

## Average example

To calculate an average, we need two values for each group: the sum of the values that we want to average and the number of values that went into the sum. These two values can be calculated on the reduce side very trivially, by iterating through each value in the set and adding to a running sum while keeping a count. After the iteration, simply divide the sum by the count and output the average. However, if we do it this way we cannot use this same reducer implementation as a combiner, because calculating an average is not an associative operation. Instead, our mapper will output two “columns” of data, count and average. For each input record, this will simply be “1” and the value of the field. The reducer will multiply the “count” field by the “average” field to add to a running sum, and add the “count” field to a running count. It will then divide the running sum with the running count and output the count with the calculated average. With this more round-about algorithm, the reducer code can be used as a combiner as associativity is preserved.

The following descriptions of each code section explain the solution to the problem.

**Problem:** Given a list of user’s comments, determine the Average age per location hour of day.

# User

```
<row Id="352268" Reputation="3313"  
CreationDate="2010-05-27T18:34:45.817"  
DisplayName="orangeoctopus"  
EmailHash="93fc5e3d9451bcd3fdb552423ceb52cd"  
LastAccessDate="2011-09-01T13:55:02.013"  
Location="Maryland" Age="26" Views="48" UpVotes="294"  
DownVotes="4" />
```