# Big Data Hadoop Training

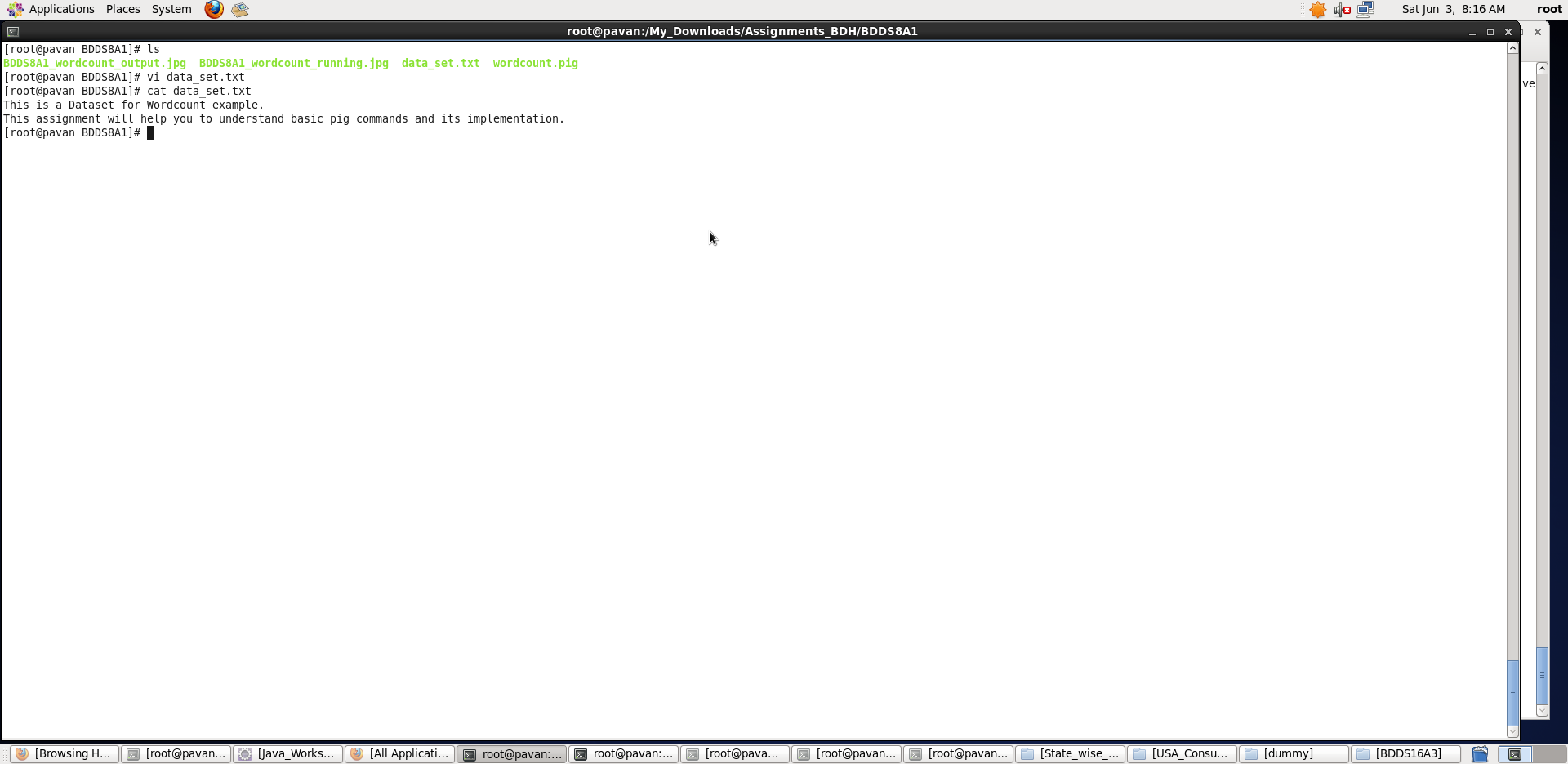
Session 8 Assignment 1 Solution:

Task:

**Write a wordcount program to implement wordcount using Pig.**

**Sample dataset (used) –**

/My\_Downloads/Assignments\_BDH/BDDS8A2/data\_set.txt



A) **Pig Latin Script**:

data\_input = LOAD './data\_set.txt' AS (line:chararray);

words = FOREACH data\_input GENERATE FLATTEN(TOKENIZE(line,' ')) AS word;

grpallwords = GROUP words BY word;

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

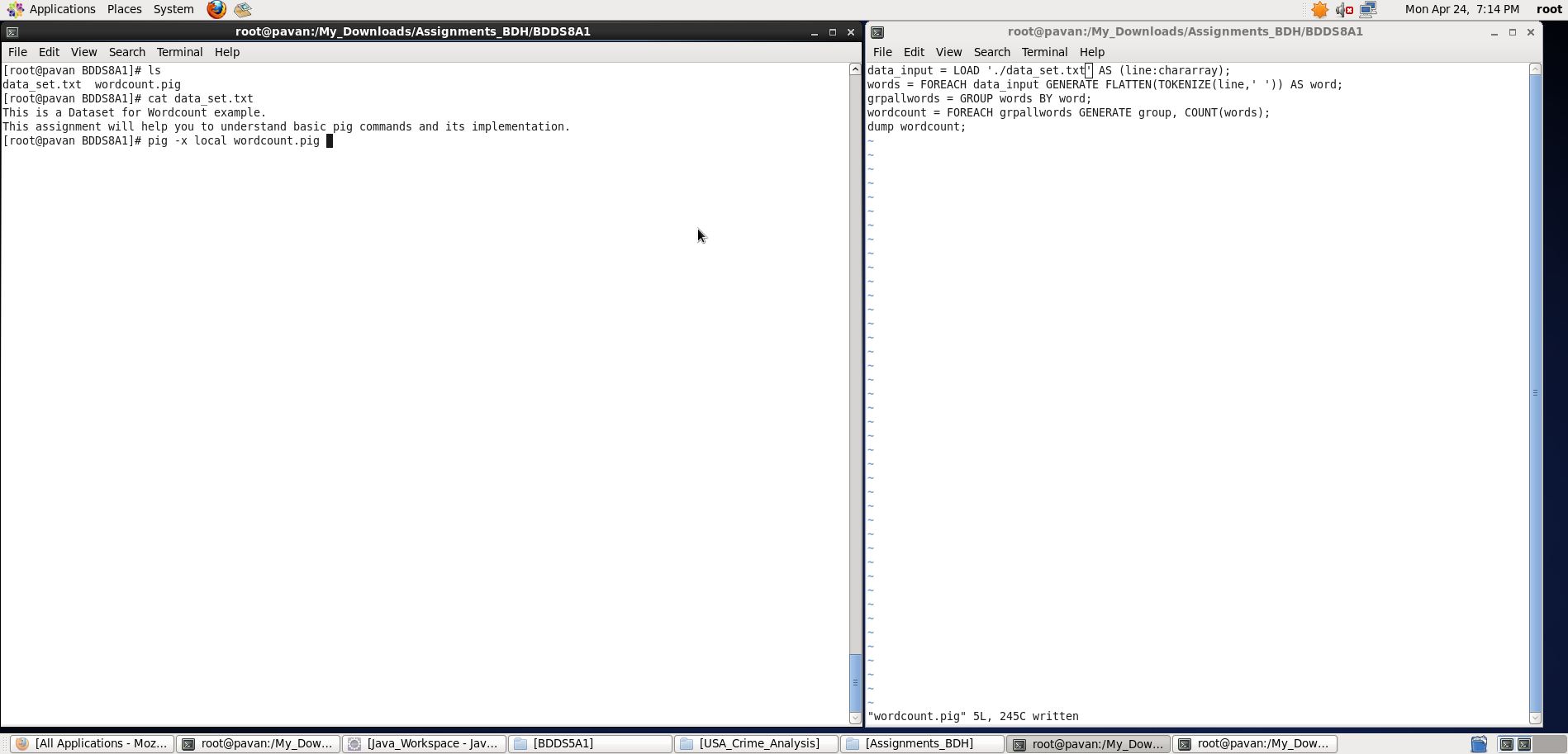
dump wordcount;

**Explanation:**

* Loaded data from data\_set.txt into “data\_input” relation. Here the entire line is taken as one record – chararray.
* Then, foreach record, I have flattened the record and tokenized it with ‘ ’(space) as delimiter into words .
* Now, I have grouped all the words based on the word as key.
* Foreach group, I have counted the number of occurrences of that particular word – key using COUNT function.
* Outputting the word and its number of occurrences in the given entire file.

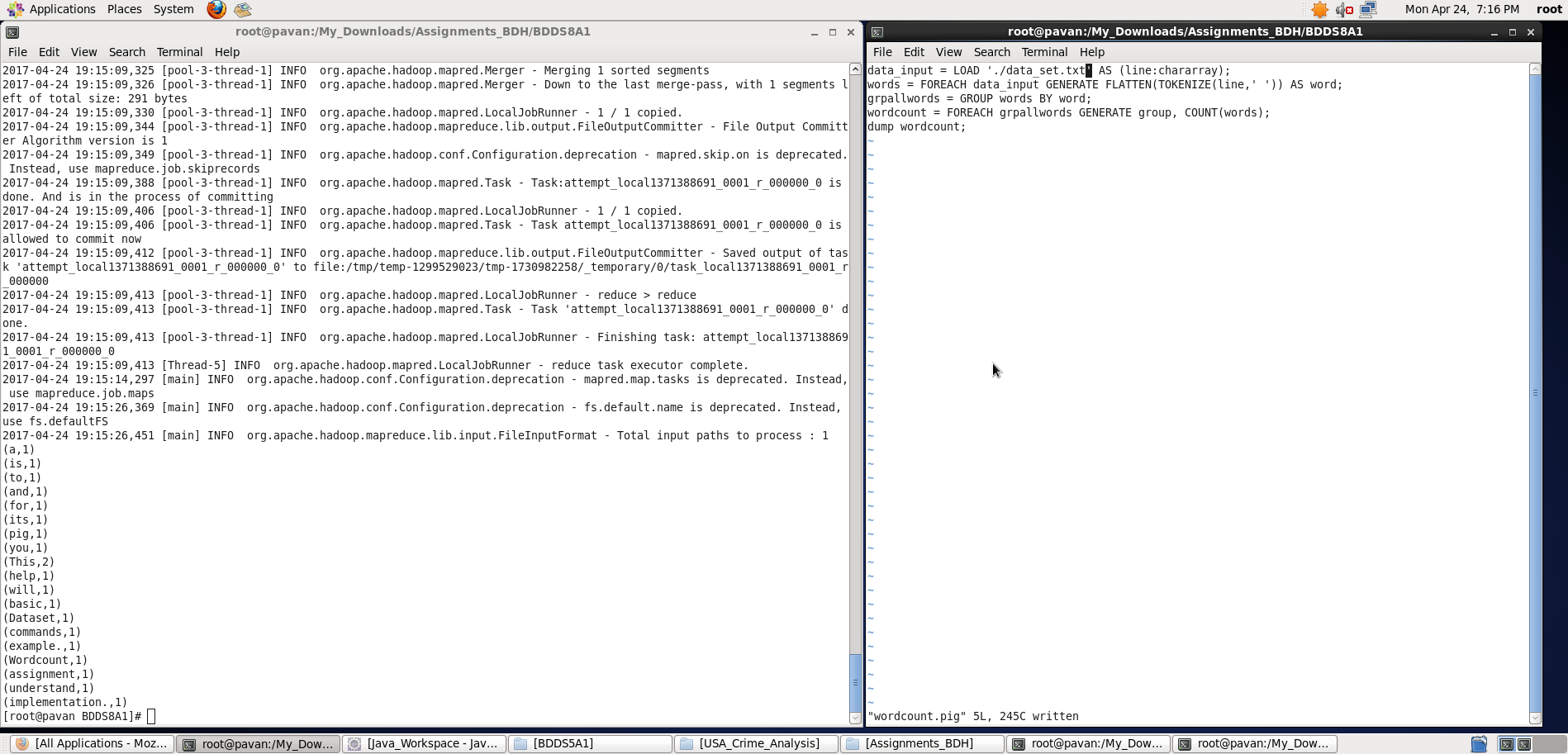
**Screenshots**

Running **Pig Latin Script in Local Mode** inside /My\_Downloads/Assignments\_BDH/BDDS8A1/ directory

****

Output displayed: **<word> <number\_of\_occurrences>**

**Note: We can store the output if required in a file using “store” command.**

****

**Thus, a wordcount program implemented using Pig.**