QGIS Application - Bug report #18342
Loading FileGDB's are really really slow

2018-03-05 04:02 PM - Casper Børgesen

Status:	Closed		
Priority:	High		
Assignee:	Even Rouault		
Category:	Data Provider/OGR		
Affected QGIS version:	3.1(master)	**Regression?:**	Yes
Operating System:	windows 10 x64	**Easy fix?:**	No
Pull Request or Patch supplied:	No	**Resolution:**	
Crashes QGIS or corrupts data:	No	**Copied to github as #:**	26232

Description

I have installed QGIS from OSGeo4W and I'm trying to load an empty ESRI File Geodatabase. The GDB has around 20 feature types each with a many attributes.

To load the GDB I drag the GDB folder into QGIS and I measured the time from dropping the folder until I got the dialog box asking me to select which feature types to load as layers.

I got the following timings:

  - QGIS 3.1 (871132eae6): 6 minutes
  - QGIS 3.0: 4½ minute
  - QGIS 2.18.17: 7 seconds

I have installed the following FileGDB drivers from OSGeo4W:

  - gdal-dev-filegdb (2.3.0-41580)
  - gdal-filegdb (2.2.3-4)

I don't know the reason for this huge increase in time consumption when using file geodatabases, but could it be because QGIS uses the Open FileGDB driver instead of the proprietary ESRI driver?

Associated revisions

Revision 55aa7a86 - 2018-06-03 01:35 AM - Even Rouault

[OGR provider] Improve performance of subLayers(), particularly on FileGDB with the proprietary driver (fixes #18342)

Revision 7f97a2f3 - 2018-06-03 12:41 PM - Even Rouault

Merge pull request #7146 from rouault/fix_18342

[OGR provider] Improve performance of subLayers(), particularly on FileGDB with the proprietary driver (fixes #18342)

History

#1 - 2018-03-06 09:32 AM - Jérôme Guélat

Confirmed here with QGIS 3.0 and 2.18.17... The opening time seems to be similar with the ESRI FileGDB driver.

Interestingly the problem doesn't happen if you use the browser to open the File Geodatabase in QGIS 3.0.

#2 - 2018-03-06 09:46 AM - Giovanni Manghi

- Status changed from Open to Feedback

- Priority changed from Normal to High

- Regression? changed from No to Yes

Can you attach sample data? thanks!

#3 - 2018-03-06 10:51 AM - Casper Børgesen

- Status changed from Feedback to Open

- File testdata.gdb.zip added

I have created a new dummy data set with 40 feature types inside a schema and each feature type has 20 attributes.

Using this data set I get the following timings:

  - QGIS 3.1 (871132eae6): 30 seconds
  - QGIS 3.0: 25 seconds
  - QGIS 2.18.17: 1 second

#4 - 2018-03-06 12:19 PM - Giovanni Manghi

- Status changed from Open to Feedback

I just tried your dataset in my Win 10 testing VM and doing what you say (" I measured the time from dropping the folder until I got the dialog box asking me to select which feature types to load as layers") and is almost instantaneous. Anyway your layers seems empty, is that right?

#5 - 2018-03-06 12:46 PM - Casper Børgesen

- Status changed from Feedback to Open

Giovanni Manghi wrote:

> *I just tried your dataset in my Win 10 testing VM and doing what you say (" I measured the time from dropping the folder until I got the dialog box asking me to select which feature types to load as layers") and is almost instantaneous. Anyway your layers seems empty, is that right?*

Yes, there are no features in the data set.

#6 - 2018-03-06 12:49 PM - Giovanni Manghi

- Status changed from Open to Feedback

> *Yes, there are no features in the data set.*

so... have you tried on another system or doing a clean install?

#7 - 2018-03-06 12:57 PM - Casper Børgesen

- Status changed from Feedback to Open

Giovanni Manghi wrote:

> *Yes, there are no features in the data set.*
>
> *so... have you tried on another system or doing a clean install?*

I have just tried on two more powerful machines and they both take around 13 seconds in QGIS 3, while still only use about 1 second in QGIS 2.18.17.

#8 - 2018-03-06 01:00 PM - Giovanni Manghi

- Status changed from Open to Feedback

> *I have just tried on two more powerful machines and they both take around 13 seconds in QGIS 3, while still only use about 1 second in QGIS 2.18.17.*

so is not instantaneous as here, but also is not minutes as in the original report. What about after adding the layers, any difference (in performance)?

#9 - 2018-03-06 01:13 PM - Casper Børgesen

- Status changed from Feedback to Open

Giovanni Manghi wrote:

> *I have just tried on two more powerful machines and they both take around 13 seconds in QGIS 3, while still only use about 1 second in QGIS 2.18.17.*
>
> *so is not instantaneous as here, but also is not minutes as in the original report. What about after adding the layers, any difference (in performance)?*

I haven't added the layers since they doesn't contain anything. I was trying to keep the issue simple, since just opening a data set doesn't include panning or zooming or other user interactions to make performance measurements more complex.

But I guess something has happened from 2.18 to 3.x that significantly increases time consumption when trying to load a GDB data set.

Do you use your own compiled QGIS or the one from OSGeo?

#10 - 2018-03-06 01:31 PM - Giovanni Manghi

- Status changed from Open to Feedback

> *But I guess something has happened from 2.18 to 3.x that significantly increases time consumption when trying to load a GDB data set.*

I can't see any difference here (between 2.18 and 3, with your dataset)

> *Do you use your own compiled QGIS or the one from OSGeo?*

osgeo4w on a clean testing Win10 VM.

#11 - 2018-03-06 01:55 PM - Jérôme Guélat

I'm having the same problem using a clean install (from OSGeo4W) on Windows 7.

#12 - 2018-03-06 01:57 PM - Giovanni Manghi

Jérôme Guélat wrote:

> *I'm having the same problem using a clean install (from OSGeo4W) on Windows 7.*

are your datasources in the local disk? network share? else?

#13 - 2018-03-06 02:14 PM - Casper Børgesen

Giovanni Manghi wrote:

> *Jérôme Guélat wrote:*
>
>> *I'm having the same problem using a clean install (from OSGeo4W) on Windows 7.*
>
> *are your datasources in the local disk? network share? else?*

Mine are located on a local SSD.

#14 - 2018-03-07 08:26 AM - Jérôme Guélat

Mine are located on a network drive.

#15 - 2018-03-07 09:58 AM - Giovanni Manghi

Jérôme Guélat wrote:

> *Mine are located on a network drive.*

and if you copy them on a local drive?

#16 - 2018-03-07 09:59 AM - Giovanni Manghi

> *I have just tried on two more powerful machines and they both take around 13 seconds in QGIS 3, while still only use about 1 second in QGIS 2.18.17.*

should we change the description (that reports several minutes, not seconds)?

#17 - 2018-03-07 10:41 AM - Casper Børgesen

Giovanni Manghi wrote:

> *I have just tried on two more powerful machines and they both take around 13 seconds in QGIS 3, while still only use about 1 second in QGIS 2.18.17.*
>
> *should we change the description (that reports several minutes, not seconds)?*

I am working on creating a data set matching the one I used in my initial timings. I will report back with my results.

#18 - 2018-03-07 10:55 AM - Jérôme Guélat

Giovanni Manghi wrote:

> *Jérôme Guélat wrote:*
>
>> *Mine are located on a network drive.*
>
> *and if you copy them on a local drive?*

Same problem with the dataset on a local SSD.

#19 - 2018-03-07 04:21 PM - Casper Børgesen

- File testdata_2.gdb.zip added
- Status changed from Feedback to Open

I have created a data set (testdata_2) that matches my initial timings more accurately. The data set has no features but a lot of feature types with corresponding attributes and some of them has domain values.
Laptop timings (corresponding to the original):
  - QGIS 3.1: 03:53
  - QGIS 3.0: 04:01
  - QGIS 2.18.17: 00:06
Workstation (corresponding to the latter timings):
  - QGIS 3.1: 01:50
  - QGIS 3.0: 01:37
  - QGIS 2.18.17: 00:03

#20 - 2018-06-02 11:23 PM - Even Rouault

- Category changed from Data Provider to Data Provider/OGR

- Assignee set to Even Rouault

#21 - 2018-06-03 12:41 PM - Even Rouault

- % Done changed from 0 to 100

- Status changed from Open to Closed

Applied in changeset commit:qgis|55aa7a86be56ce098a6c0596d4e4bb191c3a52f5.

#22 - 2018-06-04 09:49 AM - Casper Børgesen

Even Rouault wrote:

> *Applied in changeset commit:qgis|55aa7a86be56ce098a6c0596d4e4bb191c3a52f5.*

Hi Even, Thanks a lot.

I can confirm that the change really improves the loading performance! It is still 2-3 seconds slower than with QGIS 2.18 but that is negligible compared to the previous timings.

Files

testdata.gdb.zip	79.4 KB	2018-03-06	Casper Børgesen
testdata_2.gdb.zip	110 KB	2018-03-07	Casper Børgesen