Robert Rigler
Nov 2014

Map/Reduce

Question 1:

R	G	G	G
R	B	O	P
P	B	O	R
B	P	R	O

1

R G

R B

2

G G

O P

3

P B

B P

4

O R

R O

Map

1) (Red, 2)
(Green, 1)
(Blue, 1)

2) (Green, 2)
(Orange, 1)
(Purple, 1)

3) (Purple, 2)
(Blue, 2)

4) (Orange, 2)
(Red, 2)

Robert Rigler

Gather

(Red, 2)
(Red, 2)

(Blue, 1)
(Blue, 2)

(Purple, 1)
(Purple, 2)

(Green, 2)
(Green, 1)

(Orange, 2)
(Orange, 1)

Reduce

(Red, 4)

(Green,3)

(Orange,3)

(Blue, 3)

(Purple, 3)

Output

Red : 4 , Green: 3, Orange: 3, Blue: 3, Purple: 3

Question 2:

Companies will collect large amounts of unsorted data(represented by the jumbled words) and analyse it to see if there is an emergent pattern (The answer to the Jumble). They will then take many samples(four different jumbled words) of this data and sort it (de-jumbling)so that they might discern a pattern within the data. These patterns are then compared to the patterns found by the other samples to see if they are concurrent.

The data will be sorted by trying each combination of letters in the jumbled word and then comparing it to a dictionary to check whether the derived pattern is a word. One jumbled pattern may be rearranged into many words.
When all of the matches have been found, the correct letters are then extracted from the words and placed into the correct order in the answer field.
If the answer makes sense then the process is complete , else check if any of the original jumbled words had multiple word matches and use the letters from those instead, trying all combinations until the correct answer is found.

Robert Rigler

Question 3:

1)

a) To find the amount of times the word “whale” appears in the book Moby Dick manually, I would have to read each page of the book and keep record of how many times the word is mentioned.

b) With 50 friends this process would be much faster;
   i. Assign each friend 1/50 of the book and let them scan their assigned pages and keep record of how many times the word appears.
   ii. After they have all finished they would submit their results.
   iii. After submission I would gather the data and find the total.

c) With 10 friends to help sort the data:
   i. Give each friend 5 sets of the submitted data.
   ii. Let them total the data, which leave 5 sets of data.
   iii. Combine the last 5 data sets.

2) If this task were performed by hadoop.

a) Each map node would be assigned an equal share of the data.
b) Each map node will generate the key value pair of how many time the word “whale” appears in the data.
c) Each map node would send their key value pair to the reducer node. If there are multiple reducer nodes then each map node will send their key/value pairs to its assigned reducer node.
d) The reducer nodes will reduce and output the generated data.
e) The generated data can then be combined