**Activity 2**

Let us go back to the Wordcount application example. This time, we will try to execute the same MapReduce job in Pig instead of in Java.

Before we start, create this file and copy it to the HDFS.

|  |  |
| --- | --- |
|  | Hello World Bye World |

file01.txt

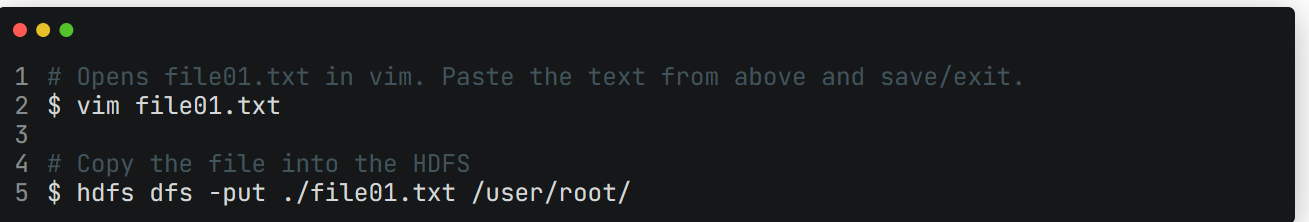
**You can do this with the following commands:**

# Opens file01.txt in vim. Paste the text from above and save/exit.

$ vim file01.txt

# Copy the file into the HDFS

$ hdfs dfs -put ./file01.txt /user/root/



Solution of above :

-- Load input file from HDFS

inputFile = LOAD 'hdfs:///user/root/deepa/file01.txt' AS (line);

-- Tokeize each word in the file (Map)

words = FOREACH inputFile GENERATE FLATTEN(TOKENIZE(line)) AS word;

-- Combine the words from the above stage

grpd = GROUP words BY word;

-- Count the occurence of each word (Reduce)

cntd = FOREACH grpd GENERATE group, COUNT(words);

-- Store the result in HDFS

STORE cntd INTO 'hdfs:///user/root/deepa/results';