## MapReduce Assignment:

**Solve the following problems using MRJob in python:**

1. Count the number of words in a text excluding stop words. Ignore the stop words “The”, “is”, “a”.
   Create a text file(word_count.txt) with this content:
   **This is a simple text file. This file is meant to test word count.**
   **The test includes a few common words and uncommon ones.**

2. Find the frequency of each character in a given text file.
   Create a text file(word_freq.txt) with this content:
   **Hello World!**
   **MapReduce is fun.**
   **Python is awesome.**

3. Output the longest word(s) in the input file. If there are multiple words with the same length, output them all.
   Create a text file(long_word.txt) with this content:
   **Sometimes supercalifragilisticexpialidocious is just a word people remember.**
   **However, pseudopseudohypoparathyroidism is another long one!**

4. Each line of the file contains <username> <tweet>. Use MapReduce to count how many tweets each user has made.
   Create a text file(tweets.txt) with this content:
   **alice I love data science.**
   **bob MapReduce is great!**
   **alice Python is fun.**
   **carol I’m learning so much!**
   **bob This is amazing.**

5. Calculate the average length of words in the file using MapReduce. (Keep track of total number of characters and total number of words)
   Create a text file(avg_length.txt) with this content:

**Data is beautiful and powerful.**
**Understanding data helps build solutions.**

6. In a list of social media posts (one per line), count how many times each hashtag (words starting with #) appears.
   Create a text file(hashtag.txt) with this content:
   **I love #Python and #MachineLearning!**
   **#AI is transforming the world. #Python rocks!**
   **Exploring #DataScience with #Python.**

7. Find the most frequent words in a document. If multiple words have the same maximum count, return them all.
   Create a text file(frequent.txt) with this content:
   **dog cat dog bird cat dog elephant cat bird**

8. Given a file with lines like - product_name, quantity_sold. Use MapReduce to find the total quantity sold per product.
   Create a text file(product.txt) with this content:
   **Apple,10**
   **Banana,5**
   **apple,4**
   **Orange,7**
   **banana,3**

9. Given a file with each line as - city, temperature. Compute the average temperature for each city.
   Create a text file(temperatures.txt) with this content:
   **Vancouver,15**
   **Toronto,10**
   **Vancouver,18**
   **Calgary,12**
   **Toronto,14**

**Submission:**
**Please submit individual python scripts and upload the zip file to classrooms. Please name the zip to MapReduce_<Group_Number>. For example, if you are from Group 1, name the zip as MapReduce_Group1**