

Passerelle - Development #47946

mise à jour d'un CSV : ne pas conserver l'ancienne version

22 octobre 2020 11:33 - Thomas Noël

Statut:	Fermé	Début:	22 octobre 2020
Priorité:	Normal	Echéance:	
Assigné à:	Lauréline Guérin	% réalisé:	0%
Catégorie:		Temps estimé:	0:00 heure
Version cible:		Planning:	Non
Patch proposed:	Oui		
Description			
Quand on met à jour un connecteur CSV, il a y création d'un nouveau fichier sur le filesystem (/media/csv/xxxx).			
Si un CSV existait déjà, il devient juste orphelin et reste en place indéfiniment.			
On pourrait faire un nettoyage régulier : supprimer tous les fichiers de /media/csv/ qui ne sont pas référencés par un CsvDataSource (et, par sécurité, qui ont plus d'une semaine)			
Demandes liées:			
Lié à Hobo - Development #48071: Le tenant n'est pas disponible dans les migr...		Fermé	27 octobre 2020

Révisions associées

Révision 9cd83a93 - 05 novembre 2020 09:27 - Lauréline Guérin

csvdatasource: change upload_to and move files (#47946)

Révision 08370b3f - 05 novembre 2020 09:27 - Lauréline Guérin

nancypoll: change upload_to and move files (#47946)

Révision 2d9bbb54 - 05 novembre 2020 09:43 - Lauréline Guérin

csvdatasource: clean old files (#47946)

Historique

#2 - 22 octobre 2020 18:36 - Benjamin Dauvergne

On peut aussi s'arranger pour ne pas conserver l'ancien fichier sur une mise à jour, ça évitera dans un premier temps de créer tant de fichiers orphelins.

#3 - 22 octobre 2020 20:17 - Thomas Noël

Benjamin Dauvergne a écrit :

On peut aussi s'arranger pour ne pas conserver l'ancien fichier sur une mise à jour, ça évitera dans un premier temps de créer tant de fichiers orphelins.

Yep, j'ai pensé à ça (à faire lors du save()) mais je me disais que c'était assez "agréable" de conserver les quelques dernières versions... une sorte de backup du pauvre. Et comme de toute façon il va falloir écrire le code pour nettoyer les /media/csv actuels...

#4 - 26 octobre 2020 14:17 - Lauréline Guérin

- Assigné à mis à Lauréline Guérin

#5 - 27 octobre 2020 15:43 - Lauréline Guérin

- Fichier 0003-csvdatasource-clean-old-files-47946.patch ajouté
- Fichier 0002-nancypoll-change-upload_to-and-move-files-47946.patch ajouté
- Fichier 0001-csvdatasource-change-upload_to-and-move-files-47946.patch ajouté
- Statut changé de Nouveau à Solution proposée
- Patch proposed changé de Non à Oui

Vu avec Thomas: nancypoll aussi stocke ses csv dans media/csv

0001 et 0002: on déplace les fichiers csv des connecteurs csvdatasource et nancypoll, pour ranger un peu

0003: pour chaque connecteur csvdatasource, on nettoie les vieux fichiers inutilisés du même répertoire, avec un setting pour juste les déplacer (default), ou les supprimer pour de bon

#6 - 27 octobre 2020 16:13 - Lauréline Guérin

- Fichier 0003-csvdatasource-clean-old-files-47946.patch ajouté
- Fichier 0002-nancypoll-change-upload_to-and-move-files-47946.patch ajouté
- Fichier 0001-csvdatasource-change-upload_to-and-move-files-47946.patch ajouté

#7 - 27 octobre 2020 16:33 - Benjamin Dauvergne

C'est bizarre je ne vois pas de migration AlterField pour la modification à upload_to.

--

Le tenant s'obtient via tenant = getattr(django.db.connection, 'tenant', None) mais en fait ça ne me paraît pas nécessaire, le storage attaché au field devrait tout seul trouver le bon répertoire (ils sont tenant "aware") donc un simple accès à instance.csv_file.path donnera le bon chemin, et pour trouver le répertoire de base model._meta.get_field('csv_file').storage.path("").

Donc grosso merdo je pense qu'on peut :

- modifier tous les upload_to sans réfléchir (1 migration)
- repasser sur tous les champs csv_file et faire

```
old_name = instance.csv_file.name
old_path = instance.csv_file.path
if not old_name.startswith('csv/'):
    continue
filename = os.path.basename(old_name)
instance.csv_file.save(filename, instance.csv_file)
unlink(old_path)
```

car upload_to n'est utilisé qu'à la première sauvegarde pour générer name, ensuite il respecte name absolument.

PS: édité.

#8 - 27 octobre 2020 17:07 - Lauréline Guérin

Non, lors d'une migration on a un FakeTenant sur connection.tenant, qui ne contient que schema_name, et ne nous permet pas de retrouver le path jusqu'au dossier media, d'où ma tambouille.

Si on essaie d'accéder à csv_file.path lors d'une migration, on a une erreur ici:

```
class TenantFileSystemStorage(FileSystemStorage, TenantStorageMixin):
    '''Lookup files first in $TENANT_BASE/<tenant.schema>/media/ then in default location'''
    def path(self, name):
        if connection.tenant:
            location = safe_join(settings.TENANT_BASE, connection.tenant.domain_url, 'media') # <= ici
        else:
            location = self.location
        try:
            path = safe_join(location, name)
        except ValueError:
            raise SuspiciousOperation("Attempted access to '%s' denied." % name)
        return os.path.normpath(path)
```

Car sur un FakeTenant il n'y a pas de domain_url
Le storage fonctionne mal sur une migration

Les AlterField sont dans les migrations 0019 (csvdatasource) et 0004 (nancypoll)

#9 - 27 octobre 2020 17:14 - Benjamin Dauvergne

Lauréline Guérin a écrit :

Non, lors d'une migration on a un FakeTenant sur connection.tenant, qui ne contient que schema_name, et ne nous permet pas de retrouver le path jusqu'au dossier media, d'où ma tambouille.

Mouais c'est tout pourri, il faudrait d'abord corriger ça.

#10 - 27 octobre 2020 17:15 - Benjamin Dauvergne

- Lié à Development #48071: Le tenant n'est pas disponible dans les migrations ajouté

#11 - 29 octobre 2020 15:03 - Lauréline Guérin

- Fichier 0003-csvdatasource-clean-old-files-47946.patch ajouté
- Fichier 0002-nancypoll-change-upload_to-and-move-files-47946.patch ajouté
- Fichier 0001-csvdatasource-change-upload_to-and-move-files-47946.patch ajouté

(ne fonctionnera qu'avec la modif de [#48071](#))

J'ai un doute: st_ctime ou st_mtime ?

On utilise déjà st_ctime dans le code pour afficher la date d'ajout du fichier, mais en local la commande de nettoyage fonctionne mieux (avec les fichiers que j'avais déjà dans mon rep media) avec st_mtime (st_mtime semble donner la même date qu'un ls -l en shell)

#12 - 29 octobre 2020 15:14 - Lauréline Guérin

- Fichier 0003-csvdatasource-clean-old-files-47946.patch ajouté

#13 - 29 octobre 2020 16:40 - Benjamin Dauvergne

Lauréline Guerin a écrit :

(ne fonctionnera qu'avec la modif de [#48071](#))

J'ai un doute: st_ctime ou st_mtime ?

On utilise déjà st_ctime dans le code pour afficher la date d'ajout du fichier, mais en local la commande de nettoyage fonctionne mieux (avec les fichiers que j'avais déjà dans mon rep media) avec st_mtime (st_mtime semble donner la même date qu'un ls -l en shell)

Je dirai mtime, c'est ce que touch modifie si jamais on veut conserver un fichier plus longtemps (dans les faits ça n'a pas beaucoup d'importance, juste que sur un unlink/link/chmod/chown ça va faire ctime <- now(), est-ce qu'on veut vraiment que ça prolonge la vie du fichier ?).

#14 - 05 novembre 2020 09:43 - Lauréline Guérin

- Fichier 0003-csvdatasource-clean-old-files-47946.patch ajouté
- Fichier 0002-nancypoll-change-upload_to-and-move-files-47946.patch ajouté
- Fichier 0001-csvdatasource-change-upload_to-and-move-files-47946.patch ajouté

#15 - 05 novembre 2020 18:10 - Benjamin Dauvergne

- Statut changé de Solution proposée à Solution validée

#16 - 06 novembre 2020 08:09 - Lauréline Guérin

- Statut changé de Solution validée à Résolu (à déployer)

```
commit 2d9bbb54ff09e615b8bb4047de5c22ee31fe9f33
Author: Lauréline Guérin <zebuline@entrouvert.com>
Date: Tue Oct 27 15:31:19 2020 +0100
```

```
csvdatasource: clean old files (#47946)
```

```
commit 08370b3fef92624fa0963953b7a932ec3655f47b
Author: Lauréline Guérin <zebuline@entrouvert.com>
Date: Tue Oct 27 10:36:10 2020 +0100
```

```
nancypoll: change upload_to and move files (#47946)
```

```
commit 9cd83a93977b5a2730d46cfea83024ee9d9d37a7
Author: Lauréline Guérin <zebuline@entrouvert.com>
Date: Tue Oct 27 10:35:38 2020 +0100
```

```
csvdatasource: change upload_to and move files (#47946)
```

#17 - 09 novembre 2020 10:16 - Frédéric Péters

- Statut changé de Résolu (à déployer) à Solution déployée

Fichiers

0002-nancypoll-change-upload_to-and-move-files-47946.patch	4,29 ko	27 octobre 2020	Lauréline Guérin
0001-csvdatasource-change-upload_to-and-move-files-47946.patch	4,55 ko	27 octobre 2020	Lauréline Guérin
0003-csvdatasource-clean-old-files-47946.patch	7,77 ko	27 octobre 2020	Lauréline Guérin
0002-nancypoll-change-upload_to-and-move-files-47946.patch	4,29 ko	27 octobre 2020	Lauréline Guérin
0003-csvdatasource-clean-old-files-47946.patch	7,77 ko	27 octobre 2020	Lauréline Guérin
0001-csvdatasource-change-upload_to-and-move-files-47946.patch	4,55 ko	27 octobre 2020	Lauréline Guérin
0003-csvdatasource-clean-old-files-47946.patch	8,24 ko	29 octobre 2020	Lauréline Guérin
0002-nancypoll-change-upload_to-and-move-files-47946.patch	3,21 ko	29 octobre 2020	Lauréline Guérin
0001-csvdatasource-change-upload_to-and-move-files-47946.patch	3,44 ko	29 octobre 2020	Lauréline Guérin
0003-csvdatasource-clean-old-files-47946.patch	8,22 ko	29 octobre 2020	Lauréline Guérin
0003-csvdatasource-clean-old-files-47946.patch	8,22 ko	05 novembre 2020	Lauréline Guérin
0002-nancypoll-change-upload_to-and-move-files-47946.patch	3,21 ko	05 novembre 2020	Lauréline Guérin
0001-csvdatasource-change-upload_to-and-move-files-47946.patch	3,44 ko	05 novembre 2020	Lauréline Guérin