Quiz

Running Hadoop MapReduce Programs Quiz

1. Download the text to Alice's Adventures in Wonderland from http://www.gutenberg.org/files/11/11-0.txt (If it redirects you to a page with a welcome popup, click on the "Plain Text UTF-8" option on that page or just download the attachment below) and run wordcount on it. This can be done by using hadoop commands. How many times does the word Cheshire occur? (Do **not** include the word 'Cheshire with an apostrophe. The string -->'Cheshire<-- does not count) 1 point

   alice.txt

   Enter a number:

   6

2. **The set of example MapReduce applications includes *wordmedian*, which computes the median length of words in a text file. If you run *wordmedian* using words.txt (the Shakespeare text) as input, what is the median word length?** 1 point

   Note that **wordmedian prints the median length to the terminal** at the end of the MapReduce job; the output file does not contain the median length.

   Enter a number:

   4

Hadoop MapReduce Programs Quiz

In order to check if we really handle the `MapReduce` technique, there are two following questions to be finished. Here I'll explain how to achieve them in ternimal shell step by step.

Question 1

Download the text to Alice's Adventures in Wonderland from http://www.gutenberg.org/cache/epub/11/pg11.txt and run wordcount on it. This can be done by using hadoop commands. **How many times does the word "Cheshire" occur?** (Do not include the word 'Cheshire with an apostrophe. The string `'Chesire` does not count)

```
# Download the text file and named "alice.txt"

# Copy file to HDFS.
hadoop fs -copyFromLocal alice.txt

# Run WordCount for alice.txt with "wordcount". As WordCount executes, the
# Hadoop prints the progress in terms of Map and Reduce. When the `WordCount`
is
# complete, both will say 100%.
hadoop jar /usr/jars/hadoop-examples.jar wordcount alice.txt count

# Copy WordCount results to local file system. Here, the file part-r-00000
# contains the results from WordCount.
hadoop fs –copyToLocal count/part-r-00000 count.txt

# View the WordCount results.
more local.txt

# Search the key "Cheshire" and find its count.
Cheshire 6
```

Question 2

The set of example MapReduce applications includes wordmedian, which computes the median length of words in a text file. **If you run *wordmedian* using words.txt (the Shakespeare text) as input, what is the median word length?**

For this question, we only change the input and use another `MapReduce` application `wordmedian`. Thus, we only need to change one line code as follows.

```
hadoop jar /usr/jars/hadoop-examples.jar wordmedian words.txt wordmedian

# Note that wordmedian prints the median length to the terminal at the end
of
# the MapReduce job; the output file does not contain the median length.
The median is: 4
```