



Use shaded guava from thirdparty

Details

Type:	Sub-task	Status:	RESOLVED
Priority:	Major	Resolution:	Fixed
Affects Version/s:	None	Fix Version/s:	3.3.1, 3.4.0
Component/s:	None		
Labels:	pull-request-available		
Hadoop Flags:	Reviewed		

Description

Use the shaded version of guava in hadoop-thirdparty

Issue Links

breaks

HADOOP-17319 Update the checkstyle config to ban some guava functions	RESOLVED
---	----------

is related to

HADOOP-17439 No shade guava in trunk	OPEN
HADOOP-17315 Use shaded guava in ClientCache.java	RESOLVED

relates to

HDFS-16240 Replace unshaded guava in HttpFS Server WebServer	RESOLVED
--	----------

links to


GitHub Pull Request #2342

Show 1 more links (1 links to)

Activity

- [Wei-Chiu Chuang](#) added a comment - 28/Sep/20 03:15
- We need to ensure no public API breakage.
There will be applications using Hadoop internal APIs and they will likely break. We need to understand the impact of change.
- [Ayush Saxena](#) added a comment - 29/Sep/20 14:19
- Thanx [weichiu](#) for following this up.
I tried checking a couple of API's, Mostly with using `immutable*` data types. Mostly they were returning the there parent Java classes only.
Couple of them were just `VisibleForTesting`
I don't have much experience handling dependencies over the entire hadoop stack, so not sure how to check everything. Any help or pointers, so as how to verify.. I can try whatever looks suspicious to you.
[vinayakumarb](#) you must have faced similar issues, while shading Google `Protobuff`, any pointers here?
In case we get too many things breaking, we may try partial relocation?
- [Michael Stack](#) added a comment - 01/Oct/20 04:43


[ayushtkn](#) What is the approach here (big patch, being lazy). guava is in thirdparty and now you are going to move all guava references in all hadoop to use the thirdparty version and remove the current included guava completely? if so, sounds good to me (big, ugly, simple patch). What you thinking of targeting as fix version? 3.4.0? Thanks.

▼  [Ayush Saxena](#) added a comment - 01/Oct/20 07:07

[stack](#) yes the intentions are like that only.

The hadoop jars would be using the shaded guava from the thirdparty. but guava as of now would be still packaged as it is part of several transitive dependencies. So need to see whether we can go ahead like this, or try shading the ones which bring in guava(they are many)

Yes, I am targeting this for 3.4.0 as of now

▼  [Michael Stack](#) added a comment - 01/Oct/20 07:26

but guava as of now would be still packaged as it is part of several transitive dependencies.

Can you say more on the above? Transitively included by Hadoop because Hadoop dependencies pull it in or are you talking downstreamers that expect Hadoop to provide guava to them (transitively?)

I'm wondering about the downstreamers whose apps use guava 11 because thats what hadoop used until 3.3.0/3.2.1. They want to upgrade to 3.4. They'll have to do the work to upgrade to guava 27 because that is what 3.2.1/3.3.0 have even though you've done all this work here. Seems a shame?

(I set fix version as 3.4.0 – thanks).

▼  [Ayush Saxena](#) added a comment - 01/Oct/20 07:47 - edited


Transitive dependency as the dependencies which hadoop pulls up, on which hadoop depends. Like `mockserver-core` is one. If you see the `hadoop-project/pom.xml`, whichever dependency has excluded Guava, requires guava.

Ideally if the downstream has to upgrade guava, then this patch has no meaning. The basic requirement itself is that the downstream should not need to upgrade guava.

If the transitive dependency, say `x` is a dependency of `hadoop-common`, which is compatible with the guava version of the downstream project they can exclude it, and I think things should work fine?

If there are dependencies with higher version of Guava, (need to analyse), which aren't compatible with guava version of downstream projects. Then we might need to shade them as well? May be `curator` can be one of those.

Let me know, if you have any idea/suggestion or different approach which may make things better

▼  [Michael Stack](#) added a comment - 01/Oct/20 16:04

Ideally if the downstream has to upgrade guava, then this patch has no meaning.

+1

Then we might need to shade them as well? May be `curator` can be one of those.

Yes. Unfortunately the tangles start to compound fast when the dependencies' dependency is also an hadoop dependency (and versions don't align). One thing to consider removing problem dependencies (like `curator`) if not heavily used. Thanks [ayushtkn](#)

▼  [Ayush Saxena](#) added a comment - 03/Oct/20 15:54

I rolled back Guava to 11, in the latest PR.

There weren't any tests failing due to this.


<https://ci-hadoop.apache.org/view/Hadoop/job/hadoop-multibranch/view/change-requests/job/PR-2342/4/>

I tried compiling Hive with this upgraded version, and ran some tests like `TestAccumuloIndexParameters` which failed with 3.3.0, due to guava conflict, that passed. Compiled Spark as well with this hadoop version, the compilation was success.

Need to handle couple of cases in the present PR, to counter the Jenkins complains, but I think the code can be used for verification/reviews. Will handle the Jenkins complains going forward, should be just a tweak in replacer plugin only for couple of sub-projects...

▼  [Michael Stack](#) added a comment - 05/Oct/20 18:41


Looking good [ayushtkn](#) . You thinking that this patch would include the roll back to guava 11 in branch-3 ? Thanks.

▼  [Ayush Saxena](#) added a comment - 06/Oct/20 09:32


[stack](#) The guava I planned to revert to 11 as of now, that would be used by the dependencies only, Not by the hadoop code, and people including hadoop dependencies will have a choice to exclude this `Gauva-11` if they themselves have guava. So, I think it should be safe? The dependencies also doesn't seems to be very specific with the version, as `curator` has already shaded `gauva` apart from 3-4 files, for which we need guava, and those files are quite stable, there are couple of more dependencies which use guava, they also doesn't seems to create much issues. All tests passed with `{gauva-11 and guava-27}` as well.

So, I prefer staying at `guava-11`. Let me know your thoughts?


Plan to address [vinayakumarb](#) comments by tomorrow, once you confirm. Thanx..

▼  [Michael Stack](#) added a comment - 07/Oct/20 18:00

[ayushtkn](#) talking out loud, 3.2.1 and 3.3.0 shipped w/ guava 27, right? This patch is for 3.4.0 and would revert the guava included by hadoop to guava 11 from 27 (though it will not use the guava 11 itself). A regression on lib version is unexpected. The argument for the revert is that it will make it easier on folks running hadoop versions older than 3.2.1/3.3.0 to migrate especially for downstreamers like Hive where guava 11 is deeply embedded. Do I have that right? If so, I agree w/ this rationale. I suggest that the revert needs to be broadcast – the dev list? – since it such an unusual move. Thanks.

▼  [Ayush Saxena](#) added a comment - 17/Oct/20 07:25

Merged to trunk. Thanx [stack](#) and [vinayakumarb](#) for the review and [weichiu](#) for initiating this.

▼  [Ayush Saxena](#) added a comment - 13/Nov/20 17:16


Planing to put up a PR to backport this to branch-3.3

Does any one has concerns doing so, Anything with respect to compatibility or so. Considering 3.3.0 is already released.

Do let me know, Thoughts on this...

▼  [Michael Stack](#) added a comment - 13/Nov/20 17:32

+1 on backport. Will ease move to 3.3.x+


▼  [Wei-Chiu Chuang](#) added a comment - 13/Nov/20 20:38

please make sure to add [HADOOP-17344](#).

I suspect we want to backport all the way to 3.1, to make future cherrypick easier. But one step at a time.

▼  [angerszhu](#) added a comment - 01/Dec/20 03:46


+1 on back port to hadoop-3.3.x.

▼  [Ayush Saxena](#) added a comment - 02/Dec/20 07:04


Thanx Everyone, Have put up a PR backporting to 3.3

Please have a look

[GitHub Pull Request #2505](#)

▼  [Ayush Saxena](#) added a comment - 10/Dec/20 00:26

Backported to branch-3.3, Thanx Everyone!!!

▼  [Lisheng Sun](#) added a comment - 18/Dec/20 14:07


hi [ayushtkn](#)

I found no shaded guava in trunk of my local.


I don't know whether it's my environment or other problems.

▼ People

Assignee:

 Ayush Saxena

Reporter:

 Ayush Saxena

Votes:

 Vote for this issue

Watchers:

 Start watching this issue

▼ **Dates**

Created:

27/Sep/20 16:34

Updated:

28/Sep/21 04:13

Resolved:

17/Oct/20 07:25

▼ **Time Tracking**

Estimated:

 Not Specified

Remaining:

 0h

Logged:

 5h 10m