QGIS Application - Feature request #8850
Oracle not looking in USER_SDO_GEOM_METADATA for extents

2013-10-15 04:17 AM - Jonathan Moules

Status:	Closed		
Priority:	Normal		
Assignee:	Jürgen Fischer		
Category:	Data Provider/Oracle		
Pull Request or Patch supplied:	No	**Resolution:**	fixed/implemented
Easy fix?:	No	**Copied to github as #:**	17532

Description

I have an Oracle table with 7.3million features in it that is spatially indexed. When I add the layer to QGIS, even though I am viewing an area with just 333 features on it, it still takes a full 3 (three!) minutes to add the layer to QGIS once I have pressed "add". During that period QGIS acts like it has crashed (Windows 7, x64).

I appreciate that QGIS is probably getting table extents (SDO_TUNE.EXTENT_OF takes anything from 60-150 seconds to run on this table), but if I have dozens of users trying to load the same data, that's going to be a massive waste of time/processing.

I have "Only look in meta data table" selected, so it should be getting the extents from USER_SDO_GEOM_METADATA. I know Oracle doesn't keep that up to date automagically, but we keep on top of that which is why that checkbox is ticked.

So I'd suggest:
a) If "Only look in meta data table" then use USER_SDO_GEOM_METADATA for extents.
b) If not selected then do the SDO_TUNE.EXTENT_OF that I suspect is being done currently.

That way folks with large datasets in corporate environments can use QGIS.

Related issues:

Duplicated by QGIS Application - Feature request # 9591: Oracle connection to...	**Closed**	**2014-02-18**

Associated revisions

Revision 45fea7b4 - 2013-11-05 06:26 PM - Jürgen Fischer

oracle provider: optionally use extents from metadata table (implements #8850)

History

#1 - 2013-10-15 04:37 AM - Jukka Rahkonen

Sounds good, knowing the exact extents of the layer is not really often so important. Actually the extents which are stored into metadata may be more important with Oracle because it is not possible to store new features outside the extents.

One detail, it would be better to use ALL_SDO_GEOM_METADATA, especially in corporate environments QGIS users do not necessarily own the tables they are using.

#2 - 2013-10-15 04:50 AM - Jonathan Moules

Good point on "ALL_SDO_GEOM_METADATA".

#3 - 2013-11-05 08:47 AM - Jonathan Moules

I believe it's a result of this, but if I have a QGIS project with multiple Oracle layers in it. When I load the project QGIS hangs for several minutes while waiting to establish all the connection/metadata stuff - even though none of the layers are enabled.

Unfortunately this basically renders QGIS un-useable with larger database Oracle sets. We're having to use our WMS as an interim solution.

#4 - 2013-11-05 09:26 AM - Jürgen Fischer

- Tracker changed from Bug report to Feature request

#5 - 2013-11-05 09:32 AM - Jürgen Fischer

- Status changed from Open to Closed

- Resolution set to fixed/implemented

implemented in commit:45fea7b46

#6 - 2013-11-06 03:01 AM - Jonathan Moules

Many thanks Jürgen - I'll test it in the next stable. :-)

#7 - 2013-11-06 03:04 AM - Jürgen Fischer

Jonathan Moules wrote:

> *Many thanks Jürgen - I'll test it in the next stable. :-)*

So there's no need for the weeklies anymore?

#8 - 2013-11-06 03:21 AM - Jonathan Moules

Hmmm. I didn't know they were still going on. I can't find them on the new website - http://www.qgis.org/en/site/forusers/alldownloads.html doesn't have them.

#9 - 2013-11-06 03:29 AM - Jürgen Fischer

Jonathan Moules wrote:

> *Hmmm. I didn't know they were still going on. I can't find them on the new website - http://www.qgis.org/en/site/forusers/alldownloads.html doesn't have them.*

No change there: http://qgis.org/downloads/weekly/

#10 - 2013-11-06 03:30 AM - Jonathan Moules

Many thanks. I'll try and test it next week then.

#11 - 2014-02-21 06:58 AM - Jonathan Moules

I don't think this has actually been implemented; or at least, it doesn't seem be working (or maybe I'm missing something).

If I add a Oracle layer to QGIS, the amount of time it takes to actually add it to QGIS (rendering = off) is around the same amount of time that it takes to perform a SDO_TUNE.EXTENT_OF query on the layer.

Looking at the linked code:

```
        if ( !mSpatialIndex.isNull() && ( mUseEstimatedMetadata || mSqlWhereClause.isEmpty() ) )
2021       {
2022         sql = QString( "SELECT SDO_TUNE.EXTENT_OF(%1,%2) FROM dual" )
2023             .arg( quotedValue( QString( "%1.%2" ).arg( mOwnerName ).arg( mTableName ) ) )
2024             .arg( quotedValue( mGeometryColumn ) );
2025
2026         ok = exec( qry, sql );
2027       }
```

Surely "mUseEstimatedMetadata" should be a "not"? But I'm rubbish at C/++ and can't follow it in depth.

#12 - 2014-02-21 07:53 AM - Jürgen Fischer

Jonathan Moules wrote:

> *I don't think this has actually been implemented; or at least, it doesn't seem be working (or maybe I'm missing something).*
> *If I add a Oracle layer to QGIS, the amount of time it takes to actually add it to QGIS (rendering = off) is around the same amount of time that it takes*
> *to perform a SDO_TUNE.EXTENT_OF query on the layer.*

Well, you were suggesting that SDO_TUNE.EXTENT_OF would be slower that retrieving the data directly from the view. Maybe it's not as bad as thought.

> *Looking at the linked code:*
> *if (!mSpatialIndex.isNull() && (mUseEstimatedMetadata || mSqlWhereClause.isEmpty()))*

That means if there is a spatialindex (prerequisite for SDO_TUNE.EXTENT_OF) and there's no filter clause (because then the extents of the subset could be smaller than the full extent) or "estimated metadata" is on (in which case that difference is ignored) the extents are retrieved by SDO_TUNE.EXTENT_OF.

But if "estimated metadata" is on, it'll retrieve the extents from the view, if successful return them in line 2013 of commit:45fea7b46 and this piece of code will not be reached.

#13 - 2014-02-21 09:14 AM - Jonathan Moules

I see what's happening here. There's some confusion about what the two metadata checkboxes do. My perception is that they should be like this:

"only look in meta data table" - I believe this checkbox should speed up loading of any single layer. Basically everything comes out of ALL_SDO_GEOM_METADATA except geometry type (which can't). This includes the MBR.

"Use estimated table metadata" - This should improve table listing speed when the user presses "connect". I don't believe it should have anything to do with

SDO_TUNE.EXTENTS_OF (as QGIS doesn't need to know the bounding box when listing layers, and doesn't currently).

The problem with the current implementation is

> or "estimated metadata" is on (in which case that difference is ignored) the extents are retrieved by SDO_TUNE.EXTENT_OF.

Which means that if I have "estimated metadata" checked (I do) then adding layers is super slow because it's doing SDO_TUNE. But if I un-check it then listing the tables becomes incredibly slow because it's getting data for every table!

I hope that's clearer. It's a little confused in my mind.
The TL;DR version is: I believe absolutely everything should come out of the ALL_SDO_GEOM_METADATA wherever possible.

#14 - 2014-02-21 11:06 AM - Jürgen Fischer

Jonathan Moules wrote:

> "only look in meta data table" - I believe this checkbox should speed up loading of any single layer. Basically everything comes out of
> ALL_SDO_GEOM_METADATA except geometry type (which can't). This includes the MBR.

No, it restricts the scan to the geometry columns listed in the meta data table. If that's not checked, it also scans all columns with geometry type (eg. views).

> "Use estimated table metadata" - This should improve table listing speed when the user presses "connect". I don't believe it should have anything to
> do with SDO_TUNE.EXTENTS_OF (as QGIS doesn't need to know the bounding box when listing layers, and doesn't currently).

That accepts unprecise information about that data (ie. scans just a few rows instead of all to determine the geometry type, extents from metadata, instead of scanning all rows and ignoring that applied filters usually make the estimations about full extent and full row count non-applicable for a layer).   So that's the *speed tops accuracy* switch.

> The problem with the current implementation is or "estimated metadata" is on (in which case that difference is ignored) the extents are retrieved by
> SDO_TUNE.EXTENT_OF.
> Which means that if I have "estimated metadata" checked (I do) then adding layers is super slow because it's doing SDO_TUNE. But if I un-check it
> then listing the tables becomes incredibly slow because it's getting data for every table!

First of all, extent doesn't matter when listing the tables and therefore doesn't matter for the table scan.

Second, as I already tried to lay out above, on layer insertion (that is after the scan is already finished), if you have checked "estimated metadata" the extent is read from the metadata table (as you requested, that's what commit:45fea7b46 does).   Only if that fails for some reason SDO_TUNE.EXTENT_OF is used - and that only if there is no filter (because SDO_TUNE.EXTENT_OF would only consider the full extent w/o a filter - and that's inaccurate if there is one) **or** "estimated metadata" is on (ie. wrong extent, but *speed tops accuracy*).

Still if you don't notice an improvement when inserting layers after commit:45fea7b46, SDO_TUNE.EXTENT_OF can't be that bad.

> The TL;DR version is: I believe absolutely everything should come out of the ALL_SDO_GEOM_METADATA wherever possible.

That's what we do - for the listing and now later to determine the extent. But we need the geometry type, that it doesn't have and we want to be able to use tables and **views** that are not listed.

An approach to solve this would be caching what we need (in the database for all or locally per user) or consider the search settings in the scan and not just to filter it's results. I think we discussed both already - not sure if there's a tickets.

#15 - 2014-02-24 04:28 AM - Jonathan Moules

Hi Jürgen,

Thanks for taking the time to write a detailed reply. I appreciate I'm not being as clear as I should be.

Following what you've said I've now investigated with our DBA's what gets run when I click "add" data. Note that the "render" button is disabled. The following is run:

    select * from osmm.osm_mm_topographic_area where rownum = 0

Which is wrong because render is off; it shouldn't be selecting anything.

Also the below (which seems to be correct):

    select DIMINFO, SRID from ALL_SDO_GEOM_METADATA WHERE owner = SYS.DBMS_ASSERT.SIMPLE_SQL_NAME(:B2) AND
    table_name = SYS.DBMS_ASSERT.SIMPLE_SQL_NAME(:B1) AND column_name =
    upper(SYS.DBMS_ASSERT.QUALIFIED_SQL_NAME(COLUMN_NAME))

So the problem is that Oracle is selecting all the data when the layer is added even though QGIS isn't doing anything with it. I'll open a new ticket.

Thanks again for your patience!
Jonathan