

w.c.s. - Development #58013

Itérer sur les ids au lieu d'utiliser les curseurs nommés

20 octobre 2021 15:57 - Benjamin Dauvergne

Statut:	Fermé	Début:	20 octobre 2021
Priorité:	Normal	Echéance:	
Assigné à:	Benjamin Dauvergne	% réalisé:	0%
Catégorie:		Temps estimé:	0:00 heure
Version cible:		Planning:	Non
Patch proposed:	Oui		
Description			
Le principe est de récupérer d'abord la totalité des id des demandes, puis de lire les lignes de tables complètes par parquets (défini par itersize). itersize est par défaut à 200, on pourra l'augmenter à plus.			
Ce mode d'itération est activé automatiquement quand on passe iterator=True à SqlMixin.select().			
Ça supprime le besoin des connections isolées, le code correspondant est donc retiré.			
Demandes liées:			
Lié à w.c.s. - Development #57623: sql: ne pas itérer ligne par ligne sur un ...		Rejeté	05 octobre 2021

Révisions associées

Révision 5389956f - 29 octobre 2021 11:03 - Benjamin Dauvergne

sql: use batch iteration on ids instead of named cursors (#58013)

Named cursors imposed the use of isolated connections and were misused resulting in reading using one SQL query by row (because of the use of .fetchone() with cursors). This commit revert to the behaviour of one connection per request and reading full SQL statement results at a time without using cursors.

Historique

#1 - 20 octobre 2021 16:01 - Benjamin Dauvergne

- Fichier 0001-sql-use-batch-iteration-on-ids-instead-of-named-curs.patch ajouté
- Statut changé de Nouveau à Solution proposée
- Patch proposed changé de Non à Oui

#3 - 20 octobre 2021 16:03 - Benjamin Dauvergne

- Lié à Development #57623: sql: ne pas itérer ligne par ligne sur un SqlKlass.select(iterator=False) ajouté

#4 - 20 octobre 2021 16:04 - Benjamin Dauvergne

- Fichier 0001-sql-use-batch-iteration-on-ids-instead-of-named-curs.patch ajouté

print qui traîne.

#5 - 20 octobre 2021 17:13 - Benjamin Dauvergne

- Statut changé de Solution proposée à En cours

#6 - 20 octobre 2021 18:16 - Benjamin Dauvergne

- Fichier 0001-sql-use-batch-iteration-on-ids-instead-of-named-curs.patch ajouté
- Statut changé de En cours à Solution proposée

warnings pylint corrigés.

#7 - 22 octobre 2021 11:17 - Frédéric Péters

Le bout qui fait :

```
else: # for AnyFormData
```

ça me semble assez indiquer que cette part de code devrait aller dans la classe en question, et j'allais demander pourquoi, sauf qu'à relire la condition,

```
if iter_size and cls._has_id:
```

ça semble être assez différent de "for anything but AnyFormData".

Je me demande dans quelle mesure il y a moyen de rétrécir cette grosse condition partager davantage, sans avoir testé du tout, dans cette idée :

```
if iter_size and cls._has_id:
    sql_statement = '''SELECT id FROM %s''' % cls._table_name
else:
    sql_statement = '''SELECT %s FROM %s''' % (select_fields, cls._table_name)
    where_clauses, parameters, func_clause = parse_clause(clause)
    if where_clauses:
        sql_statement += ' WHERE ' + ' AND '.join(where_clauses)

    sql_statement += cls.get_order_by_clause(order_by)

    if not func_clause:
        if limit:
            sql_statement += ' LIMIT %(limit)s'
            parameters['limit'] = limit
        if offset:
            sql_statement += ' OFFSET %(offset)s'
            parameters['offset'] = offset

    sql_id_statement = '''SELECT %s FROM %s WHERE id IN %%s''' % (
        select_fields,
        cls._table_name,
    )
    sql_id_statement += cls.get_order_by_clause(order_by)

if iter_size and cls._has_id:
    conn, cur = get_connection_and_cursor()
    with cur:
        cur.execute(sql_statement, parameters)
        conn.commit()
        ids = [row[0] for row in cur]
        while ids:
            cur.execute(sql_id_statement, [tuple(ids[:iter_size])])
            conn.commit()
            yield from retrieve()
            ids = ids[iter_size:]

    else:
        conn, cur = get_connection_and_cursor()
        with cur:
            cur.execute(sql_statement, parameters)
            conn.commit()
            yield from retrieve()
```

#8 - 22 octobre 2021 17:43 - Benjamin Dauvergne

- Fichier 0001-sql-use-batch-iteration-on-ids-instead-of-named-curs.patch ajouté

Voilà, la factorisation demandé, et j'ai changé les commentaires pour être plus explicite sur le pourquoi de la spécialisation qui ne sert actuellement qu'à AnyFormData.

#9 - 29 octobre 2021 11:04 - Frédéric Péters

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

Poussé et je tagguerai rapidement pour avoir ça déployé et suivre la conso.

```
commit 5389956f22fc6e7afcc2c6ac8c862cf8eb356cb
Author: Benjamin Dauvergne <bdauvergne@entrouvert.com>
Date:   Wed Oct 20 10:49:44 2021 +0200
```

```
sql: use batch iteration on ids instead of named cursors (#58013)
```

```
Named cursors imposed the use of isolated connections and were misused
resulting in reading using one SQL query by row (because of the use of
```

.fetchone() with cursors). This commit revert to the behaviour of one connection per request and reading full SQL statement results at a time without using cursors.

#10 - 29 octobre 2021 16:17 - Frédéric Péters

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

Fichiers

0001-sql-use-batch-iteration-on-ids-instead-of-named-curs.patch	8,19 ko	20 octobre 2021	Benjamin Dauvergne
0001-sql-use-batch-iteration-on-ids-instead-of-named-curs.patch	8,16 ko	20 octobre 2021	Benjamin Dauvergne
0001-sql-use-batch-iteration-on-ids-instead-of-named-curs.patch	8,38 ko	20 octobre 2021	Benjamin Dauvergne
0001-sql-use-batch-iteration-on-ids-instead-of-named-curs.patch	7,44 ko	22 octobre 2021	Benjamin Dauvergne