TP1 : Apache HBase

Rapport

Réalisé par :

Abderrahim   ALAKOUCHE
Abdellah   AGHLALOU

Encadré par :

Mme. Dounia ZAIDOUNI

PLAN

HBase est un système de gestion de base de données non-relationnelles distribué, écrit en Java, disposant d'un stockage structuré pour les grandes tables.[1]

Les objectifs de ce TP sont les suivants :

- Installation et configuration d'Apache HBase en mode « Standalone » et en mode « Pseudo distribué».
- Manipulation de « HBase » (opérations create, put , get, list, scan, disable, drop, exists)
- L'utilisation de l'API Java de Hbase.
- Traitement de données avec Spark.

[1]:https://www.formation-bigdata.com/hbase-pas-a-pas/

Etape1:Installation et configuration d'Apache Hbase :

agh@agh-VirtualBox: ~

```
agh@agh-VirtualBox:~/Downloads$ ls
hadoop-2.7.4.tar.gz    hbase-1.4.7-bin.tar.gz    spark-2.4.3-bin-hadoop2.7.tgz
hbase-1.4.7            jdk-8u71-linux-x64.tar.gz  tp_hbase.pdf
agh@agh-VirtualBox:~/Downloads$ mv hbase-1.4.7 hbase
agh@agh-VirtualBox:~/Downloads$ ls
hadoop-2.7.4.tar.gz    hbase-1.4.7-bin.tar.gz    spark-2.4.3-bin-hadoop2.7.tgz
hbase                 jdk-8u71-linux-x64.tar.gz  tp_hbase.pdf
agh@agh-VirtualBox:~/Downloads$ sudo mv hbase /usr/local/
[sudo] password for agh:
agh@agh-VirtualBox:~/Downloads$ cd
agh@agh-VirtualBox:~$ vim .bashrc
agh@agh-VirtualBox:~$ vim .bashrc
agh@agh-VirtualBox:~$ source .bashrc
agh@agh-VirtualBox:~$
```

- Le fichier de Hbase
- Renommer le fichier
- déplacement
- Ajouter le /bin dans le PATH

```
# enable programmable completion features (you don't need to enable
# this, if it's already enabled in /etc/bash.bashrc and /etc/profile
# sources /etc/bash.bashrc).
if ! shopt -oq posix; then
  if [ -f /usr/share/bash-completion/bash_completion ]; then
    . /usr/share/bash-completion/bash_completion
  elif [ -f /etc/bash_completion ]; then
    . /etc/bash_completion
  fi
fi
#hadoop variables start
export JAVA_HOME=/opt/java/jdk1.8.0_71/
export HADOOP_INSTALL=/usr/local/hadoop
export PATH=$PATH:$HADOOP_INSTALL/bin
export PATH=$PATH:$HADOOP_INSTALL/sbin
export HADOOP_MAPRED_HOME=$HADOOP_INSTALL
export HADOOP_COMMON_HOME=$HADOOP_INSTALL
export HADOOP_HDFS_HOME=$HADOOP_INSTALL
export YARN_HOME=$HADOOP_INSTALL
export HADOOP_COMMON_LIB_NATIVE_DIR=$HADOOP_INSTALL/lib/native
#export HADOOP_OPTS="-Djava.library.path=$HADOOP_INSTALL/lib"
#HADOOP VARIABLES END
export SPARK_HOME=/usr/local/spark
export PATH=$PATH:$SPARK_HOME/bin
export HBASE_HOME=/usr/local/hbase
export PATH=$PATH:$HBASE_HOME/bin
```

```
agh@agh-VirtualBox:~$ java -version
java version "1.8.0_71"
Java(TM) SE Runtime Environment (build 1.8.0_71-b15)
Java HotSpot(TM) 64-Bit Server VM (build 25.71-b15, mixed mode)
agh@agh-VirtualBox:~$ cd /usr/local/hbase/conf/
agh@agh-VirtualBox:/usr/local/hbase/conf$ vim hbase env.sh
```

```
# * See the License for the specific language governing permissions and
# * limitations under the License.
# */

# Set environment variables here.

# This script sets variables multiple times over the course of starting an hbas
e process,
# so try to keep things idempotent unless you want to take an even deeper look
# into the startup scripts (bin/hbase, etc.)

# The java implementation to use.   Java 1.7+ required.
export JAVA_HOME=/opt/java/jdk1.8.0_71/
# Extra Java CLASSPATH elements.  Optional.
# export HBASE_CLASSPATH=

# The maximum amount of heap to use. Default is left to JVM default.
# export HBASE_HEAPSIZE=1G

# Uncomment below if you intend to use off heap cache. For example, to allocate
 8G of
# offheap, set the value to "8G".
# export HBASE_OFFHEAPSIZE=1G

# Extra Java runtime options.
# Below are what we set by default.  May only work with SUN JVM.
# For more on why as well as other possible settings,
# see http://wiki.apache.org/hadoop/PerformanceTuning
```

```
agh@agh-VirtualBox:/usr/local/hbase/conf$ cd
agh@agh-VirtualBox:~$ sudo vim /etc/hosts
[sudo] password for agh:
agh@agh-VirtualBox:~$ sudo reboot
```

```
127.0.0.1       localhost
127.0.0.1       agh-VirtualBox

# The following lines are desirable for IPv6 capable hosts
::1     ip6-localhost ip6-loopback
fe00::0 ip6-localnet
ff00::0 ip6-mcastprefix
ff02::1 ip6-allnodes
ff02::2 ip6-allrouters
~
~
~
~
~
~
~
~
~
~
~
~
~
~
~
~
```

Etape2:Démarrage du cluster Hadoop configuré dans notre machine .

```
agh@agh-VirtualBox:/usr/local/hadoop_store$ rm -rf *
agh@agh-VirtualBox:/usr/local/hadoop_store$ ls
agh@agh-VirtualBox:/usr/local/hadoop_store$ ls
agh@agh-VirtualBox:/usr/local/hadoop_store$ mkdir -p /usr/local/hadoop_store/hd
fs/namenode
agh@agh-VirtualBox:/usr/local/hadoop_store$ mkdir -p /usr/local/hadoop_store/hd
fs/datanode
agh@agh-VirtualBox:/usr/local/hadoop_store$ chown -R agh  /usr/local/hadoop_sto
re/hdfs/namenode
agh@agh-VirtualBox:/usr/local/hadoop_store$ chown -R agh  /usr/local/hadoop_sto
re/hdfs/datanode
agh@agh-VirtualBox:/usr/local/hadoop_store$
```

Formatage de namenode :

Lancement de namenode,datanode,secondry namenode , yarn


```
agh@agh-VirtualBox:/usr/local/hadoop/etc/hadoop$ start-all.sh
This script is Deprecated. Instead use start-dfs.sh and start-yarn.sh
21/02/14 10:40:34 WARN util.NativeCodeLoader: Unable to load native-hadoop libr
ary for your platform... using builtin-java classes where applicable
Starting namenodes on [localhost]
localhost: starting namenode, logging to /usr/local/hadoop/logs/hadoop-agh-name
node-agh-VirtualBox.out
localhost: starting datanode, logging to /usr/local/hadoop/logs/hadoop-agh-data
node-agh-VirtualBox.out
Starting secondary namenodes [0.0.0.0]
0.0.0.0: starting secondarynamenode, logging to /usr/local/hadoop/logs/hadoop-a
gh-secondarynamenode-agh-VirtualBox.out
21/02/14 10:41:10 WARN util.NativeCodeLoader: Unable to load native-hadoop libr
ary for your platform... using builtin-java classes where applicable
starting yarn daemons
starting resourcemanager, logging to /usr/local/hadoop/logs/yarn-agh-resourcema
nager-agh-VirtualBox.out
localhost: starting nodemanager, logging to /usr/local/hadoop/logs/yarn-agh-nod
emanager-agh-VirtualBox.out
agh@agh-VirtualBox:/usr/local/hadoop/etc/hadoop$
```

Vérification

```
------------------------------------------------
Live datanodes (1):

Name: 127.0.0.1:50010 (localhost)
Hostname: localhost
Decommission Status : Normal
Configured Capacity: 31147880448 (29.01 GB)
DFS Used: 24576 (24 KB)
Non DFS Used: 11531415552 (10.74 GB)
DFS Remaining: 18010599424 (16.77 GB)
DFS Used%: 0.00%
DFS Remaining%: 57.82%
Configured Cache Capacity: 0 (0 B)
Cache Used: 0 (0 B)
Cache Remaining: 0 (0 B)
Cache Used%: 100.00%
Cache Remaining%: 0.00%
Xceivers: 1
Last contact: Sun Feb 14 10:42:06 WET 2021


agh@agh-VirtualBox:/usr/local/hadoop/etc/hadoop$
```

configuration d'Apache Hbase pour un mode « Standalone » :

```
agh@agh-VirtualBox:/usr/local/hadoop/etc/hadoop$ cd /usr/local/hbase/conf/
agh@agh-VirtualBox:/usr/local/hbase/conf$ vim hbase site.xml
agh@agh-VirtualBox:/usr/local/hbase/conf$ ls
hadoop-metrics2-hbase.properties   hbase-policy.xml   regionservers
hbase-env.cmd                      hbase-site.xml
hbase-env.sh                       log4j.properties
agh@agh-VirtualBox:/usr/local/hbase/conf$ vim hbase-site.xml
```

```xml
 * regarding copyright ownership.  The ASF licenses this file
 * to you under the Apache License, Version 2.0 (the
 * "License"); you may not use this file except in compliance
 * with the License.  You may obtain a copy of the License at
 *
 *     http://www.apache.org/licenses/LICENSE-2.0
 *
 * Unless required by applicable law or agreed to in writing, software
 * distributed under the License is distributed on an "AS IS" BASIS,
 * WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied.
 * See the License for the specific language governing permissions and
 * limitations under the License.
 */
-->
<configuration>
        <property>
                <name>hbase.rootdir</name>
                <value>file:///home/agh/hbase</value>
        </property>
        <property>
                <name>hbase.zookeeper.property.dataDir</name>
                <value>/home/agh/zookeeper</value>
        </property>
        <property>
                <name>hbase.unsafe.stream.capability.enforce</name>
                <value>false</value>
        </property>
</configuration>
```

Lancement de Hbase

```
agh@agh-VirtualBox:/usr/local/hbase$ jps
4469 Jps
4181 HMaster
2567 SecondaryNameNode
2379 DataNode
2254 NameNode
2719 ResourceManager
agh@agh-VirtualBox:/usr/local/hbase$
```

```
HBase Shell
Use "help" to get list of supported commands.
Use "exit" to quit this interactive shell.
Version 1.4.7, r763f27f583cf8fd7ecf79fb6f3ef57f1615dbf9b, Tue Aug 28 14:40:11 P
DT 2018

hbase(main):001:0> status
1 active master, 0 backup masters, 1 servers, 0 dead, 2.0000 average load

hbase(main):002:0>
```

- Création des fichiers nécessaires pour Hbase dans le HDFS :

```
agh@agh-VirtualBox:/usr/local/hbase$ hadoop fs -mkdir -p /hbase
21/02/14 11:03:09 WARN util.NativeCodeLoader: Unable to load native-hadoop libr
ary for your platform... using builtin-java classes where applicable
agh@agh-VirtualBox:/usr/local/hbase$ hadoop fs -mkdir -p /agh/zookeeper
21/02/14 11:04:43 WARN util.NativeCodeLoader: Unable to load native-hadoop libr
ary for your platform... using builtin-java classes where applicable
agh@agh-VirtualBox:/usr/local/hbase$ hadoop fs -ls /
21/02/14 11:05:20 WARN util.NativeCodeLoader: Unable to load native-hadoop libr
ary for your platform... using builtin-java classes where applicable
Found 2 items
drwxr-xr-x   - agh supergroup          0 2021-02-14 11:04 /agh
drwxr-xr-x   - agh supergroup          0 2021-02-14 11:03 /hbase
agh@agh-VirtualBox:/usr/local/hbase$ hadoop fs -ls /agh
21/02/14 11:05:56 WARN util.NativeCodeLoader: Unable to load native-hadoop libr
ary for your platform... using builtin-java classes where applicable
Found 1 items
drwxr-xr-x   - agh supergroup          0 2021-02-14 11:04 /agh/zookeeper
agh@agh-VirtualBox:/usr/local/hbase$
```

```
 * regarding copyright ownership.  The ASF licenses this file
 * to you under the Apache License, Version 2.0 (the
 * "License"); you may not use this file except in compliance
 * with the License.  You may obtain a copy of the License at
 *
 *      http://www.apache.org/licenses/LICENSE-2.0
 *
 * Unless required by applicable law or agreed to in writing, software
 * distributed under the License is distributed on an "AS IS" BASIS,
 * WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied.
 * See the License for the specific language governing permissions and
 * limitations under the License.
 */
-->
<configuration>
        <property>
                <name>hbase.rootdir</name>
                <value>hdfs://localhost:54310/hbase</value>
        </property>
        <property>
                <name>hbase.zookeeper.property.dataDir</name>
                <value>hdfs://localhost:54310/agh/zookeeper</value>
        </property>
        <property>
                <name>hbase.cluster.distributed</name>
                <value>true</value>
        </property>
</configuration>
```

Lancement de Hbase :

```
agh@agh-VirtualBox:/usr/local/hbase$ ./bin/start-hbase.sh
localhost: running zookeeper, logging to /usr/local/hbase/bin/../logs/hbase-agh
-zookeeper-agh-VirtualBox.out
running master, logging to /usr/local/hbase/logs/hbase-agh-master-agh-VirtualBo
x.out
: running regionserver, logging to /usr/local/hbase/logs/hbase-agh-regionserver
-agh-VirtualBox.out
agh@agh-VirtualBox:/usr/local/hbase$ jps
5330 Jps
2567 SecondaryNameNode
5096 HQuorumPeer
2379 DataNode
2254 NameNode
5150 HMaster
2719 ResourceManager
5215 HRegionServer
agh@agh-VirtualBox:/usr/local/hbase$
```

-Requête de création d'une base de donne avec Hbase shell

-la commande list pour lister les tables existants

```
hbase(main):001:0> create 'registre_ventes','client','ventes'
0 row(s) in 4.2580 seconds

=> Hbase::Table - registre_ventes
hbase(main):002:0> list
TABLE
registre_ventes
1 row(s) in 0.0920 seconds

=> ["registre_ventes"]
hbase(main):003:0>
```

Ajout d'un nouveau cell :

Put   'nom_table', ' num_of_cell',   'name_columnfamily:name_column ', 'valeur'

```
agh@agh-VirtualBox: /usr/local/hbase/bin

hbase(main):019:0> scan 'registre_ventes'
ROW                  COLUMN+CELL
 101                 column=client:nom, timestamp=1613303968964, value=Mohamed
                     A
 101                 column=client:ville, timestamp=1613303969162, value=Casabl
                     anca
 101                 column=ventes:montant, timestamp=1613303969530, value=2000
                     ,00 MAD
 101                 column=ventes:produit, timestamp=1613303969357, value=Chai
                     ses
 102                 column=client:nom, timestamp=1613303969673, value=Amine B
 102                 column=client:ville, timestamp=1613303969869, value=Sal\xC
                     3\xA9
 102                 column=ventes:montant, timestamp=1613303970142, value=300,
                     00 MAD
 102                 column=ventes:produit, timestamp=1613303970001, value=Lamp
                     es
 103                 column=client:nom, timestamp=1613303970289, value=Younes C
 103                 column=client:ville, timestamp=1613303970380, value=Rabat
 103                 column=ventes:montant, timestamp=1613303970687, value=1000
                     0,00 MAD
 103                 column=ventes:produit, timestamp=1613303970510, value=Bure
                     aux
 104                 column=client:nom, timestamp=1613303970774, value=Othman D
 104                 column=client:ville, timestamp=1613303970863, value=Tanger
 104                 column=ventes:montant, timestamp=1613303976845, value=1500
                     0,00 MAD
 104                 column=ventes:produit, timestamp=1613303971052, value=Lits
```

Visualisation de résultat avec la commande scan

Getting a value :
Get 'name_table', 'num_cell', {COLUMN=>'name_columnfamily:name_column'}

```
hbase(main):020:0> get 'registre_ventes','102',{COLUMN => 'ventes:produit'}
COLUMN                   CELL
 ventes:produit          timestamp=1613303970001, value=Lampes
1 row(s) in 0.1920 seconds

hbase(main):021:0> disable 'registre_ventes'
0 row(s) in 2.5210 seconds

hbase(main):022:0> drop 'registre_ventes'
0 row(s) in 2.4480 seconds

hbase(main):023:0> exists 'registre_ventes'
Table registre_ventes does not exist
0 row(s) in 0.0550 seconds
```

drop pour supprimer

Exists pour voir vérifier l'éxistance de la table

```
agh@agh-VirtualBox:/usr/local/hbase/bin$ cd
agh@agh-VirtualBox:~$ vim .bashrc
agh@agh-VirtualBox:~$ source .bashrc
agh@agh-VirtualBox:~$ cd /home/agh/Downloads
agh@agh-VirtualBox:~/Downloads$ mkdir hbase-code
agh@agh-VirtualBox:~/Downloads$ cd hbase-code/
agh@agh-VirtualBox:~/Downloads/hbase-code$ vim HelloHbase.java
agh@agh-VirtualBox:~/Downloads/hbase-code$
```

```
export CLASSPATH=/usr/local/hbase/lib/*:/usr/local/hadoop/share/hadoop/common/*

-- INSERT --                                                    134,80      Bot
```

API

agh@agh-VirtualBox: ~/Downloads/hbase-code

```java
import org.apache.hadoop.conf.Configuration;
import org.apache.hadoop.hbase.HBaseConfiguration;
import org.apache.hadoop.hbase.HColumnDescriptor;
import org.apache.hadoop.hbase.HTableDescriptor;
import org.apache.hadoop.hbase.TableName;
import org.apache.hadoop.hbase.client.*;
import org.apache.hadoop.hbase.util.Bytes;
import java.io.IOException;
public class HelloHBase {
  private Table table1;
  private String tableName = "user";
  private String family1 = "PersonalData";
  private String family2 = "ProfessionalData";
  public void createHbaseTable() throws IOException {
    Configuration config = HBaseConfiguration.create();
    Connection connection = ConnectionFactory.createConnection(config);
    Admin admin = connection.getAdmin();
    HTableDescriptor ht = new HTableDescriptor(TableName.valueOf(tableName));
  ht.addFamily(new HColumnDescriptor(family1));
  ht.addFamily(new HColumnDescriptor(family2));
  System.out.println("connecting");
  System.out.println("Creating Table");
  createOrOverwrite(admin, ht);
  System.out.println("Done......");
  table1 = connection.getTable(TableName.valueOf(tableName));
  try {
  System.out.println("Adding user: user1");
  byte[] row1 = Bytes.toBytes("user1");
-- INSERT --                                                    1,3      Top
```

```
agh@agh-VirtualBox:~/Downloads/hbase-code$ javac HelloHBase.java
agh@agh-VirtualBox:~/Downloads/hbase-code$ ls
HelloHBase.class   HelloHBase.java
agh@agh-VirtualBox:~/Downloads/hbase-code$
```

La table est bien créé , et la lecriture est parfaitement fait , grace aux
api en peux manipuler , lire , ecrire les donnes de hbase par un code
java aulieu de hbase shell

```
agh@agh-VirtualBox:~/Downloads/hbase-code$ java -cp .:/usr/local/hbase/lib/* He
lloHBase
log4j:WARN No appenders could be found for logger (org.apache.hadoop.util.Shell
).
log4j:WARN Please initialize the log4j system properly.
log4j:WARN See http://logging.apache.org/log4j/1.2/faq.html#noconfig for more i
nfo.
connecting
Creating Table
Done......
Adding user: user1
Adding user: user2
reading data...
mohamed
agh@agh-VirtualBox:~/Downloads/hbase-code$
```

Ecriture de fichier dans HDFS

```
agh@agh-VirtualBox:~/Downloads/hbase-code$ cd
agh@agh-VirtualBox:~$ hadoop fs -mkdir -p /input
21/02/14 13:31:47 WARN util.NativeCodeLoader: Unable to load native-hadoop libr
ary for your platform... using builtin-java classes where applicable
```

```
agh@agh-VirtualBox:~$ hadoop fs -put /home/agh/Downloads/purchases2.txt /input
21/02/14 13:33:54 WARN util.NativeCodeLoader: Unable to load native-hadoop libr
ary for your platform... using builtin-java classes where applicable
agh@agh-VirtualBox:~$ hadoop fs -ls /input
21/02/14 13:34:51 WARN util.NativeCodeLoader: Unable to load native-hadoop libr
ary for your platform... using builtin-java classes where applicable
Found 1 items
-rw-r--r--   1 agh supergroup   243309628 2021-02-14 13:34 /input/purchases2.txt
agh@agh-VirtualBox:~$
```

Création de la base de donnes products

```
hbase(main):001:0> create 'products', 'cf'
0 row(s) in 9.6590 seconds

=> Hbase::Table - products
hbase(main):002:0> exit
agh@agh-VirtualBox:/usr/local/hbase$
```

L'éxécution de la commande Importtssv


```
completed successfully
2021-02-15 12:32:50,450 INFO  [main] mapreduce.Job: Counters: 21
        File System Counters
                FILE: Number of bytes read=60102777
                FILE: Number of bytes written=61275114
                FILE: Number of read operations=0
                FILE: Number of large read operations=0
                FILE: Number of write operations=0
                HDFS: Number of bytes read=377535548
                HDFS: Number of bytes written=0
                HDFS: Number of read operations=7
                HDFS: Number of large read operations=0
                HDFS: Number of write operations=0
        Map-Reduce Framework
                Map input records=4138476
                Map output records=4138476
                Input split bytes=216
                Spilled Records=0
                Failed Shuffles=0
                Merged Map outputs=0
                GC time elapsed (ms)=112693
                Total committed heap usage (bytes)=170041344
        ImportTsv
                Bad Lines=0
        File Input Format Counters
                Bytes Read=243313724
        File Output Format Counters
                Bytes Written=0
agh@agh-VirtualBox:/usr/local/hbase$
```

L'utilisation de la commande scan avec Limite pour voir une partie de donnes

Lire une ligne

```
hbase(main):004:0> get 'products','2000',{COLUMN => 'cf:town'}
COLUMN                        CELL
 cf:town                      timestamp=1613311095452, value=Oklahoma City
1 row(s) in 0.1810 seconds

hbase(main):005:0>
```

Traitement de données avec Spark :

```
agh@agh-VirtualBox:/usr/local/hbase$ cd ~/Downloads
agh@agh-VirtualBox:~/Downloads$ ls
apache-maven-3.5.0-bin.tar.gz    jdk-8u71-linux-x64.tar.gz
hadoop-2.7.4.tar.gz              purchases2.txt
hbase-1.4.7-bin.tar.gz           spark-2.4.3-bin-hadoop2.7.tgz
hbase-code                       tp_hbase.pdf
agh@agh-VirtualBox:~/Downloads$ tar -zxvf apache-maven-3.5.0-bin.tar.gz
apache-maven-3.5.0/README.txt
apache-maven-3.5.0/LICENSE
apache-maven-3.5.0/NOTICE
apache-maven-3.5.0/lib/
```

```
agh@agh-VirtualBox:~/Downloads$ sudo mv /home/agh/Downloads/apache-maven-3.5.0
/opt/
[sudo] password for agh:
agh@agh-VirtualBox:~/Downloads$ ls
apache-maven-3.5.0-bin.tar.gz    jdk-8u71-linux-x64.tar.gz
hadoop-2.7.4.tar.gz              purchases2.txt
hbase-1.4.7-bin.tar.gz           spark-2.4.3-bin-hadoop2.7.tgz
hbase-code                       tp_hbase.pdf
agh@agh-VirtualBox:~/Downloads$
```

```
agh@agh-VirtualBox:~/Downloads$ sudo vim /etc/profile
agh@agh-VirtualBox:~/Downloads$ source /etc/profile
agh@agh-VirtualBox:~/Downloads$ mvn -v
Apache Maven 3.5.0 (ff8f5e7444045639af65f6095c62210b5713f426; 2017-04-03T20:39:
06+01:00)
Maven home: /opt/apache-maven-3.5.0
Java version: 1.8.0_71, vendor: Oracle Corporation
Java home: /opt/java/jdk1.8.0_71/jre
Default locale: en_US, platform encoding: UTF-8
OS name: "linux", version: "5.8.0-43-generic", arch: "amd64", family: "unix"
agh@agh-VirtualBox:~/Downloads$
```

Installation de maven

Création de projet maven Myapp

```
----
[INFO] Parameter: basedir, Value: /home/agh
[INFO] Parameter: package, Value: TP.Myapp
[INFO] Parameter: groupId, Value: TP.Myapp
[INFO] Parameter: artifactId, Value: Myapp
[INFO] Parameter: packageName, Value: TP.Myapp
[INFO] Parameter: version, Value: 1.0-SNAPSHOT
[INFO] project created from Old (1.x) Archetype in dir: /home/agh/Myapp
[INFO] ------------------------------------------------------------------------
[INFO] BUILD SUCCESS
[INFO] ------------------------------------------------------------------------
[INFO] Total time: 01:01 min
[INFO] Finished at: 2021-02-15T12:43:00Z
[INFO] Final Memory: 15M/44M
[INFO] ------------------------------------------------------------------------
agh@agh-VirtualBox:~$ 
```

```
agh@agh-VirtualBox:~/Myapp$ mvn package
[INFO] Scanning for projects...
```

```
[INFO]
[INFO] --- maven-jar-plugin:2.4:jar (default-jar) @ Myapp ---
[INFO] Building jar: /home/agh/Myapp/target/Myapp-1.0-SNAPSHOT.jar
[INFO] ------------------------------------------------------------------------
[INFO] BUILD SUCCESS
[INFO] ------------------------------------------------------------------------
[INFO] Total time: 11.876 s
[INFO] Finished at: 2021-02-15T12:44:56Z
[INFO] Final Memory: 15M/44M
[INFO] ------------------------------------------------------------------------
agh@agh-VirtualBox:~/Myapp$
```

Configuration de projet maven : ajouter les dépendences (spark,hbase) de notre projet


```xml
<project xmlns="http://maven.apache.org/POM/4.0.0" xmlns:xsi="http://www.w3.org
/2001/XMLSchema-instance"
  xsi:schemaLocation="http://maven.apache.org/POM/4.0.0 http://maven.apache.
/xsd/maven-4.0.0.xsd">
  <modelVersion>4.0.0</modelVersion>

  <groupId>TP.Myapp</groupId>
  <artifactId>Myapp</artifactId>
  <version>1.0-SNAPSHOT</version>
  <packaging>jar</packaging>

  <name>Myapp</name>
  <url>http://maven.apache.org</url>

  <properties>
        <project.build.sourceEncoding>UTF-8</project.build.sourceEncoding>

  </properties>
  <build>
<plugins>
<plugin>
<groupId>org.apache.maven.plugins</groupId>
<artifactId>maven-compiler-plugin</artifactId>
<configuration>
<source>1.8</source>
<target>1.8</target>
</configuration>
</plugin>
```

```
-- INSERT --                                        7,14     Acti
```

```xml
</plugin>
</plugins>
</build>
  <dependencies>
    <dependency>
      <groupId>junit</groupId>
      <artifactId>junit</artifactId>
      <version>3.8.1</version>
      <scope>test</scope>
    </dependency>
    <dependency>
<groupId>org.apache.hbase</groupId>
<artifactId>hbase</artifactId>
<version>2.1.3</version>
<type>pom</type>
</dependency>
<dependency>
<groupId>org.apache.hbase</groupId>
<artifactId>hbase-spark</artifactId>
<version>2.0.0-alpha4</version>
</dependency>
<dependency>
<groupId>org.apache.spark</groupId>
<artifactId>spark-core_2.11</artifactId>
<version>2.2.1</version>
</dependency>
    </dependencies>
</project>
```

```
-- INSERT --
```

```
agh@agh-VirtualBox:~/Myapp/src/main/java/TP/Myapp$ ls
App.java
agh@agh-VirtualBox:~/Myapp/src/main/java/TP/Myapp$ mv App.java
mv: missing destination file operand after 'App.java'
Try 'mv --help' for more information.
agh@agh-VirtualBox:~/Myapp/src/main/java/TP/Myapp$ mv App.java HbaseSparkProces
s.java
agh@agh-VirtualBox:~/Myapp/src/main/java/TP/Myapp$ ls
HbaseSparkProcess.java
agh@agh-VirtualBox:~/Myapp/src/main/java/TP/Myapp$
```

```
[INFO]
[INFO] --- maven-jar-plugin:2.4:jar (default-jar) @ Myapp ---
[INFO] Building jar: /home/agh/Myapp/target/Myapp-1.0-SNAPSHOT.jar
[INFO] ------------------------------------------------------------------------
[INFO] BUILD SUCCESS
[INFO] ------------------------------------------------------------------------
[INFO] Total time: 12.865 s
[INFO] Finished at: 2021-02-15T13:00:15Z
[INFO] Final Memory: 44M/105M
[INFO] ------------------------------------------------------------------------
agh@agh-VirtualBox:~/Myapp$
```

```
[INFO]
[INFO] --- maven-jar-plugin:2.4:jar (default-jar) @ Myapp ---
[INFO] ------------------------------------------------------------------------
[INFO] BUILD SUCCESS
[INFO] ------------------------------------------------------------------------
[INFO] Total time: 13.401 s
[INFO] Finished at: 2021-02-15T13:01:34Z
[INFO] Final Memory: 44M/106M
[INFO] ------------------------------------------------------------------------
agh@agh-VirtualBox:~/Myapp$
```

Le résultat de compter le nombre d'élément

```
nombre d'enregistrements: 4138476 ✓
21/02/15 13:07:28 INFO SparkContext: Invoking stop() from shutdown hook
21/02/15 13:07:29 INFO SparkUI: Stopped Spark web UI at http://10.0.2.15:4040
21/02/15 13:07:29 INFO MapOutputTrackerMasterEndpoint: MapOutputTrackerMasterEn
dpoint stopped!
21/02/15 13:07:29 INFO MemoryStore: MemoryStore cleared
21/02/15 13:07:29 INFO BlockManager: BlockManager stopped
21/02/15 13:07:29 INFO BlockManagerMaster: BlockManagerMaster stopped
21/02/15 13:07:29 INFO OutputCommitCoordinator$OutputCommitCoordinatorEndpoint:
 OutputCommitCoordinator stopped!
21/02/15 13:07:29 INFO SparkContext: Successfully stopped SparkContext
21/02/15 13:07:29 INFO ShutdownHookManager: Shutdown hook called
21/02/15 13:07:29 INFO ShutdownHookManager: Deleting directory /tmp/spark-46c3d
bb2-8739-43cd-8772-3c015169e138
21/02/15 13:07:29 INFO ShutdownHookManager: Deleting directory /tmp/spark-9d16f
e00-e410-45ae-917f-108e64d6225a
agh@agh-VirtualBox:~/Myapp/target$                          Activer V
```

- **Merci**