

Passerelle - Bug #11258

csv datasource : augmenter la taille du contenu passé au sniffer

07 juin 2016 19:06 - Josué Kouka

Statut:	Fermé	Début:	07 juin 2016
Priorité:	Normal	Echéance:	
Assigné à:	Josué Kouka	% réalisé:	0%
Catégorie:		Temps estimé:	0:00 heure
Version cible:		Planning:	
Patch proposed:	Oui		
Description			
Pour des fichiers dont les lignes > 1024 on a une traceback			
File "/home/josue/.virtualenvs/passerelle/local/lib/python2.7/site-packages/django/core/handlers/base.py" in get_response 111. response = wrapped_callback(request, *callback_args, **callback_kwargs)			
File "./passerelle/urls_utils.py" in f 52. return func(request, *args, **kwargs)			
File "/home/josue/.virtualenvs/passerelle/local/lib/python2.7/site-packages/django/views/generic/base.py" in view 69. return self.dispatch(request, *args, **kwargs)			
File "/home/josue/.virtualenvs/passerelle/local/lib/python2.7/site-packages/django/views/generic/base.py" in dispatch 87. return handler(request, *args, **kwargs)			
File "./passerelle/urls/__init__.py" in _wrapped_view 104. return view_func(instance, request, *args, **kwargs)			
File "./passerelle/urls/jsonresponse.py" in wrapper 263. return self.method(f, *args, **kwargs)			
File "./passerelle/urls/jsonresponse.py" in api_method 345. return self.api(f, args ¹ , *args, **kwargs)			
File "./passerelle/urls/jsonresponse.py" in api 350. resp = f(*args, **kwargs)			
File "./passerelle/apps/csvdatasource/views.py" in get 81. return obj.get_data(filters)			
File "./passerelle/apps/csvdatasource/models.py" in get_data 62. dialect = csv.Sniffer().sniff(content[:1024])			
File "/usr/lib/python2.7/csv.py" in sniff 188. raise Error, "Could not determine delimiter"			

Révisions associées

Révision 923937cd - 08 juin 2016 18:50 - Josué Kouka

csv data source: increase content length passed to sniffer (#11258)

Historique

#1 - 07 juin 2016 19:06 - Josué Kouka

- Lié à Development #11057: csv datasource : supporter les operateurs <eq> <ne> <lt> <le> <gt> <ge> ajouté

#2 - 07 juin 2016 23:34 - Josué Kouka

- Fichier 0001-csv-data-source-increase-content-length-passed-to-sn.patch ajouté

- Patch proposed changé de Non à Oui

#3 - 07 juin 2016 23:54 - Thomas Noël

- Lié à Development #11057: csv datasource : supporter les operateurs <eq> <ne> <lt> <le> <gt> <ge> supprimé

#4 - 07 juin 2016 23:55 - Thomas Noël

Je n'avais pas vu la trace, on pourrait éventuellement passer à 2048 seulement en cas d'erreur sur un try/except du sniff sur 1024 ?

#5 - 08 juin 2016 00:08 - Josué Kouka

- Fichier 0001-csv-data-source-increase-content-length-passed-to-sn.patch ajouté

#6 - 08 juin 2016 00:18 - Frédéric Péters

Perso je serais plus large et rapide et soit juste dire 10ko; soit prendre `index("\n")*3` vu que le ticket parle de longueur ligne; mais il me semble qu'on ne gagne rien à essayer un coup puis un deuxième, etc. (et prendre 10ko c'est pas la mort). Aussi, toute autre approche, c'est de faire cette récupération de dialect une bonne fois pour toutes lors d'upload du fichier csv plutôt que lors de l'appel à l'API.

#7 - 08 juin 2016 00:59 - Frédéric Péters

Pour info, 10000 passes sur `csv.Sniffer().sniff(content)`, selon la taille de content :

- sur 1024 octets, 0.1s
- sur 10 ko, 0.8s

(mesuré avec `import timeit`).

Sur une seule passe,

- 0.000627994537354s
- 0.000849962234497s

(i.e. faisons le patch simple avec 10ko chargés dès le début).

#8 - 08 juin 2016 14:27 - Josué Kouka

- Fichier `0001-csv-data-source-increase-content-length-passed-to-sn.patch` ajouté

Pour l'instant j'ai augmenté la taille du contenu passé au sniffer. J'ai créé un ticket afin de pouvoir détecter le dialect au moment de l'upload [#11264](#)

#9 - 08 juin 2016 18:48 - Thomas Noël

Ack pour 10240

#10 - 08 juin 2016 18:51 - Thomas Noël

```
commit 923937cdf82d092ab837050a82ba55d686a76e5b
Author: Josue Kouka <jkouka@entrouvert.com>
Date:   Wed Jun 8 14:23:14 2016 +0200
```

```
csv data source: increase content length passed to sniffer (#11258)
```

#11 - 08 juin 2016 18:52 - Thomas Noël

- Statut changé de *En cours* à *Résolu* (à déployer)

#12 - 04 août 2018 12:31 - Benjamin Dauvergne

- Statut changé de *Résolu* (à déployer) à *Fermé*

Fichiers

<code>0001-csv-data-source-increase-content-length-passed-to-sn.patch</code>	892 octets	07 juin 2016	Josué Kouka
<code>0001-csv-data-source-increase-content-length-passed-to-sn.patch</code>	1006 octets	07 juin 2016	Josué Kouka
<code>0001-csv-data-source-increase-content-length-passed-to-sn.patch</code>	903 octets	08 juin 2016	Josué Kouka