

## Arvados - Bug #18490

### Permission table bloat

11/29/2021 03:47 PM - Peter Amstutz

<b>Status:</b>	Resolved	<b>Start date:</b>	12/01/2021
<b>Priority:</b>	Normal	<b>Due date:</b>	
<b>Assigned To:</b>	Peter Amstutz	<b>% Done:</b>	100%
<b>Category:</b>	API	<b>Estimated time:</b>	0.00 hour
<b>Target version:</b>	2021-12-08 sprint		
<b>Description</b>			
<p>Postgres seems to be creating a lot of fragmentation in the materialized_permissions table in a context with a lot of users (~1000) and large groups. Reported the table size went from 200MB to 2GB to 20 GB in the span of a few days (? or even faster).</p> <p>The group sync script is very slow in this context while there are other updates ongoing.</p> <p>Normal usage (e.g. navigating a large project with lots of subprojects in a read-only arv-mount) is also very slow when the table is fragmented that way.</p> <p>Vacuum full helps but can often not complete because normal usage patterns prevent it from completing.</p> <p>This has been observed on Postgres 9.5 and 13.</p>			
<b>Subtasks:</b>			
Task # 18510: Review			<b>Resolved</b>

#### Associated revisions

##### Revision b0654eb5 - 12/01/2021 10:27 PM - Peter Amstutz

Merge branch '18490-redundant-updates' refs #18490

Arvados-DCO-1.1-Signed-off-by: Peter Amstutz <[peter.amstutz@curii.com](mailto:peter.amstutz@curii.com)>

##### Revision b97c5edf - 12/06/2021 05:13 PM - Peter Amstutz

Merge branch '18490-redundant-updates' refs #18490

Arvados-DCO-1.1-Signed-off-by: Peter Amstutz <[peter.amstutz@curii.com](mailto:peter.amstutz@curii.com)>

#### History

##### #1 - 11/29/2021 03:56 PM - Ward Vandewege

- Description updated

##### #2 - 11/29/2021 04:07 PM - Ward Vandewege

- Description updated

##### #3 - 11/29/2021 04:08 PM - Ward Vandewege

- Description updated

##### #4 - 11/29/2021 09:18 PM - Peter Amstutz

18490-redundant-updates @ [26aa25c76d3ea4285e724fe874c76aa9da03b4c9](https://github.com/Arvados/arvados/commit/26aa25c76d3ea4285e724fe874c76aa9da03b4c9)

Add a where clause to "insert on conflict update" to suppress redundant updates of rows that haven't changed.

[developer-run-tests: #2820](#) [icon?job=developer-run-tests&build=2820](#)

##### #5 - 11/30/2021 06:38 PM - Peter Amstutz

- Assigned To set to Peter Amstutz

- Status changed from New to In Progress

- Category set to API

- Subject changed from Research postgres tuning for tables with lots of updates to Permission table bloat

Feedback from customer after emailing them and they applied a hotfix from #note-4:

"This seems to have solved the issue. We had over 100 new users added to a large group, and the table is currently sitting at 273MB. I'll continue to monitor it to check for any changes but it definitely looks promising."

**#6 - 11/30/2021 09:19 PM - Peter Amstutz**

- *Description updated*

**#7 - 12/01/2021 03:09 PM - Lucas Di Pentima**

My only observation is that it would be convenient to add a comment explaining the workaround, just for future reference. Apart from that, LGTM.

**#8 - 12/01/2021 07:25 PM - Peter Amstutz**

Lucas Di Pentima wrote:

My only observation is that it would be convenient to add a comment explaining the workaround, just for future reference. Apart from that, LGTM.

Yea, that's a good idea. Will do.

**#9 - 12/02/2021 03:06 PM - Peter Amstutz**

- *Status changed from In Progress to Resolved*

**#10 - 12/02/2021 10:33 PM - Peter Amstutz**

- *Release set to 48*