QGIS Application - Bug report #21839
QgsDataSourceUri parameter checkPrimaryKeyUnicity is ignored for postgres provider

2019-04-11 11:24 PM - Tom Vanzieleghem

Status:	Closed		
Priority:	Normal		
Assignee:	Alessandro Pasotti		
Category:	Data Provider/PostGIS		
Affected QGIS version:	3.7(master)	**Regression?:**	No
Operating System:		**Easy fix?:**	No
Pull Request or Patch supplied:	Yes	**Resolution:**	fixed/implemented
Crashes QGIS or corrupts data:	No	**Copied to github as #:**	29654

Description

Initial loading time of large PostgreSQL views (> 10 million records) seems to have increased considerably compared to QGis 2. After observing the db logs I noticed the sql queries for checking primary key unicity is one of the causes for this. From the API documentation I had the impression that this behaviour could be disabled by setting the property *TrustLayerMetadata* on the project to true (in the GUI under project properties). This did not make a difference when loading new PG layers.

So I tried setting the parameter *checkPrimaryKeyUnicity* to '0' in several ways (editing an exported layer definition file and through pyqgis) but it did not make a difference.
Seems that, in case of a postgres provider, the parameter *checkPrimaryKeyUnicity* set on the QgsDatasourecUri instance is changed depending on the *layerOptions* (more specific the readExtentFromXml option) specified on the QgsVectorLayer class. The parameter also seems to disappear in the uri string after loading into the canvas and/or saving as a layer definition file.

I always was impressed with the loading of postgis layers in QGis 2 but this has considerably changed in QGis 3. Is it posssible to include an option in the GUI somewhere to prevent the checks on primaryKeyUnicity or resolve this in another way?

best regards and keep up the good work,

Tom

Associated revisions

Revision 38c91e4e - 2019-04-24 05:14 PM - Alessandro Pasotti

[postgres] Fix checkPrimaryKeyUnicity option

This provider option was linked to the project level option
"Trust layer metadata..." which was implemented
to speed up loading of large dataset by trusting extent
read from metadata to avoid costly operations to determine
the layer extent.

Check PK unicity on the other hand has only effect on views
and query layers and it is useful as an independent
option to prevent loading of layers that have no PK (or the
wrong one).

But the operation of determine unicity of a values in a column
can also be costly, so better to get control back to the user.

Legacy default is preserved (the project-level "Trust..." option).

Fixes #21839

Revision 4009b2d4 - 2019-04-24 05:14 PM - Alessandro Pasotti

[postgres] Fix checkPrimaryKeyUnicity option

This provider option was linked to the project level option
"Trust layer metadata..." which was implemented
to speed up loading of large dataset by trusting extent
read from metadata to avoid costly operations to determine
the layer extent.

Check PK unicity on the other hand has only effect on views
and query layers and it is useful as an independent
option to prevent loading of layers that have no PK (or the
wrong one).

But the operation of determine unicity of a values in a column
can also be costly, so better to get control back to the user.

Legacy default is preserved (the project-level "Trust..." option).

Fixes #21839

History

#1 - 2019-04-11 11:30 PM - Giovanni Manghi

- Easy fix? changed from Yes to No

Interesting... if this will not get traction I suggest you to raise the matter in the developers mailing list. Thanks!

#2 - 2019-04-17 04:52 PM - Regis Haubourg

> *This did not make a difference when loading new PG layers*

Giovanni Manghi wrote:

> *Interesting... if this will not get traction I suggest you to raise the matter in the developers mailing list. Thanks!*

Hi, "trust" option is only acting when loading a project file I think. The use case was to speed up opening big project with data having no metadata, either in
desktop or in server context.

#3 - 2019-04-17 05:02 PM - Regis Haubourg

And in I confirm that whatever the "select at id" value you set, QGIS will still throw a `select count(distinct(..` query to check unicity.

I have no idea why this could be slower in QGIS3 for huge tables though.

The proposal to add a column with a checkbox "don't check unicity on load" would be the way to go. however this PostgreSQL connection dialog is already overcrowded, and very few users understand the use of "select at id" option. I think we should work on tooltips and UI design here.

#4 - 2019-04-24 09:32 AM - Alessandro Pasotti

- Operating System deleted (Windows 10)

- Affected QGIS version changed from 3.4.6 to 3.7(master)

- Assignee set to Alessandro Pasotti

#5 - 2019-04-26 09:41 AM - Alessandro Pasotti

- Resolution set to fixed/implemented

- Pull Request or Patch supplied changed from No to Yes

- Status changed from Open to In Progress

PR https://github.com/qgis/QGIS/pull/9873

#6 - 2019-04-27 07:00 PM - Alessandro Pasotti

- Status changed from In Progress to Closed

- % Done changed from 0 to 100

Applied in changeset commit:qgis|38c91e4e9715321b0b1164eea3bca82b45eb830c.