

TP: Big Data Hadoop MapReduce

Part 1

Modification du Mapper:

```
while (tokenizer.hasMoreTokens()) {
    word.set(tokenizer.nextToken());
    if (word.charAt(0) == 'm' || word.charAt(0) == 'M') {
        context.write(word, one);
    }
}
```

Modification du Reducer:

```
int sum = 0;
if (key.charAt(0) != 'm' && key.charAt(0) != 'M') {
    return;
}
for (IntWritable val : values) {
    sum += val.get();
}
context.write(key, new IntWritable(sum));
```

La modification du Mapper est plus avantageuse car cela réduit l'output et par conséquent l'input du reducer. Réduire les données le plus tôt possible est le plus avantageux.