Big Data Lab Answer Sheet.

Please complete this answer sheet and turn it in at the beginning of class on the due date posted in LEARN.

Part I

|  |  |
| --- | --- |
| **Part 1:** | Answer |
| **1**  (4 pts) | The data is organized vertically with the things on the left and the values of those things on the right |
| **2**  (4 pts) | Everything is nested and looks more like html |
| **3**  (4 pts) | There were 2552 deaths in Alabama and 388 in Alaska and the reason these numbers are so different is because not many people live in Alaska compared to Alabama |
| **4**  (6 pts) | First I went to topics then I went to climate then I went to the side and looked at the format tab. Then I clicked on the link that showed the Json files. I picked USGS US Topo Map Collection. I prefer the Json format because it is easier to read and understand |
| **Part 2:** | Answer |
| **5**  (2 pts) | It shows how many times a certain word appears in the sample text |
| **6**  (2 pts) | It uses each line as a piece |
| **7**  (2 pts) | It runs like this so that many computers can work together at the same time. It runs faster as the sample gets larger because it is more efficient to have many working on small parts than having few work on big parts |
| **8**  (2 pts) | The mapper starts to count letters rather than words |
| **9**  (2 pts) | First the lines are spilt into words then an empty dictionary is created and a loop is used to check all the words. If the word is already in the dictionary the count is increased by 1 and if it is not the word is added and given the value of 1 |
| **10**  (2 pts) | The reducer uses each word and its value from the mapper as an input and emits the amount of times each word is used in the poem |
| **11**  (5 pts) | Because the reducer can simply add all the 1 values for each word |
| **12**  (15 pts) | Alice 81.6666666667  Bob 68.0  Carol 67.0  Dave 78.0  Eve 63.6666666667 |
| **13**  (20 pts) | def mapper(key, value):  school = eval(key) # automatically parses the data in JSON format  # grade\_map['Alice'] contains 95 87 or 63 (and so on) depending on which row the mapper is reading  for label in school: # student is 'Alice', 'Bob', and so on  # get each of the 4 grades in your mapper data line  #Wmr.emit("Enrollments", school["Enrollments"])  if label == 'Enrollments':  Wmr.emit (label,school[label])  if label == 'Dropouts':  Wmr.emit (label,school[label])  def reducer(key, values):  sum = 0  for value in values:  sum = sum + float(value)  Wmr.emit(key, sum)  Dropouts 2  Enrollments 183  Dropouts 8  Enrollments 57  Dropouts 0  Enrollments 26  Dropouts 0  Enrollments 24  Dropouts 5736  Enrollments 63983  Dropouts 5746.0  Enrollments 64273.0 |