



**Module :** Virtualisation/Cloud/Plateforme Big data

---

**Activité 3 : INITIATION À HADOOP ET MAPREDUCE**

**Membres du groupe :**

Babacar DIAGO

Alioune GUEYE

**Professeur :**

M. TINE

**Année Universitaire 2019 - 2020**

## DÉMARRAGE DE HADOOP APRÈS INSTALLATION :

## Formatage du système de fichier pour prise en compte de HDFS

```
Fichier Édition Affichage Rechercher Terminal Aide
alioune@alioune-virtual-machine:~$ hdfs namenode -format
2021-06-16 16:02:02,620 INFO namenode.NameNode: STARTUP_MSG:
/*****
STARTUP_MSG: Starting NameNode
STARTUP_MSG:   host = alioune-virtual-machine/127.0.1.1
STARTUP_MSG:   args = [-format]
STARTUP_MSG:   version = 3.2.2
*****/
2021-06-16 16:02:03,890 INFO namenode.NameNode: *** NAME NODE FORMATTING SUCCESSFUL ***
2021-06-16 16:02:03,900 INFO namenode.NameNode: ****
alioune/dfs/name has been successfully formatted.
2021-06-16 16:02:17,080 INFO namenode.FSImageFormatProtobuf: Saving image file
/tmp/hadoop-alioune/dfs/name/current/fsimage.ckpt_000000000000000000 using no
compression
2021-06-16 16:02:18,215 INFO namenode.FSImageFormatProtobuf: Image file /tmp/had
oop-alioune/dfs/name/current/fsimage.ckpt_000000000000000000 of size 402 byte
s saved in 1 seconds .
2021-06-16 16:02:18,397 INFO namenode.NNStorageRetentionManager: Going to retai
n 1 images with txid >= 0
2021-06-16 16:02:18,501 INFO namenode.FSImage: FSImageSaver clean checkpoint: t
xid=0 when meet shutdown.
2021-06-16 16:02:18,502 INFO namenode.NameNode: SHUTDOWN_MSG:
/*****
SHUTDOWN_MSG: Shutting down NameNode at alioune-virtual-machine/127.0.1.1
*****/
```

## Démarrage des démons NameNode et DataNode

```
alioune@alioune-virtual-machine:~$ start-dfs.sh
Starting namenodes on [localhost]
Starting datanodes
Starting secondary namenodes [alioune-virtual-machine]
```

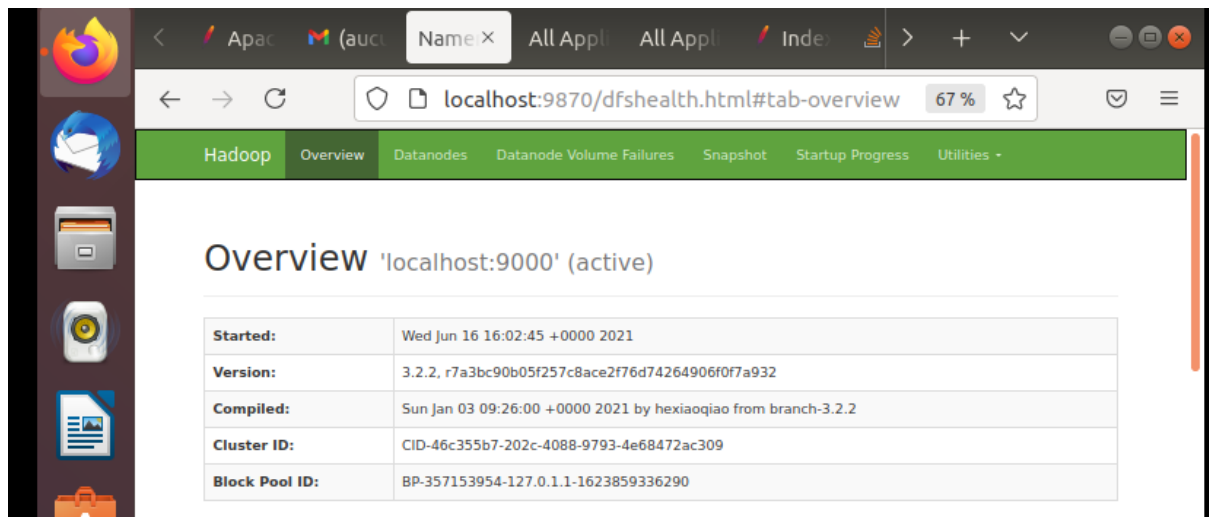
## Démarrez le démon ResourceManager et le démon NodeManager

```
alioune@alioune-virtual-machine:~$ start-yarn.sh
Starting resourcemanager
Starting nodemanagers
```

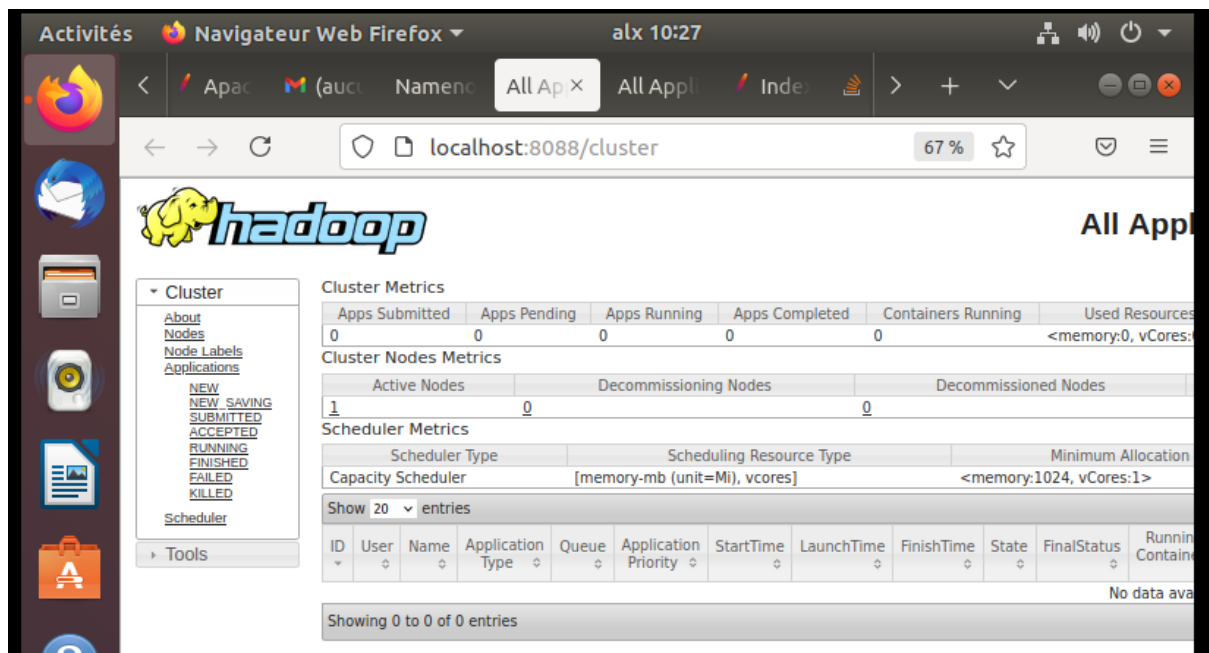
## Test sur le fonctionnement des ressource de HADOOP

```
alioune@alioune-virtual-machine:~$ jps
2832 ResourceManager
2994 NodeManager
2295 DataNode
2522 SecondaryNameNode
3149 Jps
2125 NameNode
```

Accès à l'interface Web pour le NameNode ( <http://localhost:9870/>)



Accès à l'interface Web du ResourceManager ( <http://localhost:8088/>)



## Manipulation des données de l'étude de cas : analyse de ventes:

### Organisation du répertoire de travail

```
diago@ubuntu:~/udacity_training$ ls -R
.:
code data

./code:
mapper.py reducer.py

./data:
purchases.txt
diago@ubuntu:~/udacity_training$
```

### les répertoires HDFS requis pour exécuter les tâches MapReduce

```
diago@ubuntu:~$ hdfs dfs -mkdir /user
diago@ubuntu:~$ hdfs dfs -mkdir /user/diago
```

## Activité 1

### un répertoire dans HDFS, appelé myinput

```
diago@ubuntu:~$ hdfs dfs -mkdir myinput
```

### copier le fichier purchases.txt dans HDFS sous le répertoire myinput

```
diago@ubuntu:~$ hdfs dfs -mkdir myinput
```

### affichage du contenu du répertoire myinput

```
diago@ubuntu:~$ hdfs dfs -ls myinput
Found 1 items
-rw-r--r--  1 diago supergroup  211312924 2021-06-18 02:02 myinput/purchases.t
xt
diago@ubuntu:~$
```

### Afficher les dernières lignes du fichier

```
diago@ubuntu:~$ hdfs dfs -tail myinput/purchases.txt
31      17:59  Norfolk Toys      164.34  MasterCard
2012-12-31  17:59  Chula Vista      Music   380.67  Visa
2012-12-31  17:59  Hialeah Toys     115.21  MasterCard
2012-12-31  17:59  Indianapolis     Men's Clothing 158.28  MasterCard
2012-12-31  17:59  Norfolk Garden   414.09  MasterCard
2012-12-31  17:59  Baltimore        DVDs    467.3   Visa
2012-12-31  17:59  Santa Ana        Video Games 144.73  Visa
2012-12-31  17:59  Gilbert Consumer Electronics 354.66  Discover
2012-12-31  17:59  Memphis Sporting Goods 124.79  Amex
2012-12-31  17:59  Chicago Men's Clothing 386.54  MasterCard
2012-12-31  17:59  Birmingham       CDs     118.04  Cash
2012-12-31  17:59  Las Vegas        Health and Beauty 420.46  Amex
2012-12-31  17:59  Wichita Toys     383.9   Cash
2012-12-31  17:59  Tucson Pet Supplies 268.39  MasterCard
2012-12-31  17:59  Glendale         Women's Clothing 68.05   Amex
2012-12-31  17:59  Albuquerque      Toys    345.7   MasterCard
2012-12-31  17:59  Rochester        DVDs    399.57  Amex
2012-12-31  17:59  Greensboro       Baby    277.27  Discover
2012-12-31  17:59  Arlington        Women's Clothing 134.95  MasterC
ard
2012-12-31  17:59  Corpus Christi   DVDs    441.61  Discover
diago@ubuntu:~$
```

## Activité 2

- Que permet de faire chaque ligne de ce code?
- Tester ce mapper en local sur les 50 premières lignes du fichier purchases.txt en tapant l'instruction suivante, directement à partir de votre répertoire code: `head -50 ../data/purchases.txt | mapper.py`

```
allioune@allioune-virtual-machine:~/udacity_training/code$ head -50 ../data/purch
ases.txt | python3 mapper.py
San Jose      214.05
Fort Worth    153.57
San Diego     66.08
Pittsburgh    493.51
Omaha         235.63
Stockton      247.18
Austin        379.6
New York      296.8
Corpus Christi 25.38
Fort Worth    213.88
Las Vegas     53.26
Newark        39.75
Austin        469.63
Greensboro    290.82
San Francisco 260.65
Lincoln       136.9
Buffalo       483.82
San Jose      215.82
Boston        418.94
Houston       309.16
Las Vegas     93.39
Virginia Beach 376.11
Riverside     252.88
Tulsa         205.06
Reno          88.25
```

## Activité 3

- Expliquer ce code.
- Tester ce Reducer sur le disque local, en utilisant cette instruction.

`head -50 ../data/purchases.txt |./mapper.py |sort |./reducer.py`

```
alioune@alioune-virtual-machine:~/udacity_training/code$ head -50 ../data/purchases.txt |python3 ./mapper.py |sort |python3 ./reducer.py
11259.820000000002      50
alioune@alioune-virtual-machine:~/udacity_training/code$
```

### Lancement d'un Job entier

```
2021-06-18 16:07:12,942 INFO mapreduce.Job: map 0% reduce 0%
2021-06-18 16:07:40,716 INFO mapreduce.Job: map 26% reduce 0%
2021-06-18 16:07:46,801 INFO mapreduce.Job: map 44% reduce 0%
2021-06-18 16:07:52,883 INFO mapreduce.Job: map 60% reduce 0%
2021-06-18 16:07:56,991 INFO mapreduce.Job: map 77% reduce 0%
2021-06-18 16:07:59,076 INFO mapreduce.Job: map 83% reduce 0%
2021-06-18 16:08:02,120 INFO mapreduce.Job: map 100% reduce 0%
2021-06-18 16:08:20,387 INFO mapreduce.Job: map 100% reduce 72%
2021-06-18 16:08:26,489 INFO mapreduce.Job: map 100% reduce 78%
2021-06-18 16:08:32,552 INFO mapreduce.Job: map 100% reduce 84%
2021-06-18 16:08:38,625 INFO mapreduce.Job: map 100% reduce 89%
2021-06-18 16:08:44,729 INFO mapreduce.Job: map 100% reduce 95%
2021-06-18 16:08:50,785 INFO mapreduce.Job: map 100% reduce 100%
2021-06-18 16:08:51,815 INFO mapreduce.Job: Job job_1624057360134_0001 completed successfully
2021-06-18 16:08:52,081 INFO mapreduce.Job: Counters: 55
      File System Counters
        FILE: Number of bytes read=75658848
        FILE: Number of bytes written=152031983
        FILE: Number of read operations=0
        FILE: Number of large read operations=0
        FILE: Number of write operations=0
        HDFS: Number of bytes read=211317232
        HDFS: Number of bytes written=115533092
        HDFS: Number of read operations=11
        HDFS: Number of large read operations=0
        HDFS: Number of write operations=2
        HDFS: Number of bytes read erasure-coded=0
```

```

Reduce input records=4138476
Reduce output records=4138476
Spilled Records=8276952
Shuffled Maps =2
Failed Shuffles=0
Merged Map outputs=2
GC time elapsed (ms)=1174
CPU time spent (ms)=103940
Physical memory (bytes) snapshot=917114880
Virtual memory (bytes) snapshot=7761334272
Total committed heap usage (bytes)=757596160
Peak Map Physical memory (bytes)=294944768
Peak Map Virtual memory (bytes)=2609242112
Peak Reduce Physical memory (bytes)=351744000
Peak Reduce Virtual memory (bytes)=2620739584
Shuffle Errors
BAD_ID=0
CONNECTION=0
IO_ERROR=0
WRONG_LENGTH=0
WRONG_MAP=0
WRONG_REDUCE=0
File Input Format Counters
  Bytes Read=211317020
File Output Format Counters
  Bytes Written=115533092
2021-06-18 16:08:52,084 INFO streaming.StreamJob: Output directory: joboutput
diago@ubuntu:~$

```

## Activité 5

- Donner la liste des ventes par catégorie de produits.

Mapper

```

4  import sys
5  for line in sys.stdin:
6      data = line.strip().split("\t")
7      if len(data) == 6:
8          date, time, store, item, cost, payment = data
9          print ("{}{}\t{}".format(item, cost))
10
11

```

Reducer

```

import sys
current_item = None
current_cost = 0
# l'entrée vient de STDIN
for line in sys.stdin:
    line = line.strip()
    item, cost = line.split('\t')
    try:

```



```

        cost = float(cost)
    except ValueError:
        continue
    if current_item == item:
        current_cost = cost
    else:
        if current_item != None:
            print("{0}\t{1}".format(current_item, current_cost))
        current_item = item
        current_cost = cost
if current_item == item:
    print("{0}\t{1}".format(current_item, current_cost))

```

## Exécution du job hadoop sur le fichier purchases.txt

```

diago@ubuntu:~/udacity_training/code$ mapred streaming -mapper mapper.py -reducer reducer.py -file mapper.py -file reducer.py -input myInput -
output joboutput3.1
2021-06-24 18:27:21,395 WARN streaming.StreamJob: -file option is deprecated, please use generic option -files instead.
packageJobJar: [mapper.py, reducer.py] [/home/diago/mesLogiciels/hadoop-3.2.2/share/hadoop/tools/lib/hadoop-streaming-3.2.2.jar] /tmp/streamjo
b2701914233649132343.jar tmpDir=null
2021-06-24 18:27:26,118 INFO client.RMProxy: Connecting to ResourceManager at /0.0.0.0:8032
2021-06-24 18:27:27,568 INFO client.RMProxy: Connecting to ResourceManager at /0.0.0.0:8032
2021-06-24 18:27:29,198 INFO mapreduce.JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/diago/.staging/job_162
4520759334_0002
2021-06-24 18:27:31,323 INFO mapred.FileInputFormat: Total input files to process : 1
2021-06-24 18:27:31,363 INFO net.NetworkTopology: Adding a new node: /default-rack/127.0.0.1:9866
2021-06-24 18:27:31,513 INFO mapreduce.JobSubmitter: number of splits:2
2021-06-24 18:27:33,828 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1624520759334_0002
2021-06-24 18:27:33,833 INFO mapreduce.JobSubmitter: Executing with tokens: []
2021-06-24 18:27:34,607 INFO conf.Configuration: resource-types.xml not found
2021-06-24 18:27:34,608 INFO resource.ResourceUtils: Unable to find 'resource-types.xml'.
2021-06-24 18:27:35,382 INFO impl.YarnClientImpl: Submitted application application_1624520759334_0002
2021-06-24 18:27:36,576 INFO mapreduce.Job: The url to track the job: http://ubuntu:8088/proxy/application_1624520759334_0002/
2021-06-24 18:27:36,596 INFO mapreduce.Job: Running job: job_1624520759334_0002
2021-06-24 18:28:10,927 INFO mapreduce.Job: Job job_1624520759334_0002 running in uber mode : false
2021-06-24 18:28:10,935 INFO mapreduce.Job: map 0% reduce 0%
2021-06-24 18:28:52,000 INFO mapreduce.Job: map 2% reduce 0%
2021-06-24 18:28:58,237 INFO mapreduce.Job: map 14% reduce 0%
2021-06-24 18:29:04,364 INFO mapreduce.Job: map 28% reduce 0%
2021-06-24 18:29:10,471 INFO mapreduce.Job: map 37% reduce 0%
2021-06-24 18:29:11,491 INFO mapreduce.Job: map 46% reduce 0%
2021-06-24 18:29:16,586 INFO mapreduce.Job: map 56% reduce 0%
2021-06-24 18:29:17,598 INFO mapreduce.Job: map 65% reduce 0%
2021-06-24 18:29:21,726 INFO mapreduce.Job: map 83% reduce 0%
2021-06-24 18:29:22,771 INFO mapreduce.Job: map 100% reduce 0%
2021-06-24 18:29:44,240 INFO mapreduce.Job: map 100% reduce 86%
2021-06-24 18:29:49,333 INFO mapreduce.Job: map 100% reduce 100%
2021-06-24 18:29:50,401 INFO mapreduce.Job: Job job_1624520759334_0002 completed successfully
2021-06-24 18:29:50,666 INFO mapreduce.Job: Counters: 55
File System Counters

```

## Résultat



```
diago@ubuntu:~/udacity_training/code$ hdfs dfs -cat joboutput3.1/*
Baby      57491808.439999565
Books     57450757.910000086
CDs       57410753.04000101
Cameras   57299046.64000095
Children's Clothing  57624820.93999975
Computers 57315406.31999985
Consumer Electronics 57452374.130001
Crafts    57418154.50000002
DVDs      57649212.13999978
Garden    57539833.109999985
Health and Beauty 57481589.56000032
Men's Clothing 57621279.04
Music     57495489.70000011
Pet Supplies 57197250.240000114
Sporting Goods 57599085.88999987
Toys      57463477.10999886
Video Games 57513165.579998754
Women's Clothing 57434448.969999254
diago@ubuntu:~/udacity_training/code$
```

**Quelle est la valeur des ventes pour la catégorie Toys?**

**Mapper**

```
import sys
for line in sys.stdin:
    data = line.strip().split("\t")
    if len(data) == 6:
        date, time, store, item, cost, payment = data
        print ("{}{}\t{}\t{}".format(item, cost))
```

**Reducer**

```
import sys
current_cost = 0
for line in sys.stdin:
    line = line.strip()
    item, cost = line.split('\t')
    try:
        cost = float(cost)
    except ValueError:
        continue
    if item == 'Toys':
        current_cost += cost
print("{}{}\t{}\t{}".format('Toys', current_cost))
```

## Exécution du job hadoop sur le fichier purchases.txt

```
diago@ubuntu:~/udacity_training/code$ mapred streaming -mapper mapper.py -reducer reducer.py -file mapper.py -file reducer.py -input myinput -
output joboutput3.3
2021-06-24 18:49:50,030 WARN streaming.StreamJob: -file option is deprecated, please use generic option -files instead.
packageJobJar: [mapper.py, reducer.py] [/home/diago/mesLogiciels/hadoop-3.2.2/share/hadoop/tools/lib/hadoop-streaming-3.2.2.jar] /tmp/streamjo
b4215836941498422695.jar tmpDir=null
2021-06-24 18:49:52,693 INFO client.RMProxy: Connecting to ResourceManager at /0.0.0.0:8032
2021-06-24 18:49:53,404 INFO client.RMProxy: Connecting to ResourceManager at /0.0.0.0:8032
2021-06-24 18:49:54,143 INFO mapreduce.JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/diago/.staging/job_162
4520759334_0004
2021-06-24 18:49:55,704 INFO mapred.FileInputFormat: Total input files to process : 1
2021-06-24 18:49:55,721 INFO net.NetworkTopology: Adding a new node: /default-rack/127.0.0.1:9866
2021-06-24 18:49:55,976 INFO mapreduce.JobSubmitter: number of splits:2
2021-06-24 18:49:56,375 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1624520759334_0004
2021-06-24 18:49:56,378 INFO mapreduce.JobSubmitter: Executing with tokens: []
2021-06-24 18:49:57,111 INFO conf.Configuration: resource-types.xml not found
2021-06-24 18:49:57,112 INFO resource.ResourceUtils: Unable to find 'resource-types.xml'.
2021-06-24 18:49:57,500 INFO impl.YarnClientImpl: Submitted application application_1624520759334_0004
2021-06-24 18:49:57,633 INFO mapreduce.Job: The url to track the job: http://ubuntu:8088/proxy/application_1624520759334_0004/
2021-06-24 18:49:57,638 INFO mapreduce.Job: Running job: job_1624520759334_0004
2021-06-24 18:50:13,449 INFO mapreduce.Job: Job job_1624520759334_0004 running in uber mode : false
2021-06-24 18:50:13,455 INFO mapreduce.Job: map 0% reduce 0%
2021-06-24 18:50:38,223 INFO mapreduce.Job: map 13% reduce 0%
2021-06-24 18:50:39,251 INFO mapreduce.Job: map 25% reduce 0%
2021-06-24 18:50:44,355 INFO mapreduce.Job: map 35% reduce 0%
2021-06-24 18:50:45,381 INFO mapreduce.Job: map 45% reduce 0%
2021-06-24 18:50:50,452 INFO mapreduce.Job: map 55% reduce 0%
2021-06-24 18:50:51,470 INFO mapreduce.Job: map 67% reduce 0%
2021-06-24 18:50:53,566 INFO mapreduce.Job: map 83% reduce 0%
2021-06-24 18:50:56,653 INFO mapreduce.Job: map 100% reduce 0%
2021-06-24 18:51:15,924 INFO mapreduce.Job: map 100% reduce 92%
2021-06-24 18:51:19,005 INFO mapreduce.Job: map 100% reduce 100%
2021-06-24 18:51:20,031 INFO mapreduce.Job: Job job_1624520759334_0004 completed successfully
2021-06-24 18:51:20,320 INFO mapreduce.Job: Counters: 55
File System Counters
FILE: Number of bytes read=80443816
FILE: Number of bytes written=161601928
FILE: Number of read operations=0
```

## Résultat

```
2021-06-24 18:51:20,920 INFO streaming.StreamJob: Output directory:
diago@ubuntu:~/udacity_training/code$ hdfs dfs -cat joboutput3.3/
Toys      57463477.10999886
diago@ubuntu:~/udacity_training/code$
```

## Activité 6

Donner le montant de la vente le plus élevé pour chaque magasin

### Mapper

```
import sys
for line in sys.stdin:
    data = line.strip().split("\t")
    if len(data) == 6:
        date, time, store, item, cost, payment = data
        print ("{0}\t{1}".format(item, cost))
```

### Reducer

```
import sys
import numpy as np
current_store=None
tab_current_cost= []
# 1'entrée vient de STDIN
for line in sys.stdin:
    line=line.strip()
```

```

store,cost=line.split("\t")
try:
    cost = float(cost)
except ValueError:
    continue
if current_store==store:
    tab_current_cost.append(cost)
else:
    if current_store!= None:
        print
        ("{0}\t{1}".format(current_store,np.max(tab_current_cost)))
        tab_current_cost.clear()
    current_store=store

```

## Exécution du job hadoop sur le fichier purchases.txt

```

diago@ubuntu:~/udacity_training/code$ mapred streaming -mapper mapper.py -reducer reducer.py -file mapper.py -file reducer.py -input myinp
output joboutput2
2021-06-24 22:26:41,260 WARN streaming.StreamJob: -file option is deprecated, please use generic option -files instead.
packageJobJar: [mapper.py, reducer.py] [/home/diago/mesLogiciels/hadoop-3.2.2/share/hadoop/tools/lib/hadoop-streaming-3.2.2.jar] /tmp/stre
b6842627578741276838.jar tmpDir=null
2021-06-24 22:26:43,947 INFO client.RMPProxy: Connecting to ResourceManager at /0.0.0.0:8032
2021-06-24 22:26:44,703 INFO client.RMPProxy: Connecting to ResourceManager at /0.0.0.0:8032
2021-06-24 22:26:45,337 INFO mapreduce.JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/diago/.staging/job
4572892036_0003
2021-06-24 22:26:46,256 INFO mapred.FileInputFormat: Total input files to process : 1
2021-06-24 22:26:46,279 INFO net.NetworkTopology: Adding a new node: /default-rack/127.0.0.1:9866
2021-06-24 22:26:46,459 INFO mapreduce.JobSubmitter: number of splits:2
2021-06-24 22:26:46,868 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1624572892036_0003
2021-06-24 22:26:46,871 INFO mapreduce.JobSubmitter: Executing with tokens: []
2021-06-24 22:26:47,496 INFO conf.Configuration: resource-types.xml not found
2021-06-24 22:26:47,498 INFO resource.ResourceUtils: Unable to find 'resource-types.xml'.
2021-06-24 22:26:47,700 INFO impl.YarnClientImpl: Submitted application application_1624572892036_0003
2021-06-24 22:26:47,793 INFO mapreduce.Job: The url to track the job: http://ubuntu:8088/proxy/application_1624572892036_0003/
2021-06-24 22:26:47,796 INFO mapreduce.Job: Running job: job_1624572892036_0003
2021-06-24 22:27:03,417 INFO mapreduce.Job: Job job_1624572892036_0003 running in uber mode : false
2021-06-24 22:27:03,420 INFO mapreduce.Job: map 0% reduce 0%
2021-06-24 22:27:31,390 INFO mapreduce.Job: map 22% reduce 0%
2021-06-24 22:27:37,532 INFO mapreduce.Job: map 39% reduce 0%
2021-06-24 22:27:43,667 INFO mapreduce.Job: map 55% reduce 0%
2021-06-24 22:27:49,768 INFO mapreduce.Job: map 64% reduce 0%
2021-06-24 22:27:50,908 INFO mapreduce.Job: map 81% reduce 0%
2021-06-24 22:27:55,989 INFO mapreduce.Job: map 100% reduce 0%
2021-06-24 22:28:14,356 INFO mapreduce.Job: map 100% reduce 86%
2021-06-24 22:28:19,429 INFO mapreduce.Job: map 100% reduce 100%
2021-06-24 22:28:20,513 INFO mapreduce.Job: Job job_1624572892036_0003 completed successfully
2021-06-24 22:28:20,774 INFO mapreduce.Job: Counters: 55
    File System Counters
        FILE: Number of bytes read=80443816
        FILE: Number of bytes written=161601922
        FILE: Number of read operations=0
        FILE: Number of large read operations=0
        FS: Number of files created=1

```

## Résultat

```
diago@ubuntu:~/udacity_training/code$ hdfs dfs -cat joboutput2/*
Baby      499.99
Books     499.99
CDs       499.99
Cameras   499.98
Children's Clothing    499.99
Computers      499.99
Consumer Electronics  499.99
Crafts        499.99
DVDs          499.99
Garden        499.99
Health and Beauty     499.99
Men's Clothing  499.99
Music         499.99
Pet Supplies      499.99
Sporting Goods  499.99
Toys          499.99
Video Games      499.99
diago@ubuntu:~/udacity_training/code$
```

### Quelle est cette valeur pour les magasins suivants ?

Reno

Toledo

Chandler

### Mapper

```
import sys
for line in sys.stdin:
    data = line.strip().split("\t")
    if len(data) == 6:
        date, time, store, item, cost, payment = data
        print ("{}{}\t{}\t{}".format(store, cost))
```

### Reducer

```
import sys
import numpy as np
current_store = None
current_cost = 0
tab_current_cost = []
tab_selected = ['Reno', 'Toledo', 'Chandler']
# l'entrée vient de STDIN
for line in sys.stdin:
    line = line.strip()
    store, cost = line.split('\t')
    try:
        cost = float(cost)
    except ValueError:
```

```
        continue
    if current_store in tab_selected:

print("{0}\t{1}".format(current_store.np.max(tab_current_cost)))
    tab_current_cost.clear()
    current_store = store
    current_cost = cost
```