







DLI Accelerated Data Science Teaching Kit

### Lecture 10.4 - Example MapReduce Program



The Accelerated Data Science Teaching Kit is licensed by NVIDIA, Georgia Institute of Technology, and Prairie View A&M University under the Creative Commons Attribution-NonCommercial 4.0 International License.







# Example: Compute Word Counts

#### Input

"Apple Orange Mango Orange Grapes Plum"

"Apple Plum Mango Apple Apple Plum"

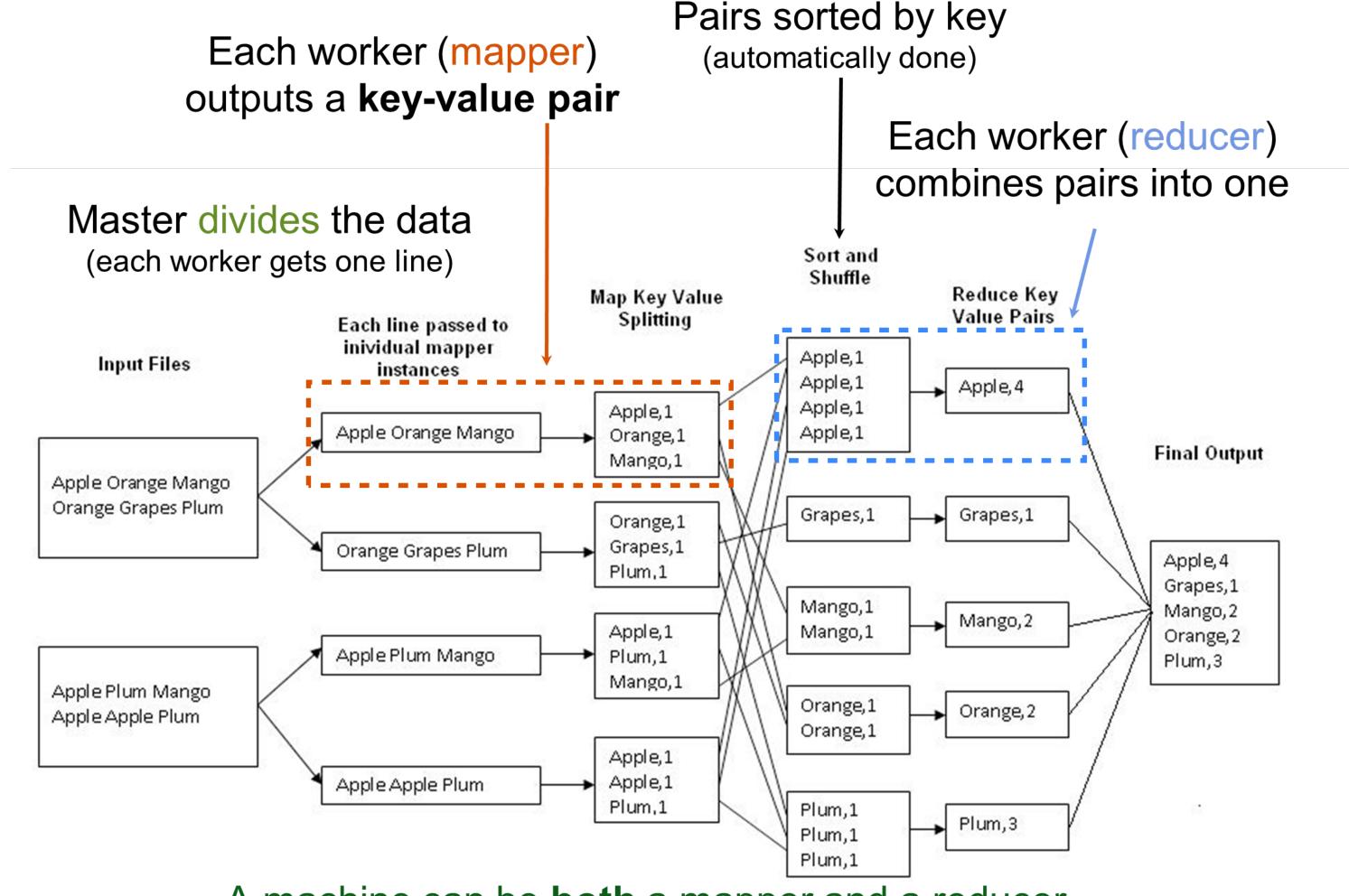
#### **Output**

Apple, 4 Grapes, 1 Mango, 2 Orange, 2 Plum, 3







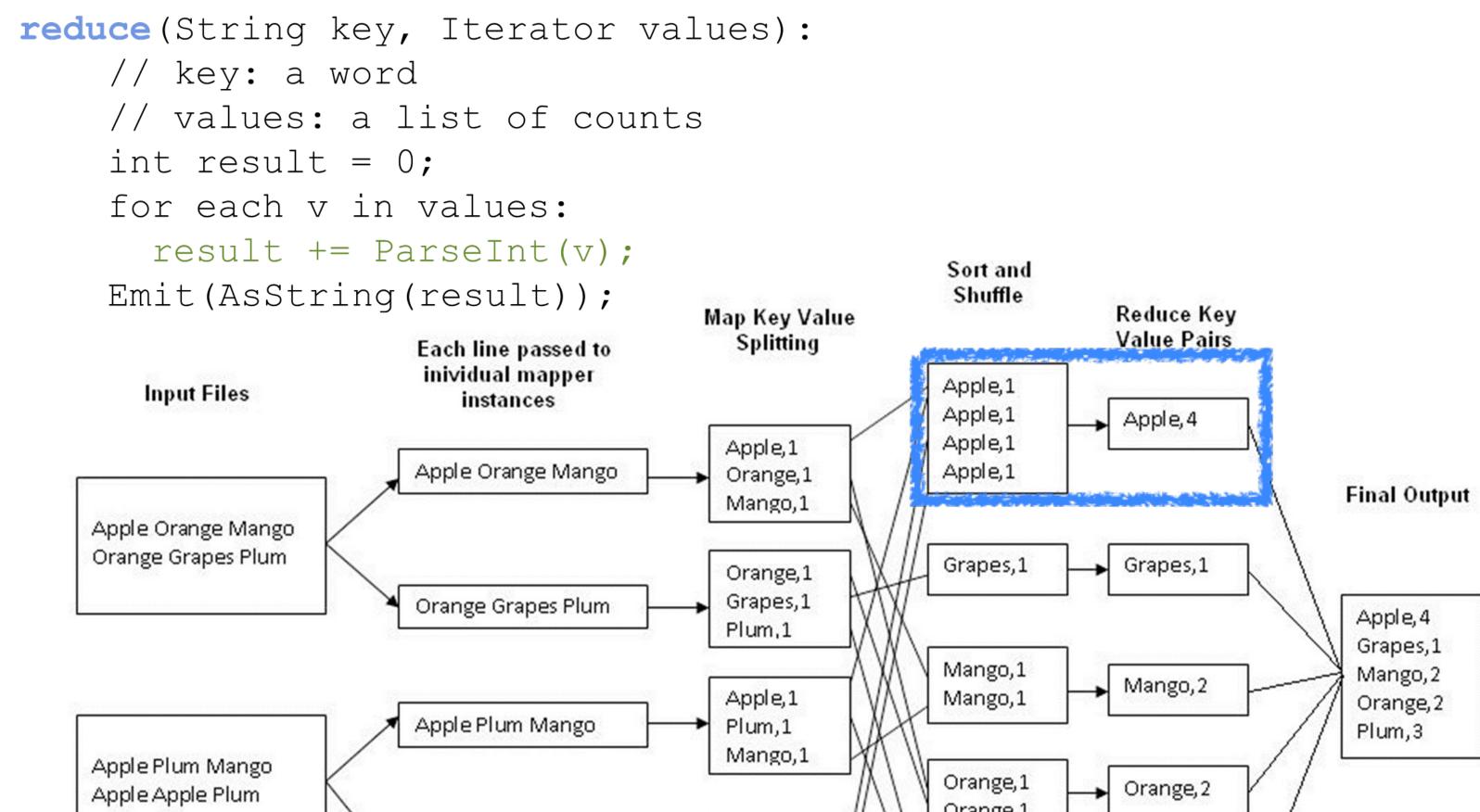


A machine can be both a mapper and a reducer

# How to Implement This?

```
map(String key, String value):
 // key: document id
 // value: document contents
 for each word w in value:
    emit(w, "1");
                                                                  Sort and
                                                                  Shuffle
                                                                               Reduce Key
                                               Map Key Value
                                                                               Value Pairs
                                                  Splitting
                         Each line passed to
                          inividual mapper
                                                                  Apple,1
    Input Files
                                                                  Apple,1
                                                                               Apple, 4
                                                                  Apple,1
                                                 Apple,1
                         Apple Orange Mango
                                                                  Apple,1
                                                 Orange,1
                                                                                                Final Output
                                                 Mango,1
Apple Orange Mango
Orange Grapes Plum
                                                                  Grapes,1
                                                                               Grapes,1
                                                 Orange,1
                                                 Grapes,1
                         Orange Grapes Plum
                                                                                                 Apple, 4
                                                 Plum,1
                                                                                                 Grapes,1
                                                                 Mango,1
                                                                                                 Mango, 2
                                                                               Mango, 2
                                                 Apple,1
                                                                  Mango,1
                                                                                                 Orange, 2
                         Apple Plum Mango
                                                 Plum,1
                                                                                                 Plum,3
                                                 Mango,1
Apple Plum Mango
                                                                  Orange, 1
                                                                               Orange, 2
Apple Apple Plum
                                                                  Orange,1
                                                 Apple,1
                                                 Apple,1
                         Apple Apple Plum
```

## How to Implement This?











DLI Accelerated Data Science Teaching Kit

### Thank You