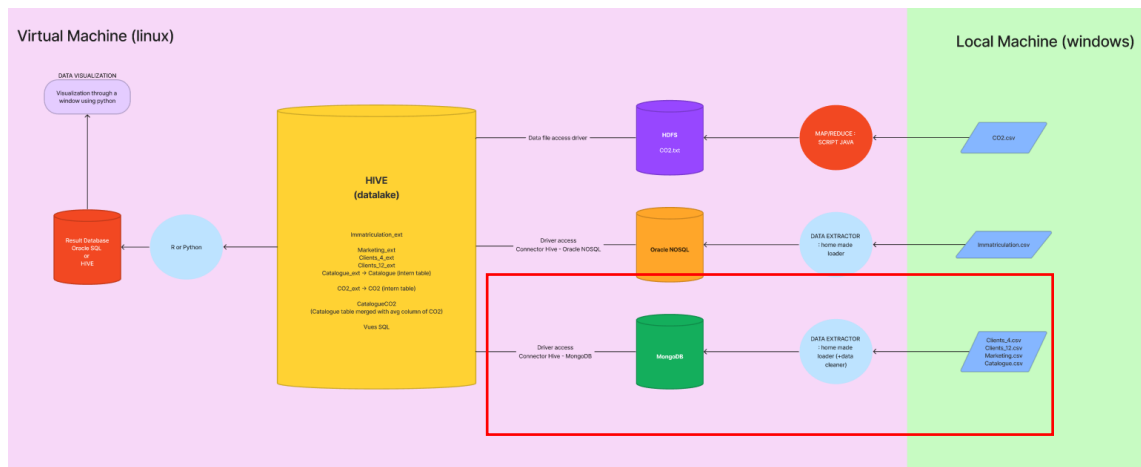


Extracteur MongoDB



Git repo

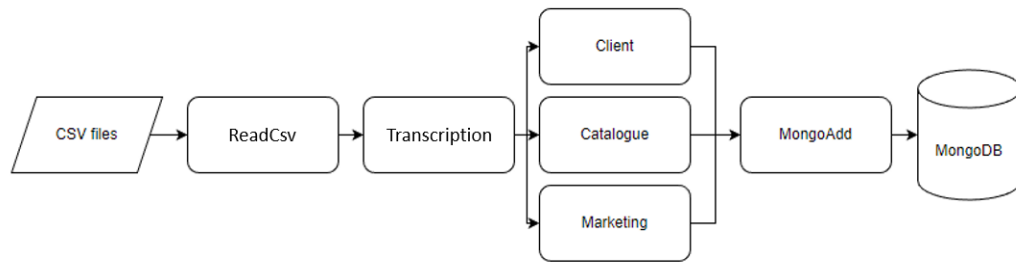
(<https://github.com/PierrickDu/ExtractorMongoDB/blob/main/ExtractorMongoDB/src/MongoAdd.java>)

L'extracteur

L'extracteur MongoDB est développé en java avec la dépendance Mogondb-jdbc. Il s'agit donc d'un extracteur qui récupère les données contenues dans les fichiers csv pour les insérer dans la base MongoDB. Il lit, nettoie et insère les données dans la base Mongo DB. Le main appelle la classe readCsv qui lit le csv et stocke chacune des lignes en tant qu'objet, soit Client, soit Catalogue, soit Marketing. Une fois les objets créés il sont insérés dans la base de données par la classe MongoAdd. La classe transcription sert à modifier des champs lus par la classe Read csv pour les uniformiser et les faire correspondre à des valeurs acceptables. La correspondance est décrite dans un fichier transcription.txt.

Cette classe possède 2 méthodes :

- La première crée un mapping du fichier transcription.txt, stocké dans un hashmap de référence.
- La deuxième méthode remplace les chaines de caractère selon le mapping et est appelée lors de la lecture du fichier csv. D'un autre coté la classe client possède une méthode check qui renvoie un booléen indiquant si les attributs de l'objet sont acceptables pour l'insertion dans la base de données. Si la méthode renvoie true, l'objet est inséré, sinon il est ignoré et inséré dans le fichier rejet.txt. Cela permet ainsi de voir les lignes qui ont été rejetées et de, si besoin, adapter le fichier transcription et ceci sans avoir besoin de modifier le code source.



Implémentation

On allume la Machine Virtuelle

```
Vagrant up
```

```
Vagrant ssh
```

On allume MongoDB

```
sudo systemctl start mongod
```

On ouvre un nouveau cmd

```
vagrant ssh
```

On va utiliser un JAR (chargeur maison + nettoyage)

On commence par déplacer le fichier de transcription (qui permet de nettoyer le csv avant d'entrer les valeurs dans MongoDB)

```
cp ../../vagrant/ProjetTPABigData/ExtractorMongoDB-main/ExtractorMongoDB/transcription.txt ~
```

On crée un fichier rejet.txt et on déplace toutes les lignes non insérées car non conformes

```
cp ../../vagrant/ProjetTPABigData/ExtractorMongoDB-main/ExtractorMongoDB/rejet.txt ~
```

On exécute le JAR afin de créer la database et remplir les collections

```
java -jar ../../vagrant/ProjetTPABigData/ExtractorMongoDB-main/ExtractorMongoDB/ExtractorMongoDB.jar ~
```

On lance ensuite MongoDB

```
mongo
```

On teste que tout fonctionne

```
show dbs
```

On vérifie que les tables sont bien remplies (sur mongo avec le cmd (lancé précédemment))

```
use projetBigData  
db.marketing.find();
```

On arrête MongoDB

```
sudo systemctl stop mongod
```