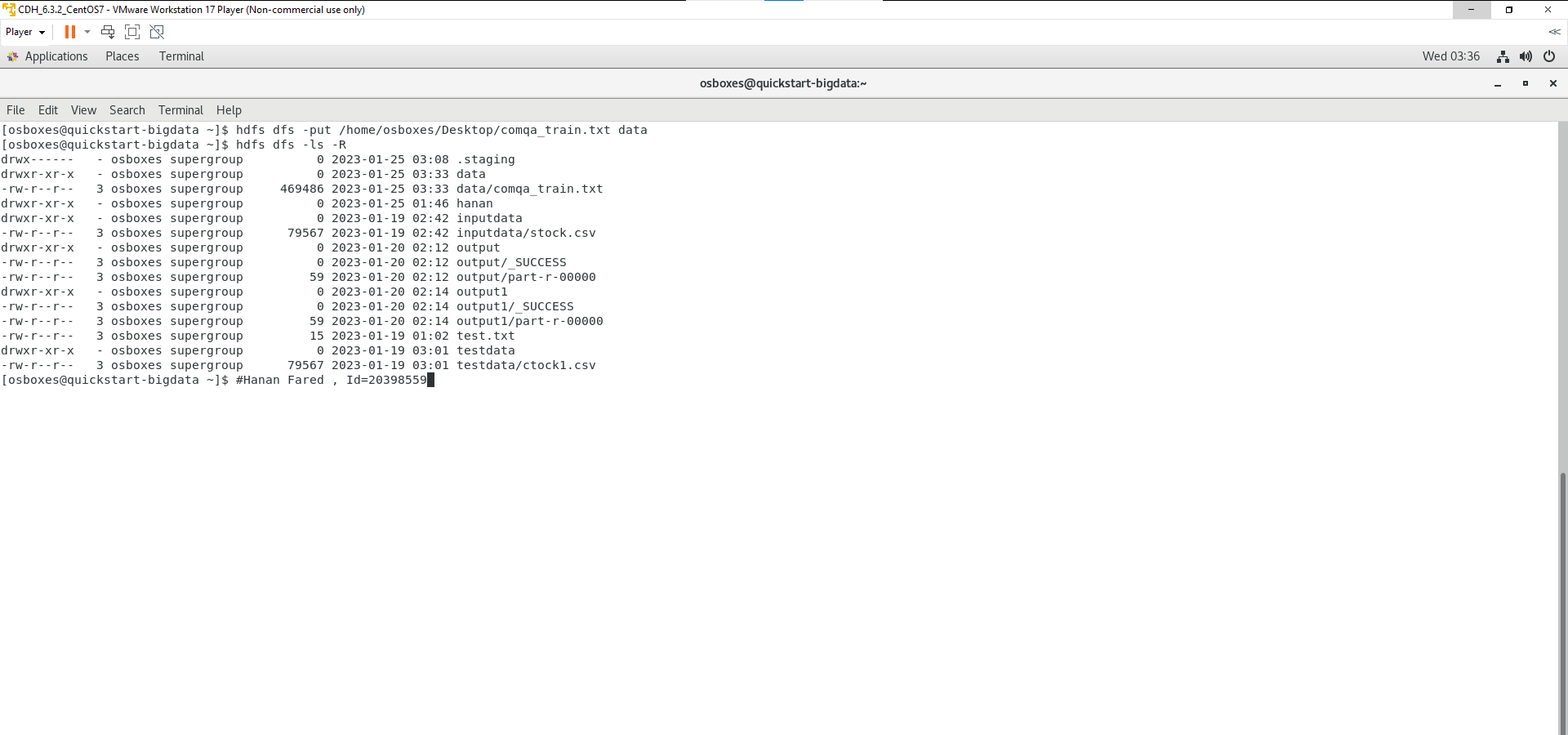
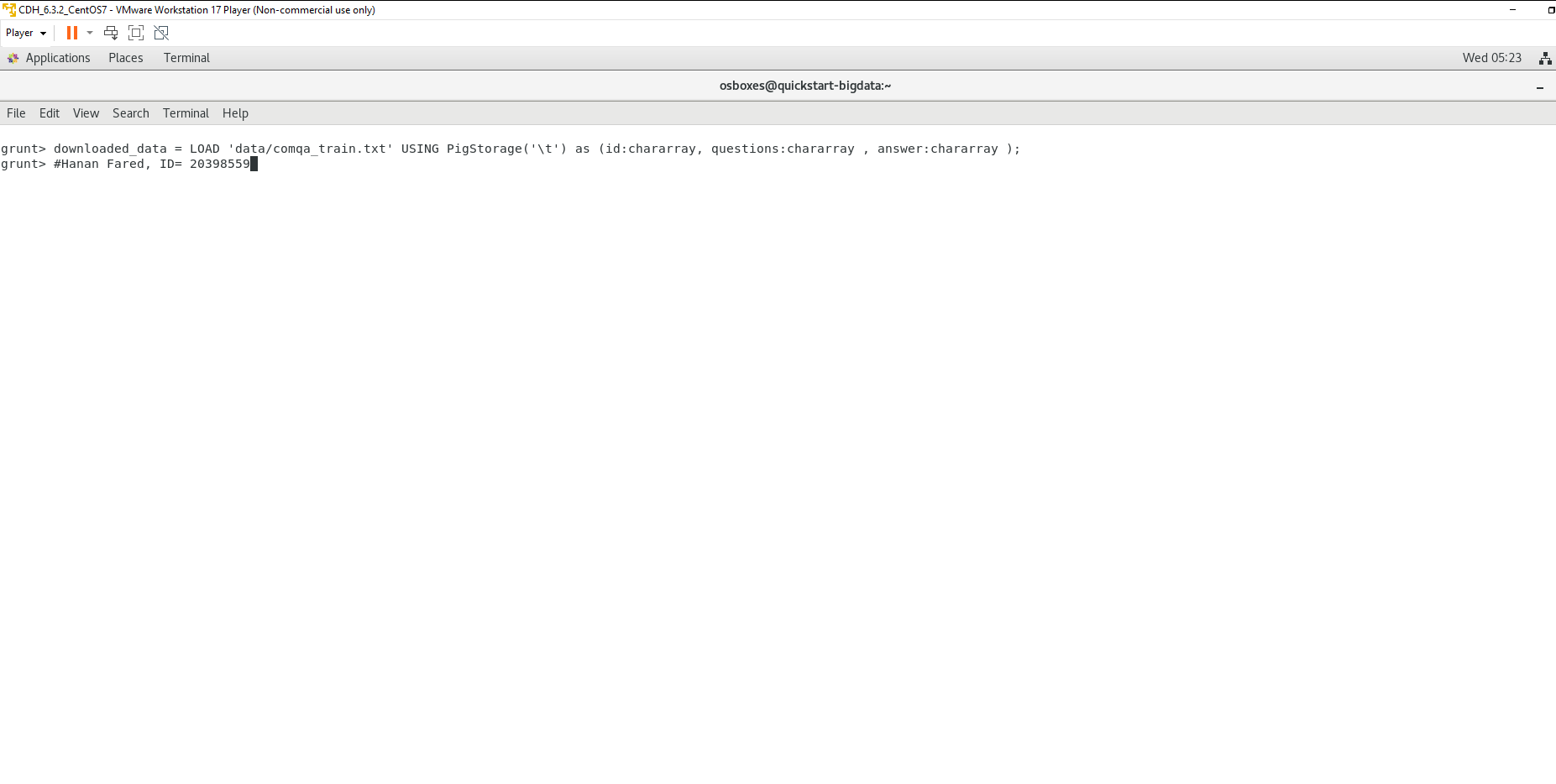
**Name: Hanan Fared Mohamed Omara.**

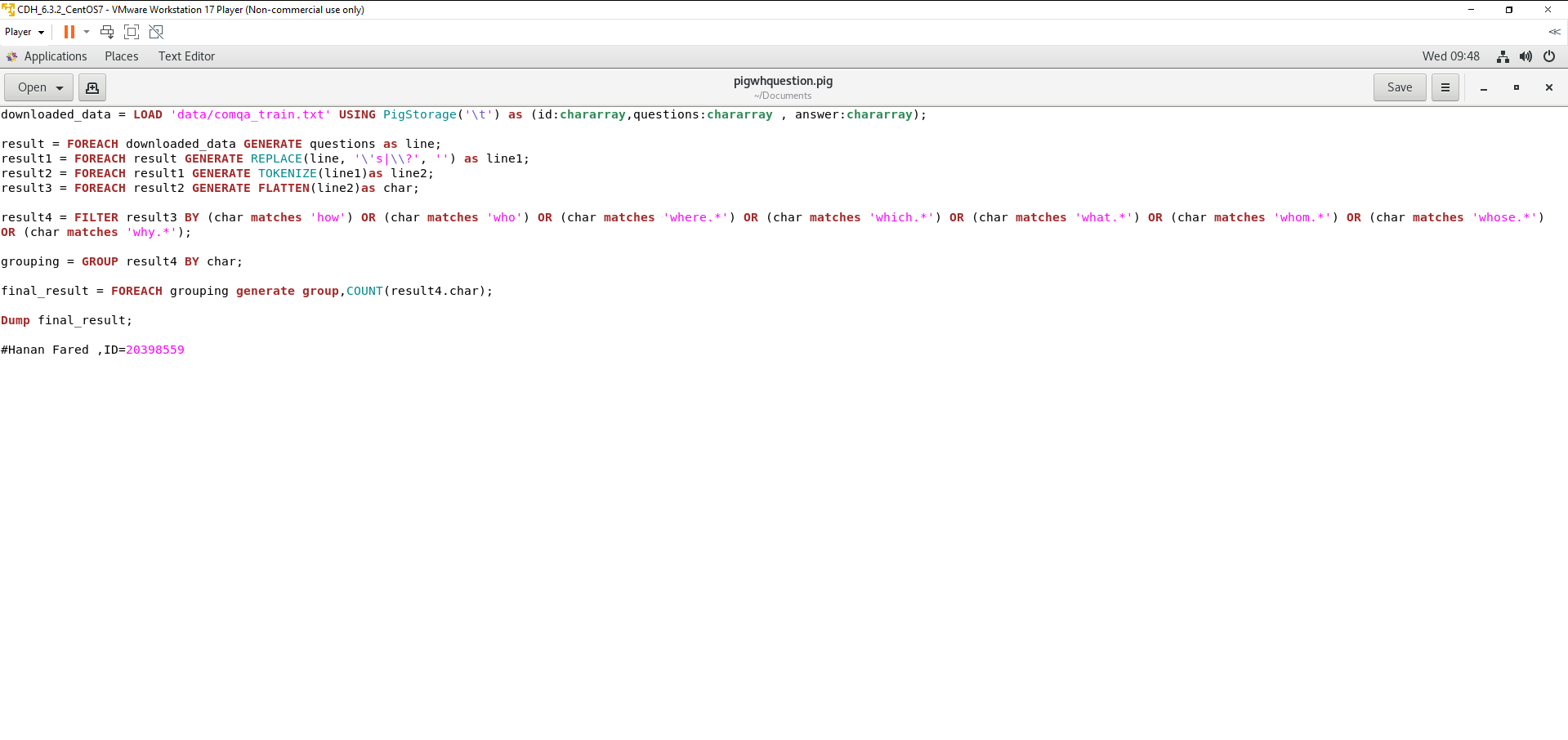
**ID: 20398559**

**Q1- file**

**Q2-**



**Q3,4-** the script pig which called pigwhquestion.pig



**The code is:**

downloaded\_data = LOAD 'data/comqa\_train.txt' USING PigStorage('\t') as (id:chararray,questions:chararray , answer:chararray);

result = FOREACH downloaded\_data GENERATE questions as line;

result1 = FOREACH result GENERATE REPLACE(line, '\'s|\\?', '') as line1;

result2 = FOREACH result1 GENERATE TOKENIZE(line1)as line2;

result3 = FOREACH result2 GENERATE FLATTEN(line2)as char;

result4 = FILTER result3 BY (char matches 'how') OR (char matches 'who') OR (char matches 'where.\*') OR (char matches 'which.\*') OR (char matches 'what.\*') OR (char matches 'whom.\*') OR (char matches 'whose.\*') OR (char matches 'why.\*');

grouping = GROUP result4 BY char;

final\_result = FOREACH grouping generate group,COUNT(result4.char);

Dump final\_result;

**Q5,6-** The output of script (count only the frequency of the WH question words from the question file that you put in hdfs location)

