QGIS Application - Bug report #19774
Missing attributes in large table

2018-09-05 03:42 PM - belg4mit -

Status:	Closed			
Priority:	Normal			
Assignee:				
Category:	Attribute table			
Affected QGIS version:	3.2.2		**Regression?:**	No
Operating System:	Windows 10		**Easy fix?:**	No
Pull Request or Patch supplied:			**Resolution:**	up/downstream
Crashes QGIS or corrupts data:	No		**Copied to github as #:**	27599

Description

I have a shapefile with a table that slightly exceeds two gigabytes. When I open the file in QGIS (2.18 and 3.2) many of the (later-loading) features have null attributes, however the file works fine in ArcMap.

The shapefile in question (97MB ZIP) is available here: [[
https://nmrgroupinc-my.sharepoint.com/:u:/p/jpierce/EYs9N_pNQ_VMnkYv_QwOrRMBKQxtrfWrh-1IeCYo3y467w?e=U8WX1p]]

History

#1 - 2018-09-05 03:53 PM - Giovanni Manghi

- Status changed from Open to Feedback

Can you make an example of how a record/feature looks like and how it should look like?
I loaded now in QGIS master/Linux and seems ok to me.

#2 - 2018-09-05 06:24 PM - belg4mit -

- File Null.png added

- File Attributes.PNG added

- File Null attributes.PNG added

Here are the requested screenshots.

My hunch is that is has something to do with the 2GB file size of the DBF.

#3 - 2018-09-06 12:38 PM - Giovanni Manghi

belg4mit - wrote:

> Here are the requested screenshots.
>
> My hunch is that is has something to do with the 2GB file size of the DBF.

Now I see the images I realize that here (QGIS master and 2.18, Windows and Linux) the points that in your case are classified in red (because of the NULLs) are never loaded. The vector metadata shows 2+ million records, but opening the attributes table is shows only a limited amount. The header says that records were filtered (but no filter was applied). I'm wondering if here there are several different issues.

#4 - 2018-09-06 01:06 PM - Giovanni Manghi

When using ogr2ogr to convert the sahepfile (to GPKG for example) it returns

ERROR 1: fread(1754) failed on DBF file.

not sure it means the DBF file is corrupted (the limit is 4gb, not 2, afaik). But also is similar to https://trac.osgeo.org/gdal/ticket/4203

I believe this issue should be raised first in the gdal/ogr mailing list and see what they say.

#5 - 2018-09-21 09:45 PM - Even Rouault

Reported to GDAL as https://github.com/OSGeo/gdal/issues/937

#6 - 2018-09-22 03:12 PM - Even Rouault

- Resolution set to up/downstream

#7 - 2018-09-24 10:26 AM - Jukka Rahkonen

Giovanni Manghi wrote:

> *When using ogr2ogr to convert the sahepfile (to GPKG for example) it returns*
>
> *ERROR 1: fread(1754) failed on DBF file.*
>
> *not sure it means the DBF file is corrupted (the limit is 4gb, not 2, afaik). But also is similar to https://trac.osgeo.org/gdal/ticket/4203*
>
> *I believe this issue should be raised first in the gdal/ogr mailing list and see what they say.*

The fread()error in https://trac.osgeo.org/gdal/ticket/4203 was a different case and for making it happen you'd need to use an ancient version of shapelib <1.2.7 together with GDAL that is more than 4 years old (prior https://trac.osgeo.org/gdal/changeset/27577). That error was a systematic one. The shx part was always written to have a different content length value than the shp part. For users that meant that the last record of the shapefile was dropped in reading.

There are no technical limits in the size of DBF file. I wrote a test case into https://github.com/OSGeo/gdal/issues/937 that creates a shapefile with 18 GB sized .dbf part and GDAL and QGIS can read it. It is still not a valid shapefile because ESRI has specified that no part in shapefile is allowed to exceed 2 GB.

I guess that what happened is:
- QGIS tried to write a shapefile with so much attribute data that DBF part exceeded 2 GB
- QGIS applies the GDAL setting "2GB_LIMIT=YES" and at 2 GB limit the writing of DBF was stopped but writing of .shp was continued
- As a result the final .shp has more geometries than .dbf has rows

If there is a bug I would say that it is in the writing of shapefile. Now it seems that writer is using the option 2GB_LIMIT=YES (see https://www.gdal.org/drv_shapefile.html) but when it meets the 2 GB limit with .dbf it does not error. But because the writer adds data sequentially into the data files of shapefile the total rollback would require deletion of the .shp, .shx, .dfb, and .prj files from the file system.

#8 - 2018-09-24 10:52 AM - Even Rouault

@Jukka QGIS doesn't mess with setting 2GB_LIMIT=YES. There is all evidence from the dataset that is was produced by ESRI software as it contains sbn, sbx, mxd, shp.xml files etc. So I suspect that the corruption comes from non-GDAL & non-QGIS side.

#9 - 2018-10-05 07:45 PM - belg4mit -

- Status changed from Feedback to Closed

Yes, it turns out the errors was ESRI's.

Files

Null.png	2.73 MB	2018-09-05	belg4mit -
Attributes.PNG	104 KB	2018-09-05	belg4mit -
Null attributes.PNG	107 KB	2018-09-05	belg4mit -

@Jukka QGIS doesn't mess with setting 2GB_LIMIT=YES. There is all evidence from the dataset that is was produced by ESRI software as it contains sbn, sbx, mxd, shp.xml files etc. So I suspect that the corruption comes from non-GDAL & non-QGIS side.