**Assignment 3**

**Question 1: Illustrate the word count algorithm for the above scenario.**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **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** | | | | | |
| **Machine 1 Reducer 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:**

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

**Question 2: How many tokens (key-value pairs) will be transferred across the network for getting the final reducer output?**

Because the reducer is running on the Machine 1, so the total transactions across network only counted on tokens from Machine 2 and Machine 3:

Tokens from Machine 2:

<banana, {1}> x 4

<kiwi, {1}> x 3

<plum, {1}> x 2

<mango, {1}> x 2

<cherry, {1}> x 1

<apple, {1}> x 1

<carrot, {1}> x 1

<radish, {1}> x 1

<olive, {1}> x 1

* Sum = 16 tokens

Tokens from Machine 3:

<olive, {1}> x 3

<banana, {1}> x 4

<radish, {1}> x 3

<kiwi, {1}> x 2

<cherry, {1}> x 3

<plum, {1}> x 1

* Sum = 16 tokens

Hence, total transactions is 32