**BiG Data**

**TP4 - Stockage de Données dans une Base NOSQL avec HBase**

Annassiri Fatima Zahra \_ MBisd2

**Introduction :**

**HBase**

En utilise Apache HBase ™ lors du besoin d'un accès aléatoire en lecture / écriture en temps réel à votre Big Data.

L'objectif de ce projet est d'héberger de très grandes tables - des milliards de lignes X des millions de colonnes - au sommet de grappes de matériel de base.

Apache HBase est une base de données open source, distribuée, versionnée et non relationnelle, inspirée de Bigtable de Google: un système de stockage distribué pour les données structurées par Chang et al. Tout comme Bigtable exploite le stockage de données distribué fourni par le système de fichiers Google, Apache HBase fournit des capacités de type Bigtable en plus de Hadoop et HDFS.



**Objectif du tp**

Manipulation de données avec HBase, et traitement co-localisé avec Spark.

## Outils et Versions

* [Apache HBase](https://hbase.apache.org/) Version 1.4.7
* [Apache Hadoop](http://hadoop.apache.org/) Version: 2.7.2
* [Apache Spark](https://spark.apache.org/) Version: 2.2.1
* [Docker](https://www.docker.com/) Version 17.09.1
* [IntelliJ IDEA](https://www.jetbrains.com/idea/download/) Version Ultimate 2016.1 (ou tout autre IDE de votre choix)
* [Java](http://www.oracle.com/technetwork/java/javase/downloads/index.html) Version 1.8
* Unix-like ou Unix-based Systems (Divers Linux et MacOS)

## Installation

docker start hadoop-master hadoop-slave1 hadoop-slave2

puis d'entrer dans le contenaire master:

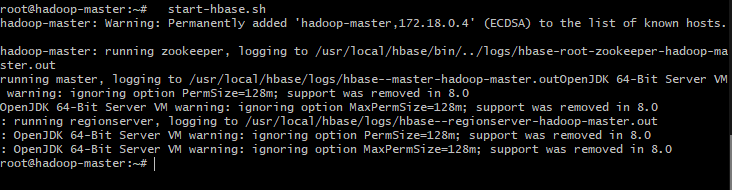
docker exec -it hadoop-master bash

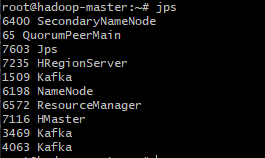
Lancer ensuite les démons yarn et hdfs:

./start-hadoop.sh

Lancer HBase en tapant :

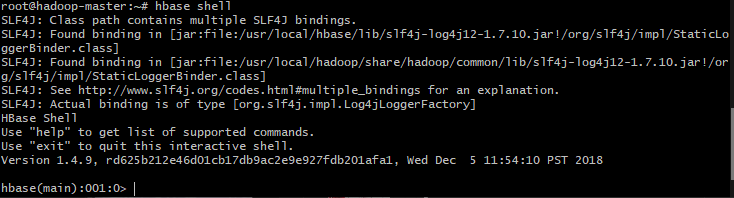
start-hbase.sh

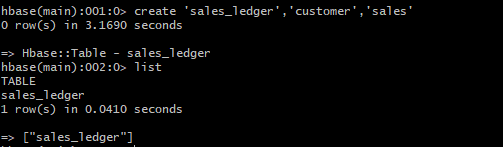




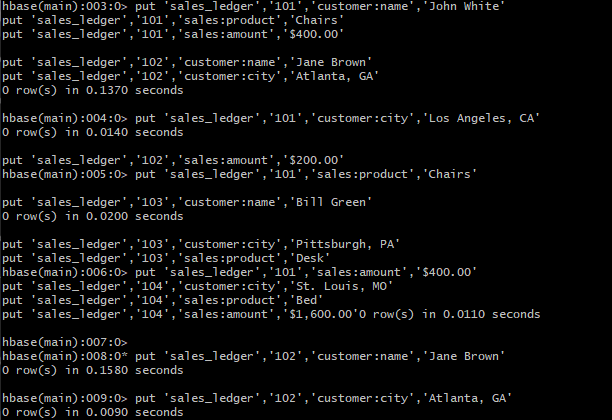
## Première manipulation de HBase

HBase Shell

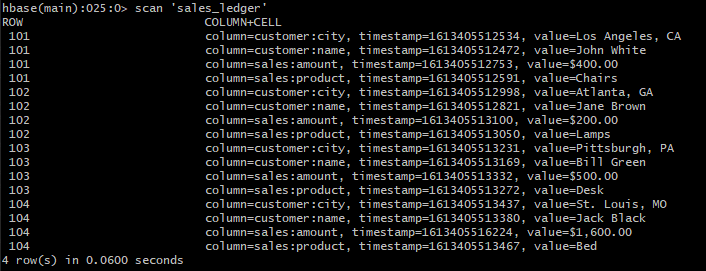




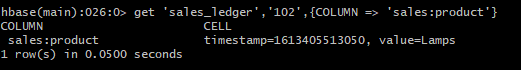
Insérer les différentes lignes:



Visualiser le résultat de l'insertion :



Afficher les valeurs de la colonne *product* de la ligne *102 :*



## HBase API



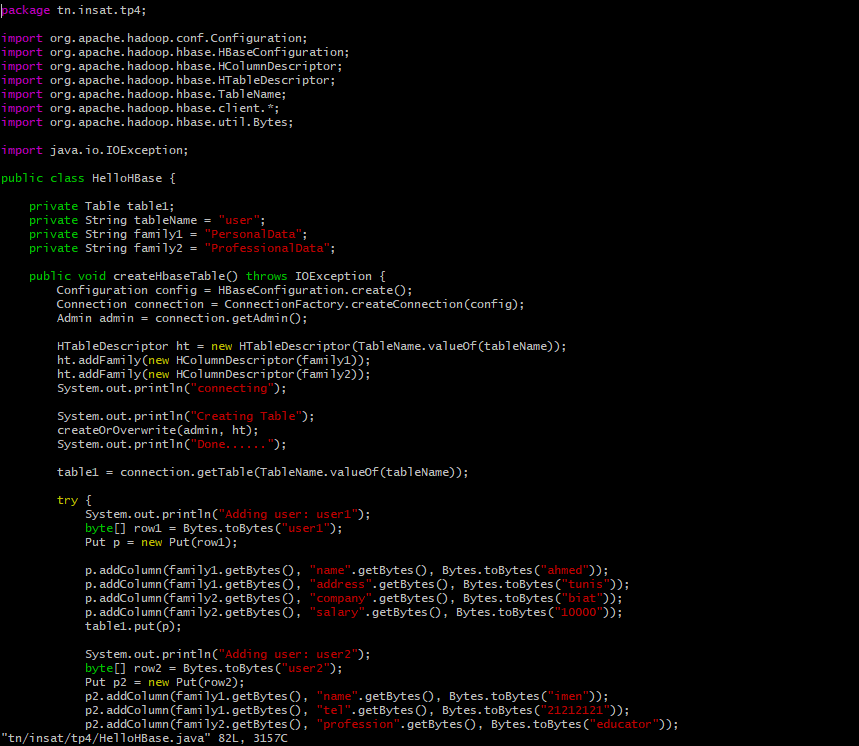
On va créer l'arborescence *tn/insat/tp4* dans ce repertoire :



Créer et ouvrir le fichier *HelloHBase.java* sous le répertoire tp4 avec la commande suivant :



Avec le code suivant :



## Chargement de fichiers

On va charger le fichier dans le répertoire *input* de HDFS (mais d'abord, créer ce répertoire s'il n'existe pas déjà):

## 

## 

hbase org.apache.hadoop.hbase.mapreduce.ImportTsv \ Dimporttsv.separator=',' \ -Dimporttsv.columns=HBASE\_ROW\_KEY,cf:date,cf:time,cf:town,cf:product,cf:price,cf:payment \ products input

## 

## Vérifier que la base a bien été créée en consultant la ville de l'enregistrement numéro 2000:

## 

## Traitement de données avec Spark

## 

## 

## 

## The end ..