

w.c.s. - Development #75594

Performances: requête lente sur wcs_all_forms

20 mars 2023 09:54 - Pierre Ducroquet

Statut:	Fermé	Début:	20 mars 2023
Priorité:	Normal	Echéance:	
Assigné à:	Frédéric Péters	% réalisé:	0%
Catégorie:		Temps estimé:	0:00 heure
Version cible:		Planning:	Non
Patch proposed:	Non		
Description			
La requête suivante (simplifiée pour ce ticket) est particulièrement lente sur l'instance de toulouse:			
<pre>SELECT * FROM wcs_all_forms WHERE status != 'draft' AND ((fts @@ plainto_tsquery('0612345678')) or (fts @@ plainto_tsquery('+3 3612345678')) AND anonymised IS NULL ORDER BY receipt_time DESC LIMIT 20;</pre>			
Le problème est simple : l'optimiseur ne peut pas savoir à quel point le critère sur fts va filtrer ou non le contenu de la table. Il estime donc qu'en la parcourant en suivant l'index sur receipt_time il rencontrera assez vite les 20 enregistrements requis pour répondre à la requête. => Perdu.			
La requête prend donc 8 secondes au lieu de moins de 3 millisecondes, et fait environ 5000 fois plus d'IOs. J'ai plusieurs solutions dans ma besace, je fais pas mal de tests pour identifier ce qui sera le mieux.			
Demandes liées:			
Lié à w.c.s. - Bug #75302: Une recherche sur un numéro dans le listing d'une ...		Fermé	10 mars 2023

Révisions associées

Révision e682f42f - 21 mars 2023 07:54 - Frédéric Péters

sql: revert looking up both raw and normalized values in FTS search (#75594)

Révision 00222aae - 21 mars 2023 07:54 - Frédéric Péters

sql: do not consider numbers not starting with 0 as phone numbers (#75594)

Historique

#2 - 20 mars 2023 10:08 - Pierre Ducroquet

Prod.

Le problème est déclenché par le OR sur le fts.
Il estime que ça va doubler le nombre de lignes renvoyées par la requête, ce qui justifie alors l'utilisation de l'index sur receipt_time.

Réécrire la requête pour transformer le OR en tsquery ne résoud pas le problème, il utilise le tsquery pour évaluer le nombre de lignes et donc on tombe sur le même problème.
Par contre un UNION permet de contourner le problème, ou éventuellement ajouter une opération vierge (coalesce, +0, ...) sur l'order by pour forcer l'utilisation du FTS...

#4 - 20 mars 2023 10:14 - Pierre Ducroquet

- Lié à Bug #75302: Une recherche sur un numéro dans le listing d'une démarche ne retourne plus rien ajouté

#5 - 20 mars 2023 10:26 - Frédéric Péters

Il y a aussi possibilité de se dire qu'on peut juste chercher une valeur; annuler [#75302](#) et le reprendre différemment.

Typiquement il est écrit pour reproduire entrer "974230200934596" et ça pourrait se gérer en incluant uniquement les numéros commençant par un 0 dans la normalisation téléphone.

#6 - 20 mars 2023 11:25 - Benjamin Dauvergne

Il faut reverter jusqu'à [#72773](#) dans ce cas, pour ne pas échanger un bug contre un souci de performance.

#7 - 20 mars 2023 11:27 - Benjamin Dauvergne

J'ai testé `set enable_indexscan=false`, pour ce cas particulier ça marche, mais ça me parait difficilement généralisable.

#8 - 20 mars 2023 11:46 - Pierre Ducroquet

Benjamin Dauvergne a écrit :

J'ai testé `set enable_indexscan=false`, pour ce cas particulier ça marche, mais ça me parait difficilement généralisable.

Pour ce cas mieux vaut écrire la requête avec un UNION ou avec un ORDER BY `receipt_time + interval '0 second'`...

#9 - 20 mars 2023 12:11 - Benjamin Dauvergne

- Fichier `fts_match_optimization.diff` ajouté

Un truc comme ça ? La pente me parait glissante.

#10 - 20 mars 2023 12:17 - Frédéric Péters

- Assigné à *changé de Pierre Ducroquet à Frédéric Péters*

Stop svp.

Je vais "annuler [#75302](#) et le reprendre différemment".

#11 - 20 mars 2023 12:34 - Benjamin Dauvergne

Frédéric Péters a écrit :

Je vais "annuler [#75302](#) et le reprendre différemment".

C'est bon on a trouvé une solution en bossant uniquement coté pg avec les statistiques.

#12 - 20 mars 2023 13:03 - Pierre Ducroquet

Ouais enfin, une solution... c'est pas non plus magique, avec l'évolution des données dans le temps je ne sais pas quand on atteindra un nouveau seuil, et là j'ai mis les stats au maximum viable, ça nous a déjà fait x10 sur le temps de planification, mettre plus ça devient trop onéreux. Au moins maintenant on n'a plus les requêtes lentes sur la prod, mais j'approuve l'idée de Fred de reprendre différemment le correctif.

#13 - 20 mars 2023 13:08 - Benjamin Dauvergne

Ça fera ça à chaque fois qu'on a un OR dans la condition ou est-ce que c'est spécifique à la recherche sur une colonne `ts_vector` ?

#14 - 20 mars 2023 13:27 - Pierre Ducroquet

Benjamin Dauvergne a écrit :

Ça fera ça à chaque fois qu'on a un OR dans la condition ou est-ce que c'est spécifique à la recherche sur une colonne `ts_vector` ?

Un OR ça change radicalement un plan d'exécution, et il est courant de voir un OR plusieurs fois plus lent que l'UNION équivalent (et c'est dommage que l'optimiseur ne sache pas aller dans cette direction). Après la colonne `ts_vector` est un facteur supplémentaire ici puisqu'elle casse encore plus les prédictions, et nous a fait tomber dans le piège de l'order by... Les joies de l'optimisation sur des stats...

#15 - 20 mars 2023 18:06 - Robot Gitea

- Statut *changé de Nouveau à Solution proposée*

Frédéric Péters (fpeters) a ouvert une pull request sur Gitea concernant cette demande :

- URL : <https://git.entrouvert.org/entrouvert/wcs/pulls/177>
- Titre : affiner la détection de numéro de téléphone pour la recherche fts ([#75594](#))
- Modifications : <https://git.entrouvert.org/entrouvert/wcs/pulls/177/files>

#16 - 20 mars 2023 18:10 - Robot Gitea

- Statut *changé de Solution proposée à Solution validée*

Pierre Ducroquet (pducroquet) a approuvé une pull request sur Gitea concernant cette demande :

- URL : <https://git.entrouvert.org/entrouvert/wcs/pulls/177>

#17 - 21 mars 2023 07:54 - Robot Gitea

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

Frédéric Péters (fpeters) a mergé une pull request sur Gitea concernant cette demande :

- URL : <https://git.entrouvert.org/entrouvert/wcs/pulls/177>
- Titre : affiner la détection de numéro de téléphone pour la recherche fts (#75594)
- Modifications : <https://git.entrouvert.org/entrouvert/wcs/pulls/177/files>

#18 - 21 mars 2023 15:14 - Transition automatique

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

#19 - 21 mai 2023 04:42 - Transition automatique

Automatic expiration

Fichiers

fts_match_optimization.diff	3,29 ko	20 mars 2023	Benjamin Dauvergne
-----------------------------	---------	--------------	--------------------