

Hobo - Development #81814

migrate_schemas: surcharger MigrationRecorder.has_table pour accélérer les migrations

29 septembre 2023 16:27 - Pierre Ducroquet

Statut:	En cours	Début:	29 septembre 2023
Priorité:	Normal	Echéance:	
Assigné à:	Pierre Ducroquet	% réalisé:	0%
Catégorie:		Temps estimé:	0:00 heure
Version cible:		Planning:	Non
Patch proposé:	Non		
Description			
Il apparait, en lisant le code du MigrationRecorder de Django, que la fonction MigrationRecorder.has_table appelle systématiquement la méthode introspection.table_names. Cette méthode fait un appel à la base pour lister toutes les tables, ce qui devient coûteux avec notre organisation en multi-tenants (environ 20ms par appel). https://github.com/django/django/blob/stable/4.2.x/django/db/migrations/recorder.py#L55 Étant donné que l'on vérifie au préalable l'existence de la table django_migrations, on pourrait tout simplement éliminer cette fonction pour obtenir un gain significatif à l'exécution des migrations.			

Historique

#1 - 29 septembre 2023 16:30 - Robot Gitea

- Tracker changé de Support à Development
- Statut changé de Nouveau à En cours

Pierre Ducroquet (pducroquet) a ouvert une pull request sur Gitea concernant cette demande :

- URL : <https://git.entrouvert.org/entrouvert/hobo/pulls/78>
- Titre : WIP: multitenant: hook MigrationRecorder::has_table for a speedup (#81814)
- Modifications : <https://git.entrouvert.org/entrouvert/hobo/pulls/78/files>

#2 - 29 septembre 2023 17:00 - Robot Gitea

- Statut changé de En cours à Solution proposée

#3 - 02 octobre 2023 14:12 - Thomas Noël

J'ai testé, avant et après le patch, une migration où je n'ai changé que la longueur d'un champ inutilisé :

```
--- a/passerelle/apps/soap/models.py
+++ b/passerelle/apps/soap/models.py
@@ -35,7 +35,7 @@ from passerelle.utils.jsonresponse import APIError

class SOAPConnector(BaseResource, HTTPResource):
    wsdl_url = models.URLField(
-        max_length=400, verbose_name=_('WSDL URL'), help_text=_('URL of the WSDL file')
+        max_length=401, verbose_name=_('WSDL URL'), help_text=_('URL of the WSDL file')
    )
    zeep_strict = models.BooleanField(default=False, verbose_name=_('Be strict with returned XML'))
    zeep_xsd_ignore_sequence_order = models.BooleanField(
```

Lancement de migrate_schemas sur 50 tenants :

- sans le patch : 1m32,351s
- avec le patch : 1m30,653s

Je ne sais pas trop calculer la part de risque du patch par rapport à ce gain de 2% qui ne va pas changer grand chose en pratique (attendre 19 minutes contre 20 actuellement).

J'ai envie d'être conservateur...

#4 - 24 janvier 2024 22:44 - Benjamin Dauvergne

Ça fera gagner du temps sur les restart (qui prendront 8 secondes de moins pour 200 tenants donc) en fait pas tellement sur les migrations du jeudi soir. Par contre ce has_table n'est pas juste inutile, migrate_schemas est aussi utilisé pour initialiser un schéma (dans Tenant.create_schema

upstream dans django-tenant-schemas) et dans ce cas on veut créer la table des migrations.

#5 - 24 janvier 2024 22:45 - Robot Gitea

- *Statut changé de Solution proposée à En cours*

Benjamin Dauvergne (bdauvergne) a relu et demandé des modifications sur une pull request sur Gitea concernant cette demande :

- URL : <https://git.entrouvert.org/entrouvert/hobo/pulls/78>