Assume there are three map tasks and three reducers such that map task i and reduce task i run on the same data node (i = 0, 1, 2). There are three input-splits

[{cat rat mat sat cat} {cat, sat, bat, sat rat}], [{bat, cat, sat, bat, sat}, {cat, bat, fat, sat cat}] and [{fat, cat, bat, fat, sat} {fat rat mat bat cat}]

and are in data nodes 0, 1 and 2.

Let the partitioner be as shown below:

int getPartition(String s) { if  (s < "f") return 0; else if  (s < "n") return 1;  else return 2; }

QUESTION 1: Illustrate word count

(a) with no combiner

(b) with combiner

QUESTION 2: Count the number of key-value pairs sent across the network in the case of

(a) word count with no combiner

(b) word count with combiner

**QUESTION 1: Illustrate word count**

**(a) with no combiner**

|  |  |  |  |
| --- | --- | --- | --- |
| {cat rat mat sat cat}  {cat, sat, bat, sat rat} | {bat, cat, sat, bat, sat}  {cat, bat, fat, sat cat} | {fat, cat, bat, fat, sat}  {fat rat mat bat cat} | input |
| (cat,1), (rat,1), (mat,1), (sat,1), (cat,1 | (bat,1), (cat,1), (sat,1), (bat,1), (sat,1 | (fat,1), (cat,1), (bat,1), (fat,1), (sat,1 | Map method on 1st record |
| (cat,1), (sat,1), (bat,1), (sat,1), (rat,1 | (cat,1), (bat,1), (fat,1), (sat,1), (cat,1 | (fat,1), (rat,1), (mat,1), (bat,1), (cat,1 | Map method on 2nd record |
| (bat, [1,1,1,1,1,1])  (cat, [1,1,1,1,1,1,1,1]) | (fat, [1,1,1,1])  (mat, [1,1]) | (rat, [1,1,1])  (sat, [1,1,1,1,1,1,1] | Reducer Input |
| (bat, 6  (cat, 8 | (fat, 4  (mat, 2 | (rat, 3  (sat, 7 | Reducer  Input |

**(b) with combiner**

|  |  |  |  |
| --- | --- | --- | --- |
| {cat rat mat sat cat}  {cat, sat, bat, sat rat} | {bat, cat, sat, bat, sat}  {cat, bat, fat, sat cat} | {fat, cat, bat, fat, sat}  {fat rat mat bat cat} | input |
| (cat,1), (rat,1), (mat,1), (sat,1), (cat,1 | (bat,1), (cat,1), (sat,1), (bat,1), (sat,1 | (fat,1), (cat,1), (bat,1), (fat,1), (sat,1 | Map method on 1st record |
| (cat, 2), (mat,1), (rat,1), (sat,1) | (bat,2), (cat,1), (sat,2) | (bat,1), (cat,1), (fat,2), (sat,1) | Combiner output  1st record |
| (cat,1), (sat,1), (bat,1), (sat,1), (rat,1 | (cat,1), (bat,1), (fat,1), (sat,1), (cat,1 | (fat,1), (rat,1), (mat,1), (bat,1), (cat,1 | Map method on 2nd record |
| (bat,1), (cat,1), (rat,1), (sat,2) | (bat,1), (cat,2), (fat,1), (sat,1) | (bat,1), (cat,1), (fat,1), (mat,1), (rat,1) | Combiner output  2nd record |
| (bat, [2,1,1,1,1])  (cat, [2,1,1,1,2,1]) | (fat, [2,1,1])  (mat, [1,1]) | (rat, [1,1,1])  (sat, [1,2,1,2,1] | Reducer Input |
| (bat, 6)  (cat, 8) | (fat, 4)  (mat, 2) | (rat, 3)  (sat, 7) | Reducer  Onput |

**QUESTION 2:**

**(a) word count with no combiner: 23**

**(b) word count with combiner: 18**