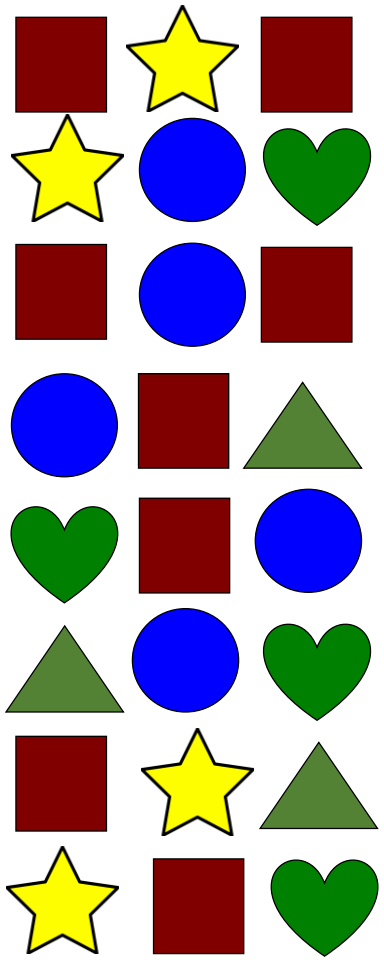
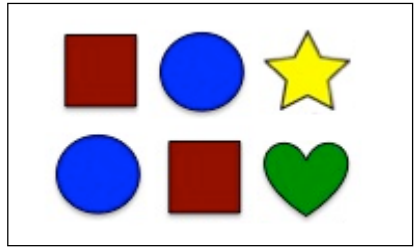
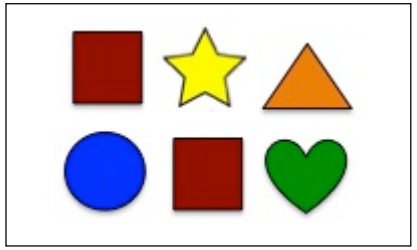
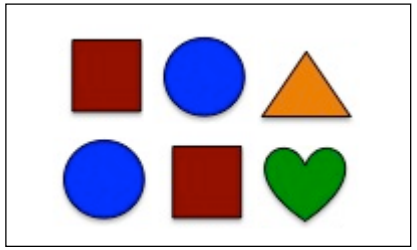


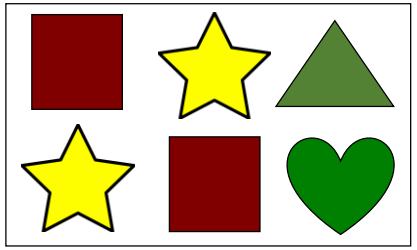
STEP 0 – STORE TO HDFS

Assume 4 data partitions.









1 - MAP

(■,1), (■,1) (▲,1)
(★,1), (★,1) (♥,1)

(■,1), (■,1) (▲,1)
(★,1), (●,1) (♥,1)

(■,1), (■,1) (▲,1)
(●,1), (●,1) (♥,1)

(■,1), (■,1) (★,1)
(●,1), (●,1) (♥,1)

2 – SHUFFLE and SORT

(■,1)(■,1)(■,1)(■,1)
(■,1)(■,1)(■,1)(■,1)

(★,1)(★,1)(★,1)(★,1)
(♥,1)(♥,1)(♥,1)(♥,1)

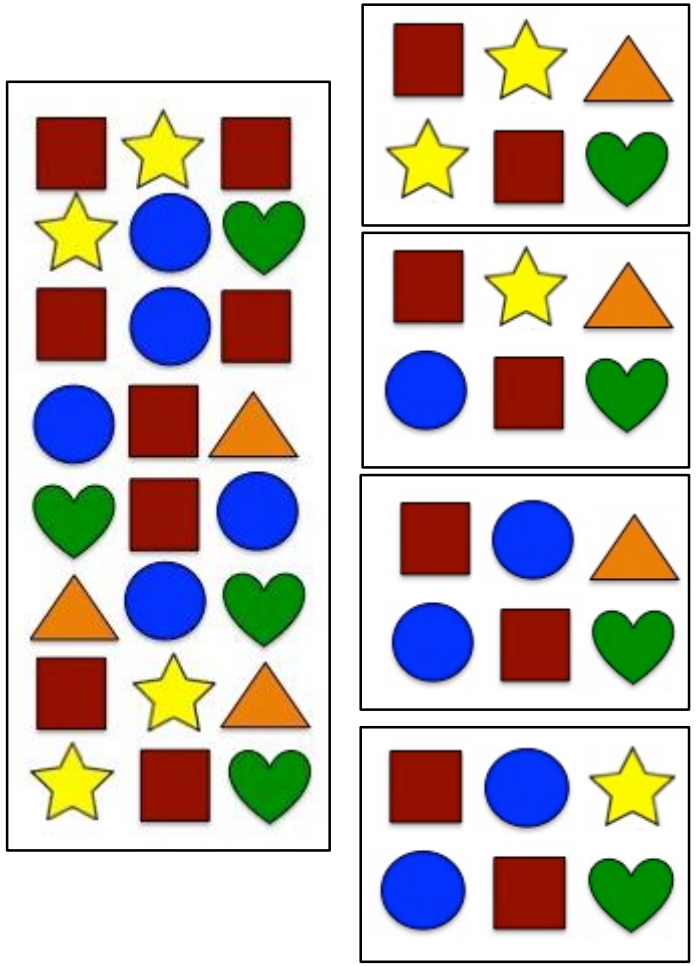
(▲,1)(▲,1)(▲,1)(●,1)
(●,1)(●,1)(●,1)(●,1)

3 - REDUCE

(■ , 8)
(★ , 4)
(● , 5)
(♥ , 4)
(▲ , 3)

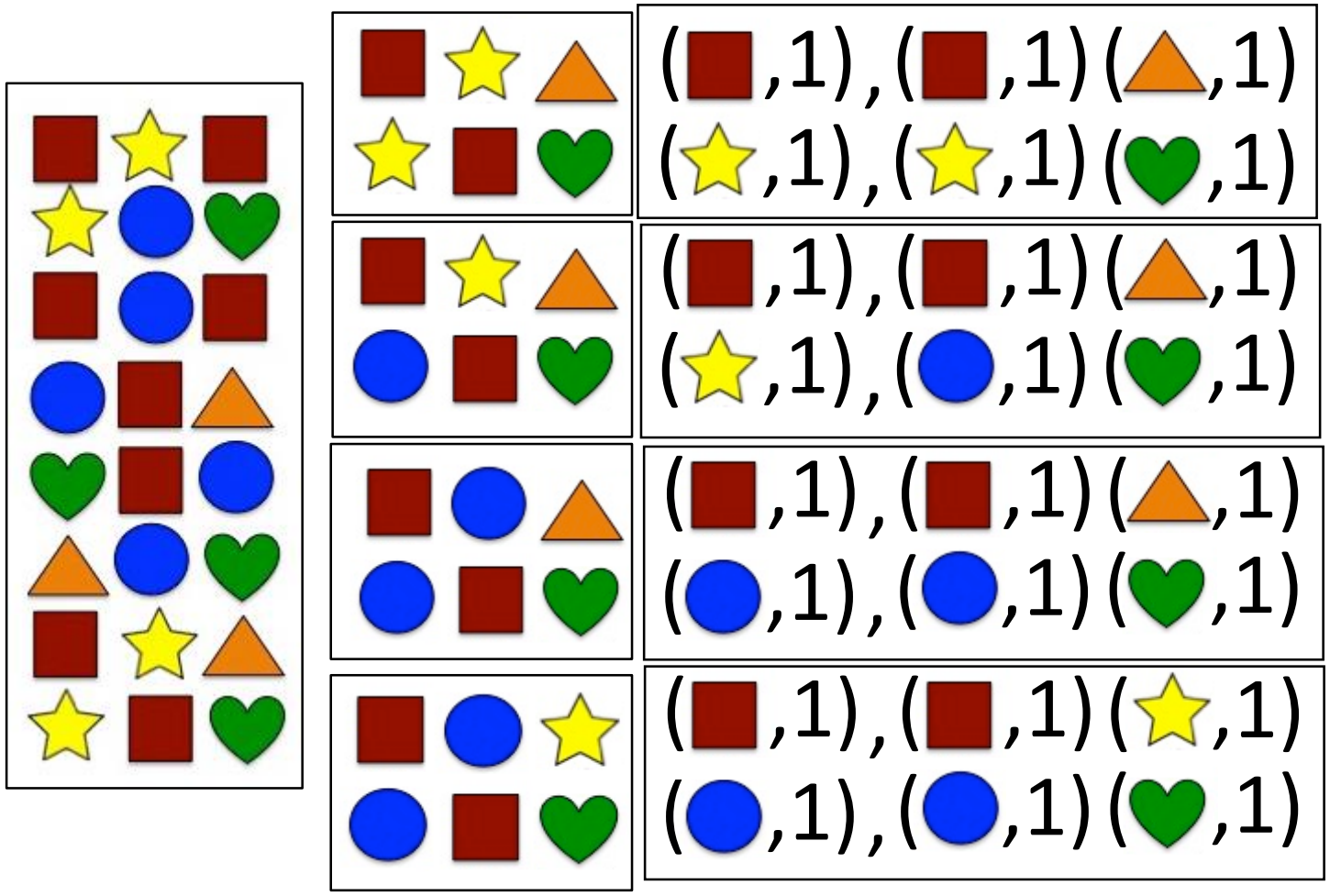
Step 0 – Store to HDFS

4 nodes



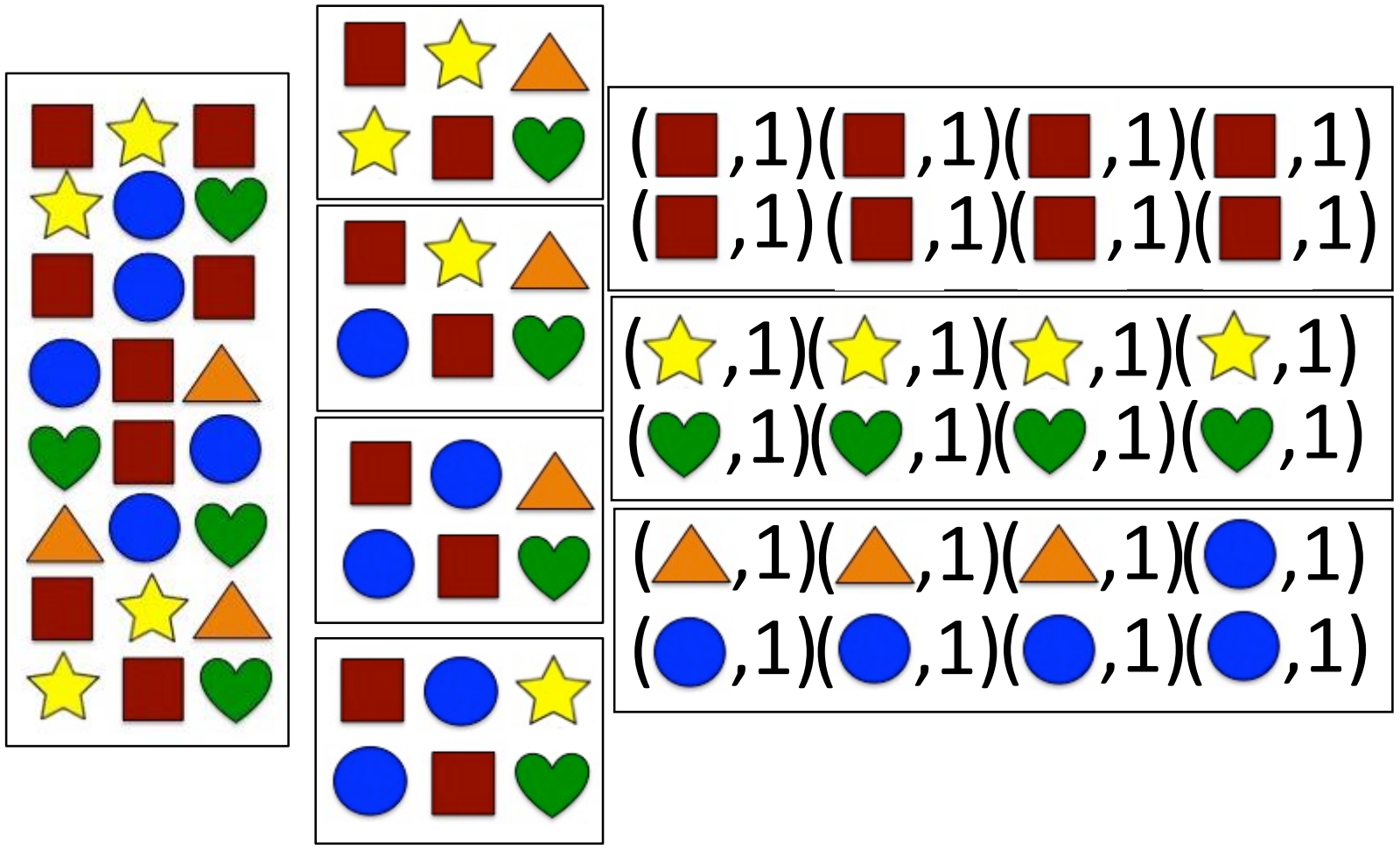
Step 1 – Map the Data

Key-Value pairs



Step 2 – Sort and Shuffle

Same Keys to same Node



Step 3 – Reduce

Add values for same Keys