Homework – Spark Streaming

Code Program

```
import sys
from pyspark import SparkContext
from pyspark.streaming import StreamingContext

if __name__ == "__main__":
        sc = SparkContext(appName="StreamingErrorCount")
        ssc = StreamingContext(sc, 10)
        ssc.checkpoint("file:///content/spark")
        lines = ssc.socketTextStream(sys.argv[1], int(sys.argv[2]))
        counts = lines.flatMap(lambda line: line.split(" "))\
        .filter(lambda word:"ERROR" in word)\
        .map(lambda word: (word, 1))\
        .reduceByKey(lambda a, b: a+b)
        counts.pprint()
        ssc.start()
        ssc.awaitTermination()
```

1. Dari program dibawah ini, apa arti angka 10 dari ssc = StreamingContext(sc, 10) ?

   Value 10 mean -> batchDuration

2. Pada program diatas, kata apa yang akan dideteksi dan akan diberi value 1 ?
   "ERROR"
3. Apa fungsi dari code lines = ssc.socketTextStream(sys.argv[1], int(sys.argv[2])) ?
   This lines DStream represents the stream of data that will be received from the data server. Each record in this DStream is a line of text. Next, we want to split the lines by space characters into words.