

## Cloud Computing

Spring 2019

Dr. Farshad Khunjush

### Homework 3: Spark

**Due date: July 20th, 23:55**

The given Datasets are the bag of words.

Using apache spark, You Should write a program for the following purposes:

1. (35 points) Receives a word from the input, count, and print the number of words containing that word. For example, if the word is "test", the following words should be printed (in dataset 1):  
"test, testament, latest, testimonial, protest"
2. (40 points) Counts the number of mutual words between dataset 1 and dataset 2. (The mutual words are the words that exist in both datasets.) (in data set 1,2)
3. (25 points) Counts the number of words with inverse letters. (Like abba) (in dataset1)

#### Reminders:

- Each HW has to be done individually (**For assignments that are similar to each other, the grade is 0.**)
- The programs should be developed using JAVA and apache spark.
- You should submit a report and explain the problem-solving method and show your results. (50% of your points belong to the report for each part)
- Your codes and the report should be submitted before the due date through email to TA.
- In-person delivery will be announced.