**Instructions**

Q1. Write an in-mapper combiner **algorithm** modifying algorithm 3.8 (That is, pairs approach)

Q2. Write an in-mapper combiner **algorithm** modifying algorithm 3.9 (That is, stripes approach)

Q3. Assume that there are two datanodes and two reducers. Note that the Reducer 1 run on datanode 1 and the Reducer 2 run on datanode 2.

Further, let the partitioner  assign all "keys" less than 25 to Reducer 1 and  everything else to Reducer 2.

Also assume that there are four input-splits:

Input-Split 1 : [{12 14 96 81 10} {10 14 81 10 12}]

Input-Split 2 : [{14 81 12 14 96} {81 7 96 12 10]

Input-Split 3 : [{10 12 10 7 96} {81 10 7 12 14}]

Input-Split 4 : [{7 96 12 7 10} {96 10 7 12 14}]

Input-Split 1, and Input-Split 2 are in datanode 1 and others to datanode 2. Let the **neighborhood of X, N(X)**be set of **all term after X and before the next X**.

Example: Let Data block be [a b c a d e]

N(a) = {b, c}, N(b) = {c, a, d, e}, N(c) = {a, d, e}, N(a) ={d, e}, N(d) = {e}, N(e) = {}.

1. Illustrate algorithm 3.8 (with no combiner, **no** in-mapper combining).
2. Illustrate algorithm 3.8 (with in-mapper combining. That is, apply your algorithm Q1).
3. Illustrate algorithm 3.9 (with no combiner, **no** in-mapper combining).
4. Illustrate algorithm 3.9 (with in-mapper combining. That is, apply your algorithm Q2).

Top of Form

Bottom of Form

**Q1**:

class Mapper

    method Initialize      
        H = new Associative Array  
      
    method Map(docid a,doc d)  
        for all term w in doc d do  
            for all term u in Neighbors(w) do  
                t = pair(w;u)  
                H{t} = H{t}+ 1

    method Close  
        for all term t in H do  
            Emit(term t; H{t})

**Q2**:

class Mapper  
      
    method Initialize  
        H = new Associative Array  
      
    method Map(docid a,doc d)  
        for all term w in doc d do  
            subH = H{w}  
            for all term u in Neighbors(w) do  
                subH{u} = subH{u} + 1  
      
    method Close  
        for all term w in H do  
            Emit(term w; stripe H{w})

**Q3:**  
  
a:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **1st Data Node** | | **2nd Data Node** | |  |
| {12 14 96 81 10}  {10 14 81 10 12} | {14 81 12 14 96}  {81 7 96 12 10} | {10 12 10 7 96}  {81 10 7 12 14} | {7 96 12 7 10}  {96 10 7 12 14} | Input splits |
| ((12,14),1) , ((12,96),1), ((12,81),1), ((12,10),1),((14,96),1), ((14,81),1), ((14,10),1), ((96,81),1), ((96,10),1), ((81,10),1 | ((14,81),1), ((14,12),1), ((81,12),1), ((81,14),1),((81,96),1), ((12,14),1), ((12,96),1), ((14,96),1), | ((10,12),1), ((12,10),1), ((12,7),1), ((12,96),1), ((10,7),1), ((10,96),1), ((7,96),1 | ((7,96),1), ((7,12),1), ((96,12),1), ((96,7),1), ((96,10),1), ((12,7),1), ((12,10),1), ((7,10),1 | 1st record mapper output |
| ((10,14),1), ((10,81),1), ((14,81),1), ((14,10),1), ((14,12,1),((81,10),1), ((81,12),1), ((10,12),1 | ((81,7),1), ((81,96),1), ((81,12),1), ((81,10),1), ((7,96),1), ((7,12),1), ((7,10),1), ((96,12),1), ((96,10),1), ((12,10),1 | ((81,10),1), ((81,7),1), ((81,12),1), ((81,14),1), ((10,7),1), ((10,12),1), ((10,14),1), ((7,12),1), ((7,14),1), ((12,14),1 | ((96,10),1), ((96,7),1), ((96,12),1), ((96,14),1), ((10,7),1), ((10,12),1), ((10,14),1), ((7,12),1), ((7,14),1), ((12,14),1 | 2nd record mapper output |
| ((7,10),[1,1]), ((7,12),[1,1,1,1]), ((7,14), [1,1]), ((10,7), [1,1,1]), ((10,12), [1,1,1,1]), ((10,14), [1,1,1]), ((12,7),[1,1]), ((12,10),[1,1,1,1]), ((12,14),[1,1,1,1]), ((14,10),[1,1]), ((14,12),[1,1] | | ((7,96),[1,1,1]), ((10,81), [1]), ((10,96), [1]), ((12,81),[1]), ((12,96),[1,1,1]), ((14,81),[1,1,1]), ((14,96),[1,1]), ((81,7),[1,1]), ((81,10),[1,1,1,1]), ((81,12),[1,1,1,1]), ((81,14),[1,1]), ((81,96),[1,1]),((96,7),[1,1]), ((96,10),[1,1,1,1]),((96,12),[1,1,1]),((96,14),[1]), ((96,81),[1] | | reducer input |
| ((7,10),2), ((7,12),4), ((7,14), 2), ((10,7), 3), ((10,12), 4), ((10,14), 3), ((12,7),2), ((12,10),4), ((12,14),4), ((14,10),2), ((14,12),2) | | ((7,96),3), ((10,81), 1), ((10,96), 1), ((12,81),1), ((12,96),3), ((14,81),3)((14,96),2), ((81,7),2), ((81,10),4),((81,12),4), ((81,14),2), ((81,96),2), ((96,7),2), ((96,10),4),((96,12),3),((96,14),1), ((96,81),1 | | reducer output |

b:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **1st Data Node** | | **2nd Data Node** | |  |  |
| {12 14 96 81 10}  {10 14 81 10 12} | {14 81 12 14 96}  {81 7 96 12 10} | {10 12 10 7 96}  {81 10 7 12 14} | {7 96 12 7 10}  {96 10 7 12 14} | Input splits |  |
| ((10,12),1), ((10,14),1), ((10,81),1), ((12,10),1) , ((12,14),1), ((12,81),1), ((12,96),1), ((14,10),2), ((14,12,1), ((14,81),2), ((14,96),1), ((81,10),2), ((81,12),1), ((96,10),1), ((96,81),1 | ((7,10),1), ((7,12),1), ((7,96),1), ((12,10),1), ((12,14),1), ((12,96),1), ((14,12),1), ((14,81),1), ((14,96),1), ((81,7),1), ((81,10),1), ((81,12),2), ((81,14),1),((81,96),2), ((96,10),1), ((96,12),1 | ((7,12),1), ((7,14),1), ((7,96),1), ((10,7),2), ((10,12),2), ((10,14),1), ((10,96),1), ((12,7),1), ((12,10),1), ((12,14),1), ((12,96),1), ((81,7),1), ((81,10),1), ((81,12),1), ((81,14),1 | ((7,10),1), ((7,12),2), ((7,14),1), ((7,96),1),((10,7),1), ((10,12),1), ((10,14),1), ((12,7),1), ((12,10),1), ((12,14),1), ((96,7),2), ((96,10),2), ((96,12),2), ((96,14),1 | mapper output |  |
|  |
| ((7,10),[1,1]), ((7,12),[1,1,2]), ((7,14), [1,1]), ((10,7), [2,1]), ((10,12), [1,2,1]), ((10,14), [1,1,1]), ((12,7),[1,1]), ((12,10),[1,1,1,1]), ((12,14),[1,1,1,1]), ((14,10),[2]), ((14,12),[1,1] | | ((7,96),[1,1,1]), ((10,81), [1]), ((10,96), [1]), ((12,81),[1]), ((12,96),[1,1,1]), ((14,81),[2,1]), ((14,96),[1,1]), ((81,7),[1,1]), ((81,10),[2,1,1]), ((81,12),[1,2,1]), ((81,14),[1,1]), ((81,96),[2]),((96,7),[2]), ((96,10),[1,1,2]),((96,12),[1,2]),((96,14),[1]), ((96,81),[1] | | reducer input |  |
| ((7,10),2), ((7,12),4), ((7,14), 2), ((10,7), 3), ((10,12), 4), ((10,14), 3), ((12,7),2), ((12,10),4), ((12,14),4), ((14,10),2), ((14,12),2) | | ((7,96),3), ((10,81), 1), ((10,96), 1), ((12,81),1), ((12,96),3), ((14,81),3)((14,96),2), ((81,7),2), ((81,10),4),((81,12),4), ((81,14),2), ((81,96),2), ((96,7),2), ((96,10),4),((96,12),3),((96,14),1), ((96,81),1 | | reducer output |  |

c:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **1st Data Node** | | **2nd Data Node** | |  |
| {12 14 96 81 10}  {10 14 81 10 12} | {14 81 12 14 96}  {81 7 96 12 10} | {10 12 10 7 96}  {81 10 7 12 14} | {7 96 12 7 10}  {96 10 7 12 14} | Input splits |
| (12,[(14,1),(96,1),(81,1),(10,1)]  (14, [(96,1),(81,1), (10,1)])  (96, [(81,1), (10,1)]  (81,[(10,1)] | (14,([81,1),(12,1)]  (81,[(12,1), (14,1),(96,1)]  (12,[(14,1), (96,1)]  (14,[(96,1)] | (10,[(12,1)]  (12,[(10,1), (7,1), (96,1)]  (10,[(7,1), (96,1)]  (7,[(96,1)] | (7,[(96,1), (12,1)]  (96,[(12,1), (7,1), (10,1)]  (12,[(7,1), (10,1)]  (7,[(10,1)] | 1st record mapper output |
| (10,[(14,1), (81,1)]  (14,[(81,1), (10,1), (12,1)]  (81,[(10,1), (12,1)]  (10,[(12,1)] | (81,[(7,1), (96,1), (12,1), (10,1)]  (7,[(96,1), (12,1), (10,1)]  (96,[(12,1), (10,1)]  (12,[(10,1)] | (81,[(10,1), (7,1), (12,1), (14,1)]  (10,[(7,1), (12,1), (14,1)]  (7,[(12,1), (14,1)]  (12,[(14,1)] | (96,[(10,1), (7,1), (12,1), (14,1)]  (10,[(7,1), (12,1), (14,1)]  (7,[(12,1), (14,1)]  (12,[14,1] | 2nd record mapper output |
| (7, [[(96,1), (12,1), (10,1)],[(96,1)], [(12,1), (14,1)], [(96,1), (12,1)], [(10,1)], [(12,1), (14,1)] ]  (10, [[(14,1), (81,1)], [(12,1)],[(12,1)], [(7,1), (96,1)], [(7,1), (12,1), (14,1)], [(7,1), (12,1), (14,1)]]  (12, [[(14,1),(96,1),(81,1),(10,1)], [(14,1), (96,1)], [(10,1)], [(10,1), (7,1), (96,1)], [(14,1)], [(7,1), (10,1)], [14,1]]  (14, [[(96,1),(81,1), (10,1)], [(81,1), (10,1), (12,1)], [81,1),(12,1)], [(96,1)]] | | (81, [[(10,1)], [(10,1), (12,1)], [(12,1), (14,1),(96,1)], [(7,1), (96,1), (12,1), (10,1)], [(10,1), (7,1), (12,1), (14,1)]]  (96, [[(81,1), (10,1)], [(12,1), (10,1)], [(12,1), (7,1), (10,1)], [(10,1), (7,1), (12,1), (14,1)]] | | reducer input |
| (7, [(96,3), (12,4), (10,2), (14,2)]  (10, [(14,3), (81,1), (12,4), (7,3), (96,1)]  (12, [(14,4),(96,3),(81,1),(10,4), (7,2)]  (14, [(96,2),(81,3), (10,2), (12,2)] | | (81, [(10,4), (12,4), (14,2), (96,2), (7,2)]  (96, [(81,1), (10,4), (12,3), (7,2), (14,1)] | | reducer output |

d:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **1st Data Node** | | **2nd Data node** | |  |
| {12 14 96 81 10}  {10 14 81 10 12} | {14 81 12 14 96}  {81 7 96 12 10} | {10 12 10 7 96}  {81 10 7 12 14} | {7 96 12 7 10}  {96 10 7 12 14} | input spits |
| (12,[(14,1),(96,1),(81,1),(10,1)]  (14,[(96,1),(81,2),(10,2),(12,1)]  (96,[(81,1),(10,1)]  (81,[(10,2),(12,1)]  (10,[(14,1),(81,1),(12,1)] | (14,[(81,1),(12,1),(96,1)]  (81,[(12,2),(14,1),(96,2),(7,1),(10,1)]  (12,[(14,1),(96,1),(10,1)]  (7,[(96,1),(12,1),(10,1)]  (96,[(12,1),(10,1)] | (10,[(12,2),(7,2),(96,1),(14,1)]  (12,[(10,1),(7,1),(96,1),(14,1)]  (7,[(96,1),(12,1),(14,1)]  (81,[(10,1),(7,1),(12,1),(14,1)] | (7,[(96,1),(12,2),(10,1),(14,1)]  (96,[(12,2),(7,2),(10,2),(14,1)]  (12,[(7,1),(10,1),(14,1)]  (10,[(7,1),(12,1),(14,1)] | mapper output |
| (7,[(96,1),(12,1),(10,1)],[(96,1),(12,1),(14,1)],[(96,1),(12,2),(10,1),(14,1)]  (10,[(14,1),(81,1),(12,1)],[(12,2),(7,2),(96,1),(14,1)],[(7,1),(12,1),(14,1)]  (12,[(14,1),(96,1),(81,1),(10,1)],[(14,1),(96,1),(10,1)],[(10,1),(7,1),(96,1),(14,1)],[(7,1),(10,1),(14,1)]  (14,[(96,1),(81,2),(10,2),(12,1)],[(81,1),(12,1),(96,1)] | | (81,[(10,2),(12,1)],[(12,2),(14,1),(96,2),(7,1),(10,1)],[(10,1),(7,1),(12,1),(14,1)]  (96,[(81,1),(10,1)],[(12,1),(10,1)],[(12,2),(7,2),(10,2),(14,1)] | | reducer input |
| (7,[(96,3),(12,4),(10,2),(14,2)]  (10,[(14,3),(81,1),(12,4),(7,3),(96,1)]  (12, [(14,4),(96,3),(81,1),(10,4), (7,2)]  (14, [(96,2),(81,3), (10,2), (12,2)] | | (81, [(10,4), (12,4), (14,2), (96,2), (7,2)]  (96, [(81,1), (10,4), (12,3), (7,2), (14,1)] | | reducer output |

Top of Form

Bottom of Form