

Cheatsheet – Configurer un remote DVC sur un stockage cloud

Par Arthur Cartel Foahom Gouabou cartelgouabou.github.io • [LinkedIn](#)

Objectif

Configurer un remote DVC pour externaliser le stockage des fichiers suivis (datasets, modèles), et permettre à tous les collaborateurs de les récupérer automatiquement avec `dvc pull`.

Prérequis

- Un projet DVC initialisé (`dvc init`)
 - Un ou plusieurs fichiers sous suivi DVC (`dvc add`)
 - Un accès configuré au service de stockage cloud (S3, Google Drive, SSH)
 - Le bon extra DVC installé (voir ci-dessous)
-

1. Installer les modules nécessaires

Remote	Installation
Amazon S3	<code>pip install 'dvc[s3]'</code>
Google Drive	<code>pip install 'dvc[gdrive]'</code>
SSH	<code>pip install 'dvc[ssh]'</code>

2. Configurer un remote DVC par type de stockage

A. Amazon S3

Étapes

1. Créer un bucket S3 (ou utiliser un existant)
2. Configurer les credentials AWS :
 - Via `~/.aws/credentials`
 - Ou via variables d'environnement :
 - `AWS_ACCESS_KEY_ID`
 - `AWS_SECRET_ACCESS_KEY`
3. Ajouter et configurer le remote :

```
dvc remote add -d s3_remote s3://nom-du-bucket/chemin/dvc
dvc remote modify s3_remote region eu-west-1
```

```
dvc remote modify s3_remote access_key_id <AWS_ACCESS_KEY_ID>
dvc remote modify s3_remote secret_access_key <AWS_SECRET_ACCESS_KEY>
```

⚠ Ne jamais commit les clés d'accès dans Git.

B. Google Drive

Étapes

1. Créer un dossier dans Google Drive
2. Récupérer l'ID du dossier dans l'URL :
Exemple : <https://drive.google.com/drive/folders/1A2B3C4D5E6F>
→ ID = **1A2B3C4D5E6F**
3. Ajouter le remote :

```
dvc remote add -d gdrive_remote gdrive://1A2B3C4D5E6F
```

- Lors du premier **dvc push**, DVC ouvrira une page web pour autoriser l'accès à ton compte Google via OAuth.

C. SSH (Serveur distant / VPS)

Étapes

1. S'assurer que l'utilisateur distant a un accès SSH au serveur
2. Le chemin doit pointer vers un dossier accessible en lecture/écriture

```
dvc remote add -d ssh_remote ssh://user@serveur.com:/chemin/absolu/dvc_storage
dvc remote modify ssh_remote port 22
dvc remote modify ssh_remote user user
```

- Authentification via mot de passe ou clé SSH (**~/*.ssh/id_rsa***)
- DVC se connectera au serveur distant pour y stocker les fichiers versionnés

3. Pousser les fichiers vers le remote

```
dvc push
```

- Envoie tous les fichiers suivis (présents dans le cache local) vers le remote par défaut
- DVC utilise les hashes stockés dans les fichiers **.dvc** ou **dvc.lock** pour savoir quels fichiers transférer

4. Pour les autres collaborateurs

Cloner le projet et récupérer les données

```
git clone <repo_git>
cd <repo_git>
pip install -r requirements.txt
dvc pull
```

- `dvc pull` télécharge tous les fichiers nécessaires depuis le remote cloud, selon les informations des fichiers `.dvc` et `dvc.lock`
- Aucun fichier lourd ne transite par Git

Bonnes pratiques

Bonne pratique	Action recommandée
Séparer code et données	Stocker les gros fichiers dans le remote DVC
Sécuriser les accès	Utiliser des variables d'environnement pour les clés
Documenter le remote	Committer <code>.dvc/config</code> mais jamais les credentials
Partager proprement	<code>git clone</code> + <code>dvc pull</code> suffit pour tout récupérer

Document rédigé par Arthur Cartel Foahom Gouabou – Formateur & Data Scientist • <https://cartelgouabou.github.io/> • [LinkedIn](#)