









lines = LOAD '/user/hadoop/HDFS\_File.txt' AS (line:chararray);

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

grouped = GROUP words BY word;

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

DUMP wordcount;

The above pig script, first splits each line into words using the **TOKENIZE** operator. The tokenize function creates a bag of words. Using the **FLATTEN** function, the bag is converted into a tuple. In the third statement, the words are grouped together so that the count can be computed which is done in fourth statement.