

Rafraîchir les données de développement avec anonymisation dans CloudNativePG

Julien Acroute (*Camp2Camp*) · Alain Lesage (*Dalibo*)

En bref

La mise en production du code est aujourd'hui un processus maîtrisé et bien rodé (Dev → Staging → Prod). Pourtant, dès qu'on parle de **données**, le chemin inverse (Prod → Dev) reste une difficulté à automatiser. Du fait des problématiques de désensibilisation et anonymisation des données, rafraîchir une base de développement avec des données de prod revient encore souvent à un processus artisanal : ouverture de ticket, moulinettes et scripts maison à la documentation variable.

Pour faire mieux que ce processus `dump / restore`, plus ou moins bien anonymisé, cet atelier propose de construire, en direct, un **workflow de rafraîchissement à la demande** reposant sur deux piliers :

- **CloudNativePG (CNPG)** pour la gestion du cycle de vie des instances PostgreSQL dans Kubernetes;
- **postgresql_anonymizer** pour garantir la confidentialité des données sensibles, qui ne doivent pas sortir de production.

À l'issue de la session, vous saurez offrir à vos équipes de développement des environnements **fidèles à la production, anonymisés par design et disponibles en quelques secondes**.

Déroulé

L'atelier est essentiellement basé sur la mise en pratique. Les participants disposent d'un accès à un cluster Kubernetes pré-existant et réalisent eux-mêmes les étapes :

1. Création d'un cluster « prod » avec CloudNativePG;
2. Création d'un second cluster en répllication;
3. Injection de données et définition des règles d'anonymisation;
4. Anonymisation des données sensibles par différentes méthodes;
5. Détachement du second cluster par promotion, pour obtenir un environnement de développement indépendant, anonymisé et prêt à l'emploi.

Animateurs

Julien Acroute — Camp2Camp. Auteur de l'atelier, il accompagne au quotidien la mise en œuvre de PostgreSQL sur des plateformes Kubernetes.

Alain Lesage — Dalibo. Co-animateur, DBA consultant chez Dalibo, société spécialisée dans PostgreSQL, éditrice de PostgreSQL Anonymizer et support de l'opérateur CNPG.

Prérequis

- Niveau **intermédiaire en PostgreSQL** : être à l'aise avec le SQL et l'administration courante, les notions d'instance, schéma et rôle;
- Connaissances concernant la réplication et les sauvegardes (`pg_dump`, `pg_restore`, `pg_basebackup` etc.);
- **Bases de Kubernetes** : comprendre ce qu'est un pod, un cluster, un manifeste YAML, et savoir lancer une commande `kubectl` ;
- Être à l'aise en **lignes de commandes** Linux.

Logistique

- **Votre ordinateur portable.** L'atelier se fait clavier en main.
- Un **client `kubectl`** installé et fonctionnel (les manifestes seront fournis pendant la session).
- Un **terminal** ainsi qu'un **navigateur web**.
- **Le cluster Kubernetes est fourni** : aucun déploiement d'infrastructure n'est à prévoir de votre côté avant l'atelier. Les accès vous seront communiqués sur place.