Assume the following MapReduce program:

```
public void map(LongWritable key, Text value) {
  String line = value.toString();
  StringTokenizer tokenizer = new StringTokenizer(line);
  while (tokenizer.hasMoreTokens()) {
    write(new Text(tokenizer.nextToken()), new IntWritable(1));
  }
}

public void reduce(Text key, Iterable<IntWritable> values) {
  int sum = 0;
  for (IntWritable val : values) {
   sum += val.get();
  }
  write(key, new IntWritable(sum));
}
```

Consider the following data set:

- Block0: “a b b a c | c d c e a”
- Block1: “a b d d a | b b c c f”

Simmulate the execution of the MapReduce code given the following configuration:

- The map and reduce functions are those of the wordcount
    - The combine function shares the implementation of the reduce
- One Split is one block
- The “|” divides the records inside each block
    - We have two records per block
- Hadoop is configured with the parameter *dfs.replication=1*
- We can keep four pairs [key,value] per spill
- We have two mappers and two reducers
    - Machine0, contains block0, runs mapper0 and reducer0
    - Machine1, contains block1, runs mapper1 and reducer1
- The hash function used to shuffle data to the reducers uses the correspondence:
    - {b,d,f}->0
    - {a,c,e}->1

Given Name: ……………………………… Family Name: ………………………………………………………………….

Given Name: ……………………………… Family Name: ………………………………………………………………….

Given Name: ……………………………… Family Name: ………………………………………………………………….

Fill the gaps in each step (numbers correspond to the phase in the MapReduce algorithm):

1) Machine0 contains …. blocks.
   Machine1 contains …. blocks.
2) We keep …… replicas (including the master copy) per block.
3) We have …… splits per machine.

4) Mapper0 reads ……. records.
   Mapper1 reads ……. records.

5) Memory in Machine0:

Memory 1	Memory 2	Memory 3	Memory 4
[,][,] [,][,]	[,][,] [,][,]	[,][,] [,][,]	[,][,] [,][,]

Memory in Machine1:

Memory 1	Memory 2	Memory3	Memory 4
[,][,] [,][,]	[,][,] [,][,]	[,][,] [,][,]	[,][,] [,][,]

6.1) Partitions in machine0:

Memory 1		Memory 2		Memory 3	
Partition 0	Partition 1	Partition 0	Partition 1	Partition 0	Partition 1
[,][,] [,][,]	[,][,] [,][,]	[,][,] [,][,]	[,][,] [,][,]	[,][,] [,][,]	[,][,] [,][,]

Partitions in Machine1:

Memory 1		Memory 2		Memory 3	
Partition 0	Partition 1	Partition 0	Partition 1	Partition 0	Partition 1
[,][,] [,][,]	[,][,] [,][,]	[,][,] [,][,]	[,][,] [,][,]	[,][,] [,][,]	[,][,] [,][,]

6.2) Partitions in machine0:

Memory 1		Memory 2		Memory 3	
Partition 0	Partition 1	Partition 0	Partition 1	Partition 0	Partition 1
[,][,] [,][,]	[,][,] [,][,]	[,][,] [,][,]	[,][,] [,][,]	[,][,] [,][,]	[,][,] [,][,]

Partitions in Machine1:

Memory 1		Memory 2		Memory 3	
Partition 0	Partition 1	Partition 0	Partition 1	Partition 0	Partition 1
[,][,] [,][,]	[,][,] [,][,]	[,][,] [,][,]	[,][,] [,][,]	[,][,] [,][,]	[,][,] [,][,]

6.3) Files in machine0:

File_0_0	File_0_1	File_0_2
[,][,][,][,]	[,][,][,][,]	[,][,][,][,]
File_0_3	File_0_4	File_0_5
[,][,][,][,]	[,][,][,][,]	[,][,][,][,]

Files in Machine1:

File_1_0	File_1_1	File_1_2
[,][,][,][,]	[,][,][,][,]	[,][,][,][,]
File_1_3	File_1_4	File_1_5
[,][,][,][,]	[,][,][,][,]	[,][,][,][,]

7) Merges in machine0:

Merge0	Merge1
[,][,][,][,][,][,]	[,][,][,][,][,][,]

Merges in Machine1:

Merge0	Merge1
[,][,][,][,][,][,]	[,][,][,][,][,][,]

8) Files in machine0:

File_0_0	File_0_1
[,][,][,][,][,][,]	[,][,][,][,][,][,]

Files in Machine1:

File_1_0	File_1_1
[,][,][,][,][,][,]	[,][,][,][,][,][,]

9) Reducer0 reads …………………………….…. files from machine0

and ……………….…….. files from machine1. (answer which files)

Reducer1 reads ……………………………….. files from machine0

and ……………….…….. files from machine1. (answer which files)

10) Merge in machine0:

Merge0	Merge1
[,][,][,]	[,][,][,]

Merge in Machine1:

Merge0	Merge1
[,][,][,]	[,][,][,]

11) Reduce function is executed ……. times in machine0.

Reduce function is executed ……. times in machine1.

12) Files in machine0:

File_0_0	File_0_1
[,][,][,][,][,][,]	[,][,][,][,][,][,]

Files in Machine1:

File_1_0	File_1_1
[,][,][,][,][,][,]	[,][,][,][,][,][,]