RhodeCode CE/EE - Support #5395

Svn protocols and performance

04.10.2017 23:36 - Nicholas Lebrun

Status:	Resolved		**Start date:**	04.10.2017
Priority:	Normal		**Due date:**	
Assignee:			**% Done:**	0%
Category:			**Estimated time:**	0.00 hour
Target version:			**Spent time:**	0.00 hour
Sorting:			**Affected Version:**	
Commit Number:				

Description

I am having issues with the performance of my svn respos served over http via apache, as described in the documentation here :
https://docs.rhodecode.com/RhodeCode-Enterprise/admin/svn-http.html

I am trying to commit a few thousand files (around 4k) totaling 350 MiB, and it's taking over 75 minutes. Everything runs on my local windows machine, I am using Tortoise client 1.9.5 and my Rhode Code community instance is installed in a VM running Ubuntu Server. My VM has 4 cpus allocated and more than enough ram and disk space. I increased my gunicorn worker counts as described in the documentation here : https://docs.rhodecode.com/RhodeCode-Enterprise/admin/tuning-gunicorn.html

In an attempt to identify where is my bottleneck I would like to use the svn protocol instead of http, I know the performance issues related to this should be fixed in newer versions of svn and http should work just fine, but I'm running out of ideas.

My question is two fold, 1) Would you have an idea why I am seeing such horrible performance over http? and 2) Does RhodeCode support the svn protocol and if so, how would I go about enabling it?

Thanks!

History

#1 - 04.10.2017 23:39 - Marcin Kuzminski [CTO]

Hi Nicolas,

This might be related to slow authentication because committing 4K files means 4k+ requests and each one needs to be authenticated. What kind of auth do you use LDAP, or RhodeCode builtin auth ? In both cases please consider enabling AUTH_CACHE_TTL settings in auth plugin settings.

I think this might be the cause of slowdown, can you try the cache ttl setting before we try other things ?

#2 - 05.10.2017 00:02 - Nicholas Lebrun

Hi Marcin,

Thank you for your reply. I am using the builtin RhodeCode auth for now. I enabled the Auth cache TTL and set it to 120 seconds, unfortunately, there is no immediately apparent increase in commit speeds.

While I am committing I am monitoring my VM's performance with top, and I can see that my cpus are mostly idle. My load average is roughly 0.8/4. Could that indicate the bottleneck would be with my client sending the requests?

Thank you!
-Nic

#3 - 05.10.2017 00:05 - Marcin Kuzminski [CTO]

I'd recommend enabling debug log-level and checking the time on log entries, it's usually a good way to track where the slowndown is occurring.

https://docs.rhodecode.com/RhodeCode-Enterprise/admin/enable-debug.html#debug-and-logging-configuration

Answering the second question, we don't support SVN protocol. We heavily rely on HTTP because of things like IP restrictions etc.

#4 - 05.10.2017 05:13 - Nicholas Lebrun

I enabled the logs and was able to identify that each http request takes about 700 ms to complete. Here are my logs for one request :

2017-10-04 23:02:27.006 INFO [rhodecode.lib.middleware.simplevcs] Access for IP:192.168.4.92 allowed
2017-10-04 23:02:27.197 INFO [rhodecode.lib.middleware.request_wrapper] IP: 192.168.4.92 Request to /test_repo_group/test_repo time: 0.257s

[04/Oct/2017:23:02:27 -0400] GNCRN   192.168.4.92   rqt:0.257832 401 260 "OPTIONS:/test_repo_group/test_repo " usr:- "-" "SVN/1.9.5 (x64-microsoft-windows) serf/1.3.9 TortoiseSVN-1.9.5.27581"
2017-10-04 23:02:27.238 INFO [rhodecode.lib.middleware.simplevcs] Access for IP:192.168.4.92 allowed
2017-10-04 23:02:27.434 INFO [rhodecode.authentication.base] Authenticating user nicholasl using egg:rhodecode-enterprise-ce#rhodecode plugin
2017-10-04 23:02:27.504 INFO [rhodecode.authentication.plugins.auth_rhodecode] user authenticated correctly
2017-10-04 23:02:27.533 INFO [rhodecode.lib.middleware.simplevcs] Access for IP:192.168.4.92 allowed
2017-10-04 23:02:27.737 INFO [rhodecode.lib.middleware.simplevcs] pull action on svn repo "test_repo_group/test_repo" by "nicholasl" from 192.168.4.92 SVN/1.9.5 (x64-microsoft-windows) serf/1.3.9 TortoiseSVN-1.9.5.27581

I tried hosting a svn repo with svnserve on the same vm and committing the same files using the svn protocol, it took about 2 minutes.
I tried exposing another repo through apache and using the basic subversion auth as authentication, committing the same set of files took about the same time as through the svn protocol, 2 minutes.

Would you have an idea on what could be causing the issue?

#5 - 05.10.2017 07:31 - Marcin Kuzminski [CTO]

ok, so this looks like a case that SVN somehow uses an unoptimized HTTP protocol, this means that it does a http request for each commited files.

i.e 4000 HTTP requests that each takes long ~700ms to complete.

Can you check those things:

- client compatibility. does your svn client work in 1.9 mode?
- repo compatibility, please check the repo > settings > maintenance to verify if the repository is in latest svn format

Based on this article: http://blogs.wandisco.com/2011/10/12/top-new-features-in-subversion-1-7-httpv2-svnrdump/

httpv2 was introduced in SVN 1.7 and there's a switch that allows to disable that to use old protocol, maybe this is the case here ?

#6 - 05.10.2017 07:59 - Marcin Kuzminski [CTO]

Last thing to check is if your mod_dav_svn is actually in latest version. You can obtain the latest binaries for your OS from wandisco source repositories.

#7 - 05.10.2017 16:42 - Nicholas Lebrun

I have verified that my client is working in 1.9 mode, and I ran a maintenance on the repo through the repo settings -> Maintenance, here is the output :

Path: /home/nicholasl/repos/test_repo_group/test_repo
UUID: d173df14-a90c-11e7-ac7e-83efe1718f84
Repository Format: 5
Compatible With Version: 1.9.0
Repository Capability: mergeinfo
Filesystem Type: fsfs
Filesystem Format: 7
FSFS Sharded: yes
FSFS Shard Size: 1000
FSFS Shards Packed: 0/0
FSFS Logical Addressing: yes
Configuration File: /home/nicholasl/repos/test_repo_group/test_repo/db/fsfs.conf

I tried making sure my mod_dav_svn version was up to date by following the instructions here :
https://docs.rhodecode.com/RhodeCode-Enterprise/admin/svn-http.html?highlight=svn%20protocol

Unfortunately I get the following error when running apt-get update :

W: The repository 'http://opensource.wandisco.com/ubuntu zesty Release' does not have a Release file.
N: Data from such a repository can't be authenticated and is therefore potentially dangerous to use.
N: See apt-secure(8) manpage for repository creation and user configuration details.
E: Failed to fetch http://opensource.wandisco.com/ubuntu/dists/zesty/svn19/binary-i386/Packages 404 Not Found
E: Some index files failed to download. They have been ignored, or old ones used instead.

http://opensource.wandisco.com/ubuntu/dists/zesty indeed doesn't seem to exist.

I tried verifying my mod_dav_svn version by running the following command :
strings /usr/lib/apache2/modules/mod_dav_svn.so | grep 'Powered by'

I got the following output :
Powered by Apache Subversion version 1.9.5 (r1770682).

Thank you for your help,
-Nic

#8 - 05.10.2017 16:58 - Marcin Kuzminski [CTO]

All seems to be as expected. Can you do one thing ?

There's a flag called

```
SVNAdvertiseV2Protocol On;
```

Could you manually add it to the generated mod_dav_svn.conf file ? If that doesn't help i'll let my team run some tests to see if we can reproduce this.

#9 - 05.10.2017 19:33 - Nicholas Lebrun

Unfortunately, it doesn't seem to have made a difference, I get the same response time.

If you are interested I don't mind sending you my virtual machine if you want to use it to attempt to reproduce this issue. I can clean my sensitive data from the VM and somehow send it over to you. Unfortunately, the vdi file is quite big at 20 Gigs, but it's still an option.

#10 - 05.10.2017 19:35 - Marcin Kuzminski [CTO]

Let us try various things to test it out on our env.

Btw have you tested pure apache over http ? I know you mentioned svn protocol, but we're wondering what's the response time for you bypassing RhodeCode but using http protocol to do the commit.

#11 - 05.10.2017 19:37 - Nicholas Lebrun

Yes I have, both with no authentication and with the basic apache authentication, I get response times comparable to the svn protocol, around 2 minutes for the commit.

#12 - 06.10.2017 13:26 - Marcin Kuzminski [CTO]

Hi Nicholas,

Here's our summary of investigation.

The httpv2 protocol is working as expected, we confirmed that by setting the SVNAdvertiseV2Protocol Off and this is causing errors with svn 1.9.
We also compared performance on RhodeCode vs Pure Apache via HTTP. It turns out doing a commit of 10 files does 12 HTTP calls.
1st is auth
2-11 ones are PUT calls storing the file
12 one is a MERGE call.

In our test Time for apache is sub arround 1.5s to perform a commit of 10 files. Time for RhodeCode is 4.5s (based on 4.10dev unreleased yet)
After tracking this we're now sure it's the overhead of RhodeCode system that makes the difference. In 4.10 we tunned the http calls that each takes roughly 0.150s but if you have lots of those calls it adds up to bigger amount. This is mainly because of how SVN via HTTP works, it makes a lot of HTTP calls, and each of those needs to be authenticated, permissions checked+ things like IP restrictions, hooks etc adds up to some overhead.

Because of stateless nature of HTTP we're not sure if we can tune this further. For now we manage to speed up the calls by 30% compared to 4.9 release (one you're testing) but now we'd probably need to investigate working with SVN clients on some sort of transaction indicators to verify if we could skip some of the checks for the middle 10 commands. But again security is top priority for us so we're not sure it's possible without creating some problematic security holes in our system.

Is the slower than expected performance a big problem for you ?

Best,

#13 - 06.10.2017 19:34 - Nicholas Lebrun

Thank you for putting this much effort into investigating this, it is much appreciated.

It is hard to tell if the performance is a problem or not. For day to day use I don't think it matters much, but if we have a whole team check out a branch of one of our projects it might put a heavy strain on our host and the full checkouts will be pretty long.

It is definitely something we take into consideration for deciding if we want to use Rhode in production or not. We are still unsure if it will make or break the decision.

#14 - 07.10.2017 12:54 - Marcin Kuzminski [CTO]

I think when using with multiple users the performance shouldn't be affected in this case. It's a case of "overhead" that RhodeCode puts because of all the auth and advanced checks. In any case we'd gladly help you with any issues if you consider deploying even the free CE edition. SVN usage is becoming more important for us and we always want to work with companies that could provide valuable feedback to our product to make it better.

Best,

#15 - 04.04.2019 17:52 - Mike Lanteros

Has this changed ever since this was opened?

When I create a SVN repository on my RhodeCode instance, I get the following warning:

> SVN Protocol is disabled. To enable it, see the documentation here.

That page brings us to https://docs.rhodecode.com/RhodeCode-Enterprise/admin/svn-http.html which is about enabling svn over HTTP, not enabling the SVN protocol.

Is the svn:// protocol supported at all?

#16 - 04.04.2019 18:08 - Marcin Kuzminski [CTO]

- Status changed from New to Resolved

Hi,

So when it comes to SVN performance we did few iterations and it's significantly faster now then e.g version 4.10 that this related to.

We only support http:// and ssh:// protocols for VCS communication.

I'll close this ticket, as resolved.

We worked with few pure SVN customers to get this to good speed and we're not getting any complaints at that point.