Import des données SINP-PACA pour GeoNature et l'Atlas

Ré-initialisation de la base de données GeoNature



Cette manipulation est dangereuse car elle supprime tout le contenu de la base de données de GeoNature !

- Suivre les étapes de mise à jour d'une installation de GeoNature, au niveau de l'intervention sur "*db-srv*" suivre ceci:
  - Sur l'instance "db-srv" connecté en tant que "geonat", suivre les étapes suivantes
  - Mettre à `true` le paramètre `drop_apps_db` du fichier `~/geonature/config/settings.ini`.
  - Se placer dans le dossier `~/geonature/install/` avec la commande : `cd ~/geonature/install/`
  - Lancer une session détachable car la manipulation peut être longue et une perte de connexion SSH possible : `screen -S install-db`
  - Lancer le script d'installation de la base de données de GeoNature : `./install_db.sh`
  - Voir ci-dessous les commandes à utiliser pour quitter puis se reconnecter à une session "Screen"
  - Vérifier l'absence d'erreurs de l'installation dans le fichier '' ~/geonature/var/log/install_db.log ''
  - **ATTENTION** : remettre à `false` le paramètre `drop_apps_db` du fichier `~/geonature/config/settings.ini` !

Récupération en local du dépôt "data" et transfert sur le serveur

- En local, sur votre machine, récupérer le dépôt Github "sinp-paca-data" : `git clone git@github.com:cbn-alpin/sinp-paca-data.git`
- Suivre les éventuelles étapes indiquées dans les fichiers README.md (CBNA) et README.md (CEN-PACA)
- **ATTENTION** : sur le serveur, sauvegarder le dossier "data" : voir ci-dessous
- Se placer à la racine du dossier *sinp-paca-data*
- Transférer les scripts : `rsync -av --copy-unsafe-links --exclude var --exclude .git --exclude .editorconfig --exclude .vscode --exclude .gitignore --exclude settings.ini --exclude "data/raw/*" --exclude venv --exclude .venv --exclude "import-parser/data/*" ./ geonat@db-paca-sinp:~/data/ --dry-run`
  - Supprimer l'option `--dry-run` si tout semble ok pour effectuer le transfert réel
- Se connecter au serveur
  - Créer les fichiers *settings.ini* à partir des fichiers *settings.sample.ini* pour : *area*, *cbna-cbnmed*, *cenpaca* et *shared*

    - Préparer l'environnement du script *import-parser* en suivant les indications du fichier README.md (import-parser)
        - **Notes** : il est nécessaire de redonner les droits d'execution à GCC pour tout le monde si l'on veut pouvoir installer correctement le venv avec `sudo chmod o+x /usr/bin/gcc`. Une fois l'installation terminée, retirer les à nouveau avec '' chmod o-x /usr/bin/gcc ''.
- **Notes** : les données brutes nécessaires aux scripts sont automatiquement téléchargées depuis Dropbox

Sauvegarde du dossier "data"

- Se connecter à l'utilisateur "admin" sur le serveur : `ssh geonat@db-paca-sinp`
- Créer le dossier de sauvegarde : '' mkdir -p backups/data ''
- Sauvegarder le dossier "data" : '' mv data backups/data/$(date +%F) ''
- Recréer le dossier "data" vide : '' mkdir data ''

Commande Screen

- Le script d'import peut prendre plusieurs heures pour s'éxecuter. Afin de pouvoir revenir sur l'état du script, nous allons utiliser la commande `screen`. Elle permet de créer un session qui peut être récupéré à tout moment.
- Principe générale d'utilisation de `Screen` :
    1. Pour lancer une nouvelle session : `screen -S <nom-de-la-session>`
    2. Lancer la commande/script désiré
    3. Utiliser les séquences de touches suivante pour se détacher de la session : `Ctrl-a + Ctrl-d`
    4. Pour voir la liste des sessions en cours : `screen -ls`
        1. Le numéro de session correspond au nombre présent devant ".pts-0"
    5. Pour se connecter à nouveau à une session : `screen -r <numero-session>`
- Autres commandes :
    - Détacher une session toujours attaché pour s'y attacher à nouveau : `screen -D <numero-session>.<nom-session>`
    - Quitter et supprimer une session en cours : `CTRL+a` et `:quit`

Ajout/Correction du référentiel géographique dans la base de données GeoNature

Il est nécessaire d'ajouter dans le référentiel géographique de GeoNature le contour du territoire du SINP. Pour ce faire, nous allons utiliser le script `area` du dépôt "data" :

- Se connecter au serveur "db" en tant qu'admin : `ssh geonat@db-paca-sinp`
- Se placer dans le dossier // ~/data/area/ // de l'utilisateur *geonat* : `cd ~/data/area`
- Créer le fichier de configuration : `cp config/settings.sample.ini config/settings.ini`
- Modifier le fichier *settings.ini* en fonction de l'installation locale : `vi config/settings.ini`
- Lancer une session détachable : `screen -S import-area`

- Lancer le script : ''cd bin/ ; ./initialize.sh -v ''
- Voir ci-dessus les commandes à utiliser pour quitter puis se reconnecter à une session
- Vérifier l'absence d'erreurs dans la session et vérifier le contenu de la base de données

Import des données "cen-paca" au format d'échange dans GeoNature module Synthèse

- Se connecter au serveur "db" en tant qu'admin : `ssh geonat@db-paca-sinp`
- Se placer dans le dossier // ~/data/cenpaca/ // de l'utilisateur *geonat* : `cd ~/data/cenpaca`
- Créer le fichier de configuration : `cp config/settings.sample.ini config/settings.ini`
- Modifier le fichier *settings.ini* en fonction de l'installation locale : `vi config/settings.ini`
- Lancer une session détachable : `screen -S import-cenpaca`
- Lancer le script : ''cd bin/ ; ./import_initial.sh -v ''
- Voir ci-dessus les commandes à utiliser pour quitter puis se reconnecter à une session
- Vérifier l'absence d'erreurs dans la session et vérifier le contenu de la base de données

Import des données "cbna-cbnmed" au format d'échange dans GeoNature module Synthèse

- Se connecter au serveur "db" en tant qu'admin : `ssh geonat@db-paca-sinp`
- Se placer dans le dossier // ~/data/cbna-cbnmed/ // de l'utilisateur *geonat* : `cd ~/data/cbna-cbnmed`
- Créer le fichier de configuration : `cp config/settings.sample.ini config/settings.ini`
- Modifier le fichier *settings.ini* en fonction de l'installation locale : `vi config/settings.ini`
- Lancer une session détachable : `screen -S import-cbna-cbnmed`
- Lancer le script : `cd bin/ ; ./import_initial.sh -v '' * Voir ci-dessus les commandes à utiliser pour quitter puis se reconnecter à une session * Vérifier l'absence d'erreurs dans la session et vérifier le contenu de la base de données ===== Mise à jour des données GeoNature Atlas ===== *` **`Ressources :`** https://github.com/PnX-SI/GeoNature-atlas/blob/master/docs/vues_materialisees_maj.rst `* Au préalable, ajouter une authentification HTTP au site nature.silene.eu, le temps des tests avant mise en prod. * Se connecter au serveur : ''ssh admin@<sinp-db>`
- Lancer une session détachable : `screen -S update-atlas`
- Lancer la commande : `psql`
  - Sélectionner la base de l'atlas : `\c gnatlas`
  - Exécuter les requête suivante suivant le besoin
    - Pour mettre à jour uniquement les données d'observation (table *synthese*) : SELECT `atlas.refresh_materialized_view_data() ;`
    - Si le territoire du SINP à changé et qu'il faut remettre à jour le référentiel géographique : '' SELECT atlas.refresh_materialized_view_ref_geo() ''

Exécution du script d'import des images de l'INPN

- Se connecter sur le serveur : `ssh geonat@sinp-<region>-web`
- Se placer dans le dossier du script dans TaxHub : `cd ~/www/taxhub/data/scripts/import_inpn_media '' * Créer un environnement virtuel : '' python3 -m venv venv '' * Activer l'environnement virtuel : '' source venv/bin/activate '' * Redonner les droits d'exécution à GCC pour tout le monde si l'on veut pouvoir installer correctement les paquets Python dans le venv : '' sudo chmod o+x /usr/bin/gcc '' * Installer les paquets suivant : '' pip install psycopg2 requests '' * Retirer les droits d'exécution à GCC pour tout le monde : '' sudo chmod o-x /usr/bin/gcc ''. * Créer le fichier de configuration : '' cp config.py.sample config.py '' * Modifier les paramètres : * '' SQLALCHEMY_DATABASE_URI = "postgresql://geonatadmin:<mot-de-passe>@10.0.1.10:5432/geonature2db"`
    - `QUERY_SELECT_CDREF = """SELECT DISTINCT cd_ref FROM taxonomie.bib_noms ORDER BY cd_ref LIMIT 100"""` : supprimer le '' LIMIT 100 '' une fois un premier test effectué
- Lancer le script : '' python import_inpn_media.py ''
- Si tout c'est bien passé, désactiver l'environnement virtuel : '' deactivate ''
- Le script ajoute les photos en tant que "secondaire" pour désigner celle avec le plus petit id_media (= au hasard) comme "principale", utiliser la requête :

```
WITH first_media AS (
    SELECT MIN(id_media) AS first_id_media_founded, cd_ref
    FROM taxonomie.t_medias
    GROUP BY cd_ref
)
UPDATE taxonomie.t_medias AS tm
    SET id_type = 1
    FROM first_media AS fm
    WHERE tm.id_media = fm.first_id_media_founded
        AND tm.cd_ref = fm.cd_ref ;
```

- Pour afficher les images sur l'Atlas, il est nécessaire de rafraichir les données des vues matérialisées *atlas.vm_medias* et *atlas.vm_taxons_plus_observes* :

```
REFRESH MATERIALIZED VIEW atlas.vm_medias WITH DATA ;
REFRESH MATERIALIZED VIEW atlas.vm_taxons_plus_observes WITH DATA ;
```

From:
https://sinp-wiki.cbn-alpin.fr/ - **CBNA SINP**

Permanent link:
https://sinp-wiki.cbn-alpin.fr/serveurs/sinp-paca/import-donnees?rev=1614621930

Last update: **2021/03/01 18:05**