

OLAP / Business Intelligence pour Publik - Development #30752

Conserver les données des tables de dimensions

19 février 2019 11:23 - Emmanuel Cazenave

Statut:	Fermé	Début:	19 février 2019
Priorité:	Normal	Echéance:	
Assigné à:	Emmanuel Cazenave	% réalisé:	0%
Catégorie:		Temps estimé:	0:00 heure
Version cible:		Planning:	
Patch proposed:	Oui		
Description			
Première partie du plan élaboré dans #25982 : plutôt que de détruire les tables de dimensions et les ré-créeer à partir de ce que renvoie wcs, reprendre les données déjà existences de la dimension et compléter avec ce que renvoie wcs.			
Ce qui doit produire entre autres effets bénéfiques une stabilité des identifiants des enregistrements des dimensions.			
Demandes liées:			
Lié à OLAP / Business Intelligence pour Publik - Bug #25982: rendre les ident...		Fermé	31 août 2018

Révisions associées

Révision 17bce30b - 01 mars 2019 12:29 - Emmanuel Cazenave

coding style: move class attributes (#30752)

Révision d8bd9cc8 - 01 mars 2019 12:30 - Emmanuel Cazenave

enable pyscopg2 unicode type caster (#30752)

Facilite SELECT query with a WHERE clause on a varchar column.

Révision f53b4bf9 - 01 mars 2019 13:58 - Emmanuel Cazenave

remove dead code (#30752)

An 'item' type field has no 'options'.

Révision 90cc6318 - 01 mars 2019 14:15 - Emmanuel Cazenave

re-use dimension tables from previous run (#30752)

Historique

#1 - 19 février 2019 11:23 - Emmanuel Cazenave

- Lié à Bug #25982: rendre les identifiants stables sinon les visualisation ne peuvent pointer sur des valeurs ajouté

#2 - 21 février 2019 16:16 - Emmanuel Cazenave

- Assigné à mis à Emmanuel Cazenave

#3 - 22 février 2019 15:41 - Emmanuel Cazenave

- Fichier 0001-tests-move-run_wcs_script-to-utils-module-30752.patch ajouté

- Fichier 0002-keep-data-form-previous-run-in-labeled-table-30752.patch ajouté

- Statut changé de Nouveau à Solution proposée

- Patch proposed changé de Non à Oui

Petit patch aux grands effets, il faut que je teste ça grandeur nature d'une façon ou d'une autre, mais si quelqu'un veut déjà jeter un œil.

#4 - 22 février 2019 16:07 - Frédéric Péters

Dans la structure j'aurais plutôt eu le code organisé ainsi :

```
if not table_exists:
    # la création de la table

# le code qui remplit la table, identique peu importe si la table existait ou pas avant.
```

Oui ça fait des requêtes SQL plutôt inutiles quand la table vient d'être créée, et oui au final ça concerne juste l'élimination de quatre lignes d'ajout des entrées, donc c'est vraiment un commentaire de style qui peut être ignoré.

#5 - 25 février 2019 10:48 - Benjamin Dauvergne

Faut que je relise avec le reste du code mais je me demande si ça marche pour les valeurs insérées à la volée (je ne vois pas de mise à jour de la séquence pour l'id), est-ce que l'INSERT INTO bouge la séquence dans le cas serial=True (je vois qu'il est traité implicitement via pg_get_serial_sequence()).

#6 - 28 février 2019 20:07 - Emmanuel Cazenave

- Fichier 0001-re-use-data-from-previous-run-in-dimension-tables-30.patch ajouté

Ma première tentative était bien naïve : effectivement problème sur les séquences mais aussi il ne suffit pas de toucher aux tables, il faut aussi s'occuper des pseudo caches des ces tables qui sont constitués pendant l'import.

Je suis embêté avec ce bout :

```
elif field.options:
    options = enumerate(field.options)
    tmp_options_map = self.create_labeled_table(
        table_name, [(i, o['label']) for i, o in options], comment=comment)
    self.items_mappings[field.varname] = dict(
        (o['value'], tmp_options_map[o['value']]) for i, o in options)
```

Pas compris quelle différence il ya coté wcs entre un champ qui a des items vs des options.

Un bon gros patch (j'ai découpé plus finement au début mais de retouches en retouches ...).

#7 - 01 mars 2019 00:31 - Benjamin Dauvergne

En relisant le code de w.c.s. j'ai la vague impression de m'être fourvoyé, field.options n'arrive jamais, c'est toujours items.

#8 - 01 mars 2019 14:31 - Emmanuel Cazenave

- Fichier 0001-coding-style-move-class-attributes-30752.patch ajouté

- Fichier 0002-enable-psycpg2-unicode-type-caster-30752.patch ajouté

- Fichier 0003-remove-dead-code-30752.patch ajouté

- Fichier 0004-re-use-dimension-tables-from-previous-run-30752.patch ajouté

Benjamin Dauvergne a écrit :

En relisant le code de w.c.s. j'ai la vague impression de m'être fourvoyé, field.options n'arrive jamais, c'est toujours items.

Même lecture de mon côté, et à coup de pdb sur des 'vraies' données, code mort.

J'ai retravaillé la substitutions des variables dans les requêtes pour être bien raccord avec les recommandations psycpg2 (et toujours en testant sur de vraies données, ça s'avère tout à fait nécessaire).

Redécoupage en plusieurs patches, je n'y touche plus avant une review.

#9 - 02 mars 2019 17:09 - Benjamin Dauvergne

- Statut changé de Solution proposée à Solution validée

Ack. Je suis pour pousser et voir ce que ça donne immédiatement, faire un revert/corriger après si il faut.

Je pense que dans un deuxième temps il faudra revoir ça pour ne reprendre que les valeurs "vivantes" et vraisemblablement mélanger le cas serial/non-serial, je vois d'ici le cas où on enlève une valeur d'un champ avec field.items mais qu'il y a encore des données correspondantes; voir aussi comment conserver l'ordre dans field.items s'il change avec une colonne order.

#10 - 04 mars 2019 11:39 - Emmanuel Cazenave

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

commit 90cc6318b9c534078b20cb9c5fffa5495f19b184
Author: Emmanuel Cazenave <ecazenave@entrouvert.com>
Date: Fri Mar 1 14:11:15 2019 +0100

re-use dimension tables from previous run (#30752)

commit f53b4bf9d0180f947c9bffe38e5675cadca9da0c
Author: Emmanuel Cazenave <ecazenave@entrouvert.com>
Date: Fri Mar 1 13:58:15 2019 +0100

remove dead code (#30752)

An 'item' type field has no 'options'.

commit d8bd9cc87be8f00e0f35949b354949c81b4b20ae
Author: Emmanuel Cazenave <ecazenave@entrouvert.com>
Date: Fri Mar 1 12:30:50 2019 +0100

enable pycscpg2 unicode type caster (#30752)

Facilitate SELECT query with a WHERE clause
on a varchar column.

commit 17bce30be942a81ac3179b2f07ee90c7e3c033eb
Author: Emmanuel Cazenave <ecazenave@entrouvert.com>
Date: Fri Mar 1 12:29:44 2019 +0100

coding style: move class attributes (#30752)

#11 - 14 mars 2019 10:57 - Emmanuel Cazenave

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

Fichiers

0001-tests-move-run_wcs_script-to-utils-module-30752.patch	4,32 ko	22 février 2019	Emmanuel Cazenave
0002-keep-data-form-previous-run-in-labeled-table-30752.patch	5,03 ko	22 février 2019	Emmanuel Cazenave
0001-re-use-data-from-previous-run-in-dimension-tables-30.patch	21,3 ko	28 février 2019	Emmanuel Cazenave
0002-enable-pycscpg2-unicode-type-caster-30752.patch	874 octets	01 mars 2019	Emmanuel Cazenave
0001-coding-style-move-class-attributes-30752.patch	2,24 ko	01 mars 2019	Emmanuel Cazenave
0003-remove-dead-code-30752.patch	1,22 ko	01 mars 2019	Emmanuel Cazenave
0004-re-use-dimension-tables-from-previous-run-30752.patch	19,3 ko	01 mars 2019	Emmanuel Cazenave