

Map Reduce Tutorials

- #1 Word Counts

Mappers and Reducers

[Here's](#) a quick but comprehensive introduction to the idea of splitting tasks into a MapReduce model. The four important functions involved are:

Map (the mapper function)
EmitIntermediate(the intermediate key,value pairs emitted by the mapper functions)
Reduce (the reducer function)
Emit (the final output, after summarization from the Reduce functions)

We provide you with a single system, single thread version of a basic MapReduce implementation.

Task

The task is to compute the word-frequency-counts of the different words in an input chunk of text. The code for the MapReduce class, reading and splitting the text into words, parts of IO etc. has already been provided. However certain parts of the mapper and reducer functions are incomplete. You need to replace the questionmarks (?). Your task is to fill up these question marks appropriately, such that the program works, and outputs the number of times each word occurs.

Also, this program outputs certain information to the error stream. This information has been logged to help beginners gain a better understanding of the the intermediate steps in a map-reduce process.

Languages Supported

Currently, we provide the base code in Ruby, java and python.

Input Format

A given fragment of text. The code for IO as well as word tokenization has already been provided by us. Only make additions to the mapper and reducer functions to complete them, as suitable.

Output Format

Key-value pairs, in JSON format. The 'key' is a word from the input text and the 'value' is the number of times it occurs.

Sample Input

Military general and first emperor of France, Napoleon Bonaparte was born on August 15, 1769, in Ajaccio, Corsica, France. One of the most celebrated leaders in the history of the West, he revolutionized military organization and training, sponsored Napoleonic Code, reorganized education and established the long-lived Concordat with the papacy. He died on May 5, 1821, on the island of St. Helena in the South Atlantic Ocean.

Considered one of the world's greatest military leaders, Napoleon Bonaparte was born on August 15, 1769, in Ajaccio, Corsica, France. He was the fourth, and second surviving, child of Carlo Buonaparte, a lawyer, and his wife, Letizia Ramolino.

By the time around Napoleon's birth, Corsica's occupation by the French had drawn considerable local resistance. Carlo Buonaparte had at first supported the nationalists siding with their leader, Pasquale Paoli. But after Paoli was forced to flee the island, Carlo switched his allegiance to the French. After doing so he was appointed assessor of the judicial district of Ajaccio in 1771, a plush job that eventually enabled him to enroll his two sons, Joseph and Napoleon, in France's College d'Autun.

Eventually, Napoleon ended up at the military college of Brienne, where he studied for five years, before moving on to the military academy in Paris. In 1785, while Napoleon was at the academy, his father died of stomach cancer. This propelled Napoleon to take the reins as the head of the family. Graduating early from the military academy, Napoleon, now second lieutenant of artillery, returned to Corsica in 1786.

Sample Output

{ "key": "Military", "value": 1 }
{ "key": "general", "value": 1 }
{ "key": "and", "value": 6 }
{ "key": "first", "value": 2 }
{ "key": "emperor", "value": 1 }
{ "key": "of", "value": 12 }
{ "key": "France", "value": 1 }
{ "key": "Napoleon", "value": 5 }
{ "key": "Bonaparte", "value": 2 }
{ "key": "was", "value": 6 }
{ "key": "born", "value": 2 }
{ "key": "on", "value": 5 }
{ "key": "August", "value": 2 }
{ "key": "15", "value": 2 }
{ "key": "1769", "value": 2 }
{ "key": "in", "value": 8 }
{ "key": "Ajaccio", "value": 2 }
{ "key": "Corsica", "value": 2 }
{ "key": "France.", "value": 2 }
{ "key": "One", "value": 1 }
{ "key": "the", "value": 22 }
{ "key": "most", "value": 1 }
{ "key": "celebrated", "value": 1 }
{ "key": "leaders", "value": 1 }
{ "key": "history", "value": 1 }
{ "key": "West", "value": 1 }
{ "key": "he", "value": 3 }
{ "key": "revolutionized", "value": 1 }
{ "key": "military", "value": 5 }
{ "key": "organization", "value": 1 }
{ "key": "training", "value": 1 }
{ "key": "sponsored", "value": 1 }
{ "key": "Napoleonic", "value": 1 }
{ "key": "Code", "value": 1 }
{ "key": "reorganized", "value": 1 }
{ "key": "education", "value": 1 }
{ "key": "established", "value": 1 }
{ "key": "long-lived", "value": 1 }
{ "key": "Concordat", "value": 1 }
{ "key": "with", "value": 2 }
{ "key": "papacy.", "value": 1 }
{ "key": "He", "value": 2 }
{ "key": "died", "value": 2 }
{ "key": "May", "value": 1 }
{ "key": "5", "value": 1 }
{ "key": "1821", "value": 1 }
{ "key": "island", "value": 1 }
{ "key": "St.", "value": 1 }
{ "key": "Helena", "value": 1 }
{ "key": "South", "value": 1 }
{ "key": "Atlantic", "value": 1 }
{ "key": "Ocean.", "value": 1 }
{ "key": "Considered", "value": 1 }
{ "key": "one", "value": 1 }
{ "key": "world's", "value": 1 }
{ "key": "greatest", "value": 1 }
{ "key": "leaders", "value": 1 }
{ "key": "fourth", "value": 1 }
{ "key": "second", "value": 2 }
{ "key": "surviving", "value": 1 }
{ "key": "child", "value": 1 }
{ "key": "Carlo", "value": 3 }
{ "key": "Buonaparte", "value": 1 }
{ "key": "a", "value": 2 }
{ "key": "lawyer", "value": 1 }
{ "key": "his", "value": 4 }
{ "key": "wife", "value": 1 }
{ "key": "Letizia", "value": 1 }
{ "key": "Ramolino.", "value": 1 }
{ "key": "By", "value": 1 }
{ "key": "time", "value": 1 }
{ "key": "around", "value": 1 }
{ "key": "Napoleon's", "value": 1 }
{ "key": "birth", "value": 1 }
{ "key": "Corsica's", "value": 1 }
{ "key": "occupation", "value": 1 }

```
{ "key": "by", "value": 1 }
{ "key": "French", "value": 1 }
{ "key": "had", "value": 2 }
{ "key": "drawn", "value": 1 }
{ "key": "considerable", "value": 1 }
{ "key": "local", "value": 1 }
{ "key": "resistance.", "value": 1 }
{ "key": "Buonaparte", "value": 1 }
{ "key": "at", "value": 3 }
{ "key": "supported", "value": 1 }
{ "key": "nationalists", "value": 1 }
{ "key": "siding", "value": 1 }
{ "key": "their", "value": 1 }
{ "key": "leader,", "value": 1 }
{ "key": "Pasquale", "value": 1 }
{ "key": "Paoli.", "value": 1 }
{ "key": "But", "value": 1 }
{ "key": "after", "value": 1 }
{ "key": "Paoli", "value": 1 }
{ "key": "forced", "value": 1 }
{ "key": "to", "value": 6 }
{ "key": "flee", "value": 1 }
{ "key": "island,", "value": 1 }
{ "key": "switched", "value": 1 }
{ "key": "allegiance", "value": 1 }
{ "key": "French.", "value": 1 }
{ "key": "After", "value": 1 }
{ "key": "doing", "value": 1 }
{ "key": "so", "value": 1 }
{ "key": "appointed", "value": 1 }
{ "key": "assessor", "value": 1 }
{ "key": "judicial", "value": 1 }
{ "key": "district", "value": 1 }
{ "key": "Ajaccio", "value": 1 }
{ "key": "1771,", "value": 1 }
{ "key": "plush", "value": 1 }
{ "key": "job", "value": 1 }
{ "key": "that", "value": 1 }
{ "key": "eventually", "value": 1 }
{ "key": "enabled", "value": 1 }
{ "key": "him", "value": 1 }
{ "key": "enroll", "value": 1 }
{ "key": "two", "value": 1 }
{ "key": "sons,", "value": 1 }
{ "key": "Joseph", "value": 1 }
{ "key": "Napoleon,", "value": 2 }
{ "key": "France's", "value": 1 }
{ "key": "College", "value": 1 }
{ "key": "d'Autun.", "value": 1 }
{ "key": "Eventually,", "value": 1 }
{ "key": "ended", "value": 1 }
{ "key": "up", "value": 1 }
{ "key": "college", "value": 1 }
{ "key": "Brienne,", "value": 1 }
{ "key": "where", "value": 1 }
{ "key": "studied", "value": 1 }
{ "key": "for", "value": 1 }
{ "key": "five", "value": 1 }
{ "key": "years,", "value": 1 }
{ "key": "before", "value": 1 }
{ "key": "moving", "value": 1 }
{ "key": "academy", "value": 1 }
{ "key": "Paris.", "value": 1 }
{ "key": "In", "value": 1 }
{ "key": "1785,", "value": 1 }
{ "key": "while", "value": 1 }
{ "key": "academy,", "value": 2 }
{ "key": "father", "value": 1 }
{ "key": "stomach", "value": 1 }
{ "key": "cancer.", "value": 1 }
{ "key": "This", "value": 1 }
{ "key": "propelled", "value": 1 }
{ "key": "take", "value": 1 }
{ "key": "reins", "value": 1 }
{ "key": "as", "value": 1 }
{ "key": "head", "value": 1 }
{ "key": "family.", "value": 1 }
```

```
{ "key": "Graduating", "value": 1 }  
{ "key": "early", "value": 1 }  
{ "key": "from", "value": 1 }  
{ "key": "now", "value": 1 }  
{ "key": "lieutenant", "value": 1 }  
{ "key": "artillery", "value": 1 }  
{ "key": "returned", "value": 1 }  
{ "key": "Corsica", "value": 1 }  
{ "key": "1786.", "value": 1 }
```

Explanation

We have computed the word-count via the Mapper and Reducer functions.