**1. Algorithm for calcuating the average mark of a student using map & reduce**

**Class Mapper**

Method Map(integer StudentId, Integer point)

Emit(integer studentId, integer point)

**Class Reducer**

Method Reducer(integer StudentId, integer [point1,point2,point3,….])

sum <- 0

count <- 0

for all integer point e integer [point1,point2,point3,.….]do

sum <- sum + point

count <- count +1

avg\_point <- sum/count

Emit(integer StudentId, float avg\_point)

**3A, With no Combiner**

|  |  |  |
| --- | --- | --- |
| **Machine 1** | **Machine 2** | **Machine 3** |
| **Mapper 1- Input Split 1- Record1 output** | **Mapper 2- Input Split 2- Record1 output** | **Mapper 3- Input Split 3- Record1 output** |
| <cherry,1>  <mango,1>  <radish,1>  <cherry,1> | <cherry,1>  <banana,1>  <radish,1>  <radish,1> | <banana,1>  <banana,1>  <plum,1>  <banana,1> |
| **Mapper 1- Input Split 1- Record2 output** | **Mapper 2- Input Split 2- Record2 output** | **Mapper 3- Input Split 3- Record2 output** |
| <radish,1>  <cherry,1>  <banana,1>  <cherry,1> | <cherry,1>  <cherry,1>  <banana,1>  <cherry,1> | <banana,1>  <cherry,1>  <banana,1>  <banana,1> |
| **sort** | **sort** | **sort** |
| **Mapper 1-output file saved on local disk** | **Mapper 2-output file saved on local disk** | **Mapper 3-output file saved on local disk** |
| <banana,1>  <cherry,1>  <cherry,1>  <cherry,1>  <cherry,1>  <mango,1>  <radish,1>  <radish,1> | <banana,1>  <banana,1>  <cherry,1>  <cherry,1>  <cherry,1>  <cherry,1>  <radish,1>  <radish,1> | <banana,1>  <banana,1>  <banana,1>  <banana,1>  <banana,1>  <banana,1>  <cherry,1>  <plum,1> |
| **Shuffle/sort** | **Shuffle/sort** | **Shuffle/sort** |
| **Reducer 1 input** | **Reducer 2 input** | **Reducer 3 input** |
| **<banana,[1,1,1,1,1,1,1,1,1]>**  **<cherry,[1,1,1,1,1,1,1,1,1]>** | **<mango,[1]>**  **<plum,[1]>** | **<radish,[1,1,1,1]>** |
| **Reducer 1 output** | **Reducer 2 output** | **Reducer 3 output** |
| **<banana,9>**  **<cherry,9>** | **<mango,1>**  **<plum,1>** | **<radish,4>** |

**3A, With Combiner**

|  |  |  |
| --- | --- | --- |
| **Machine 1** | **Machine 2** | **Machine 3** |
| **Mapper 1- Input Split 1- Record1 output** | **Mapper 2- Input Split 2- Record1 output** | **Mapper 3- Input Split 3- Record1 output** |
| <cherry,1>  <mango,1>  <radish,1>  <cherry,1> | <cherry,1>  <banana,1>  <radish,1>  <radish,1> | <banana,1>  <banana,1>  <plum,1>  <banana,1> |
| **Mapper 1- Input Split 1- Record2 output** | **Mapper 2- Input Split 2- Record2 output** | **Mapper 3- Input Split 3- Record2 output** |
| <radish,1>  <cherry,1>  <banana,1>  <cherry,1> | <cherry,1>  <cherry,1>  <banana,1>  <cherry,1> | <banana,1>  <cherry,1>  <banana,1>  <banana,1> |
| **combiner sort** | **combiner sort** | **combiner sort** |
| **Mapper 1-output file saved on local disk** | **Mapper 2-output file saved on local disk** | **Mapper 3-output file saved on local disk** |
| <banana,1>  <cherry,4>  <mango,1>  <radish,2> | <banana,2>  <cherry,4>  <radish,2> | <banana,6>  <cherry,1>  <plum,1> |
| **Shuffle/sort** | **Shuffle/sort** | **Shuffle/sort** |
| **Reducer 1 input** | **Reducer 2 input** | **Reducer 3 input** |
| **<banana,[1,2,6]>**  **<cherry,[4,4,1]>** | **<mango,[1]>**  **<plum,[1]>** | **<radish,[2,2]>** |
| **Reducer 1 output** | **Reducer 2 output** | **Reducer 3 output** |
| **<banana,9>**  **<cherry,9>** | **<mango,1>**  **<plum,1>** | **<radish,4>** |