Your job is to perform the steps of MapReduce to calculate a count of the number of squares, stars, circles, hearts and triangles in the dataset shown in the picture above.

**Step 0:** Store the dataset across 4 partitions in HDFS. **Note: we have already done one partition for you. Hint: Balance the load, but there is more than on possible “correct” partitioning.**

**Step 1:** Map the data. **Hint: Mapping involves clustering like keys together. Show this in the visual placement of keys within a partition.**

**Step 2:** Sort and Shuffle. **Note: as mentioned in lecture, you don’t have to use the same number of nodes in this step as you did before. Let’s use three instead. Hint: Balance the load.**

**Step 3:** Reduce to calculate the final counts. **Hint: Fill in the blank lines to finalize the key-value pairs**

**Modification: Simplify drawing the key-value pair**

The “Map” stage of MapReduce generates key-value pairs. For example :

1

my, my -> (my, 1), (my,1)

Showing that two instances of the word “my” would get mapped to two key-value pairs. You might have noticed that until the Reduce step, the value in all key-value pairs is 1. To make this activity less cluttered visually, we will have you leave out the “,1” part of each key-value pair, and just represent a key-value pair with the appropriate image.