

Passerelle - Bug #15855

connecteur famille: import orléans zip

12 avril 2017 21:47 - Frédéric Péters

| | | | |
|------------------------|---------------|----------------------|---------------|
| Statut: | Fermé | Début: | 12 avril 2017 |
| Priorité: | Normal | Echéance: | |
| Assigné à: | Serghei Mihai | % réalisé: | 0% |
| Catégorie: | | Temps estimé: | 0:00 heure |
| Version cible: | | Planning: | |
| Patch proposed: | Oui | | |

Description

Alerte nagios sur des fichiers modifiés sur passerelle.test; à y regarder, import_orleans_data.py :

```
-         with zipfile.ZipFile(archive_path, 'a') as archive:
+         with zipfile.ZipFile(archive_path, 'a', allowZip64=True) as archive:
```

Si ça correspond à un réel bug, ça doit vraiment être signalé (bon beinh voilà, ce ticket), surtout si ça doit être traité dans passerelle dans ce cycle-ci.

Révisions associées

Révision e9980c07 - 14 août 2017 15:30 - Serghei Mihai

family: orleans connector uses invoices dir path (#15855)

Instead of copying pdf files into the archive.

Historique

#1 - 13 avril 2017 08:45 - Serghei Mihai

- Assigné à mis à Serghei Mihai

Ce n'est pas un bug mais erreur de conception de la synchro.

Actuellement la commande d'import recherche l'archive avec les 4 fichiers nécessaires pour remplir la base et le répertoire "factures" dans lequel la ville pousse les pdf, elle rajoute les pdfs à l'archive et ensuite l'attribue au connecteur qui avec son loader remplit la base et mets les fichiers pdf au bon endroit.

Sauf que l'ensemble des fichiers pdfs fait environ 5.3G, et ça augmente la taille de l'archive. D'ou le allowZip64=True que j'ai rajouté pour pouvoir rajouter des données pour une première fois.

Sauf que l'archive de 5.3G est copie autant de fois sur le disque que des appels à save.

Patch en cours pour revoir cela.

#2 - 13 avril 2017 11:08 - Serghei Mihai

- Fichier 0001-family-always-create-invoices-with-orleans-loader-15.patch ajouté

- Statut changé de Nouveau à En cours

- Patch proposed changé de Non à Oui

Je suis pour créer les objets factures même si les pdfs ne sont pas présents dans l'archive. Et le boulot de la commande serait de les copier au bon endroit depuis le répertoire dans lequel la ville nous les envoie.

#3 - 13 avril 2017 13:58 - Thomas Noël

J'ai un peu de mal à saisir. Dans le travail que nous avons fait à l'époque à Orléans, Arnaud a accès au répertoire des PDF : il les ajoute (ou les retire) à sa guise. Quand un PDF est présent, la facture du CSV est considérée existante (publiée, payable, whatever). Quand le PDF n'est pas là... la facture n'est pas là.

La raison : le PDF c'est la "vraie" facture, celle qui est émise par la ville (en papier) et donc Arnaud les envoie au système seulement quand le train de facturation est fait, terminé, validé par le service. Le CSV des factures contient des factures qui ne sont pas forcément encore validées. C'est le PDF qui "active" vraiment une facture.

Bref.

En tout cas selon moi, on ne touche pas au PDF, y'a pas de truc comme `< invoice.write_pdf(file(invoice_file_path).read()) >` à faire.

(mais j'ai pas tout lu du système d'import, je relève juste ça, le système d'import Orléans ne devrait pas jouer avec les PDF, uniquement vérifier leur présence pour activer les factures correspondantes)

#4 - 13 avril 2017 14:29 - Serghei Mihai

Thomas Noël a écrit :

J'ai un peu de mal à saisir. Dans le travail que nous avons fait à l'époque à Orléans, Arnaud a accès au répertoire des PDF : il les ajoute (ou les retire) à sa guise. Quand un PDF est présent, la facture du CSV est considérée existante (publiée, payable, whatever). Quand le PDF n'est pas là... la facture n'est pas là.

La raison : le PDF c'est la "vraie" facture, celle qui est émise par la ville (en papier) et donc Arnaud les envoie au système seulement quand le train de facturation est fait, terminé, validé par le service. Le CSV des factures contient des factures qui ne sont pas forcément encore validées. C'est le PDF qui "active" vraiment une facture.

C'est le cas dans le code actuellement (sans ce patch): un objet facture est créé uniquement si le fichier correspondant existe dans l'archive. Sauf que créer l'archive avec des milliers des factures sa pose des problèmes d'espace disque et des perms.

Je vais réfléchir à une alternative pour garder l'ancien fonctionnement.

#5 - 13 avril 2017 15:14 - Frédéric Péters

Je vais réfléchir à une alternative pour garder l'ancien fonctionnement.

Inclure dans le zip le nom du répertoire où se trouvent les PDF, end of story ?

#6 - 14 avril 2017 16:12 - Serghei Mihai

Mais le fichier PDF est recherché dans l'archive même: `invoice_filename` in `archive.namelist()` et non sur l'espace disque. Je ne suis pas sûr de comprendre.

#7 - 14 avril 2017 16:21 - Frédéric Péters

... Sauf que créer l'archive avec des milliers des factures sa pose des problèmes d'espace disque. ...

Je propose donc de ne pas « créer l'archive avec des milliers des factures » ; à la place de mettre dans l'archive le nom du répertoire qui sera utilisé par le code d'import pour vérifier l'existence de la facture, et alors si `os.path.exists(le pdf)` alors la facture est créée et le pdf copié au bon endroit. (tout cela étant dit de loin, je n'ai pas suivi ce projet).

#8 - 19 avril 2017 14:57 - Serghei Mihai

- Fichier `0001-family-orleans-connector-defines-invoices-dir-path-1.patch` ajouté

Ok. Et loader ira chercher les fichiers correspondants dans le répertoire défini dans l'archive.

#9 - 27 avril 2017 10:48 - Thomas Noël

J'avoue m'y perdre. Pourquoi tripoter le zip ou déplacer les PDF ?

J'avais pensé décrire le truc existant dans mon commentaire 3, qui est l'aboutissement d'un long dialogue avec Orléans à l'époque : c'est une procédure qui ne doit pas changer. Orléans gère un répertoire avec des PDF d'un côté, et envoie des CSV zippé d'un autre côté. Et ca doit vraiment nous suffire pour faire un loader à la con.

Dans ce que dit Fred « alors si `os.path.exists(le pdf)` alors la facture est créée et le pdf copié au bon endroit » je dis que non. On ne doit pas du tout toucher aux PDF gérés directement par Orléans, pas les déplacer, rien. Une fois l'import des factures effectué (lecture du CSV), on active les factures du CSV qui ont des PDF, on désactive les autres, on supprime les PDF qui ne sont pas dans le CSV. Comme "avant".

Y'a pas à avoir dans le zip le nom du répertoire : il ne peut pas y être, Orléans ne connaît pas l'emplacement réel des PDF. C'est un argument à passer lors de l'import des CSV, genre.

En dehors de ça :

- c'est sans doute pas la peine de mettre 25 faux PDF dans les tests, je pense que 3 ou 4 fichiers devraient suffire ?
- attention aux print : c'est une commande qui sera lancée par cron, elle ne doit rien dire si tout va bien (si besoin, gérer un mode verbose)
- le `"raise CommandError('Command already running.')` doit arriver juste après le `fcntl.lockf` ; sinon il peut y avoir eu d'autres IOError dans le reste du `try/except`

- last but not least, le lock ne doit pas être fait sur un fichier tempfile mais toujours sur le même fichier (si possible dépendant du tenant, genre via un truc dans /media si y'a pas d'autre méthode tenantisée)

#11 - 02 mai 2017 17:24 - Serghei Mihai

- Fichier 0001-family-orleans-connector-uses-invoices-dir-path-1585.patch ajouté

Ok.

Pas de copie des fichiers PDF dans le répertoire de l'instance du connecteur. On posera un lien symbolique vers le répertoire ou la ville envoie ses pdfs (/var/lib/synchro-orleans/factures) et ainsi les objets des factures seront créés et les fichiers seront trouvés pour leur téléchargement.

#13 - 18 juin 2017 03:29 - Thomas Noël

(putain de retard, désolé)

Je suis pas à l'aise avec un gros « try:...except:pass » dans un test. Y'a une raison valable, en dehors que "comme ça le test marche toujours" ?

LOCK_FILENAME n'est pas une constante, pas de majuscule

Pour le reste il faut qu'on relise ça ensemble parce que je ne comprends pas la logique (genre "--data-directory: Directory containing archive and invoice files" alors que les invoices semblent devoir être dans DefaultStorage)

#14 - 20 juin 2017 10:52 - Serghei Mihai

- Fichier 0001-family-orleans-connector-uses-invoices-dir-path-1585.patch ajouté

Ok, donc --data-directory ne sert à rien vu qu'on point uniquement l'archive avec les données et le lien vers le repertoire des factures sera posé à la main.

Plus de gros try... except dans le test.

#16 - 16 août 2017 14:20 - Josué Kouka

Ack

#17 - 16 août 2017 14:25 - Serghei Mihai

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

```
commit e9980c07b1225a4379036c01cc249422dc2d0d55 (origin/master, origin/HEAD)
```

```
Author: Serghei Mihai <smihai@entrouvert.com>
```

```
Date: Thu Apr 13 09:51:19 2017 +0200
```

```
family: orleans connector uses invoices dir path (#15855)
```

```
Instead of copying pdf files into the archive.
```

#18 - 11 décembre 2017 14:25 - Serghei Mihai

- Statut changé de Résolu (à déployer) à Fermé

Fichiers

| | | | |
|---|---------|---------------|---------------|
| 0001-family-always-create-invoices-with-orleans-loader-15.patch | 4,5 ko | 13 avril 2017 | Serghei Mihai |
| 0001-family-orleans-connector-defines-invoices-dir-path-1.patch | 88 ko | 19 avril 2017 | Serghei Mihai |
| 0001-family-orleans-connector-uses-invoices-dir-path-1585.patch | 54,7 ko | 02 mai 2017 | Serghei Mihai |
| 0001-family-orleans-connector-uses-invoices-dir-path-1585.patch | 55,7 ko | 20 juin 2017 | Serghei Mihai |