QGIS Application - Bug report #6483
Qgis crashes when trying to classify huge vector (postgis db) with Natural Breaks(Jenks)

2012-10-08 08:01 AM - Nikos Ves

Status:	Closed		
Priority:	High		
Assignee:	Vincent Mora		
Category:	Symbology		
Affected QGIS version:	master	**Regression?:**	No
Operating System:		**Easy fix?:**	No
Pull Request or Patch supplied:	Yes	**Resolution:**	fixed
Crashes QGIS or corrupts data:	No	**Copied to github as #:**	15711

Description

When trying to classify a vector from a postgis DB, with natural breaks symbology, Qgis crashes without a warning.

Feature type is multipolygon with row count about ~20k entries and NULL values filtered out. The column type is of type "double precision". Values range from 0 to 1.

The same function does not crash for columns of type "integer"

I'm using linux (ubuntu 12.04) and git qgis version "dff6b84"

uname -a > Linux 3.2.0-31-generic-pae #50-Ubuntu SMP Fri Sep 7 16:39:45 UTC 2012 i686 i686 i386 GNU/Linux

video showing the crash -> http://tinyurl.com/8b9gvkm (35mb)

Associated revisions

Revision 7e171ef7 - 2013-09-15 09:18 AM - Tim Sutton

Merge pull request #720 from Oslandia/issue6483

Fix #6483 - Thanks Vincent!

History

#1 - 2012-12-30 09:46 AM - Giovanni Manghi

- Priority changed from Normal to High

#2 - 2013-07-08 02:04 AM - Vincent Mora

I'd like a dataset to reproduce the bug.

#3 - 2013-07-08 09:51 AM - Giovanni Manghi

- Status changed from Open to Feedback

Just tested a 50k polygon layer, double precision column, values between 0 and 1, natural breaks, no crash.

Please test the latest master and if it happens again then attach sample data. Thanks!

#4 - 2013-07-11 05:01 AM - Vincent Mora

- Assignee set to Vincent Mora

#5 - 2013-07-11 05:58 AM - Vincent Mora

- File test_qgsissue6483.py added

Created a SL DB with provided script.

In qgis (today's master 56210eb0b5c),

- Add SpatiaLite Layer

- Properties->Style->Graduated and Mode->Natural Breaks (Jenks) OK

No problem.

#6 - 2013-07-11 07:57 AM - Giovanni Manghi

Vincent Mora wrote:

> *Created a SL DB with provided script.*
>
> *In qgis (today's master 56210eb0b5c),*
> *- Add SpatiaLite Layer*
> *- Properties->Style->Graduated and Mode->Natural Breaks (Jenks) OK*
>
> *No problem.*

you have to test with a very large vector. I tested with a 3 million record one, and it freezes with natural breaks, while it is ok with equal interval (don't know about the other modes). Not sure what is the number of records that start to be trouble.

#7 - 2013-07-11 08:34 AM - Vincent Mora

- Pull Request or Patch supplied changed from No to Yes

https://github.com/qgis/Quantum-GIS/pull/717

#8 - 2013-07-11 08:36 AM - Giovanni Manghi

Vincent Mora wrote:

> *https://github.com/qgis/Quantum-GIS/pull/717*

this patch does not seems related to this ticket :)

#9 - 2013-07-11 08:40 AM - Vincent Mora

- Pull Request or Patch supplied changed from Yes to No

Sorry for that, it was for #7244. Working on this one right now.

#10 - 2013-07-11 09:31 AM - Vincent Mora

Giovanni: **what do you mean by "freezes"** ? Is it still computing or crashes ?

I'm trying with a few millions polygons (qgis compiled in debug), it hangs (computing) but no crash yet (15min in the process).

Apparently this classification (Jenks) is an iterative optimization process... The **complexity is k x n^2**, where k is the nb of classes (the default is 5) and **n is the number of records** (http://wiki.objectvision.nl/index.php/Fisher%27s_Natural_Breaks_Classification).

If it crashed it's a problem, but classifying 2M records with this method can surely take a long time. I'll give qgis a chance to compute a result before tomorrow morning, but I'm not really holding my breath.

#11 - 2013-07-11 10:35 AM - Giovanni Manghi

Vincent Mora wrote:

> Giovanni: **what do you mean by "freezes"** ? Is it still computing or crashes ?

freezes means that qgis become not responsive for a long time, and killing the program is the only exit

> I'm trying with a few millions polygons (qgis compiled in debug), it hangs (computing) but no crash yet (15min in the process).
> Apparently this classification (Jenks) is an iterative optimization process... The **complexity is k x n^2**, where k is the nb of classes (the default is 5) and **n is the number of records** (*http://wiki.objectvision.nl/index.php/Fisher%27s_Natural_Breaks_Classification*).
>
> If it crashed it's a problem, but classifying 2M records with this method can surely take a long time. I'll give qgis a chance to compute a result before tomorrow morning, but I'm not really holding my breath.

My 2 cents:

we need to understand what is the (more or less) number of records that makes qgis freeze.

These days datasets of >100k records are pretty usual, so I would say that at least up to 200/300k records the classification should work. For more records than the problematic limit a warning should be issued before computing the classes, so it would be a user decision to wait a lot of time, or not.

#12 - 2013-07-11 01:24 PM - Vincent Mora

It took a couple hours (more or less, I wasn't waiting in from of the machine) but my 2M records where eventually sorted out, **no crash**. But it surely wasn't responsive for this time.

Giovanni, I agree with you: an acceptable solution would be to pop-up a warning window with a time estimate which should be a function of the nb of records (*a priori* n^2).

I can try a few datasets with increasing nb of records tomorrow to try and model this processing time in order to validate the n^2 hypotheses.

Of course processing time will depend of your machines config, and maybe of your record length (not sure). So if we want an accurate enough estimate we may need to pre-process a subset of the data before popping-up the warning (and use the model to extrapolate for the complete set).

Another solution (simpler I guess) is to use the size of the dataset and modify the Natural Breaks entry of the drop-down-list accordingly:

- long processing for more than 10k records,

- very long for more than 100k,

- do that before lunch break for more than 1M,

- do that before going home for more than 10M,

- don't think about it, consider yourself lucky you got this far for 100M and above.

Of course, since I only tried in debug, I may be completely off with the dataset sizes, but I guess you got the idea :)

#13 - 2013-07-12 01:08 AM - Vincent Mora

- File timing_jenks.pdf added

So it's n^2 complexity all right ! (see the attached plot).

I'm going to try and implement the proposed warning.

#14 - 2013-07-12 02:06 AM - Vincent Mora

- Pull Request or Patch supplied changed from No to Yes

https://github.com/qgis/Quantum-GIS/pull/720

#15 - 2013-07-12 02:52 AM - Vincent Mora

- Crashes QGIS or corrupts data changed from Yes to No

BTW, since a crash was observed with a **postgis DB**, I just tryied with 80k and 640k datasets, for the latter it sure takes a few minutes, but it classifies ok, **no crash**.

#16 - 2013-07-16 03:18 AM - Hugo Mercier

- Resolution set to fixed

- Status changed from Feedback to Resolved

#17 - 2013-07-17 04:10 AM - Giovanni Manghi

- Resolution deleted (fixed)

- Status changed from Resolved to Open

Vincent Mora wrote:

> *BTW, since a crash was observed with a **postgis DB**, I just tryied with 80k and 640k datasets, for the latter it sure takes a few minutes, but it classifies ok, **no crash**.*

The patch that solves this has not yet been committed

https://github.com/qgis/Quantum-GIS/pull/720

re-close if I'm wrong :)

#18 - 2013-07-17 04:39 AM - Hugo Mercier

The patch has not yet been committed, you're right.

I thought "resolved" represented a state where a patch had been supplied (a pull request here), but waiting for commiters to accept it or discuss it.

The strict workflow is not very clear for me. Can we agree on something like https://issues.qgis.org/wiki/quantum-gis/Tickets_workflow ?

#19 - 2013-07-17 07:48 AM - Giovanni Manghi

- Status changed from Open to Resolved

- Resolution set to fixed

Hugo Mercier wrote:

> *The patch has not yet been committed, you're right.*
>
> *I thought "resolved" represented a state where a patch had been supplied (a pull request here), but waiting for commiters to accept it or discuss it.*
>
> *The strict workflow is not very clear for me. Can we agree on something like https://issues.qgis.org/wiki/quantum-gis/Tickets_workflow ?*

Hi Hugo, that's fine for me.

We can eventually discuss how to simplify the workflow at the next dev meeting.

#20 - 2013-09-15 12:18 AM - Tim Sutton

- Status changed from Resolved to Closed

Fixed in changeset commit:"7e171ef7b0eec2101c60671e7625cfea52688a7e".

#21 - 2013-09-15 02:10 AM - Vincent Mora

Bug fixed thanks to fundings from Agence de l'Eau Adour-Garonne.

Files

test_qgsissue6483.py	2.76 KB	2013-07-11	Vincent Mora
timing_jenks.pdf	13.9 KB	2013-07-11	Vincent Mora