**Exercise 3** - SKIP THE STOP WORDS

input.txt

Hadoop is great

Hadoop is scalable

Hadoop is open-source

**mapper.py**

#!/usr/bin/env python3

import sys

# Define stop words as a list directly in the script

stopwords = ["is", "a", "the", "for", "and", "of", "to", "in", "on", "with", "by", "it"]

def mapper():

for line in sys.stdin:

line = line.strip().lower() # Convert to lowercase for consistency

words = line.split() # Split the line into words

for word in words:

if word not in stopwords: # Skip the stop words

print("%s\t%d" % (word, 1)) # Output word with a count of 1

if \_\_name\_\_ == "\_\_main\_\_":

mapper()

**reducer.py**

#!/usr/bin/env python3

import sys

def reducer():

current\_word = None

current\_count = 0

for line in sys.stdin:

line = line.strip()

word, count = line.split('\t')

try:

count = int(count)

except ValueError:

continue

if current\_word == word:

current\_count += count

else:

if current\_word:

print("%s\t%d" % (current\_word, current\_count))

current\_word = word

current\_count = count

if current\_word == word:

print("%s\t%d" % (current\_word, current\_count))

if \_\_name\_\_ == "\_\_main\_\_":

reducer()