Map Reduce Tutorials - #3 Composite Keys

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 given input has a number of rows, each with four fields from a table, containing:

```
Country, State, City, Population of the city
```

You are required to output:

```
Country, State, Population of the state (obtained by summing up the population of each city in that state)
```

The code for the MapReduce class, reading and splitting the text, parts related to 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.

Language Support
Java, Python and Ruby.

Input Format

A tab separated file with four columns:
Country, State, City, Population of the city (in millions)
The code for handling IO has already been provided.

Output Format

JSONs, where the key is [country],[state] and the value is the population obtained by adding up the population values for each of the cities provided - **and rounded off to the nearest integer** (in millions).

Sample Input

```
India    Tamil Nadu  Chennai 4.344
India    Maharashtra Mumbai  11.98
India    Maharashtra Pune    2.538
India    Tamil Nadu  Coimbatore  0.931
USA Washington  Seattle 0.652
USA Washington  Tacoma  0.20
```

Sample Output

```
{"key":"India,Tamil Nadu","value":5}
{"key":"India,Maharashtra","value":15}
{"key":"USA,Washington","value":1}
```

Explanation

The population of India,Tamil Nadu is obtained by adding the population of Chennai and Coimbator and rounding it off to the nearest integer (these values are in millions). This process is repeated for India,Maharashtra and USA,Washington.