**BDT – cs523**

**Assignment 3 – Day 3 (By Sujiv Shrestha)**

---------------------------------------------------------------------------------------------------------------------------------

* Submit your *own work* on time. No credit will be given if the assignment is submitted after the due date.
* Note that the completed assignment should be submitted in .doc, .docx, .rtf or .pdf format only.

Assume that there are six input splits. Input splits 1,2 are on Machine 1, input splits 3,4 are on Machine 2 and input splits 5,6 are on Machine 3.

Input split1 : [cherry mango olive cherry]  
 [plum cherry banana cherry]

Input split2 : [cherry banana radish radish]  
 [carrot banana mango cherry]

Input split3 : [banana kiwi plum banana]  
 [mango cherry kiwi banana]

Input split4 : [apple mango carrot plum]  
 [radish kiwi banana olive]

Input split5 : [olive banana radish kiwi]  
 [cherry kiwi olive cherry]

Input split6 : [banana radish plum banana]  
 [olive cherry banana radish]

Also assume that there’s only one reducer which is running on machine 1.

1. **Illustrate the word count algorithm for the above scenario.**Write your answer on the next page. (A table is already created for you)  
   *Pls refer to the table in next page.*
2. **How many tokens (key-value pairs) will be transferred across the network for getting the final reducer output?**

Ans: Since there is only one reducer and that is in machine 1 and the input splits 1 and 2 are in machine 1, the tokens from mappers for the input splits 3,4,5,6 has to be transferred across the network to be available for reducer in Machine 1. Hence total number of tokens that will be transferred across the network is (8\*4)=32 tokens.

**Answer 1:**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **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 |