This is a step-by-step guide for running this project. Follow each step as outlined here.

1.Prerequisites and dataset

> **Note**
>
> The project root directory refers to: *C:\Users\XXX\Desktop\realtime_fraud_detection* or similar to this

Device requiement

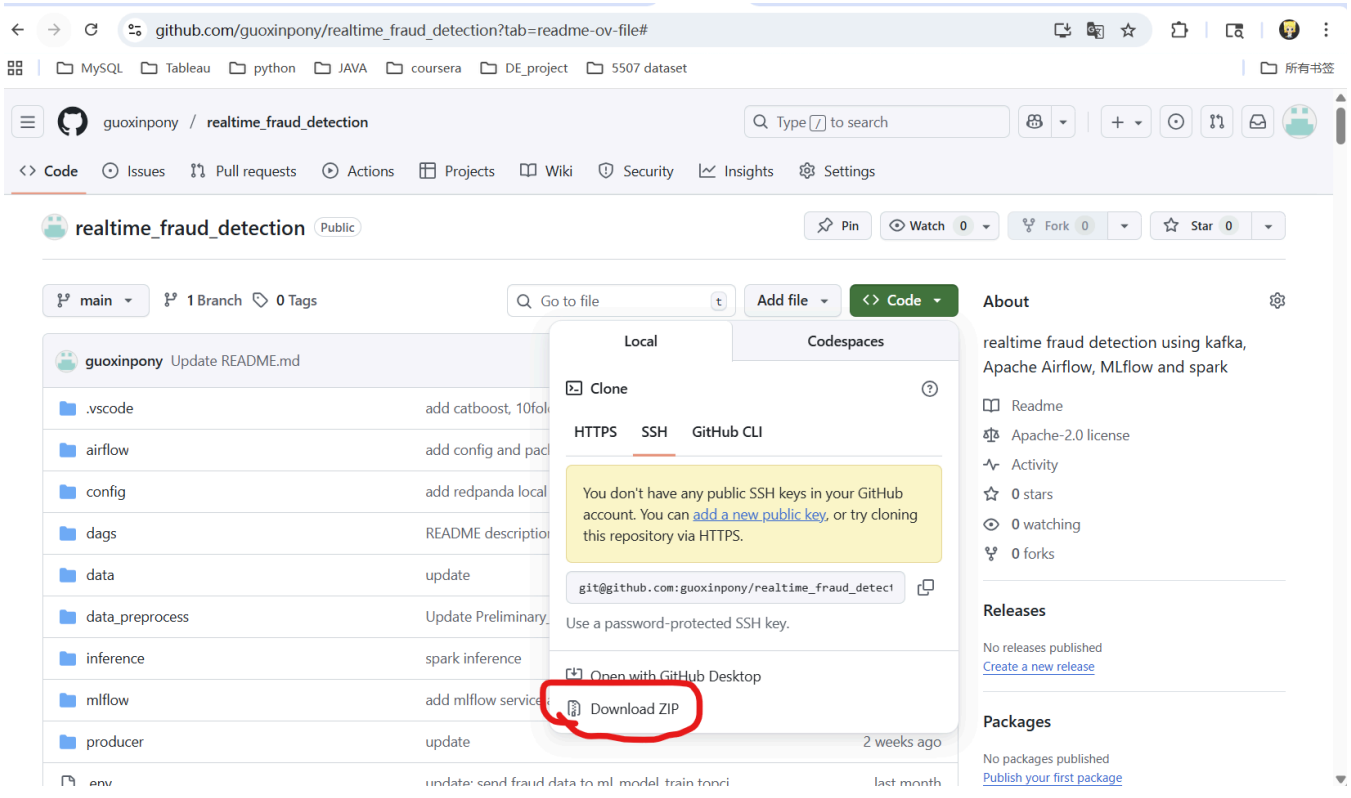
- **CPU**: 4+ cores recommended
- **Memory**: 8GB minimum, 16GB recommended
- **OS**: Linux, macOS, or Windows with WSL2

Clone or download project:

```
git clone https://github.com/guoxinpony/realtime_fraud_detection.git
```

OR

Download ZIP:



Install Docker Desktop:

1. download package from: https://www.docker.com/products/docker-desktop/
2. Install docker desktop. If you are using Windows system, the installation process will prompt you to enable WSL2. Follow the instructions to enable WSL2 and restart your computer.
3. Launch Docker Desktop

Download Dataset:

1. Download from: https://drive.google.com/file/d/1y1QqL1BdJKMpEu4dOB5OKANPeUxIkl3X/view?usp=drive_link, and place the dataset in the data directory within the project.

OR

2. Download from Kaggle: https://www.kaggle.com/datasets/kartik2112/fraud-detection, including two datasets: fraudTest.csv and fraudTrain.csv; and place the two datasets in the data directory within the project; and RUN in location of project root directory:

```
python ./data_preprocess/merge_data.py
```

MAC OS/ Linux required: Change permission of Script in location of project root directory:

```
chmod +x wait-for-it.sh
```

2.Docker image build

The initial build takes some time, which includes downloading the official image, necessary packages, and the build process itself. The exact duration depends on your computer's performance and network connection.

In the project root directory, build the following four images airflow-webserver, mlflow-server, producer, and inference. Input in the terminal:

```
docker compose build airflow-webserver
```

```
docker compose build mlflow-server
```

```
docker compose build producer
```

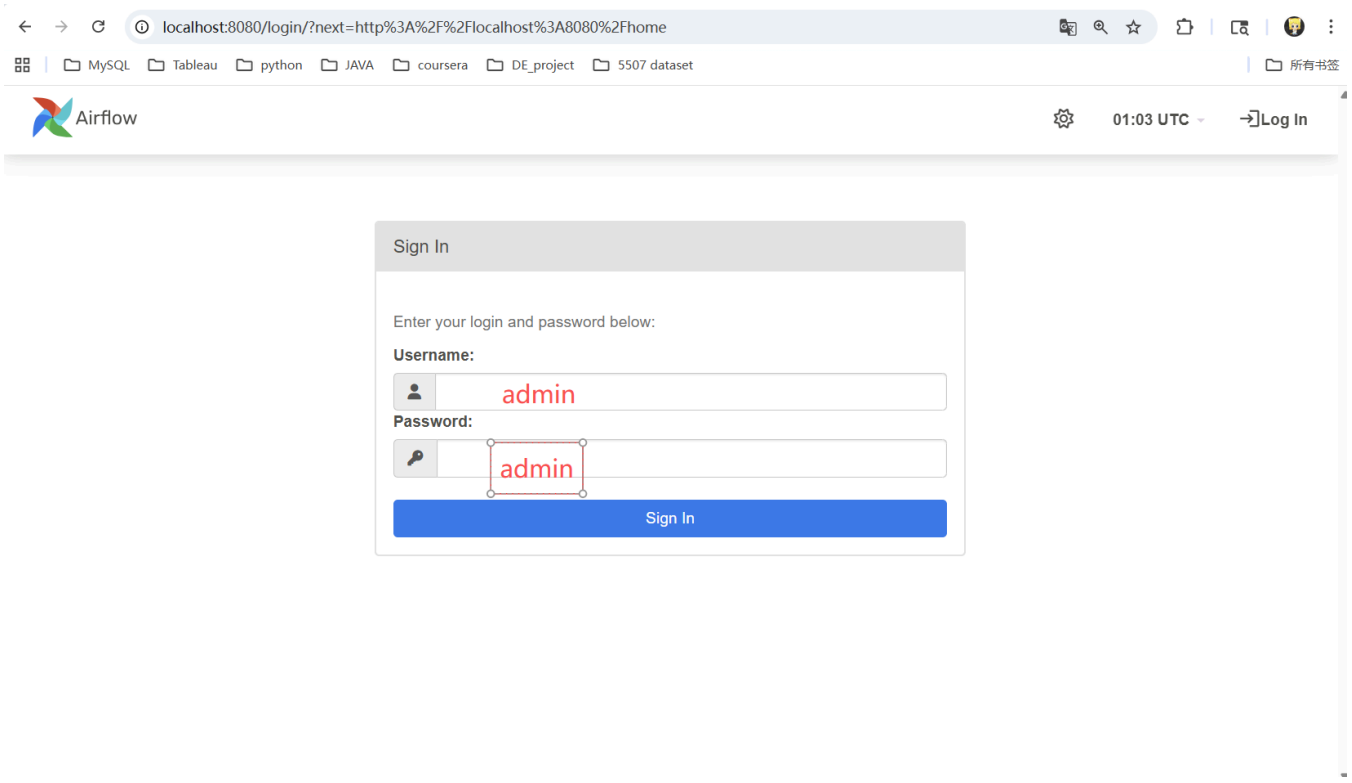
```
docker compose build inference
```

3.Start Services and Check

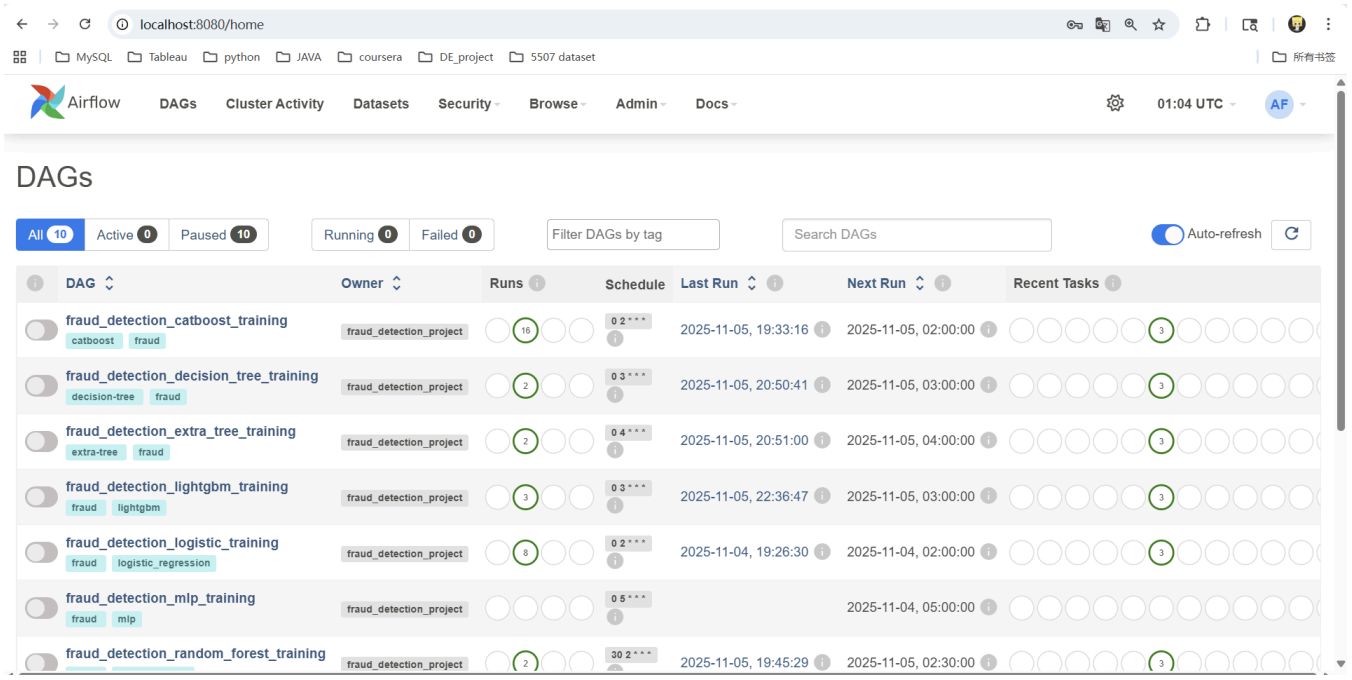
(1)Use the following command to start all services(containers), and waiting for all container run successfully:

```
docker compose up -d
```

(2) Open http://localhost:8080/home in your browser, then enter the username admin and password admin; If the page fails to open, it indicates that the service has failed to run:

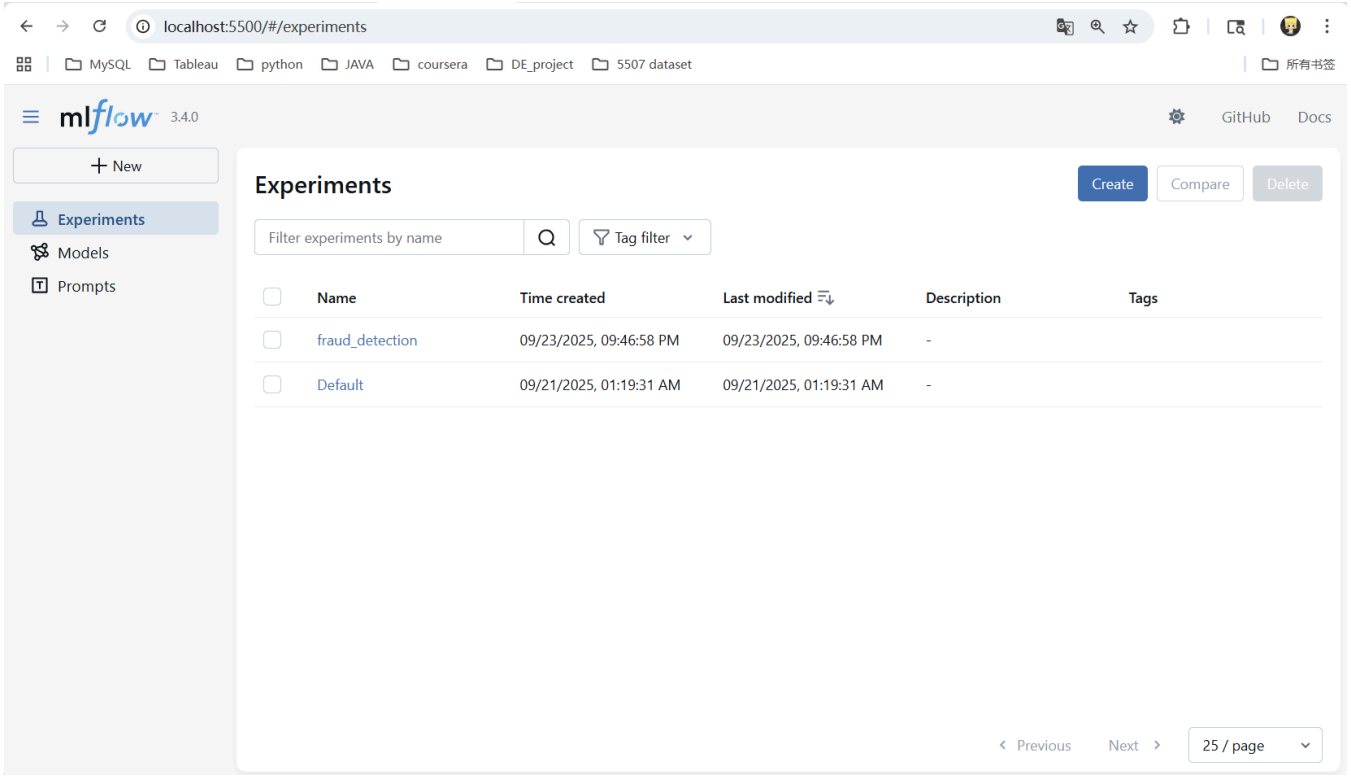


You should be able to see:

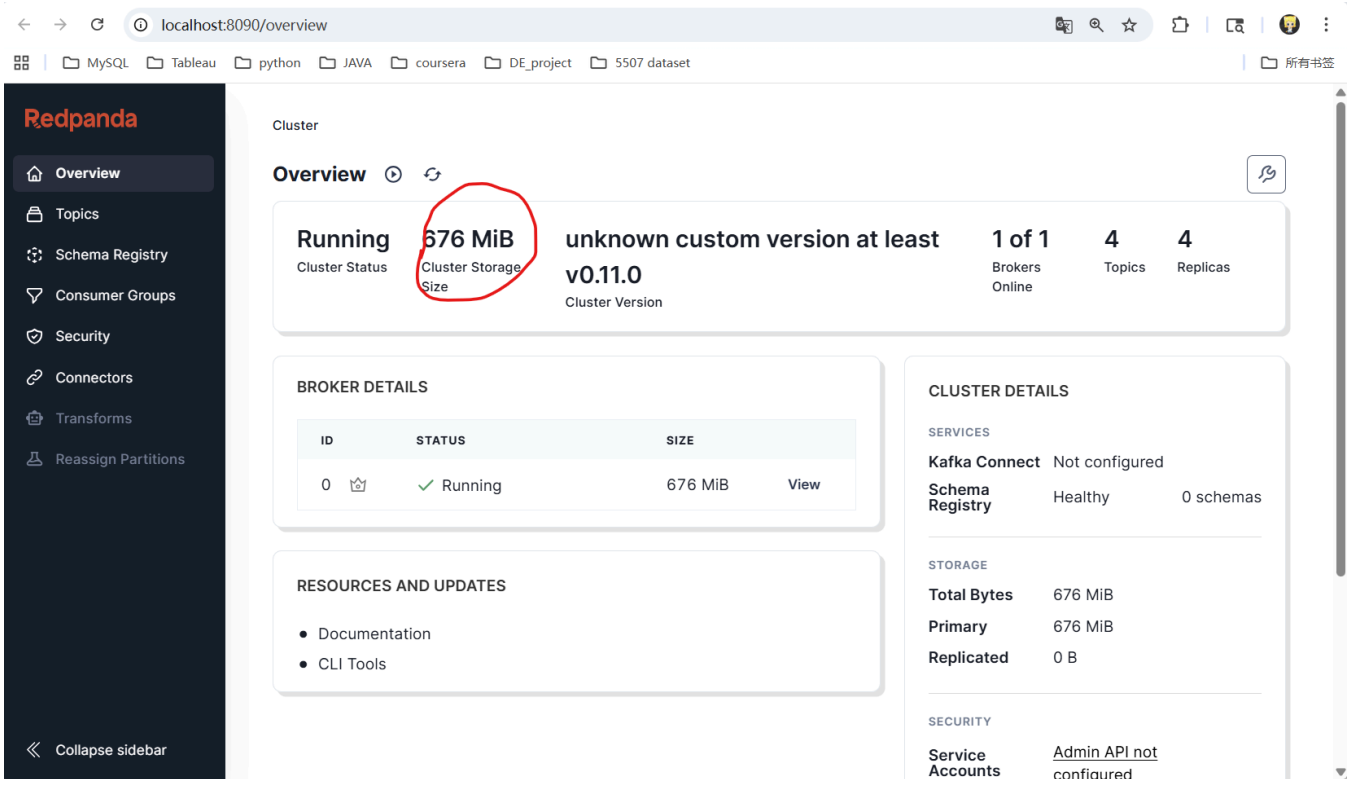


(3) Open http://localhost:5500/ in your browser, If this is the first time opening it, this interface should be blank, and the `fraud_detection` directory does not exist:

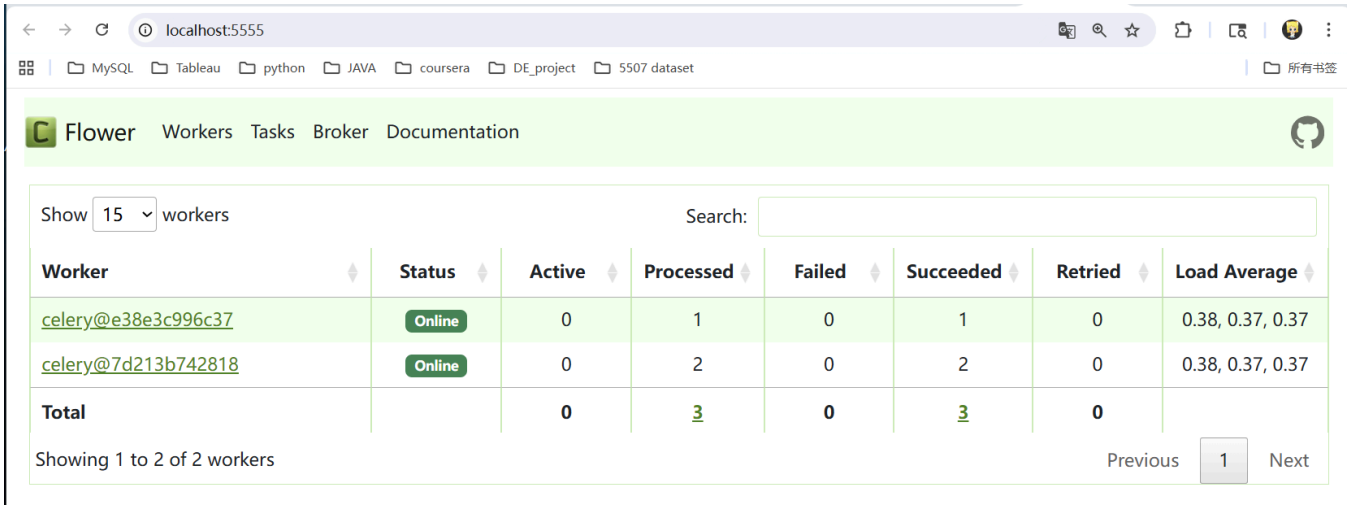
> **Note**
>
> After running a task that trains a model in Airflow, a fraud_detection directory will appear.



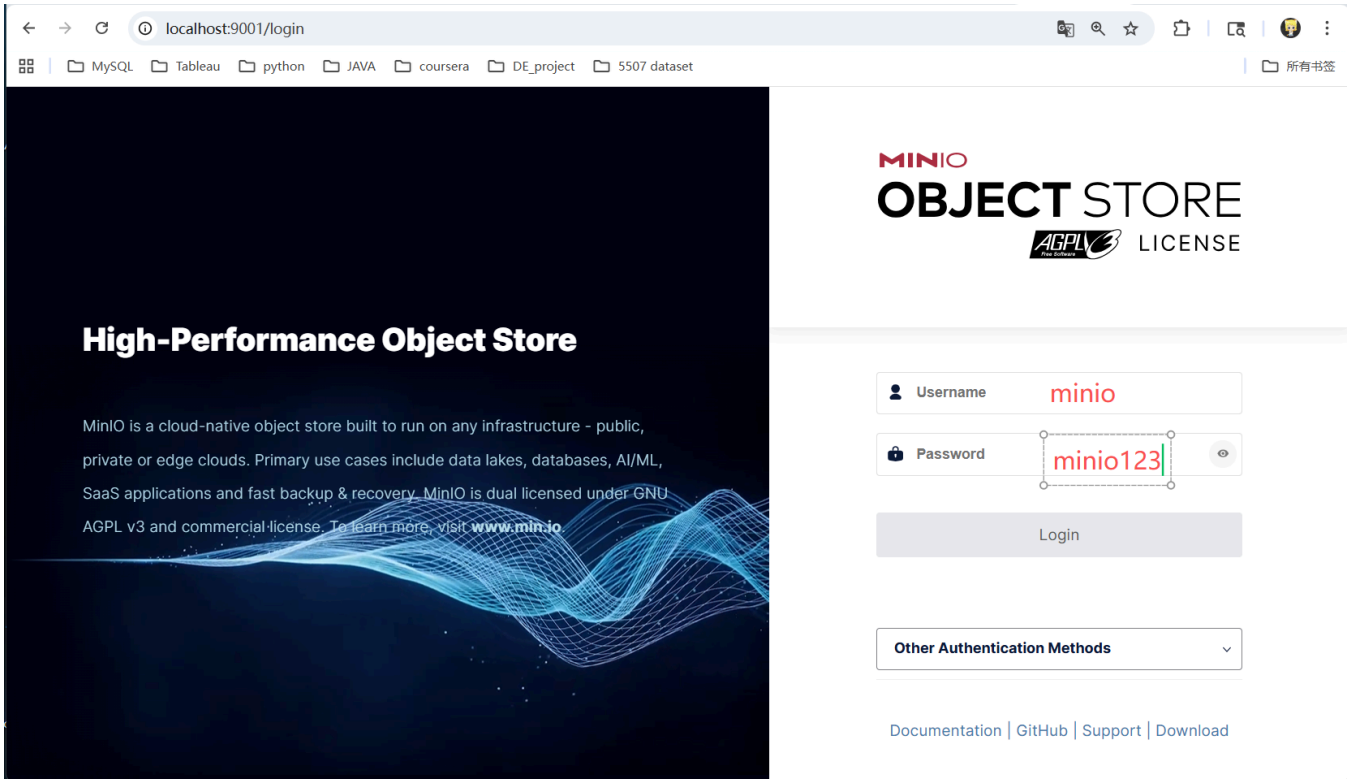
(4) Open http://localhost:8090/overview in your browser, you may notice the growing volume of data.



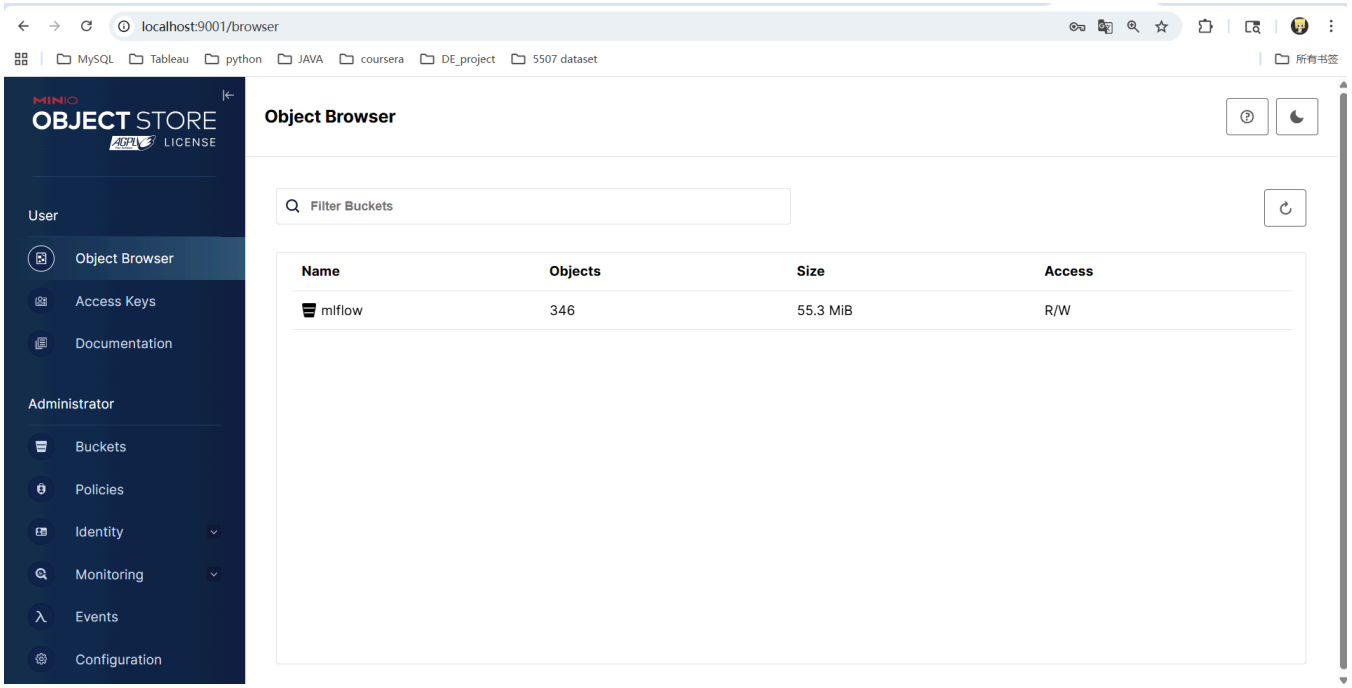
(5) Open http://localhost:5555/ in your browser, You should be able to see two workers online.



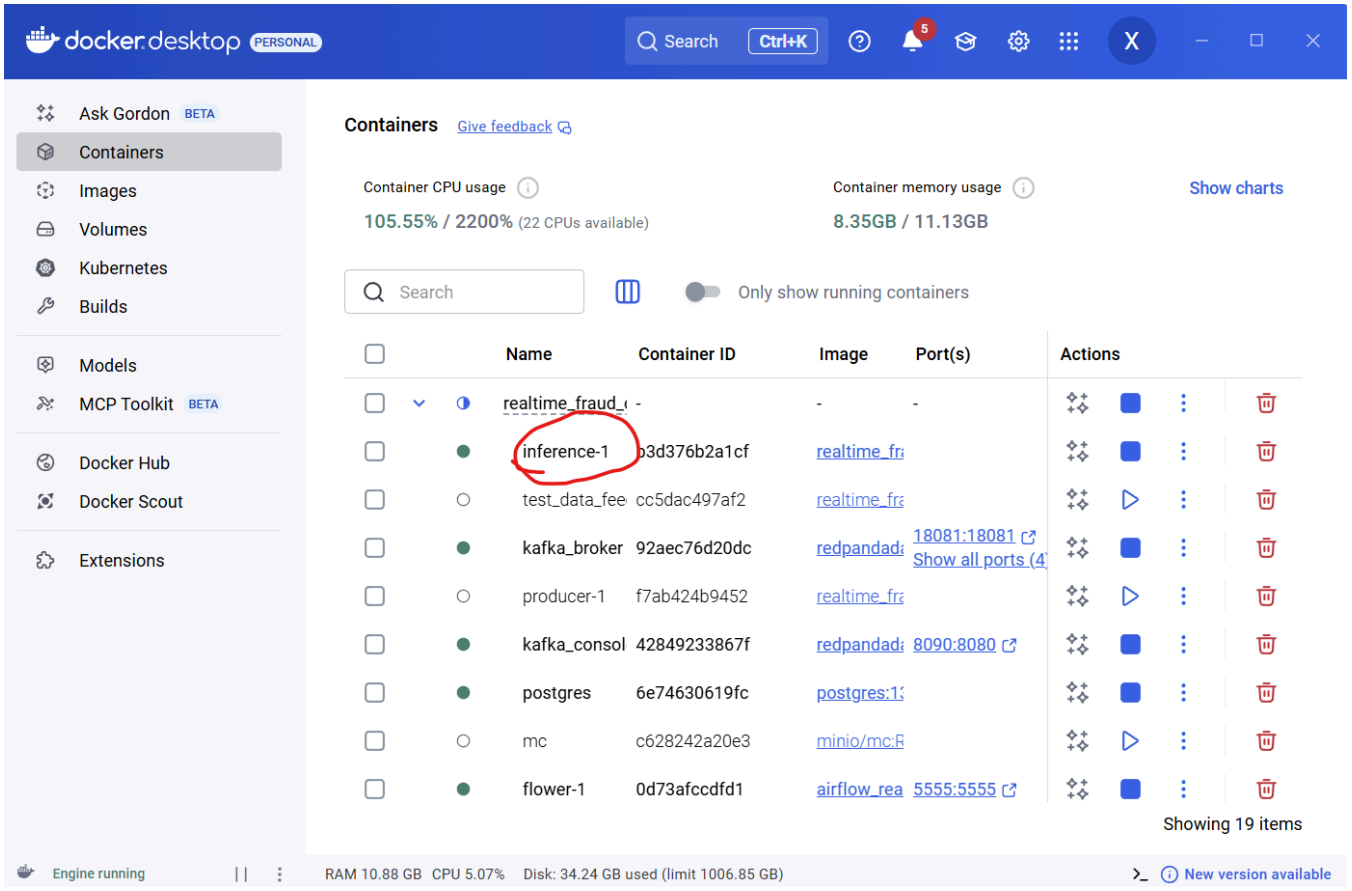
(5) Open http://localhost:9001/ in your browser, Username is minio, password is minio123:



After logging in, you will see:



(6) In Docker Desktop, click the Containers panel on the left and navigate to the inference container:



> **Note**

> Under the logs bar, if you see the log entry `“INFO MicroBatchExecution: Streaming query has been idle and waiting for new data more than 10000 ms,”` it indicates that the system's inference component is ready and waiting for new credit card data to be pushed, utilizing the machine learning model to make inferences!