



**UNIVERSITE GASTON BERGER**

*L'excellence au service du développement*

---

INSTITUT  POLYTECHNIQUE  
DE SAINT-LOUIS

Evaluation : **TP2 MAP\_REDUCE**

**Michel Nassalang**

## 1. Map Reduce en ligne de commande

```
[vagrant@10 ~]$ mkdir codes
[vagrant@10 ~]$ ls
codes                mysql-5.7.33          mysql-connector-java-5.1.48.tar.gz  shareFolder
derby.log            mysql-5.7.33.tar.gz   mysql-connector-java-5.1.49.zip     sqoop-1.4.7.bin__hadoop-2.6.0.tar.gz
itversity-books      mysql57-community-release-el7-10.noarch.rpm  nassalang                          wget-log
metastore_db         mysql-connector-java-5.1.48  README.md
[vagrant@10 ~]$ cp shareFolder/scripts/datasets/input.txt /home/vagrant/codes
[vagrant@10 ~]$ cp shareFolder/scripts/reduce.pl /home/vagrant/codes
[vagrant@10 ~]$ cp shareFolder/scripts/map.pl /home/vagrant/codes
[vagrant@10 ~]$ ls codes
input.txt  map.pl  reduce.pl
[vagrant@10 ~]$ cd codes/
[vagrant@10 codes]$ cat input.txt | perl map.pl |perl reduce.pl
-----City      : Total Cost
-----Bangalore : 270
Kolkata         : 390
NewDelhi        : 440
[vagrant@10 codes]$ |
```

## 2. Map Reduce dans Hadoop

```
[vagrant@10 codes]$ hdfs dfs -copyFromLocal input.txt /user/vagrant/datasets
2023-05-21 19:12:40,727 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted = false
[vagrant@10 codes]$ hdfs dfs -ls /user/vagrant/datasets
Found 2 items
-rw-r--r-- 1 vagrant supergroup      226 2023-05-21 19:12 /user/vagrant/datasets/input.txt
-rw-r--r-- 1 vagrant supergroup 118784 2021-11-27 10:39 /user/vagrant/datasets/mlflow.db
[vagrant@10 codes]$ mapred streaming -input /user/vagrant/datasets/input.txt -output /user/vagrant/datasets/output -mapper "perl $(pwd)/map.pl" -reducer "perl $(pwd)/reduce.pl"
packageJobJar: [] [/opt/hadoop-3.2.1/share/hadoop/tools/lib/hadoop-streaming-3.2.1.jar] /tmp/streamjob6572232605135205332.jar tmpDir=null
2023-05-21 19:12:54,530 INFO client.RMPProxy: Connecting to ResourceManager at /0.0.0.0:8032
2023-05-21 19:12:54,891 INFO client.RMPProxy: Connecting to ResourceManager at /0.0.0.0:8032
2023-05-21 19:12:55,434 INFO mapreduce.JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/vagrant/.staging/job_1684692012304_0002
2023-05-21 19:12:55,630 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted = false
2023-05-21 19:12:56,326 INFO mapred.FileInputFormat: Total input files to process : 1
2023-05-21 19:12:56,451 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted = false
2023-05-21 19:12:56,589 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted = false
2023-05-21 19:12:56,667 INFO mapreduce.JobSubmitter: number of splits:2
2023-05-21 19:12:57,131 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted = false
2023-05-21 19:12:57,349 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1684692012304_0002
2023-05-21 19:12:57,349 INFO mapreduce.JobSubmitter: Executing with tokens: []
2023-05-21 19:12:57,873 INFO conf.Configuration: resource-types.xml not found
2023-05-21 19:12:57,874 INFO resource.ResourceUtils: Unable to find 'resource-types.xml'.
2023-05-21 19:12:58,822 INFO impl.YarnClientImpl: Submitted application application_1684692012304_0002
2023-05-21 19:12:58,953 INFO mapreduce.Job: The url to track the job: http://10.0.2.15:8088/proxy/application_1684692012304_0002/
2023-05-21 19:12:58,956 INFO mapreduce.Job: Running job: job_1684692012304_0002
2023-05-21 19:13:16,019 INFO mapreduce.Job: Job job_1684692012304_0002 running in uber mode : false
2023-05-21 19:13:16,020 INFO mapreduce.Job:  map 0% reduce 0%
2023-05-21 19:13:27,601 INFO mapreduce.Job:  map 100% reduce 0%
2023-05-21 19:13:35,822 INFO mapreduce.Job:  map 100% reduce 100%
2023-05-21 19:13:37,896 INFO mapreduce.Job: Job job_1684692012304_0002 completed successfully
2023-05-21 19:13:38,043 INFO mapreduce.Job: Counters: 54
```

```
2023-05-21 19:13:38,043 INFO streaming.StreamJob: Output directory: /user/vagrant/datasets/output
[vagrant@10 codes]$ hdfs dfs -ls /user/vagrant/datasets/output
Found 2 items
-rw-r--r-- 1 vagrant supergroup      0 2023-05-21 19:13 /user/vagrant/datasets/output/_SUCCESS
-rw-r--r-- 1 vagrant supergroup 127 2023-05-21 19:13 /user/vagrant/datasets/output/part-000000
[vagrant@10 codes]$ |
```

```
[vagrant@10 codes]$ hdfs dfs -cat /user/vagrant/datasets/output/part-000000
2023-05-21 19:21:56,527 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted = false
-----City      : Total Cost
-----Bangalore : 270
Kolkata         : 390
NewDelhi        : 440
[vagrant@10 codes]$ |
```

### 3. Exercice wordcount

Programme MapReduce en java, ou python, ou perl qui permet le faire comptage des mots contenues dans le fichier Senegal.txt :

```
import java.io.BufferedReader;
import java.io.FileReader;
import java.io.IOException;
import java.util.Scanner;

public class Wordcount {
    public static Scanner scanner = new Scanner(System.in);

    public static void main(String[] args) {
        System.out.println("Veuillez spécifier le chemin d'accès du fichier en tant qu'argument.");
        String filePath = scanner.next();
        int wordCount = 0;
        try (BufferedReader br = new BufferedReader(new FileReader(filePath))) {
            String line;
            while ((line = br.readLine()) != null) {
                String[] words = line.split("[\\s,.;!?!?]+"); // Divise en utilisant les identificateurs comme séparateurs
                wordCount += words.length;
            }
        } catch (IOException e) {
            System.out.println("Erreur lors de la lecture du fichier : " + e.getMessage());
            System.exit(1);
        }

        System.out.println("Nombre de mots : " + wordCount);
    }
}
```