

Passerelle - Development #51136

Mocker une méthode dans ElementTree dans xlrd avec python>=3.8

15 février 2021 17:11 - Benjamin Dauvergne

Statut:	Fermé	Début:	15 février 2021
Priorité:	Normal	Echéance:	
Assigné à:	Benjamin Dauvergne	% réalisé:	0%
Catégorie:		Temps estimé:	0:00 heure
Version cible:		Planning:	Non
Patch proposed:	Oui		
Description			
Comme #51135 c'est un poil irritant sur debian testing.			

Révisions associées

Révision 43805e1c - 02 mars 2021 16:44 - Benjamin Dauvergne

csvdatasource: work around bug in xlrd with python 3.8 (#51136)

Historique

#1 - 15 février 2021 17:12 - Benjamin Dauvergne

- Statut changé de Nouveau à Solution proposée

- Patch proposed changé de Non à Oui

#2 - 15 février 2021 17:15 - Benjamin Dauvergne

- Fichier 0001-compat-add-missing-method-for-xlrd-on-python-3.8.patch ajouté

#3 - 16 février 2021 11:58 - Emmanuel Cazenave

Il y a pas moyen de circonscrire ça au connecteur concerné ?

Ou une version de pyexcel-xls ou xlrd qui corrigerait ?

Là ça fait un peu banzai sur tout la satck python.

#4 - 16 février 2021 12:46 - Benjamin Dauvergne

Malheureusement on utilise pyexcel-xls qui est limité à la version < 2 de xlrd :

```
Requires-Dist: pyexcel-io (>=0.6.2)
```

```
Requires-Dist: xlrd (<2)
```

```
Requires-Dist: xlwt
```

T'as le choix soit on se prive de pyexcel-xls soit d'un xlrd compatible py39 :

- xlrd changelog 2.x.x : <https://github.com/python-excel/xlrd/blob/master/CHANGELOG.rst#200-11-december-2020>
- pyexcel-xls, souci avec xlrd 2.x.x : <https://github.com/pyexcel/pyexcel-xls/pull/41https://github.com/pyexcel/pyexcel-xls/issues/45>

Le truc c'est que xlrd a abandonné le support de OpenXML (.xlsx) et conseille d'utiliser openpyxl dans ce cas mais pyexcel-xls n'a pas fait la migration.

#5 - 16 février 2021 15:59 - Emmanuel Cazenave

Je comprends pas pourquoi ça marche pas en fait, il y a un commit dans xlrd qui essaie de gérer ça <https://github.com/python-excel/xlrd/commit/93673f98637359f1be41181e7d9ca00f48c99624>.

Tu pourrais mettre la trace ?

#6 - 16 février 2021 17:02 - Benjamin Dauvergne

Emmanuel Cazenave a écrit :

Je comprends pas pourquoi ça marche pas en fait, il y a un commit dans xlrd qui essaie de gérer ça <https://github.com/python-excel/xlrd/commit/93673f98637359f1be41181e7d9ca00f48c99624>.

C'est un commit de xlrd 2.x.x, je t'ai pointé la contrainte de pyexcel-xls qui dit xlrd<2. Comment veux-tu installer xlrd>=2 si pyexcel-xls l'interdit (à foriori à raison) ? Pour information (c'est peut-être l'information qui te manque) on n'utilise pas xlrd directement dans passerelle, c'est une dépendance de pyexcel-xls.

Tu pourrais mettre la trace ?

Ben non j'ai expliqué déjà.

PS: le problème vient bien de l'absence de getiterator() utilisé par xlrd<2, et sur python>3.7 mais vu le patch et le commentaire qu'il contient ça me paraissait évident.

#7 - 16 février 2021 18:02 - Emmanuel Cazenave

Benjamin Dauvergne a écrit :

C'est un commit de xlrd 2.x.x,

Non, 0.9.4.

#8 - 18 février 2021 16:28 - Benjamin Dauvergne

Emmanuel Cazenave a écrit :

Benjamin Dauvergne a écrit :

C'est un commit de xlrd 2.x.x,

Non, 0.9.4.

Ok et bien leur commit ne marche pas :

```
self = <xlrd.xlsx.X12Book object at 0x7f32542c1280>, stream = <zipfile.ZipExtFile name='xl/workbook.xml' mode='r' compress_type=deflate>, heading = 'Workbook'
```

```
def process_stream(self, stream, heading=None):
    if self.verbosity >= 2 and heading is not None:
        fprintf(self.logfile, "\n=== %s ===\n", heading)
    self.tree = ET.parse(stream)
    getmethod = self.tag2meth.get
> for elem in self.tree.iter() if Element_has_iter else self.tree.getiterator():
E AttributeError: 'ElementTree' object has no attribute 'getiterator'
```

```
/tmp/tox-bdauvergne/passerelle/py3-django111-coverage/lib/python3.9/site-packages/xlrd/xlsx.py:266: AttributeError
```

```
>>> entering PDB >>>
```

```
>>> PDB post_mortem (IO-capturing turned off) >>
```

```
> /tmp/tox-bdauvergne/passerelle/py3-django111-coverage/lib/python3.9/site-packages/xlrd/xlsx.py (266)process_s
tream()
-> for elem in self.tree.iter() if Element_has_iter else self.tree.getiterator():
(Pdb) Element_has_iter
False
(Pdb)
```

et ça ne marche parce que ça utilise finalement defusedxml au lieu du ElementTree natif :

```
>>> PDB set_trace (IO-capturing turned off) >>>
> /tmp/tox-bdauvergne/passerelle/py3-django111-coverage/lib/python3.9/site-packages/xlrd/xlsx.py (61)ensure_ele
menttree_imported()
-> Element_has_iter = hasattr(ET, 'ElementTree') and hasattr(ET.ElementTree, 'iter')
(Pdb) hasattr(ET, 'ElementTree')
False
(Pdb) ET.Element
*** AttributeError: module 'defusedxml.cElementTree' has no attribute 'Element'
(Pdb) dir(ET)
['DefusedXMLParser', 'XML', 'XMLParse', 'XMLParser', 'XMLTreeBuilder', '_TreeBuilder', '__all__', '__builtins__']
```

```
_, '__cached__', '__doc__', '__file__', '__loader__', '__name__', '__origin__', '__package__', '__spec__', '_  
generate_etree_functions', '_iterparse', '_parse', 'absolute_import', 'fromstring', 'iterparse', 'parse', 'tos  
tring']
```

Et je vais arrêter là mes investigations parce que le code est juste affreux.

#9 - 18 février 2021 16:28 - Benjamin Dauvergne

- Fichier 0001-csvdatasource-work-around-bug-in-xlrd-with-python-3..patch ajouté

Version plus "locale" si on peut dire.

#10 - 18 février 2021 16:29 - Benjamin Dauvergne

- Fichier 0001-csvdatasource-work-around-bug-in-xlrd-with-python-3..patch ajouté

suppression de la modif à compat.py.

#11 - 02 mars 2021 14:43 - Emmanuel Cazenave

- Statut changé de Solution proposée à Solution validée

#12 - 02 mars 2021 16:45 - Benjamin Dauvergne

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

```
commit 43805e1c42717e84c1c71cc26830ed44623aedd5  
Author: Benjamin Dauvergne <bdauvergne@entrouvert.com>  
Date: Mon Feb 8 19:15:35 2021 +0100
```

```
csvdatasource: work around bug in xlrd with python 3.8 (#51136)
```

#13 - 02 mars 2021 18:17 - Frédéric Péters

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

Fichiers

0001-compat-add-missing-method-for-xlrd-on-python-3.8.patch	898 octets	15 février 2021	Benjamin Dauvergne
0001-csvdatasource-work-around-bug-in-xlrd-with-python-3..patch	1,68 ko	18 février 2021	Benjamin Dauvergne
0001-csvdatasource-work-around-bug-in-xlrd-with-python-3..patch	1,38 ko	18 février 2021	Benjamin Dauvergne