**Assignment 4**

**Question 1: Write an in-mapper combiner algorithm for the “average problem”. (Pseudo code only; show reducer method too.)**

**class Mapper**

**method setup:**

hmSum <- new HashMap

hmCnt <- new HashMap

**method map (String key, int value):**

hmSum(key) <- hmSum(key) + value

hmCnt(key) <- hmCnt(key) + 1

**method cleanup:**

for all key in hmSum do

Emit( key, pair (hmSum(key), hmCnt(key)) );

**class Reducer**

**method reduce (String key, pairs [(s1, c1), (s2, c2), …]):**

sum <- 0

cnt <- 0

for all pair (s, c) in pairs [(s1,c1), (s2,c2),…] do

sum <- sum + s

cnt <- cnt + c

avg <- sum/cnt

Emit( key, avg );

**Question 2:**

1. No combiner, no in mapper combining

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Machine 1** | | **Machine 2** | | **Machine 3** | |
| **Mapper 1 – Input Split 1- output** | | **Mapper 3 – Input Split 3- output** | | **Mapper 5 – Input Split 5- output** | |
| cherry 1  mango 1  olive 1  cherry 1 | plum 1  cherry 1  banana 1  cherry 1 | banana 1  kiwi 1  plum 1  banana 1 | mango 1  cherry 1  kiwi 1  banana 1 | olive 1  banana 1  radish 1  kiwi 1 | cherry 1  kiwi 1  olive 1  cherry 1 |
| **Mapper 1 – output file** | | **Mapper 3 – output file** | | **Mapper 5 – output file** | |
| cherry 1  mango 1  olive 1  cherry 1  plum 1  cherry 1  banana 1  cherry 1 | | banana 1  kiwi 1  plum 1  banana 1  mango 1  cherry 1  kiwi 1  banana 1 | | olive 1  banana 1  radish 1  kiwi 1  cherry 1  kiwi 1  olive 1  cherry 1 | |
| **Mapper 2–Input Split 2 –output** | | **Mapper 4 – Input Split 4- output** | | **Mapper 6 – Input Split 6- output** | |
| cherry 1  banana 1  radish 1  radish 1 | carrot 1  banana 1  mango 1  cherry 1 | apple 1  mango 1  carrot 1  plum 1 | radish 1  kiwi 1  banana 1  olive 1 | banana 1  radish 1  plum 1  banana 1 | olive 1  cherry 1  banana 1  radish 1 |
| **Mapper 2 – output file** | | **Mapper 4 – output file** | | **Mapper 6 – output file** | |
| cherry 1  banana 1  radish 1  radish 1  carrot 1  banana 1  mango 1  cherry 1 | | apple 1  mango 1  carrot 1  plum 1  radish 1  kiwi 1  banana 1  olive 1 | | banana 1  radish 1  plum 1  banana 1  olive 1  cherry 1  banana 1  radish 1 | |
| **Shuffle & Sort** | | | | | |
| **Reducer 1 input** | | **Reducer 2 input** | | **Reducer 3 input** | |
| apple [1]  banana [1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1]  carrot [1, 1]  cherry [1, 1, 1, 1, 1, 1, 1, 1, 1, 1] | | kiwi [1, 1, 1, 1, 1]  mango [1, 1, 1, 1]  olive [1, 1, 1, 1, 1]  plum [1, 1, 1, 1] | | radish [1, 1, 1, 1, 1, 1] | |

Reducer output is the same for all the cases:

|  |  |  |
| --- | --- | --- |
| **Reducer 1 output** | **Reducer 2 output** | **Reducer 3 output** |
| apple 1  banana 11  carrot 2  cherry 10 | kiwi 5  mango 4  olive 5  plum 4 | radish 6 |

1. With combiner, no in mapper combining (assume that the combiner will work all the time)

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Machine 1** | | **Machine 2** | | **Machine 3** | |
| **Mapper 1 – Input Split 1- output** | | **Mapper 3 – Input Split 3- output** | | **Mapper 5 – Input Split 5- output** | |
| cherry 1  mango 1  olive 1  cherry 1 | plum 1  cherry 1  banana 1  cherry 1 | banana 1  kiwi 1  plum 1  banana 1 | mango 1  cherry 1  kiwi 1  banana 1 | olive 1  banana 1  radish 1  kiwi 1 | cherry 1  kiwi 1  olive 1  cherry 1 |
|  | |  | |  | |
| **Combiner 1 output– saved locally as mapper1 output file** | | **Combiner 3 output– saved locally as mapper3 output file** | | **Combiner 5 output– saved locally as mapper 5 output file** | |
| cherry 4  mango 1  olive 1  plum 1  banana 1 | | banana 3  kiwi 2  plum 1  mango 1  cherry 1 | | olive 2  banana 1  radish 1  kiwi 2  cherry 2 | |
|  | |  | |  | |
| **Mapper 2–Input Split 2 –output** | | **Mapper 4 – Input Split 4- output** | | **Mapper 6 – Input Split 6- output** | |
| cherry 1  banana 1  radish 1  radish 1 | carrot 1  banana 1  mango 1  cherry 1 | apple 1  mango 1  carrot 1  plum 1 | radish 1  kiwi 1  banana 1  olive 1 | banana 1  radish 1  plum 1  banana 1 | olive 1  cherry 1  banana 1  radish 1 |
|  | |  | |  | |
| **Combiner 2 output– saved locally as mapper2 output file** | | **Combiner 4 output– saved locally as mapper4 output file** | | **Combiner 6 output– saved locally as mapper6 output file** | |
| cherry 2  banana 2  radish 2  carrot 1  mango 1 | | apple 1  mango 1  carrot 1  plum 1  radish 1  kiwi 1  banana 1  olive 1 | | banana 3  radish 2  plum 1  olive 1  cherry 1 | |
| **Shuffle & Sort** | | | | | |
| **Reducer 1 input** | | **Reducer 2 input** | | **Reducer 3 input** | |
| apple [1]  banana [1, 2, 3, 1, 1, 3]  carrot [1, 1]  cherry [4, 2, 1, 2, 1] | | kiwi [2, 1, 2]  mango [1, 1, 1, 1]  olive [1, 1, 2, 1]  plum [1, 1, 1, 1] | | radish [2, 1, 1, 2] | |

1. With in-mapper combining

|  |  |  |
| --- | --- | --- |
| **Machine 1** | **Machine 2** | **Machine 3** |
| **Mapper 1 – Input Split 1- output file** | **Mapper 3 – Input Split 3- output file** | **Mapper 5 – Input Split 5- output file** |
| cherry 4  mango 1  olive 1  plum 1  banana 1 | banana 3  kiwi 2  plum 1  mango 1  cherry 1 | olive 2  banana 1  radish 1  kiwi 2  cherry 2 |
|  |  |  |
| **Mapper 2–Input Split 2 –output file** | **Mapper 4 – Input Split 4- output file** | **Mapper 6 – Input Split 6- output file** |
| cherry 2  banana 2  radish 2  carrot 1  mango 1 | apple 1  mango 1  carrot 1  plum 1  radish 1  kiwi 1  banana 1  olive 1 | banana 3  radish 2  plum 1  olive 1  cherry 1 |
| **Shuffle & Sort** | | |
| **Reducer 1 input** | **Reducer 2 input** | **Reducer 3 input** |
| apple [1]  banana [1, 2, 3, 1, 1, 3]  carrot [1, 1]  cherry [4, 2, 1, 2, 1] | kiwi [2, 1, 2]  mango [1, 1, 1, 1]  olive [1, 1, 2, 1]  plum [1, 1, 1, 1] | radish [2, 1, 1, 2] |