

D4Science Infrastructure - Task #13050

Try to improve CKAN 2.6.X performance by adding fixes released for CKAN 2.7.X

Dec 14, 2018 05:34 PM - Francesco Mangiacrapa

Status:	Closed	Start date:	Dec 14, 2018
Priority:	Normal	Due date:	
Assignee:	_InfraScience Systems Engineer	% Done:	100%
Category:	Other	Estimated time:	0.00 hour
Target version:	Data Catalogue		
Infrastructure:	Development		
Description			
<p>We know that the CKAN system 2.6.X (currently installed for the gCube Catalogues) degrades its performance after a certain number of dataset updating.</p> <p>The source of the issue (as investigated in the past) should be SQL queries badly written.</p> <p>I noticed some fixes in this sense released for CKAN 2.7.X (at https://docs.ckan.org/en/2.8/changelog.html#v2-7-0-2017-08-02) and in particular the following ones:</p> <ul style="list-style-type: none">• "Query performance improvements" at https://github.com/ckan/ckan/pull/3430• "Performance improvements in group_activity_query" at https://github.com/ckan/ckan/pull/3466 <p>We should investigate on how to integrate them with CKAN 2.6.X</p>			

History

#1 - Jan 09, 2019 04:10 PM - Andrea Dell'Amico

- Status changed from New to In Progress

I'm going to duplicate the sources on the dev instance and I'll try to produce a patch.

#2 - Jan 09, 2019 04:40 PM - Andrea Dell'Amico

- % Done changed from 0 to 20

I can anticipate that the changes here: <https://github.com/ckan/ckan/pull/3430/commits/5d6949ca7602eb5039227229a0113269d213f856#diff-b45b33cec4c7aad7f1308caf3174bccb> cannot be applied other than the one to resource_item.html because the files are completely different. We could try to use the new files, but I'd try the other modifications first.

#3 - Jan 09, 2019 04:52 PM - Andrea Dell'Amico

The changes are active on ckan-d-d4s.d4science.org. @francesco.mangiacrapa@isti.cnr.it can you test if the involved parts are working as expected?

Let me also know if we want to try the modifications to the template files.

#4 - Jan 09, 2019 04:56 PM - Andrea Dell'Amico

- File ckan_performance_fixes_13050.diff added

I'm attaching the diff file.

#5 - Jan 09, 2019 05:06 PM - Andrea Dell'Amico

- File ckan_performance_fixes_13050-1.diff added

A fixed one, the first ad a syntax error in it.

#6 - Jan 09, 2019 05:07 PM - Andrea Dell'Amico

- File deleted (ckan_performance_fixes_13050.diff)

#7 - Jan 09, 2019 05:59 PM - Andrea Dell'Amico

- File ckan_performance_fixes_plus_templates13050.diff added

New patch that includes the changes to the template snippets.

#8 - Jan 10, 2019 04:07 PM - Francesco Mangiacrapa

Andrea Dell'Amico wrote:

New patch that includes the changes to the template snippets.

Hi Andrea, thanks for your work.
Analyzing the attached file https://support.d4science.org/attachments/download/2997/ckan_performance_fixes_plus_templates13050.diff I think we can ignore the changes regarding the template snippets. They could cause malfunctions...
I would try to update just one CKAN 2.6.X of production by adding the above patches without the template changes (e.g. the CKAN <https://ckan-ecoevo.d4science.org/>). If it is OK then a good benchmark could be the CKAN SBD (<https://ckan-sobigdata.d4science.org/dataset>), as you can see by selecting a dataset (e.g. https://ckan-sobigdata.d4science.org/dataset/forecasting_the_market_value_of_soccer_players_from_soccer-logs_and_social_media_data) it returns in 6-7 sec, but I don't know if the changes will have a good effect on this.

#9 - Jan 10, 2019 08:21 PM - Andrea Dell'Amico

- % Done changed from 20 to 70

OK. The playbook is ready, let's coordinate us to install the patch when a restart of the server is not a problem.

#10 - Jan 11, 2019 12:23 PM - Andrea Dell'Amico

- Status changed from In Progress to Closed

- % Done changed from 70 to 100

Tried on both ckan-ecoevo and ckan-sobigdata, both the version with and without the template changes. They do not break anything, but also do not solve the problem with the datasets query. I'll leave the installation of the patch that includes the template changes as it still optimises a couple of queries, but there's no urgency to install it everywhere.

Files

ckan_performance_fixes_13050-1.diff	5.37 KB	Jan 09, 2019	Andrea Dell'Amico
ckan_performance_fixes_plus_templates13050.diff	7.25 KB	Jan 09, 2019	Andrea Dell'Amico