

PROJET 5

Maintenez et documentez un système de stockage des données
sécurisé et performant

PROJET 5

CONTEXTE



CONTEXTE

Stagiaire chez DataSoluTech

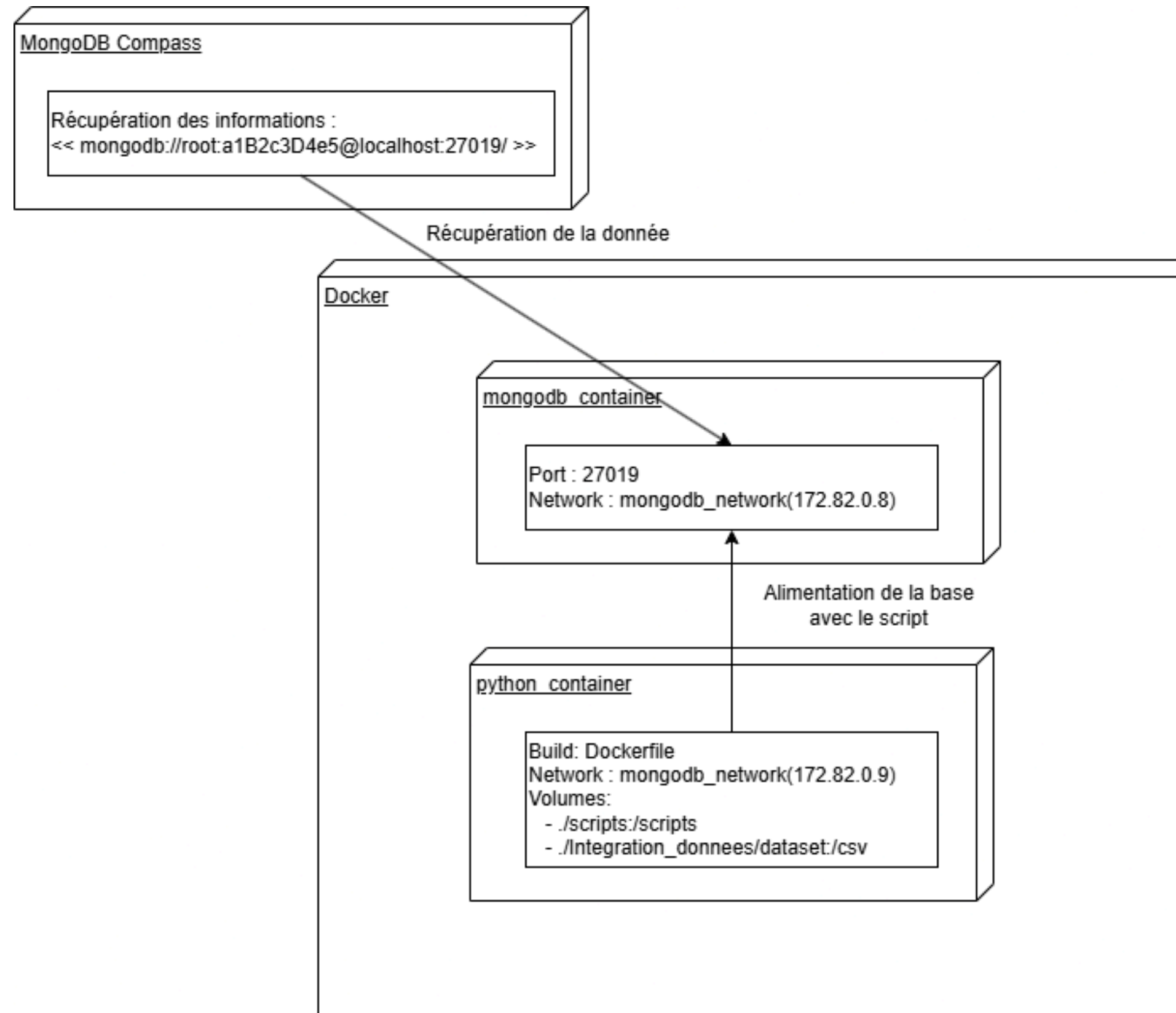
Sous la supervision de mon chef de projet Boris, je me suis penché sur la migration de données depuis un dataset CSV vers une instance MongoDB sous Docker.

Pour ce faire, il faut écrire un script d'intégration de données et expliquer le processus dans un README.md qui sera sur un dépôt GitHub.

Les différentes étapes vont être :

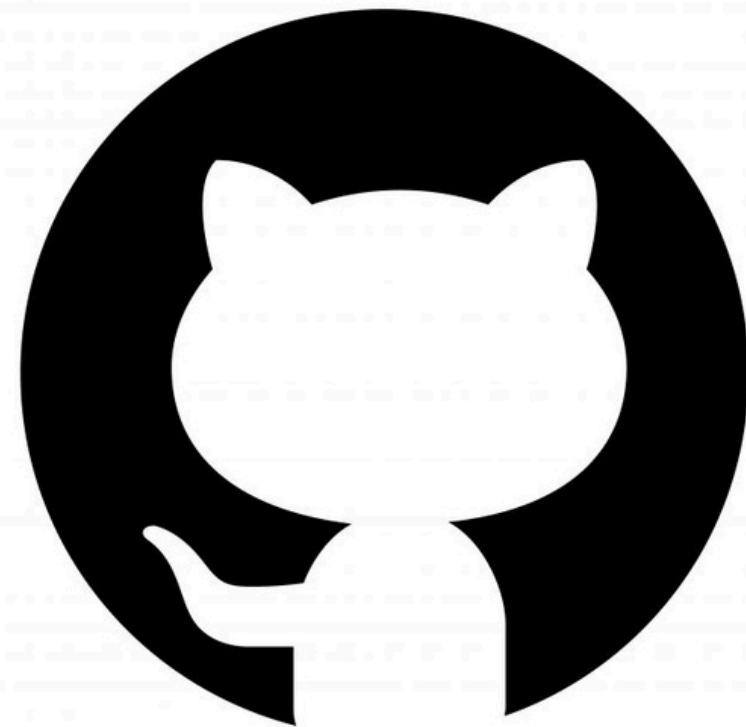
- Schéma de l'architecture souhaitée
- Création du dépôt GitHub et clonage de celui-ci en local
- Construction du docker-compose.yml et du Dockerfile
- Ecriture du script d'intégration
- Connexion de la base de données sur Compass
- Test du processus en suivant la documentation du README

ARCHITECTURE



PROJET 5

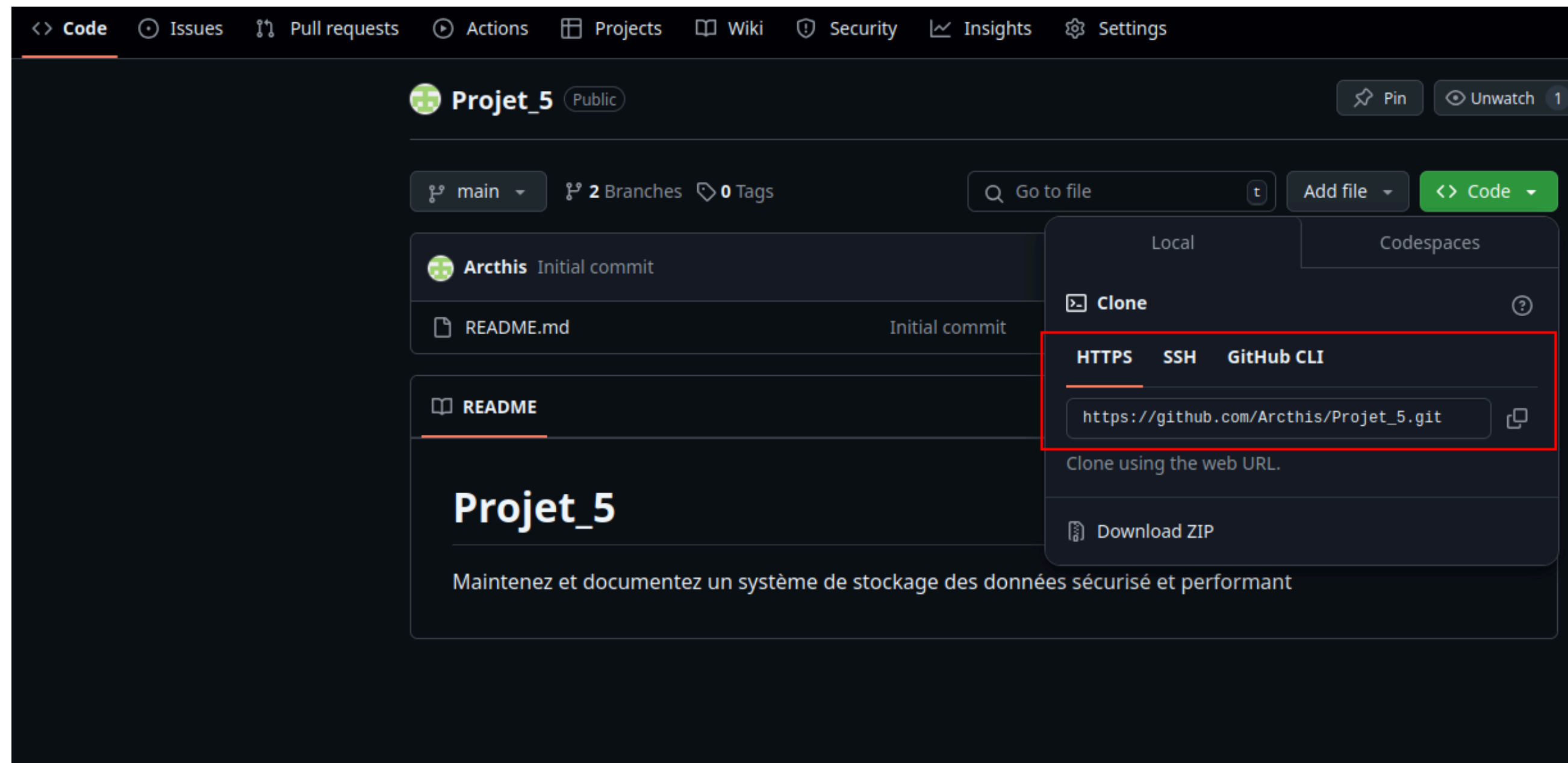
GITHUB



GITHUB

Clone du dépôt

- git clone (https://github.com/Arcthis/Projet_5.git)
- Branche DEV (git checkout dev)
- Mise à jour du code sur GitHub :
 - git add .
 - git commit -m "message du commit"
 - git push
- README.md en première page



PROJET 5

DOCKER



DOCKER COMPOSE

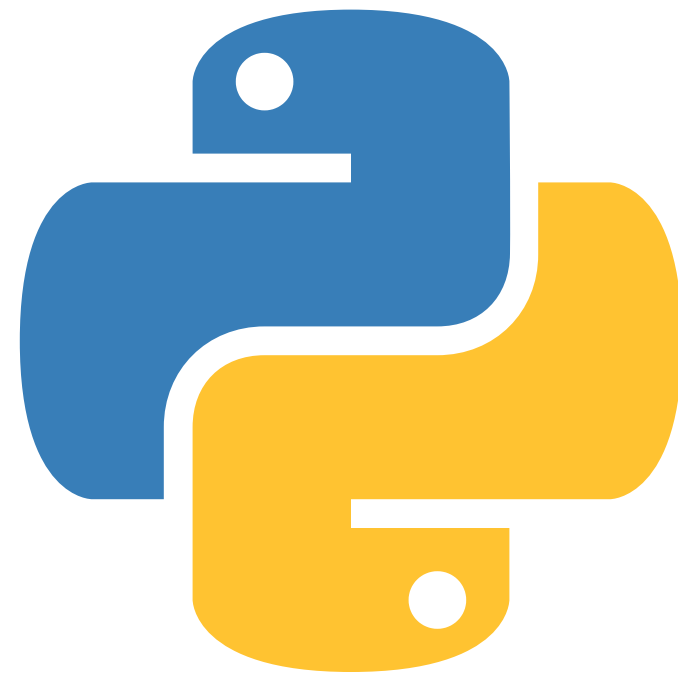
```
1  version: "3.8"
2  services:
3    db:
4      container_name: mongodb_container
5      image: mongo
6      restart: always
7      environment:
8        MONGO_INITDB_ROOT_USERNAME: root
9        MONGO_INITDB_ROOT_PASSWORD: a1B2c3D4e5
10     ports:
11       - "27019:27017"
12     networks:
13       mongodb_network:
14         ipv4_address: 172.82.0.8
15
16     python_app:
17       container_name: python_container
18       build:
19         context: .
20         dockerfile: Dockerfile
21       networks:
22         mongodb_network:
23           ipv4_address: 172.82.0.9
24       volumes:
25         - ./scripts:/scripts
26         - ./Integration_donnees/dataset:/csv
27
28     networks:
29       mongodb_network:
30         driver: bridge
31         ipam:
32           config:
33             - subnet: 172.82.0.0/24
34
```


DOCKERFILE

```
1 FROM python:3.9-slim
2
3 RUN apt-get update && apt-get install -y \
4     build-essential \
5     && rm -rf /var/lib/apt/lists/*
6
7 WORKDIR /scripts
8
9 RUN pip install --no-cache-dir pandas pymongo
10
11 CMD ["tail", "-f", "/dev/null"]
```

PROJET 5

SCRIPT PYTHON



SCRIPT PYTHON

```
1  import pandas as pd
2  from pymongo import MongoClient
3
4  mongo = MongoClient('mongodb://root:a1B2c3D4e5@db:27017/')
5  db = mongo['local']
6  collection = db['data']
7
8  df = pd.read_csv('/csv/healthcare_dataset.csv')
9
10 data = df.to_dict(orient='records')
11
12 collection.insert_many(data)
13
14 print("Importation terminée")
```

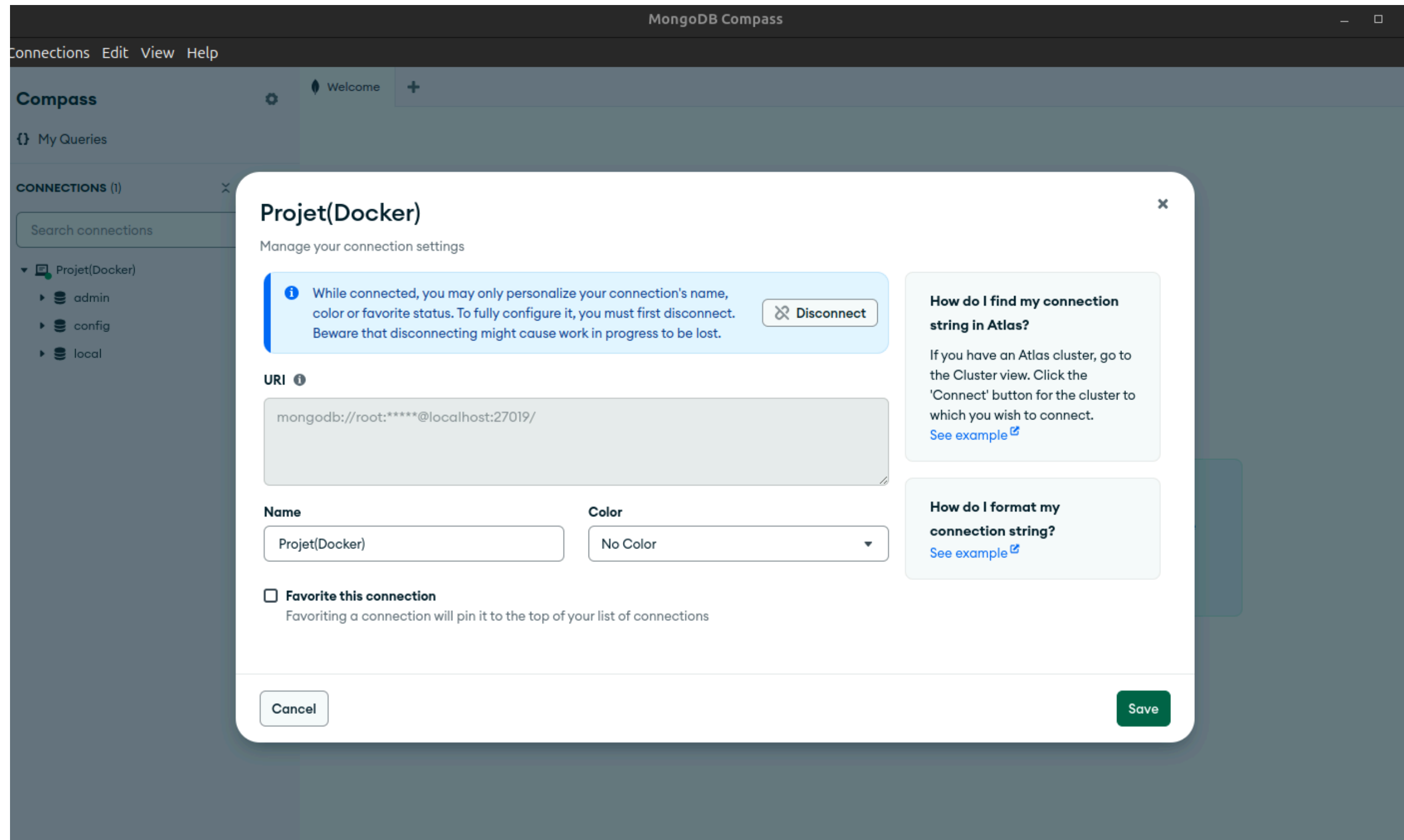
PROJET 5

MONGODB | COMPASS



MongoDB®

CONNEXION BDD



DONNEES MIGREES

 ADD DATA ▾

 EXPORT DATA ▾

 UPDATE

 DELETE

25 ▾

1 - 25 of 55500



`_id: ObjectId('677ffd26716adee50e1682d2')`
`Name : "Bobby JacksOn"`
`Age : 30`
`Gender : "Male"`
`Blood Type : "B-"`
`Medical Condition : "Cancer"`
`Date of Admission : "2024-01-31"`
`Doctor : "Matthew Smith"`
`Hospital : "Sons and Miller"`
`Insurance Provider : "Blue Cross"`
`Billing Amount : 18856.281305978155`
`Room Number : 328`
`Admission Type : "Urgent"`
`Discharge Date : "2024-02-02"`
`Medication : "Paracetamol"`
`Test Results : "Normal"`

PROJET 5

CONCLUSION