NAME: Xiang Xu 3/22/2017

605.741.31
Module 9 Quiz

Suppose that our Hadoop system contains a large terabyte-sized file with π written out to 10^{12} places. HDFS divides the file into many shards. Write MapReduce pseudocode to determine the number of 3 digit combinations of digits contained in the decimal portion of π. For example, given π written as:

3.14159265358979323846264338327950288419716939937510582097494459230781640628620899862803482534211706798214808651328230664709384460955058223172535940812848111745028410270193852110555964462294895493038196442881097566593344612847564823378678316527120190914564856692346034861045432664821339360726024914127372458700660631558817488152092096282925409171536436789259036001133053054882046652138414695194151160943305727036575959195309218611738193261179310511854807446237996274956735188575272489122793818301194912

The result file should consist of <key, value> pairs that look like the following:

<141, 2>
<415, 2>
<159, 1>
<592, 2>
<926, 1>
…

Write the Map and Reduce pseudocode. Do not concern yourselves with three digit combinations that span shards, where for example one digit is at the end of one shard and two digits are at the beginning of the next shard.

a) Map algorithm

```
Map(key:keyword, value:1):
 Write(keyword, 1)
```

b) Reduce algorithm

```
Reduce(Key: keyword, Value: <list of value>)
For each value in <list of values>:
        Sum += value;
Write(keyword, sum);
```