

# OLAP / Business Intelligence pour Publik - Bug #33009

## conflit de hash

11 mai 2019 10:14 - Frédéric Péters

<b>Statut:</b>	Fermé	<b>Début:</b>	11 mai 2019
<b>Priorité:</b>	Normal	<b>Echéance:</b>	
<b>Assigné à:</b>	Frédéric Péters	<b>% réalisé:</b>	0%
<b>Catégorie:</b>		<b>Temps estimé:</b>	0:00 heure
<b>Version cible:</b>		<b>Planning:</b>	Non
<b>Patch proposed:</b>	Oui		
<b>Description</b>			
<pre>&gt;&gt;&gt; hashlib.md5('formdata_saisie_des_frais_des_deplacements_des_conseiller5edf36_field_heure_depart_15').hexdigest()[:6] '9fe42a' &gt;&gt;&gt; hashlib.md5('formdata_saisie_des_frais_des_deplacements_des_conseiller5edf36_heure_depart_12_json_idx').hexdigest()[:6] '9fe42a'</pre>			

### Révisions associées

#### Révision 89bfb6 - 13 mai 2019 19:49 - Frédéric Péters

misc: increase json index table name hash size (#33009)

### Historique

#### #1 - 11 mai 2019 10:19 - Frédéric Péters

- Fichier 0001-misc-increase-table-name-hash-size-33009.patch ajouté
- Statut changé de Nouveau à Solution proposée
- Patch proposed changé de Non à Oui

Ça me semble simple ainsi, mais ça ignore tout à fait l'existant (genre un indicateur qui aurait été posé en se basant sur une des tables qui vont changer de nom) parce que je n'ai pas d'idée de transition.

#### #2 - 11 mai 2019 10:34 - Frédéric Péters

- Fichier 0001-misc-increase-json-index-table-name-hash-size-33009.patch ajouté

J'ai regardé la prod de Grenoble qui était le site m'inquiétant le plus et ça me semble passer sans perte, mais ce n'est pas le cas à d'autres endroits, je vois au moins Nancy,

```
{"cube": "formdata_demande_d_equipement_dotation_ou_changement_de_bf176d1"}
```

Je me dis que hash\_table\_name pourrait gagner un paramètre, genre hash\_length=6, et qu'uniquement dans create\_formdata\_json\_index on y passerait un hash\_length=8, mais 1/ c'est uniquement retarder une autre collision, 2/ il y a quand même peut-être des indicateurs qui vont sauter quelque part.

#### #3 - 11 mai 2019 11:17 - Benjamin Dauvergne

On peut laisser les noms de cube tranquille, peu de chance d'avoir une collision sur le préfixe, le souci est principalement sur les objets en base (index, table pour les champs à choix multiple) mais ceux-ci ne sont pas référencés dans les objets de configuration.

Donc oui augmenter le hash à 8 caractère pour tout le monde mais utiliser l'ancien hash pour la clé name ici :

```
887         cube.update({
888             'name': self.cube_name,
889             'label': self.formdef.schema.name,
890             'fact_table': self.table_name,
891             'key': 'id',
892         })
```

et

```
@property
```

```

def cube_name(self):
    # hash_length is 6 for retro-compatibility with existing visualizations which references cube by their
    name
    return self.hash_table_name('formdata_%s' % self.formdef.slug.replace('-', '_'), hash_length=6)

```

En fait on pourrait même abandonner le système des hashes pour les noms des objets, j'ai suivi la façon de faire de Django mais le contexte est en fait différent, Django a besoin d'une génération déterministe quand on rejoue les migrations, pas nous. On pourrait donc s'en sortir avec une fonction qui mémorise les noms raccourcis pendant la génération et un compteur.

```

table_counter = 0
table_map = None

def adapt_table_name_to_pg(self, table_name):
    '''Truncate long table name, adding suffix identifier to prevent collisions.
    '''
    table_name = table_name.format(**self.default_ctx)
    if self.table_map is None:
        self.table_map = {}
    if table_name in self.table_map:
        return self.table_map[table_name]
    if len(table_name) >= 58:
        table_name = self.table_map[table_name] = '%s_%04d' % (table_name[:58], self.table_counter)
        self.table_counter += 1
    return table_name

```

Et en finissant d'écrire je me dis que ce n'est pas forcément une bonne idée si on vise à moyen terme un chargement incrémental, on reviendra dans la situation de Django où on doit pouvoir retrouver les mêmes noms d'un chargement sur l'autre; idéalement w.c.s. nous servirait des UUIDs sur ses objets de configuration, ça nous aiderait même en cas de changement de slug.

#### #4 - 11 mai 2019 11:53 - Frédéric Péters

(...) le souci est principalement sur les objets en base (index, table pour les champs à choix multiple) mais ceux-ci ne sont pas référencés dans les objets de configuration.

Mais il y a aussi [#30752](#) qui assure des identifiants stables pour les listes de choix, qui a besoin de retrouver la table créée à l'exécution précédente; il me semble.

--

J'ai poussé une branche wip/33009-hash-length en suivant ton commentaire précédent, mais pas sûr.

#### #5 - 11 mai 2019 12:11 - Frédéric Péters

Le hash\_table\_name() et le self.cube\_name créé, ils ne sont pas dans la même classe, patch légèrement adapté.

#### #6 - 11 mai 2019 12:16 - Benjamin Dauvergne

Frédéric Péters a écrit :

Mais il y a aussi [#30752](#) qui assure des identifiants stables pour les listes de choix, qui a besoin de retrouver la table créée à l'exécution précédente; il me semble.

--

J'ai poussé une branche wip/33009-hash-length en suivant ton commentaire précédent, mais pas sûr.

Tu as raison pour [#30752](#), parrons au plus pressé, laisse le hash\_length à 6 par défaut, laisse cube\_name comme il est avec la répétition du 6 et on passe à 8 seulement pour les index, on pourra figoler pour les tables de fait et les tables filles plus tard, pour l'instant on a eu une collision qu'entre un index et une table.

#### #7 - 11 mai 2019 13:03 - Frédéric Péters

Je viens de pousser 0001-misc-increase-json-index-table-name-hash-size-33009.patch dans la branche wip.

#### #8 - 24 mai 2019 07:52 - Frédéric Péters

Je me rends compte (recevant à nouveau la même trace) que j'ai zappé le suivi ici; le patch minimaliste proposé dans la branche, ok ?

#### #10 - 24 mai 2019 11:10 - Benjamin Dauvergne

- Statut changé de Solution proposée à Solution validée

- Assigné à mis à Frédéric Péters

Oui c'est parfait.

#### #11 - 24 mai 2019 11:12 - Frédéric Péters

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

```
commit 89bfafb6dd2126d2efc13ef4d47f4f5f5b8a18b8
Author: Frédéric Péters <fpeters@entrouvert.com>
Date: Sat May 11 10:14:08 2019 +0200
```

```
misc: increase json index table name hash size (#33009)
```

#### #12 - 03 juin 2019 12:16 - Frédéric Péters

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

#### Fichiers

0001-misc-increase-table-name-hash-size-33009.patch	802 octets	11 mai 2019	Frédéric Péters
0001-misc-increase-json-index-table-name-hash-size-33009.patch	1,66 ko	11 mai 2019	Frédéric Péters