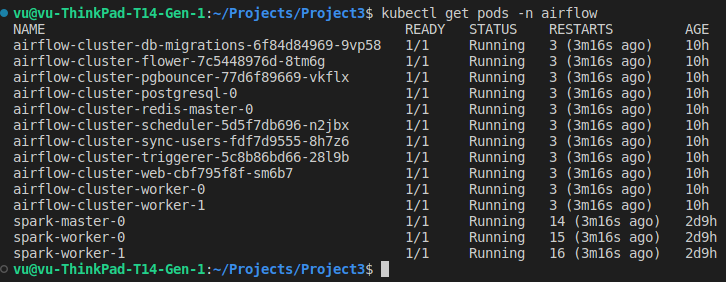
1. Tiến hành cài đặt PostGreSQL, TimeScaleDB, Docker, KinD, Kubectl
   1. Postgresql: <https://www.postgresql.org/download/>
   2. TimeScaleDB: <https://docs.timescale.com/install/latest/>
   3. Docker Desktop: <https://docs.docker.com/desktop/install/ubuntu/>
   4. Kubernetes: <https://kubernetes.io/docs/setup/>
   5. kubectl,KinD: <https://kubernetes.io/docs/tasks/tools/>
2. Hướng dẫn sẽ coi:
   1. PostgreSQL username: airflow-username
   2. PostgreSQL password: airflow-password
   3. PostgresSQL database: airflow-database
   4. Sử dụng bảng
   5. Mọi câu lệnh được xử lí trong folder source code có đường dẫn đầy đủ: /home/user/project
3. Giải nén file source code
4. Thêm “/home/user/project/Airflow” vào hostPath của 2 worker trong file k8s/kind-cluster.yaml
5. Sử dụng Terminal, truy cập vào folder source code giải nén rồi thực hiện những lệnh sau:

|  |
| --- |
| #tạo 1 cluster với config như file kind-cluster.yaml gồm 1 control plane và 2 worker kind create cluster --name airflow-cluster --config home/user/project/k8s/kind-cluster.yaml  #tạo 1 namespace (ở dây là airflow)  kubectl create ns airflow #add repo của airflow về máy helm repo add airflow-stable https://airflow-helm.github.io/charts #pull image của airflow về docker pull apache/airflow:2.2.5-python3.8  #build 1 image được sửa lại từ image airflow official docker build -t airflow-custom:1.0.0 /home/user/project/k8s/Airflow  #load image vào từng node trên cluster  kind load docker-image airflow-custom:1.0.0 --name airflow-cluster  #tạo ra persistent volume để các pods để logs/file dags/spark app của dags vào kubectl apply -f pv-logs.yaml  kubectl apply -f pvc-logs.yaml  kubectl apply -f pv-dags.yaml  kubectl apply -f pvc-dags.yaml  #đặt variable cho lệnh cài đặt airflow vào cluster  export AIRFLOW\_NAME="airflow-cluster"  export AIRFLOW\_NAMESPACE="airflow"  #cài đặt airflow vào cluster  helm install \  "$AIRFLOW\_NAME" \  airflow-stable/airflow \  --namespace "$AIRFLOW\_NAMESPACE" \  --version "8.6.1" \  --values /home/user/project/k8s/Airflow/celery-values.yaml  ### Optional  #add repo của spark về máy helm repo add bitnami https://charts.bitnami.com/bitnami #build 1 image được sửa lại từ image airflow official docker build -t spark-custom:1.0.0 /home/user/project/k8s/Spark  #load image vào từng node trên cluster  kind load docker-image airflow-custom:1.0.0 --name airflow-cluster  #cài đặt chart spark helm install spark bitnami/spark -n airflow --values .Spark/spark-values.yaml |

1. Kiểm tra các pod của airflow (có tiền tố airflow) có hoạt động, nếu các pods đã hoạt động hết, sang bước tiếp

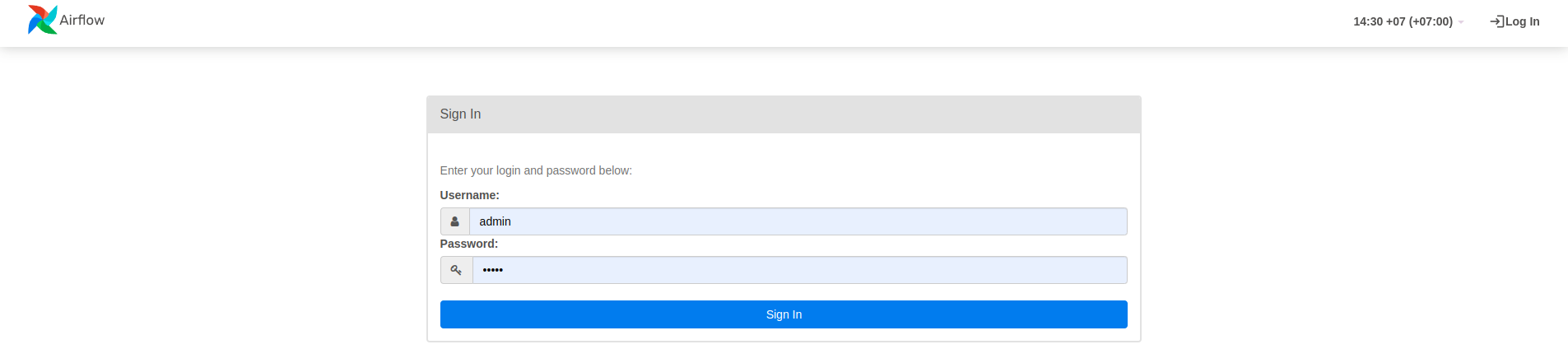
|  |
| --- |
| kubectl get pods -n airflow |

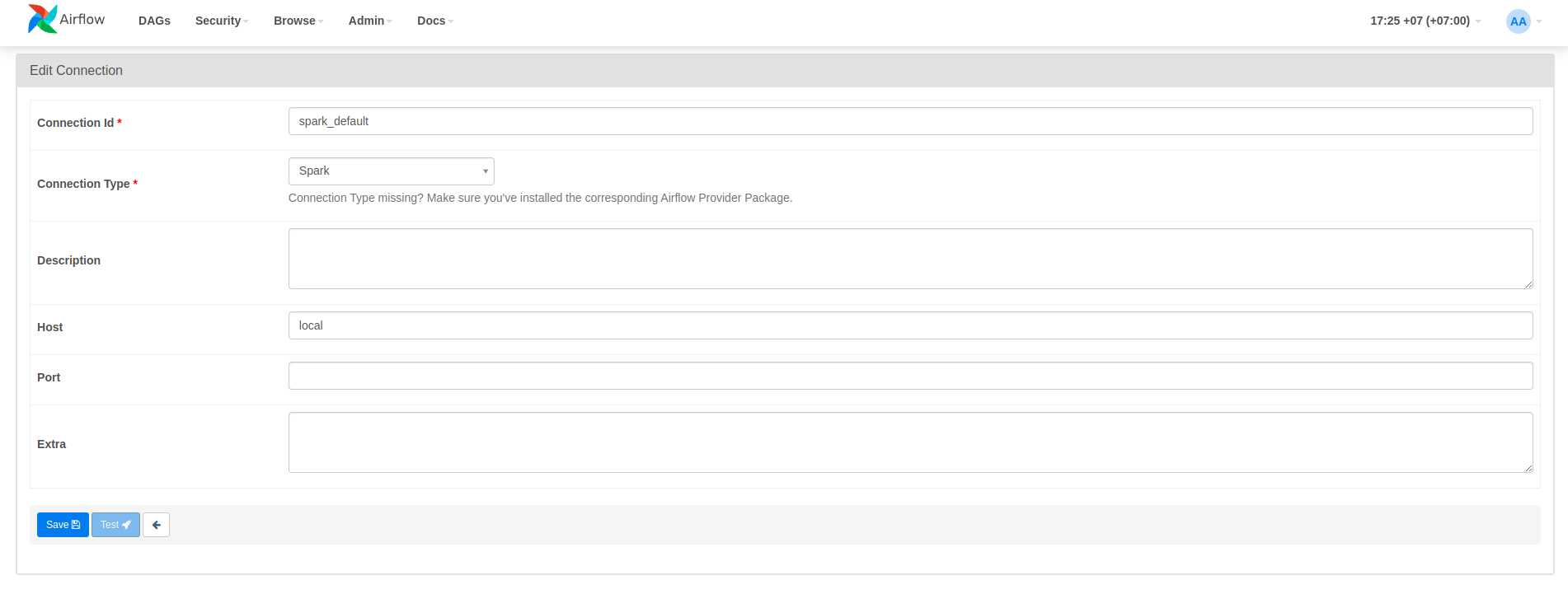


1. Bật terminal khác lên chạy

|  |
| --- |
| kubectl port-forward svc/airflow-cluster-web 8080:8080 --namespace airflow  kubectl port-forward svc/spark-master-svc 8081:80 --namespace airflow |

1. Truy cập bằng browser localhost:8080 hoặc 127.0.0.1:8080 với Username và Password đều là admin



1. Để kết nối dễ hơn cho Airflow, ta thiết lập connection tới Spark cho Airflow, truy cập vào phần Admin/Connections, thêm một connection như sau
2. Thiết lập Database và các bảng trong database

|  |
| --- |
| kubectl get pods -n airflow |

1. Sửa trong Airflow/dag/data\_to\_db.py,
   1. biến URL: “jdbc:postgresql://host.docker.internal:5432/airflow-database\*"
   2. biến USERNAME: “airflow-username”
   3. biến PASSWORD: “airflow-password”
2. Khi thích hợp, bật DAG import\_init, import\_daily, import\_quarterly
3. (Optional) Sửa lại thứ tự và số lượng core sẽ dùng của các task trong dag
4. (Optional) Cài đặt dbeaver:
   1. <https://dbeaver.io/download/>
   2. Thiết lập connection đến Database PostgreSQL đã thiết lập