

Soutenance Stage U IRIS

Lucas Gelgon - BUT Informatique Février 2024 - Mars 2024





U IRIS

- Gère l'informatique de Système-U
- Fondée en 2011
- 550 collaborateurs
- Située à Carquefou



Présentation de l'entreprise

Présentation de l' équipe

Contexte et sujet du stage

Partie Technique

Documentation et accompagnement

Ligne Data

Trois équipes:

- Décisonnel
- Analyse Data
- Socle Data



Présentation de l'entreprise

Présentation de l'équipe

Contexte et sujet du stage

Partie Technique

Documentation et accompagnement

Contexte

Bascule de l'écosystème Data sur le Cloud

Besoin d'environnements de développement et de recette à l'image de celui de production



Présentation de l'entreprise

Présentation de l' équipe

Contexte et sujet du stage

Partie Technique

Documentation et accompagnement

Contexte et Sujet

Automatisation du rafraîchissement d'Univers

Univers = Données + Traitements

Un **Rafraîchissement Data** : Copie des données depuis l'environnement de production

⇒ vers un-sous environnement de développement ou de recette

Outil de rafraîchissement simple déjà existant

Présentation de l'entreprise

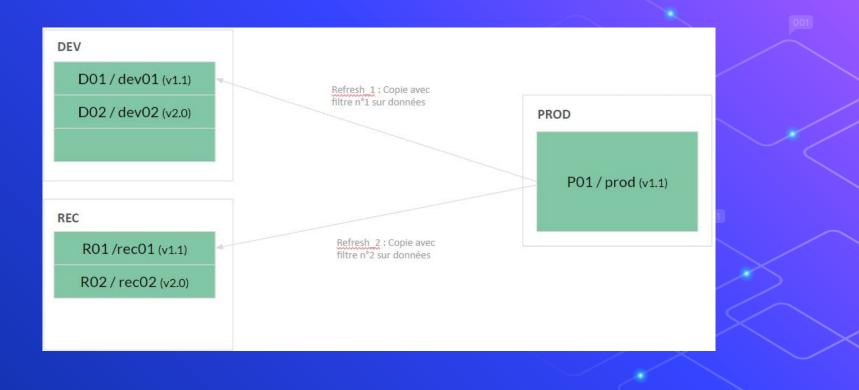
Présentation de l' équipe

Contexte et sujet du stage

Partie Technique

Documentation et accompagnement

Contexte et Sujet

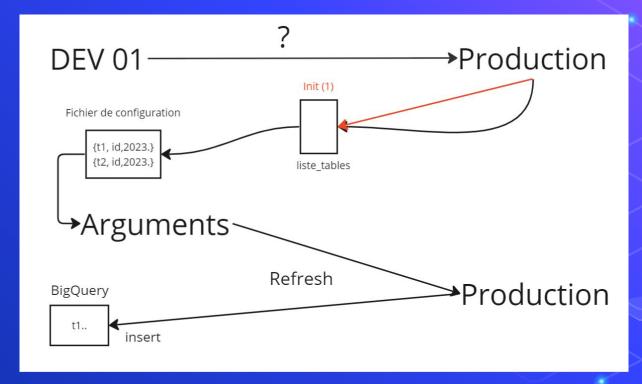


Présentation de l'entreprise

Présentation de l' équipe Contexte et sujet du stage

Partie Technique

Documentation et accompagnement

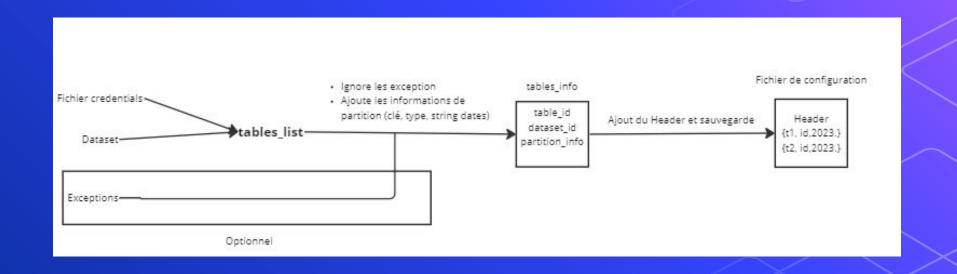


Présentation de l'entreprise

Présentation de l' équipe

Contexte et sujet du stage Partie Technique

Documentation et accompagnement

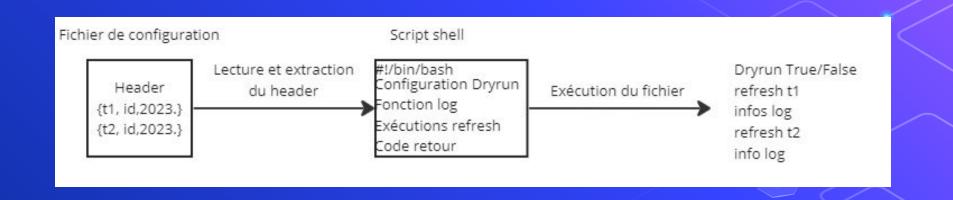


Présentation de l'entreprise

Présentation de l' équipe

Contexte et sujet du stage Partie Technique

Documentation et accompagnement

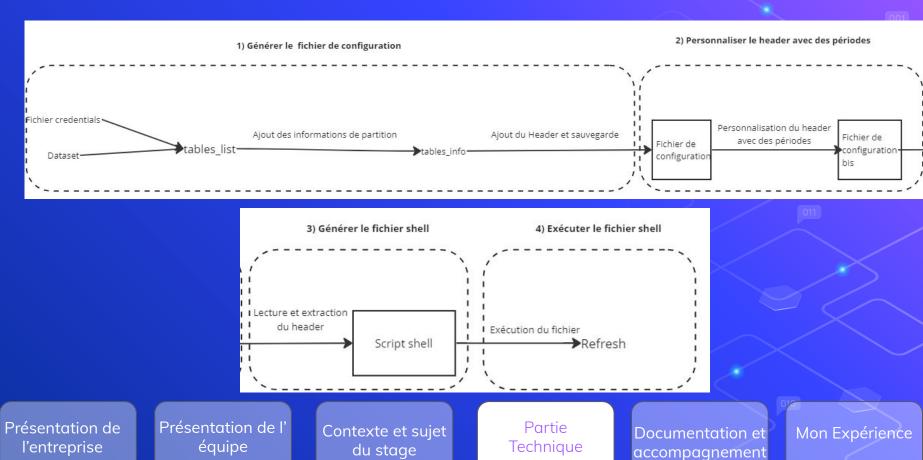


Présentation de l'entreprise

Présentation de l' équipe

Contexte et sujet du stage Partie Technique

Documentation et accompagnement



11

Documentation et Accompagnement

Fichier Backlog pour avoir les retours d'utilisation

А	В	С	D	E	F	G	Н
Qui	Quand	Outil refresh officiellement présenté ?	Cas d'usage / Projet	Fonctionnalité testée	Statut	Amélioration / anomalie rencontrée	Revue PO
Stéphane GAUDIN	16/10/2023	Oui	Validation Socle Data	Refresh des données d'une table SIDE	Remarque	Afficher nombre de lignes répliquées	Oui
						177 111 7	
	ĺ í						

Présentation de l'entreprise

Présentation de l' équipe Contexte et sujet du stage

Partie Technique

Documentation et Accompagnement



Documentation et Accompagnement

Présentation Accompagnement



Présentation de l'entreprise

Présentation de l' équipe

Contexte et sujet du stage

Partie Technique

Documentation et Accompagnement

Mon Expérience

Très bonne expérience



Beaucoup de nouvelles connaissances



Facilité de proposition d'idées



Partage et entraide



Flexibilité



Présentation de l'entreprise

Présentation de l' équipe

Contexte et sujet du stage

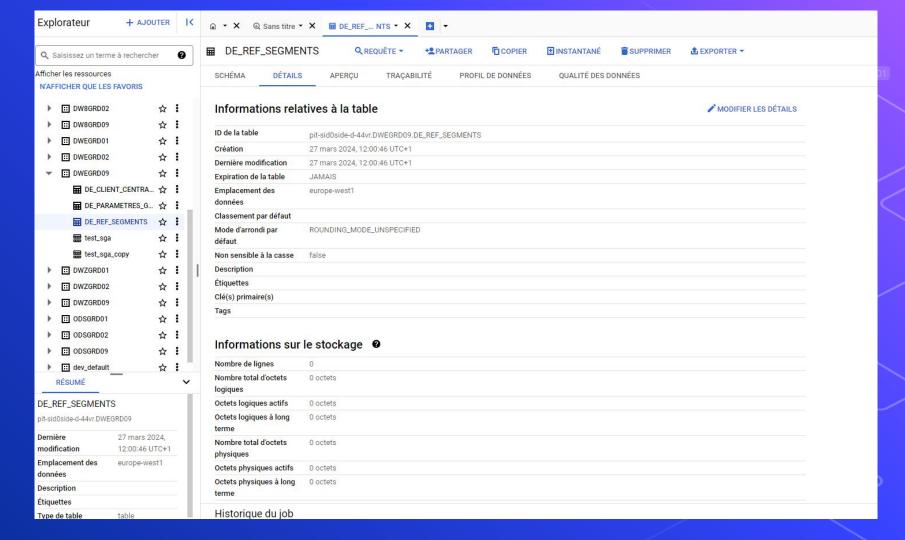
Partie Technique

Documentation et accompagnement

MERCI

CREDITS: This presentation template and the icons were created by SlidesCarnival

```
[lucas gelgon systeme u fr@g0016dedge0 ~]$ sudo su - svc edh deploy-dev
Dernière connexion : mercredi 27 mars 2024 à 10:05:11 CET sur pts/0
Ce système est configuré pour l'usage de 'sudo'. Plus d'infos dans la FAQ:
https://docs.google.com/document/d/1j0hYIvbxZvypGQQNj0HV0JxrTUyKqvx 9-GsqTUizZ4
[svc edh deploy-dev@g0016dedge0 ~]$ cd /u01/app/esg refresh univers
[svc edh deploy-dev@g0016dedge0 esg refresh univers]$ source ./venv/bin/activate
(venv) [svc edh deploy-dev@g0016dedge0 esg refresh univers]$ 11
total 48
lrwxrwxrwx. 1 svc edh deploy-dev utilisateurs du domaine
                                                        67 Mar 15 14:58 autoRefresh.py -> /u01/app/edh/edh-current/tools/esg/esg refreshSubEnv/autoRefresh.py
lrwxrwxrwx. 1 svc edh deploy-dev utilisateurs du domaine
                                                        79 Mar 15 14:59 createRefreshConfigFile.py -> /u01/app/edh/edh-current/tools/esg/esg refreshSubEnv/createRefreshConfigFile.py
drwxr-xr-x. 3 svc edh deploy-dev gl app edh admin
                                                     20480 Mar 27 10:44 logs
drwxr-xr-x. 2 svc edh deploy-dev gl app edh admin
                                                      4096 Mar 11 16:16 outputs
lrwxrwxrwx. 1 svc edh_deploy-dev_utilisateurs_du_domaine 77 Mar 25 09:58 prepareRefreshUnivers -> /u01/app/edh/edh-current/tools/esg/esg_refreshSubEnv/prepareRefreshUnivers.py
lrwxrwxrwx. 1 svc edh deploy-dev gl app edh admin
                                                        62 Feb 12 14:47 README.md -> /u01/app/edh/edh-current/tools/esg/esg refreshSubEnv/README.md
lrwxrwxrwx. 1 svc edh deploy-dev gl app edh admin
                                                        69 Feb 12 14:47 refreshSubEnv.py -> /u01/app/edh/edh-current/tools/esg/esg refreshSubEnv/refreshSubEnv.py
drwxr-xr-x. 5 svc edh deploy-dev utilisateurs du domaine 4096 Feb 21 14:51 venv
(venv) [svc edh deploy-dev@g0016dedge0 esg refresh univers]$ python ./refreshSubEnv.py --project side --target env dev --subenv 09 --schema only True --datasets DWEGRP01 --tables DE REF SEGMENTS
2024-03-27 11:00:14,457 - modules.arg parser - INFO - Log level: INFO
2024-03-27 11:00:14,457 - modules.arg parser - INFO - Target project: dev
2024-03-27 11:00:14,457 - modules.arg parser - INFO - List of tables to refresh: DE REF SEGMENTS
2024-03-27 11:00:14,457 - modules.arg parser - INFO - Starting partition date: 2024-02-26
2024-03-27 11:00:14,457 - modules.arg parser - INFO - Ending partition date: 2024-03-27
2024-03-27 11:00:14,458 - modules.arg parser - INFO - First import of data into a table: False
2024-03-27 11:00:14.458 - root - INFO - Refreshing project side in dev environment
2024-03-27 11:00:14,850 - modules.repository - INFO - Retrieving tables for dataset DWEGRP01
```



Informations sur le stockage @

Nombre de lignes 63

Nombre total d'octets 5,7 Ko
logiques

Octets logiques actifs 5,7 Ko

⊞ DE	_REF_SEGM	ENTS QREQUÊTE ▼	+⊈PARTAGER	R 🖢 INSTANTANÉ 😈 SU	PPRIMER				CACTUAL						
SCHÉN	SCHÉMA DÉTAILS APERÇU TRAÇABILITÉ PROFIL DE DONNÉES QUALITÉ DES DONNÉES														
Ligne	ID_SEGMENT	CD_SEGMENT	LB_SEGMENT	CD_SEGMENT_PARENT	LB_SEGMENT_PARENT	NR_SEQUENCE_	CD_SEGMENTATION	ID_SEGMENT_P/	NR_SEQUENCE_						
1	68	00	Non calculé	00	Non calculé	9	SGSDC	68	9						
2	11	00	Non calculé	00	Non calculé	12	SGCDV	11	12						
3	31	00	Non calculé	00	Non Calculé	12	SGSPC	31	6						
4	53	00	Non calculé	00	Non calculé	8	SGVCA	53	6						
5	1	00	Non calculé	00	Non calculé	10	SGRFM	1	10						
6	13	01	Jeune célibataire	01	Jeune célibataire	1	SGCDV	13	1						
7	77	01	Gros	01	Gros	1	SGVAL	77	1						
8	2	01	Fidèle	01	Fidèle	1	SGRFM	2	1						
9	60	01	Les Standards	01	Les Standards	7	SGSDC	60	7						

```
(venv) [svc_edh_deploy-dev@g0016dedge0 esg_refresh_univers]$ python createRefreshConfigFile.py --dataset 'DWEGRP01' --credpath '/u01/app/gcp/p/sid0/side/conf/sa/sa-replication.json' --repertoire_config 'config' --fichier_config file 'config_file'
2024-03-27 12:03:39,432 createRefreshConfigFile.py INFO Lecture GCP credential
2024-03-27 12:03:39,433 createRefreshConfigFile.py INFO Récupération des tables
2024-03-27 12:03:41,537 createRefreshConfigFile.py INFO Obtention des informations sur le partitionnement pour chaque table
2024-03-27 12:03:41,540 createRefreshConfigFile.py INFO Création du fichier configuration
2024-03-27 12:03:41,540 createRefreshConfigFile.py INFO Création du répertoire configuration
2024-03-27 12:03:41,540 createRefreshConfigFile.py INFO Création du fichier configuration
2024-03-27 12:03:41,540 createRefreshConfigFile.py INFO Comparaison des fichiers de production et de configuration
2024-03-27 12:03:41,542 createRefreshConfigFile.py INFO Comparaison des fichiers de production et de configuration
```

```
urwartarta. J syc_eun depioytuev ucilisaceurs du domaine (4050 feb 21 14.31)
(venv) [svc edh deploy-dev@g0016dedge0 esg refresh univers]$ cat config/config file
---HEADER---
debut DAY=
fin DAY=
debut NUM=
fin NUM=
{'table id': 'DE AFFINITES CLIENTS', 'dataset id': 'DWEGRP01', 'partition info': {'partitioned': False}}
 'table id': 'DE AFFINITES CLIENTS H0', 'dataset id': 'DWEGRP01', 'partition info': {'partitioned': False}}
 'table id': 'DE AFFINITES PRODUITS', 'dataset id': 'DWEGRP01', 'partition info': {'partitioned': False}}
 'table id': 'DE AGGA PDV EAN FILIERE FIM', 'dataset id': 'DWEGRP01', 'partition info': {'partitioned': False}}
 'table_id': 'DE_AGGA_PDV_EAN_FIL_PROD_FIM', 'dataset_id': 'DWEGRP01', 'partition_info': {'partitioned': False}}
 'table id': 'DE AGGA PDV EAN FIM', 'dataset id': 'DWEGRP01', 'partition info': { partitioned': False}}
 'table id': 'DE AGGA PDV EAN SECTEUR FIM', 'dataset id': 'DWEGRP01', 'partition info': {'partitioned': False}}
 'table id': 'DE AGGA STATION CA SEC FIM', 'dataset id': 'DWEGRP01', 'partition info': { 'partitioned': False}}
 'table id': 'DE AGGA STATION VO FIM', 'dataset id': 'DWEGRP01', 'partition info': {'partitioned': False}}
 'table_id': 'DE_AGGJ_MARGE_THEO_PDV_EAN', 'dataset_id': 'DWEGRP01', 'partition_info': {'partition_ed': True, 'partition_key': 'DT_VENTE', 'partition_type': 'DAY', 'date_debut': 'debut_DAY', 'date_fin': 'fin_DAY'}}
 'table id': 'DE AGGJ MARGE THEO PDV EAN 3ANS', 'dataset id': 'DWEGRP01', 'partition info': {'partitioned': False}}
 'table id': 'DE AGGJ PDV', 'dataset id': 'DWEGRP01', 'partition_info': {'partitioned': True, 'partition_key': 'DT_EMISSION_TICKET', 'partition_type': 'DAY', 'date_debut': 'debut_DAY', 'date_fin': 'fin_DAY'}}
 'table id': 'DE AGGJ PDV CAISSE', 'dataset id': 'DWEGRP01', 'partition info': {'partitioned': True, 'partition key': 'DT EMISSION TICKET', 'partition type': 'DAY', 'date debut': 'debut DAY', 'date fin': 'fin DAY'}}
 'table id': 'DE AGGJ PDV CLIENT', 'dataset id': 'DMEGRP01', 'partition_info': {'partitioned': True, 'partition_key': 'DT_EMISSION_TICKET', 'partition_type': 'DAY', 'date_debut': 'debut_DAY', 'date_fin': 'fin_DAY'}}
 table id': 'DE AGGJ PDV CLIENT TEMP', 'dataset id': 'DWEGRP01', 'partition info': {'partitioned': False}'
 'table id': 'DE AGGJ PDV EAN', 'dataset id': 'DNEGRP01', 'partition info': {'partitioned': True, 'partition key': 'DT EMISSION TICKET', 'partition type': 'DAY', 'date debut': 'debut DAY', 'date fin': 'fin DAY'}}
('table id': 'DE AGGJ PDV EAN 3ANS', 'dataset id': 'DWEGRP01', 'partition info': {'partitioned': False}}
 'table id': 'DE AGGJ PDV EAN CLIENT', 'dataset id': 'DWEGRP01', 'partition info': {'partitioned': True, 'partition key': 'DT EMISSION TICKET', 'partition type': 'DAY', 'date debut': 'debut DAY', 'date fin': 'fin DAY'}}
 'table id': 'DE AGGJ PDV EAN CLIENT 3ANS', 'dataset id': 'DWEGRP01', 'partition info': {'partitioned': False}}
 'table_id': 'DE_AGGJ_TICKET_FLAG', 'dataset_id': 'DWEGRP01', 'partition_info': {'partitioned': True, 'partition_key': 'DT_EMISSION_TICKET', 'partition_type': 'DAY', 'date_debut': 'debut_DAY', 'date_fin': 'fin_DAY'}}
 'table id': 'DE AGGJ TICKET FLAG 3ANS', 'dataset id': 'DWEGRP01', 'partition info': {'partitioned': False}}
 'table id': 'DE AGGM CEN NOM 01'. 'dataset id': 'DWEGRP01'. 'partition info': {'partitioned': False}
 'table id': 'DE AGGM CEN NOM 02', 'dataset id': 'DWEGRP01', 'partition info': { 'partitioned': False}
 'table id': 'DE AGGM CEN NOM 03', 'dataset id': 'DWEGRP01', 'partition info': {'partitioned': False}
 'table id': 'DE AGGM CEN NOM 04', 'dataset id': 'DWEGRP01', 'partition info': {'partitioned': False}}
 'table_id': 'DE_AGGM_CEN_NOM_05', 'dataset_id': 'DWEGRP01', 'partition_info': {'partitioned': False}}
 'table id': 'DE AGGM CEN NOM 06', 'dataset id': 'DWEGRP01', 'partition info': {'partitioned': False}}
 'table id': 'DE AGGM CEN NOM 07', 'dataset id': 'DWEGRP01', 'partition info': {'partitioned': False}}
 table_id': 'DE_AGGM_CEN_NOM_08_EVO12', 'dataset_id': 'DWEGRP01', 'partition_info': {'partitioned': False}'
 'table id': 'DE AGGM CEN NOM 08 EVO12PC', 'dataset id': 'DWEGRP01', 'partition info': {'partitioned': False}}
 'table id': 'DE AGGM CEN NOM 08 VAL', 'dataset id': 'DWEGRP01', 'partition info': {'partitioned': False}}
```

vim config/config_file

```
---HEADER---
debut_DAY=2024-01-01
fin_DAY=2024-01-31
debut_NUM=202401
fin_NUM=20240
```

```
(venv) [svc_edn_deploy-dev@g0016dedge0 esg_refresh_univers]$ vim config_file (venv) [svc_edn_deploy-dev@g0016dedge0 esg_refresh_univers]$ python autoRefresh.py --project 'side' --subenv '09' --target_env 'dev' --emplacement_config 'config_file' --repertoire_bash 'shell_scripts' 2024-03-27 12:06:32,350 autoRefresh.py INFO Lecture du fichier de configuration 2024-03-27 12:06:32,351 autoRefresh.py INFO Extraction des valeurs du header 2024-03-27 12:06:32,351 autoRefresh.py INFO Création et remplissage du script bash 2024-03-27 12:06:32,351 autoRefresh.py INFO Création du repertoire Bash (venv) [svc_edh_deploy-dev@g0016dedge0 esg_refresh_univers]$
```

```
# Teste si mode dryrun est actif ou non : si oui => affichage seulement de la commande sans exécuter le refresh
if [ "$vDryRun" = "True" ]; then
   echo "python3 ./refreshSubEnv.py --project side --subenv 09 --target env dev --datasets DWEGRP01 --tables test sga copy --partition date start 2024-01-01 --partition date end 2024-01-31 "
   python3 ./refreshSubEnv.py --project side --subenv 09 --target env dev --datasets DWEGRP01 --tables test sga copy --partition date start 2024-01-01 --partition date end 2024-01-31
   if [ $? -ne 0 ]; then
       log "ERR" "Erreur de connexion au projet side, au sous-environnement 09 ou à l'environnement cible dev"
       error=$((error + 1))
   else
       log "INFO" "Connexion au projet side, le sous-environnement est 09 et l'environnement cible est dev"
if [ $error -eq 0 ]; then
   # Code retour signifiant que tous les rafraîchissements ont fonctionné
   exit 0
else
    # Code retour signifiant la présence d'erreurs dans un ou plusieurs rafraîchissements
   echo "Nombre d'erreurs : $error"
   exit 3
```

```
(venv) [svc edh deploy-dev@g0016dedge0 esg refresh univers]$ . ./shell scripts/refresh DWEGRP01 202401 202401.sh
Mode Dryrun activé : Test à blanc, affiche uniquement les informations
python3 ./refreshSubEnv.py --project side --subenv 09 --target env dev --datasets DWEGRP01 --tables DE AFFINITES CLIENTS
python3 ./refreshSubEnv.py --project side --subeny 09 --target env dev --datasets DWEGRP01 --tables DE AFFINITES CLIENTS H0
python3 ./refreshSubEnv.py --project side --subenv 09 --target env dev --datasets DWEGRP01 --tables DE AFFINITES PRODUITS
python3 ./refreshSubEnv.py --project side --subenv 09 --target env dev --datasets DWEGRP01 --tables DE AGGA PDV EAN FILIERE FIM
python3 ./refreshSubEnv.py --project side --subenv 09 --target env dev --datasets DWEGRP01 --tables DE AGGA PDV EAN FIL PROD FIM
python3 ./refreshSubEnv.py --project side --subenv 09 --target env dev --datasets DWEGRP01 --tables DE AGGA PDV EAN FIM
python3 ./refreshSubEnv.py --project side --subenv 09 --target env dev --datasets DWEGRP01 --tables DE AGGA PDV EAN SECTEUR FIM
python3 ./refreshSubEnv.py --project side --subenv 09 --target env dev --datasets DWEGRP01 --tables DE AGGA STATION CA SEC FIM
python3 ./refreshSubEnv.py --project side --subenv 09 --target env dev --datasets DWEGRP01 --tables DE AGGA STATION VO FIM
python3 ./refreshSubEnv.py --project side --subeny 09 --target eny dey --datasets DWEGRP01 --tables DE AGGJ MARGE THEO PDV EAN --partition date start 2024-01-01 --partition date end 2024-01-31
python3 ./refreshSubEnv.py --project side --subenv 09 --target env dev --datasets DWEGRP01 --tables DE AGGJ MARGE THEO PDV EAN 3ANS
python3 ./refreshSubEnv.py --project side --subenv 09 --target env dev --datasets DWEGRP01 --tables DE AGGJ PDV --partition date start 2024-01-01 --partition date end 2024-01-31
python3 ./refreshSubEnv.py --project side --subenv 09 --target env dev --datasets DWEGRP01 --tables DE AGGJ PDV CAISSE --partition date start 2024-01-01 --partition date end 2024-01-31
python3 ./refreshSubEnv.py --project side --subenv 09 --target env dev --datasets DWEGRP01 --tables DE AGGJ PDV CLIENT --partition date start 2024-01-01 --partition date end 2024-01-31
python3 ./refreshSubEnv.py --project side --subenv 09 --target env dev --datasets DWEGRP01 --tables DE AGGJ PDV CLIENT TEMP
python3 ./refreshSubEnv.py --project side --subenv 09 --target env dev --datasets DWEGRP01 --tables DE AGGJ PDV EAN --partition date start 2024-01-01 --partition date end 2024-01-31
python3 ./refreshSubEnv.py --project side --subenv 09 --target env dev --datasets DWEGRP01 --tables DE AGGJ PDV EAN 3ANS
python3 ./refreshSubEnv.py --project side --subeny 09 --target eny dey --datasets DWEGRP01 --tables DE AGGJ PDV EAN CLIENT --partition date start 2024-01-01 --partition date end 2024-01-31
python3 ./refreshSubEnv.py --project side --subenv 09 --target env dev --datasets DWEGRP01 --tables DE AGGJ PDV EAN CLIENT 3ANS
python3 ./refreshSubEnv.py --project side --subenv 09 --target env dev --datasets DWEGRP01 --tables DE AGGJ TICKET FLAG --partition date start 2024-01-01 --partition date end 2024-01-31
python3 ./refreshSubEnv.py --project side --subenv 09 --target_env dev --datasets DWEGRP01 --tables DE_AGGJ_TICKET_FLAG_3ANS
python3 ./refreshSubEnv.py --project side --subenv 09 --target env dev --datasets DWEGRP01 --tables DE AGGM CEN NOM 01
python3 ./refreshSubEnv.py --project side --subenv 09 --target env dev --datasets DWEGRP01 --tables DE AGGM CEN NOM 02
python3 ./refreshSubEnv.py --project side --subenv 09 --target env dev --datasets DWEGRP01 --tables DE AGGM CEN NOM 03
python3 ./refreshSubEnv.py --project side --subenv 09 --target env dev --datasets DWEGRP01 --tables DE AGGM CEN NOM 04
python3 ./refreshSubEnv.py --project side --subenv 09 --target env dev --datasets DWEGRP01 --tables DE AGGM CEN NOM 05
```