

Passerelle - Development #67375

BAN: reprendre l'optimisation de l'import, qu'on a dû revert

13 juillet 2022 21:03 - Lauréline Guérin

Statut:	Rejeté	Début:	13 juillet 2022
Priorité:	Normal	Echéance:	
Assigné à:		% réalisé:	0%
Catégorie:		Temps estimé:	0:00 heure
Version cible:		Planning:	Non
Patch proposed:	Non		
Description			
cf #66342 + #66572			
Reprendre #65843 , #66118 , ajouter des tests, corriger le bug #66342			
Demandes liées:			
Lié à Passerelle - Bug #66342: base adresse, le cron qui met à jour les rues ...		Fermé	16 juin 2022
Lié à Passerelle - Bug #66572: IntegrityError: ERREUR: la valeur d'une clé d...		Fermé	24 juin 2022
Lié à Passerelle - Bug #65843: BAN : l'import des rues est trop gourmand en SQL		Fermé	01 juin 2022
Lié à Passerelle - Bug #66118: TypeError: can't subtract offset-naive and off...		Fermé	10 juin 2022

Historique

#1 - 13 juillet 2022 21:03 - Lauréline Guérin

- Lié à Bug #66342: base adresse, le cron qui met à jour les rues ne tourne plus ? ajouté

#2 - 13 juillet 2022 21:03 - Lauréline Guérin

- Lié à Bug #66572: IntegrityError: ERREUR: la valeur d'une clé dupliquée rompt la contrainte unique « base_adresse_streetmodel_reso... ajouté

#3 - 13 juillet 2022 21:03 - Lauréline Guérin

- Lié à Bug #65843: BAN : l'import des rues est trop gourmand en SQL ajouté

#4 - 13 juillet 2022 21:03 - Lauréline Guérin

- Lié à Bug #66118: TypeError: can't subtract offset-naive and offset-aware datetimes ajouté

#5 - 13 juillet 2022 21:38 - Lauréline Guérin

- Statut changé de Nouveau à Rejeté

- Assigné à Lauréline Guérin supprimé

ou pas, ça a été complètement revu dans [#66625](#)

#6 - 14 juillet 2022 09:48 - Benjamin Dauvergne

Ou pas.

Je me suis attelé à rendre carré au propre et optimiser la partie département/région/ville, mais il reste la partie rue qui n'est pas optimale.

Actuellement on utilise citycode, name comme clé alors que ce n'est pas une clé déclarée du modèle et qu'effectivement on a des doublons sur name/ban_id qu'on a résolu en enlevant l'unicité sur ban_id, pour moi on aurait du la remettre sur name/ban_id. Mais il y a un truc à creuser, la clé ne peut pas être name ou inclure name, ce sont possiblement des problèmes temporaire ou la base n'est pas correcte. Pour vérifier il faudrait télécharger tous les fichiers ban et voir combien il y a de doublon name (tronqué à 150 caractères, mais je mettrai bien name en TextField pour virer ça)/ban_id (ou trouver une instance qui charge tout et regarder là), ou alors si le doublon est au niveau de la paire name/citycode (et donc créer une clé citycode/ban_id dans ce cas).

Au niveau optimisation le update_or_create individuel sur les rues est un peu violent et serait bien amélioré par une récupération par lot des rues avant comparaison et éventuellement update par lot.

À long terme la bonne solution c'est de déprécier ce connecter et d'avoir un service unique à nous ban.entrouvert.com, ne plus dépendre de la BAN pour avoir un uptime meilleur et arrêter de recopier les mêmes données partout et de générer du bloat dans les backups et les WAL.