

Arvados - Story #15093

Work with Ops to decide best DB migration strategy for collection file count & size

04/10/2019 04:05 PM - Tom Morris

Status:	Resolved	Start date:	
Priority:	Normal	Due date:	
Assigned To:	Ward Vandewege	% Done:	100%
Category:	API	Estimated time:	0.00 hour
Target version:	2019-06-05 Sprint		
Description			
Related issues:			
Related to Arvados - Story #14484: [API Server] Return collection size and nu...		Resolved	03/28/2019
Blocked by Arvados - Story #15180: [Spike] Test migration on production database		Resolved	

Associated revisions

Revision 0d8adf9b - 05/31/2019 06:51 PM - Ward Vandewege

Move the population of the new columns on the collections table to a standalone script that should be run separate from the migration. Add a note to the upgrade documentation along those lines. Make the script not blow up on collections with invalid manifests, but rather just skip them.

refs #15093
refs #14484

Arvados-DCO-1.1-Signed-off-by: Ward Vandewege <wvandewege@veritasgenetics.com>

Revision 6817a411 - 05/31/2019 07:27 PM - Ward Vandewege

Address review comments.

refs #15093

Arvados-DCO-1.1-Signed-off-by: Ward Vandewege <wvandewege@veritasgenetics.com>

Revision d7775a8a - 05/31/2019 07:30 PM - Ward Vandewege

Merge branch '15093-move-collection-update-out-of-band' into 1.4-dev

closes #15093

Arvados-DCO-1.1-Signed-off-by: Ward Vandewege <wvandewege@veritasgenetics.com>

Revision 10c07dad - 05/31/2019 07:42 PM - Ward Vandewege

Update the 'upgrading' documentation to reflect the v1.4.0 release, and warn about the db migration that can take some time during upgrade.

refs #15093

Arvados-DCO-1.1-Signed-off-by: Ward Vandewege <wvandewege@veritasgenetics.com>

Revision 6f17bfca - 06/03/2019 08:26 PM - Ward Vandewege

Move the population of the new columns on the collections table to a standalone script that should be run separate from the migration. Add a note to the upgrade documentation along those lines. Make the script not blow up on collections with invalid manifests, but rather just skip them.

refs #15093
refs #14484

Arvados-DCO-1.1-Signed-off-by: Ward Vandewege <wvandewege@veritasgenetics.com>

Revision 7f68130d - 06/03/2019 08:26 PM - Ward Vandewege

Address review comments.

refs #15093

Arvados-DCO-1.1-Signed-off-by: Ward Vandewege <wvandewege@veritasgenetics.com>

Revision 65c51513 - 06/03/2019 08:27 PM - Ward Vandewege

Update the 'upgrading' documentation to reflect the v1.4.0 release, and warn about the db migration that can take some time during upgrade.

refs #15093

Arvados-DCO-1.1-Signed-off-by: Ward Vandewege <wvandewege@veritasgenetics.com>

History

#1 - 04/10/2019 04:05 PM - Tom Morris

- Status changed from New to In Progress

#2 - 04/10/2019 04:05 PM - Tom Morris

- Assigned To set to Eric Biagiotti

#3 - 04/10/2019 04:06 PM - Tom Morris

- Related to Story #14484: [API Server] Return collection size and number of files in collection record added

#4 - 04/10/2019 04:21 PM - Tom Morris

- Status changed from In Progress to New

- Assigned To deleted (Eric Biagiotti)

- Target version changed from 2019-04-24 Sprint to To Be Groomed

#5 - 05/01/2019 02:55 PM - Tom Morris

- Blocked by Story #15180: [Spike] Test migration on production database added

#6 - 05/31/2019 06:48 PM - Ward Vandewege

- Category set to API

- Status changed from New to In Progress

- Assigned To set to Ward Vandewege

- Target version changed from To Be Groomed to 2019-06-05 Sprint

#7 - 05/31/2019 06:54 PM - Ward Vandewege

15093-move-collection-update-out-of-band up for review at [0d8adf9bb22e169fb47d64ca19045283ff5fafd1](#)

#8 - 05/31/2019 07:21 PM - Eric Biagiotti

Ward Vandewege wrote:

15093-move-collection-update-out-of-band up for review at [0d8adf9bb22e169fb47d64ca19045283ff5fafd1](#)

Just a minor note. Because of the addition of the file_count=0 in the where clause, the distinct_pdh_count query could potentially return a number greater than the total number of pdhs processed and the logging in group_pdhs_for_multiple_transactions would be inaccurate. This doesn't affect actual collection processing though. Could just add the same where clause to distinct_pdh_count query.

Otherwise, this LGTM.

#9 - 05/31/2019 08:09 PM - Ward Vandewege

- Status changed from In Progress to Resolved

- % Done changed from 0 to 100

Applied in changeset [arvados|d7775a8a4f85771a4bf2ac6165561f0d4393060a](#).