

w.c.s. - Bug #38863

conversion pickle python3 et listes répétées

09 janvier 2020 16:30 - Frédéric Péters

Statut:	Fermé	Début:	09 janvier 2020
Priorité:	Normal	Echéance:	
Assigné à:		% réalisé:	0%
Catégorie:		Temps estimé:	0:00 heure
Version cible:		Planning:	Non
Patch proposed:	Oui		

Description

Cas qui arrive, dans un formdef avec roles = ['abc'] et backoffice_submission_roles = ['abc'].

Les deux listes peuvent se trouver avec un id() identique. Et du coup la deuxième ne sera pas convertie, en pensant qu'elle l'a déjà été.

Le plus simple il me semble c'est de ne pas tenir compte des objets de type liste dans le suivi des id(), qui est là pour éviter une récursion dans des objets/dictionnaires déjà convertis.

Révisions associées

Révision 2e05d710 - 11 janvier 2020 09:31 - Frédéric Péters

storage: don't track bytes/lists when converting pickles (#38863)

Historique

#1 - 09 janvier 2020 16:30 - Frédéric Péters

- Fichier 0001-storage-don-t-track-bytes-lists-when-converting-pick.patch ajouté
- Statut changé de Nouveau à Solution proposée
- Patch proposed changé de Non à Oui

#2 - 09 janvier 2020 17:08 - Thomas Noël

Le truc de l'id identique je vois mal d'où il vient, et donc, est-il aussi impossible sur les bytes ? (auquel cas il faudrait passer le isinstance(obj, bytes) au dessus de la détection du seen, aussi)

#3 - 09 janvier 2020 17:17 - Frédéric Péters

Le truc de l'id identique je vois mal d'où il vient,

Je n'ai pas creusé mais j'imagine une optimisation côté pickle, qui pour deux listes au contenu identique se trouve la stocker une seule fois, genre :

```
>>> pickle.dumps({'a': ['bbb'], 'c': ['bbb']}).count(b'bbb')
1
```

mais dans des tests basiques, je ne me retrouve in fine pas avec la situation :

```
>>> d2 = pickle.loads(pickle.dumps({'a': ['bbb'], 'c': ['bbb']}))
>>> id(d2['a'])
139631731647584
>>> id(d2['c'])
139631731647904
```

reste que la situation s'est bel et bien présentée au chargement d'un formdef (parce que plus long ? parce que pickle créé par une précédente version de Python ?).

Pour les bytes qui ne sont plus non plus ajoutés à seen, c'est que ça n'avait de toute façon pas de sens, il ne peut pas y avoir de cycle à casser en les chargeant.

#4 - 09 janvier 2020 17:20 - Thomas Noël

- Statut changé de Solution proposée à Solution validée

Frédéric Péters a écrit :

reste que la situation s'est bel et bien présentée au chargement d'un formdef (parce que plus long ? parce que pickle créé par une précédente version de Python ?).

Ok.

Pour les bytes qui ne sont plus non plus ajoutés à seen, c'est que ça n'avait de toute façon pas de sens, il ne peut pas y avoir de cycle à casser en les chargeant.

Mais je suis tellement bête, j'avais mal relu ton patch et tu as bien passé la gestion des bytes avant le "if id(obj) in seen:" donc c'est parfait. J'eusse pu valider plus tôt.

#5 - 11 janvier 2020 09:33 - Frédéric Péters

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

```
commit 2e05d710c1a11f6237805acb547bfbbaa37f7b271
Author: Frédéric Péters <fpeters@entrouvert.com>
Date: Thu Jan 9 16:29:02 2020 +0100
```

```
storage: don't track bytes/lists when converting pickles (#38863)
```

#6 - 13 janvier 2020 22:15 - Frédéric Péters

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

Fichiers

0001-storage-don-t-track-bytes-lists-when-converting-pick.patch	1,47 ko	09 janvier 2020	Frédéric Péters
---	---------	-----------------	-----------------