Error with Fetch commits with Mercurial repository when log has invalid char

When I add a mercurial repository, redmine start to scan and add to Mysql. This take several minutes. When redmine finish to index .Redmine show a error page. In production.log show this message. It happen with a huge repository about 70335 commits

production.log

Started GET "/projects/iset/repository" for 192.168.1.164 at 2016-06-13 10:56:04 +0000
Processing by RepositoriesController#show as HTML
Parameters: {"id"=>"iset"}
  Current user: admin (id=1)
Completed 500 Internal Server Error in 2039ms (ActiveRecord: 4.8ms)

NoMethodError (undefined method '[]' for nil:NilClass):
  lib/redmine/scm/adapters/mercurial_adapter.rb:194:in `each_revision'
  app/models/repository/mercurial.rb:187:in `block in fetch_changesets'
  app/models/repository/mercurial.rb:186:in `step'
  app/models/repository/mercurial.rb:186:in `fetch_changesets'
  app/controllers/repositories_controller.rb:114:in `show'
  lib/redmine/sudo_mode.rb:63:in `sudo_mode'

If you exec a manual auto_fetch to refresh data with rake -f Rakefile RAIL_ENV=production redmine:fetch_changesets --trace
Console shows:

/usr/local/bundle/gems/htmlentities-4.3.1/lib/htmlentities/mappings/expanded.rb:465: warning: duplicated key at line 466 ignored: "inodot"
** Invoke redmine:fetch_changesets (first_time)
** Invoke environment (first_time)
** Execute environment
** Execute redmine:fetch_changesets
rake aborted!

NoMethodError: undefined method '[]' for nil:NilClass
  /usr/src/redmine/lib/redmine/scm/adapters/mercurial_adapter.rb:194:in `each_revision'
  /usr/src/redmine/app/models/repository/mercurial.rb:187:in `block in fetch_changesets'
  /usr/src/redmine/app/models/repository/mercurial.rb:186:in `step'
  /usr/src/redmine/app/models/repository/mercurial.rb:186:in `fetch_changesets'
  /usr/src/redmine/app/models/repository.rb:344:in `block (2 levels) in fetch_changesets'
  /usr/local/bundle/gems/activerecord-4.2.5/lib/active_record/relation/delegation.rb:46:in `each'
  /usr/local/bundle/gems/activerecord-4.2.5/lib/active_record/relation/delegation.rb:46:in `each'
Associated revisions
Revision 19606 - 2020-03-22 17:15 - Toshi MARUYAMA

scm: mercurial: change template "escape" to "urlescape" for "author" and "desc" (#23055)

Revision 19607 - 2020-03-22 17:34 - Toshi MARUYAMA
Rev 19608 - 2020-03-22 17:35 - Toshi MARUYAMA

scm: mercurial: update test repository (#23055)

scm: mercurial: add test of author and message which have control character (#23055)

History

#1 - 2016-06-13 15:48 - Toshi MARUYAMA
- Status changed from New to Needs feedback

What output of scm_stderr_log_file?
source:tags/3.2.3/config/configuration.yml.example#L138

#2 - 2016-06-13 19:11 - mole mole

I haven't any exit on scm_stderr_log_file.

But in production.log when do the first scan logged this:

```
Changeset Load (0.8ms)  SELECT  `changesets`.* FROM `changesets` WHERE `changesets`.repository_id = 11 AND `changesets`.scmid = '6e892cac598ab919f853ca54c7745e783b70dbb8' ORDER BY changesets.id DESC LIMIT 1
Changeset Exists (0.6ms)  SELECT 1 AS one FROM `changesets` WHERE (`changesets`.revision = BINARY '10555' AND `changesets`.id != 37510 AND `changesets`.repository_id = 11) LIMIT 1
Changeset Exists (0.5ms)  SELECT 1 AS one FROM `changesets` WHERE (`changesets`.scmid = BINARY '6e892cac598ab919f853ca54c7745e783b70dbb8' AND `changesets`.id != 37510 AND `changesets`.repository_id = 11) LIMIT 1
Changeset Exists (0.4ms)  SELECT 1 AS one FROM `changesets` WHERE (`changesets`.revision = BINARY '10598' AND `changesets`.repository_id = 11) LIMIT 1
Changeset Exists (0.4ms)  SELECT 1 AS one FROM `changesets` WHERE (`changesets`.scmid = BINARY '6289f36f5adf23615f6fc249ab5828beff29b' AND `changesets`.repository_id = 11) LIMIT 1

What output of scm_stderr_log_file?
source:tags/3.2.3/config/configuration.yml.example#L138

#2 - 2016-06-13 19:11 - mole mole

I haven't any exit on scm_stderr_log_file.

But in production.log when do the first scan logged this:

```
Changeset Load (0.8ms)  SELECT  `changesets`.* FROM `changesets` WHERE `changesets`.repository_id = 11 AND `changesets`.scmid = '6e892cac598ab919f853ca54c7745e783b70dbb8' ORDER BY changesets.id DESC LIMIT 1
Changeset Exists (0.6ms)  SELECT 1 AS one FROM `changesets` WHERE (`changesets`.revision = BINARY '10555' AND `changesets`.id != 37510 AND `changesets`.repository_id = 11) LIMIT 1
Changeset Exists (0.5ms)  SELECT 1 AS one FROM `changesets` WHERE (`changesets`.scmid = BINARY '6e892cac598ab919f853ca54c7745e783b70dbb8' AND `changesets`.id != 37510 AND `changesets`.repository_id = 11) LIMIT 1
Changeset Exists (0.4ms)  SELECT 1 AS one FROM `changesets` WHERE (`changesets`.revision = BINARY '10598' AND `changesets`.repository_id = 11) LIMIT 1
Changeset Exists (0.4ms)  SELECT 1 AS one FROM `changesets` WHERE (`changesets`.scmid = BINARY '6289f36f5adf23615f6fc249ab5828beff29b' AND `changesets`.repository_id = 11) LIMIT 1

```

2020-10-25
Shelling out: "hg" '-R' '/home/share/mirror/hg/all' '--encoding' 'utf-8' '--config' 'extensions.redminehelper=/usr/src/redmine/lib/redmine/scm/adapters/mercurial/redminehelper.py' '--config' 'diff.git=false' 'log' '--debug' '-C' '--style' '//usr/src/redmine/lib/redmine/scm/adapters/mercurial/hg-template-1.0.tmpl' '-r' '10600:10699'

Completed 500 Internal Server Error in 893414ms (ActiveRecord: 465576.7ms)

NoMethodError (undefined method `[]' for nil:NilClass):
  lib/redmine/scm/adapters/mercurial_adapter.rb:194:in `each_revision'
  app/models/repository/mercurial.rb:187:in `block in fetch_changesets'
  app/models/repository/mercurial.rb:186:in `step'
  app/models/repository/mercurial.rb:186:in `fetch_changesets'
  app/controllers/repositories_controller.rb:114:in `show'
  lib/redmine/sudo_mode.rb:63:in `sudo_mode'

#3 - 2016-06-14 05:59 - Toshi MARUYAMA

mole mole wrote:

I haven't any exit on scm_stderr_log_file.

How about "#{Rails.env}.scm.stderr.log"?

source:tags/3.2.3/lib/redmine/scm/adapters/abstract_adapter.rb#L221

But in production.log when do the first scan logged this:

... 
  hg" '-R' '/home/share/mirror/hg/all' '--encoding' 'utf-8' '--config' 'extensions.redminehelper=/usr/src/redmine/lib/redmine/scm/adapters/mercurial/redminehelper.py' '--config' 'diff.git=false' 'log' '--debug' '-C' '--style' '/usr/src/redmine/lib/redmine/scm/adapters/mercurial/hg-template-1.0.tmpl' '-r' '10600:10699'

What output of above "hg" on command line?

#4 - 2016-06-14 15:33 - mole mole

- File redmine_scm_stderr.log added

I haven't any exit on scm_stderr_log_file. -> I try to say: I haven't any error related with this issue. Only have this message:
not trusting file /home/share/mirror/hg/all/.hg/hgrc from untrusted user root, group www-data

It can be solve with adding in hgrc:

    [trusted]
    users =
hg in command line:

Mercurial Distributed SCM

basic commands:

add           add the specified files on the next commit
annotate      show changeset information by line for each file
clone         make a copy of an existing repository
commit        commit the specified files or all outstanding changes
diff          diff repository (or selected files)
export        dump the header and diffs for one or more changesets
forget        forget the specified files on the next commit
init          create a new repository in the given directory
log           show revision history of entire repository or files
merge         merge working directory with another revision
pull          pull changes from the specified source
push          push changes to the specified destination
remove        remove the specified files on the next commit
serve         start stand-alone webserver
status        show changed files in the working directory
summary       summarize working directory state
update        update working directory (or switch revisions)

use "hg help" for the full list of commands or "hg -v" for details

#5 - 2016-06-14 16:00 - Toshi MARUYAMA

What is output of following command?
'hg' '-R' '/home/share/mirror/hg/all' '--encoding' 'utf-8' '--config'
'extensions.redminehelper=/usr/src/redmine/lib/redmine/scm/adapters/mercurial/redminehelper.py' '--config' 'diff.git=false' 'log' '--debug' '-C' '--style'
'/usr/src/redmine/lib/redmine/scm/adapters/mercurial/hg-template-1.0.tmpl' '-r' '10600:10699'

#6 - 2016-06-14 16:18 - mole mole

- File hgcommand.log added

The command 'hg' '-R' '/home/share/mirror/hg/all' '--encoding' 'utf-8' '--config'
'extensions.redminehelper=/usr/src/redmine/lib/redmine/scm/adapters/mercurial/redminehelper.py' '--config' 'diff.git=false' 'log' '--debug' '-C' '--style'
'/usr/src/redmine/lib/redmine/scm/adapters/mercurial/hg-template-1.0.tmpl' '-r' '10600:10699'

Can be executed without errors. Attach output.

#7 - 2016-06-14 16:48 - Toshi MARUYAMA
What is output?

select revision from changesets where id =
(select max(id) from changesets where repository_id = 11)

#8 - 2016-06-14 17:09 - mole mole

revision
10599

#9 - 2016-06-14 17:22 - Toshi MARUYAMA

Try reduce FETCH_AT_ONCE.
source:tags/3.1.3/app/models/repository/mercurial.rb#L30

I think a revision from 10600 to 10699 has problem.

#10 - 2016-06-14 17:31 - Toshi MARUYAMA

I think your revision has problem.
- no date
- no commit log

If you find this revision, try this patch.

```diff
diff --git a/lib/redmine/scm/adapters/mercurial_adapter.rb b/lib/redmine/scm/adapters/mercurial_adapter.rb
--- a/lib/redmine/scm/adapters/mercurial_adapter.rb
+++ b/lib/redmine/scm/adapters/mercurial_adapter.rb
@@ -212,8 +212,8 @@ module Redmine
 yield Revision.new({
     :revision => le['revision'],
     :scmid => le['node'],
-    :time => Time.parse(le['date']['__content__']),
-    :message => le['msg']['__content__'],
+    :time => (Time.parse(le['date']['__content__']) rescue Time.at(0)),
+    :message => (le['msg']['__content__'] rescue ''),
     :paths => paths,
     :parents => parents_ary)
 end
```

#11 - 2016-06-14 18:02 - mole mole

- File issue.jpg added

It didn't works. Same result.

Maybe a wrong code of EOL?. I see in sublime editor a DC3 code.
I think that double `!!` can break it.

The message in mercurial is:

```plaintext
-dimensiones de los geosets de m3Font funcionando bien con shader.!!
```

and the output of script have a DC3 code when in original have a `!!`:

```xml
<msg>-dimensiones de los geosets de m3Font funcionando bien con shader.!!</msg>
```

Try upgrading Mercurial. Latest version is 3.8.3.

It didn't works. Same result.

Please report to Mercurial BTS. Broken XML is Mercurial bug.

https://bz.mercurial-scm.org/

The character is `^S` I try to reproduce in small repo

You can reproduce this issue with this repo.

It is obviously Mercurial bug.

Please report to Mercurial BTS.

```
$ LANG=C /WEB-DOWN/hg-repo/hg-crew/hg log --version
Mercurial Distributed SCM (version 3.8.3)
(see https://mercurial-scm.org for more information)
```
Copyright (C) 2005-2016 Matt Mackall and others
This is free software; see the source for copying conditions. There is NO
warranty; not even for MERCHANTABILITY or FITNESS FOR A PARTICULAR PURPOSE.

$ /WEB-DOWN/hg-repo/hg-crew/hg log --debug --style
/REDMINE/MY-REDMINE/test-trunk/lib/redmine/scm/adapters/mercurial/hg-template-1.0.tmpl > hg.xml

$ xmllint hg.xml
hg.xml:10: parser error : PCDATA invalid Char value 19
<msg>-Test</msg>
^
hg.xml:17: parser error : Premature end of data in tag log line 2

#19 - 2016-06-15 13:16 - Toshi MARUYAMA

Sorry, we need add "</log>".

$ /WEB-DOWN/hg-repo/hg-crew/hg log --debug --style
/REDMINE/MY-REDMINE/test-trunk/lib/redmine/scm/adapters/mercurial/hg-template-1.0.tmpl > hg.xml

$ echo "</log>" >> hg.xml

$ xmllint hg.xml
hg.xml:23: parser error : PCDATA invalid Char value 19
<msg>-Test</msg>
^

#20 - 2016-06-15 13:28 - mole mole

Using Mercurial xml template generate a correct xml

$ hg log --debug --style=xml > hg.xml

$ xmllint hg.xml

It runs correctly

PS: --style is a deprecated parameter in mercurial

#21 - 2016-06-15 13:44 - mole mole

I changed in hg-template-1.0.tmpl {desc|escape} by {desc|revscape} and xmllint runs correctly...
Sorry, its probably we must use xmlscape

I see this in this file sample of mercurial

/usr/share/mercurial/templates/map-cmdline.xml

#23 - 2016-06-15 14:08 - mole mole
- File hg-template-1.0.tmpl added

Attach hg-template-1.0.tmpl modified

#24 - 2016-06-15 14:11 - mole mole
- File hg-template-1.0.tmpl added

Sorry, is this file

#25 - 2016-06-15 14:39 - Toshi MARUYAMA
- Subject changed from Error with Fetch commits automatically with mercurial repository to Error with Fetch commits with Mercurial repository when log has invalid char
- Status changed from Needs feedback to Confirmed
- Assignee set to Toshi MARUYAMA

#26 - 2016-06-15 14:41 - mole mole

Thanks for your help.

#27 - 2016-11-28 13:58 - Oscar Edvardsson

I can also confirm this. Any workaround?

Environment:
Redmine version 3.2.0.stable
Ruby version 2.3.1-p112 (2016-04-26) [x86_64-linux-gnu]
Rails version 4.2.5
Environment production
Database adapter PostgreSQL
SCM: Mercurial 3.7.3

#28 - 2020-03-26 02:41 - Toshi MARUYAMA
- Status changed from Confirmed to Closed
I choose 'urlescape' because 'paths' uses 'urlescape' and decodes it.

source: trunk/lib/redmine/scm/adapters/mercurial_adapter.rb@19606#L204

<table>
<thead>
<tr>
<th>Files</th>
<th>Size</th>
<th>Date</th>
<th>Author</th>
</tr>
</thead>
<tbody>
<tr>
<td>redmine_scm_stderr.log</td>
<td>19.8 KB</td>
<td>2016-06-14</td>
<td>mole mole</td>
</tr>
<tr>
<td>hgcommand.log</td>
<td>70.2 KB</td>
<td>2016-06-14</td>
<td>mole mole</td>
</tr>
<tr>
<td>issue.jpg</td>
<td>295 KB</td>
<td>2016-06-14</td>
<td>mole mole</td>
</tr>
<tr>
<td>all.tar.gz</td>
<td>1.14 KB</td>
<td>2016-06-15</td>
<td>mole mole</td>
</tr>
<tr>
<td>hg-template-1.0.tmpl</td>
<td>864 Bytes</td>
<td>2016-06-15</td>
<td>mole mole</td>
</tr>
<tr>
<td>hg-template-1.0.tmpl</td>
<td>869 Bytes</td>
<td>2016-06-15</td>
<td>mole mole</td>
</tr>
</tbody>
</table>