**Documentation sur Apache Superset**

**Introduction**

Apache Superset est une plateforme de visualisation et d'exploration de données open-source. Conçue pour être intuitive, puissante et évolutive, elle permet aux utilisateurs de créer et partager des tableaux de bord interactifs, explorer des ensembles de données et exécuter des requêtes SQL. Apache Superset est compatible avec une large gamme de bases de données SQL et d'entrepôts de données.

**Caractéristiques principales**

* **Support multi-base de données** : Compatibilité avec les bases de données courantes telles que PostgreSQL, MySQL, MariaDB, BigQuery, Snowflake, et bien d'autres.
* **Exploration sans code** : Interface utilisateur intuitive permettant de créer des visualisations sans avoir besoin d'écrire du code.
* **Requêtes SQL avancées** : Interface pour l'écriture et l'exécution de requêtes SQL, avec des outils d'analyse des résultats.
* **Extensibilité** : Une architecture modulaire permettant des intégrations personnalisées et l'ajout de nouveaux types de visualisations.
* **Sécurité et gestion des accès** : Contrôle granulé des permissions pour les utilisateurs et les rôles.
  1. **Installation et configuration d'Apache Superset dans Kubernetes**

Pour installer Apache Superset sur Kubernetes, vous pouvez suivre ces étapes. Apache Superset est une application Web de business intelligence moderne et adaptée aux entreprises que vous pouvez déployer à l'aide de graphiques Helm ou en créant manuellement des ressources Kubernetes. Vous trouverez ci-dessous un guide pour la méthode manuel.

Installer Superset dans Kubernetes implique plusieurs étapes, notamment la préparation du cluster, la configuration de la base de données backend et l'intégration des composants nécessaires.

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Etape 1: Déployer Apache Superset manuellement :**

Préparer les ressources Kubernetes Créez des ressources Kubernetes telles que Deployment, Service et PersistentVolumeClaim (PVC).

Le super-ensemble de configuration de base de données nécessite une base de données de métadonnées (par exemple, PostgreSQL ou MySQL).

Déploiement d’une base de données PostgreSQL dans :

apiVersion: apps/v1

kind: Deployment

metadata:

name: superset-postgres

spec:

replicas: 1

selector:

matchLabels:

app: superset-postgres

template:

metadata:

labels:

app: superset-postgres

spec:

containers:

- name: postgres

image: postgres:13

env:

- name: POSTGRES\_DB

value: superset

- name: POSTGRES\_USER

value: superset\_user

- name: POSTGRES\_PASSWORD

value: superset\_password

ports:

- containerPort: 5432

**Déploiement de Superset : Créer un manifeste de déploiement de Superset**:

apiVersion: apps/v1

kind: Deployment

metadata:

name: superset

spec:

replicas: 1

selector:

matchLabels:

app: superset

template:

metadata:

labels:

app: superset

spec:

containers:

- name: superset

image: apache/superset:latest

env:

- name: SUPERSET\_SECRET\_KEY

value: "your\_secret\_key"

- name: SUPERSET\_DATABASE\_URI

value: "postgresql+psycopg2://superset\_user:superset\_password@superset-postgres/superset"

ports:

- containerPort: 8088

command:

- "/bin/sh"

- "-c"

- |

superset db upgrade &&

superset init &&

gunicorn -w 2 -b 0.0.0.0:8088 "superset.app:create\_app()"

**Créer un Service : Exposer le déploiement de Superset à l'aide d'un Service:**

apiVersion: v1

kind: Service

metadata:

name: superset

spec:

selector:

app: superset

ports:

- protocol: TCP

port: 80

targetPort: 8088

type: LoadBalancer

**Appliquer les ressources : Déployer les fichiers YAML:**

kubectl apply -f superset-postgres.yaml

kubectl apply -f superset.yaml

kubectl apply -f superset-service.yaml

**Accéder à Superset : Obtenir l'IP externe du service Superset:**

kubectl get svc superset

**Ouvrez votre navigateur et accédez à l'IP externe**. **\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Par défaut, lorsque vous installez Apache Superset, il n'y a pas d'utilisateur préconfiguré. Vous devez créer un utilisateur administrateur avant de pouvoir vous connecter. Voici les étapes pour créer un utilisateur administrateur et configurer le mot de passe :**

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Étape 2 : Créer un utilisateur administrateur**

**1. Exécutez une commande pour accéder à votre pod Superset :**

kubectl exec -it <nom-du-pod-superset> -- bash

**Remplacez <nom-du-pod-superset> par le nom de votre pod Superset. Vous pouvez lister les pods avec :**

kubectl get pods

**2. Une fois dans le conteneur, créez un utilisateur administrateur avec la commande suivante :**

superset fab create-admin

**Cette commande vous demandera les informations suivantes :**

**Username :** Entrez un nom d'utilisateur (ex. admin).

**User first name :** Prénom de l'utilisateur.

**User last name :** Nom de famille de l'utilisateur.

**Email :** Adresse e-mail.

**Password :** Définissez un mot de passe.

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Etape 3: Lancement de l’application:**

Si vous exécutez Kubernetes localement (par exemple, avec Minikube, Kind ou MicroK8s), l'utilisation du type de service **LoadBalancer** ne fonctionnera pas, sauf si vous avez un fournisseur de load balancer configuré. La plupart des configurations Kubernetes locales n'ont pas de load balancer natif, donc vous devrez trouver une autre méthode pour exposer le service.

Voici quelques options pour exécuter votre service localement :

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**1. Utiliser l'émulation du LoadBalancer de Minikube :**

Si vous utilisez Minikube, vous pouvez exposer le service LoadBalancer en exécutant la commande suivante :

**minikube service superset**

Cela ouvrira l'interface utilisateur de Superset dans votre navigateur. Minikube émule le comportement du LoadBalancer pour le développement local.

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**2. Utiliser Port Forwarding**

Vous pouvez rediriger le port du service vers votre machine locale :

**kubectl port-forward svc/superset 8088:80**

**Vous pouvez maintenant accéder à Superset à l'adresse** [**http://localhost:8088**](http://localhost:8088)**.**

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Étape 4 : Vérifier et initialiser la base de données (si nécessaire)**

**Si ce n'est pas déjà fait, assurez-vous d'avoir initialisé la base de données de Superset :**

superset db upgrade

**Puis, appliquez les rôles par défaut et configurez l'application :**

superset init

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Étape 5 : Connexion à Superset**

**1. Une fois l'utilisateur créé, accédez à l'interface web de Superset à l'adresse correspondante, par exemple :**

[**http://192.168.49.2:30775/superset/welcome/**](http://192.168.49.2:30775/superset/welcome/)

**ou**

**http://<adresse-ip-du-service-superset>:<port>**

**2. Connectez-vous avec le username et le password que vous avez définis.**

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Étape 6 (Optionnel) : Réinitialiser le mot de passe**

**Si vous oubliez votre mot de passe, vous pouvez le réinitialiser en exécutant la commande suivante dans le pod Superset :**

superset fab reset-password --username <votre-nom-utilisateur>

**Ensuite, entrez un nouveau mot de passe.**

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

* 1. **Utilisation détaillée d'Apache Superset après l'installation**

**Étape 1 : Connexion initiale**

1. **Accéder à Superset** :
   * Accédez à l’URL configurée dans votre ingress, par exemple :

http://192.168.49.2:30775/superset/welcome/

* + Connectez-vous avec les identifiants administrateurs créés lors de l'étape d'initialisation (superset fab create-admin).

1. **Interface utilisateur** :
   * Explorez l’interface, qui est divisée en plusieurs sections : **Dashboards**, **Charts**, **Datasets**, et **Databases**.

**Étape 2 : Configuration d’une connexion à une base de données**

Pour commencer à explorer et visualiser des données, connectez une base de données source.

1. **Aller dans la section Bases de données** :
   * Accédez à **Settings > Database Connections**.
   * Cliquez sur **+ Add Database**.
2. **Configurer la connexion** :
   * Sélectionnez le type de base de données (PostgreSQL, MySQL, Oracle, etc.).
   * Fournissez les détails de connexion :
     + **URI de la base de données** : Par exemple :

postgresql+psycopg2://username:password@hostname:port/database

* + - **Nom de la connexion** : Un nom descriptif de la base de données.

1. **Tester la connexion** :
   * Cliquez sur **Test Connection**.
   * Si la connexion est réussie, sauvegardez.
2. **Configurer des paramètres avancés (optionnel)** :
   * Activez des fonctionnalités comme la mise en cache ou la personnalisation des paramètres de pool de connexions.

**Étape 3 : Création d’un dataset**

Un dataset est une vue ou une table provenant de votre base de données connectée.

1. **Créer un dataset** :
   * Accédez à **Datasets > + Dataset**.
   * Sélectionnez la base de données source et la table ou requête SQL.
2. **Configurer le dataset** :
   * Ajoutez des métadonnées comme des colonnes, des filtres par défaut, ou des labels.
   * Sauvegardez.

**Étape 4 : Création de graphiques (Charts)**

Les graphiques sont créés à partir des datasets.

1. **Créer un graphique** :
   * Allez dans **Charts > + Chart**.
   * Choisissez une source de données (dataset) et un type de visualisation (barres, lignes, cartes, etc.).
2. **Configurer les paramètres** :
   * Définissez les axes, filtres, mesures, et autres paramètres de visualisation.
   * Sauvegardez le graphique.

**Étape 5 : Création de tableaux de bord (Dashboards)**

Les tableaux de bord regroupent plusieurs graphiques pour une vue unifiée.

1. **Créer un tableau de bord** :
   * Accédez à **Dashboards > + Dashboard**.
   * Donnez un nom et ajoutez une description (optionnelle).
2. **Ajouter des graphiques au tableau de bord** :
   * Glissez-déposez des graphiques existants dans le tableau de bord.
   * Organisez les graphiques en fonction de vos besoins.
3. **Configurer et publier** :
   * Ajoutez des filtres globaux, des annotations ou des informations contextuelles.
   * Publiez le tableau de bord pour le rendre accessible aux autres utilisateurs.

**Étape 6 : Gestion des utilisateurs et des permissions**

Superset offre des rôles et permissions pour gérer l'accès.

1. **Créer des utilisateurs** :
   * Utilisez l’interface ou la CLI pour ajouter de nouveaux utilisateurs :

kubectl exec -it <superset-pod-name> -n superset -- superset fab create-user \

--username <username> --firstname <firstname> --lastname <lastname> \

--email <email> --role <role> --password <password>

1. **Configurer les rôles** :
   * Accédez à **Settings > Roles**.
   * Créez ou modifiez des rôles pour restreindre l’accès à certains tableaux de bord, bases de données ou graphiques.

**Étape 7 : Fonctionnalités avancées**

**a) Extensions et plugins**

1. Installez des extensions ou développez des plugins pour ajouter des types de visualisation ou des fonctionnalités.
2. Compilez les plugins et redéployez Superset.

**b) Monitoring et logs**

1. Accédez aux logs via Kubernetes :

kubectl logs -f <superset-pod-name> -n superset

1. Intégrez avec Prometheus ou Grafana pour surveiller les métriques de Superset.

**Étape 8 : Maintenance**

**a) Mise à jour de Superset**

Ouvrez le manifeste du Deployment de Superset (le fichier YAML où vous avez défini le déploiement).

Modifiez la section image pour pointer vers la dernière version de l'image de Superset. Par exemple :

spec:

containers:

- name: superset

image: apache/superset:latest # Mettez à jour cette ligne avec la version souhaitée

Appliquez cette modification à votre cluster Kubernetes :

kubectl apply -f superset.yaml

Cette commande va redéployer Superset avec la nouvelle version de l'image Docker.

 **Vérification après la mise à jour :**

Pour vous assurer que l'application fonctionne correctement après la mise à jour, vous devez vérifier l'état du pod et des logs :

1. **Vérifiez que les pods sont en cours d'exécution :**

kubectl get pods -l app=superset

1. **Vérifiez les logs pour toute erreur potentielle :**

kubectl logs <nom-du-pod-superset>

**b) Sauvegarde des données**

1. Planifiez des sauvegardes de la base de données backend (PostgreSQL/MySQL) à l’aide de CronJobs Kubernetes.
2. Sauvegardez les volumes persistants (PVC) utilisés par Superset.

Ces étapes couvrent l’utilisation courante et avancée de Superset après son installation dans Kubernetes.

**Ressources supplémentaires**

* [Site officiel Apache Superset](https://superset.apache.org/)
* [Documentation GitHub](https://github.com/apache/superset/tree/master/docs)
* [Communauté et support](https://github.com/apache/superset/discussions)