

Ruby - Bug #5978

YAML.load_stream should process documents as they are read

02/08/2012 01:47 AM - trans (Thomas Sawyer)

Status:	Closed	Backport:
Priority:	Normal	
Assignee:	tenderlovmaking (Aaron Patterson)	
Target version:	2.0.0	
ruby -v:	ruby 1.9.3p0 (2011-10-30 revision 33570) [x86_64-linux]	
Description		
Psych say YAML.load_documents is deprecated and say to use YAML.load_stream instead.		
Looking at the implementation for load_stream(), looks to me as if it waits for all documents in the stream to load before anything can be done with it.		
# File 'lib/psych.rb', line 221		
def self.load_stream yaml parse_stream(yaml).children.map { child child.to_ruby } end		
I don't think this should be the case. Ideally load_stream() would take a block, and if an IO object is given, read a document, yield it and then read the next document, and so on.		
I imagine an Enumerator might be applicable to this as well.		

Associated revisions

Revision a2e3de1b3f876dd9b14fa3a6291a202719405382 - 03/08/2012 09:21 PM - tenderlovmaking (Aaron Patterson)

- ext/psych/lib/psych.rb (parse_stream, load_stream): if a block is given, documents will be yielded to the block as they are parsed. [ruby-core:42404] [Bug #5978]
- ext/psych/lib/psych/handlers/document_stream.rb: add a handler that yields documents as they are parsed
- test/psych/test_stream.rb: corresponding tests.

git-svn-id: svn+ssh://ci.ruby-lang.org/ruby/trunk@34953 b2dd03c8-39d4-4d8f-98ff-823fe69b080e

Revision a2e3de1b - 03/08/2012 09:21 PM - tenderlovmaking (Aaron Patterson)

- ext/psych/lib/psych.rb (parse_stream, load_stream): if a block is given, documents will be yielded to the block as they are parsed. [ruby-core:42404] [Bug #5978]
- ext/psych/lib/psych/handlers/document_stream.rb: add a handler that yields documents as they are parsed
- test/psych/test_stream.rb: corresponding tests.

git-svn-id: svn+ssh://ci.ruby-lang.org/ruby/trunk@34953 b2dd03c8-39d4-4d8f-98ff-823fe69b080e

Revision 7d984d76bac857a54c1775b203c944da6cb20ac4 - 03/29/2012 01:25 AM - tenderlovmaking (Aaron Patterson)

merge revision(s)

32578,33401,33403,33404,33531,33655,33679,33809,33900,33965,34067,34069,34087,34328,34330,34527,34772,34783,34839,34914,34953,349

- * ext/psych/lib/psych.rb: updating version to match gem
- * ext/psych/psych.gemspec: ditto
- * ext/psych/lib/psych/visitors/to_ruby.rb: fixing deprecation warning
- * ext/psych/lib/psych.rb: define a new BadAlias error class.
- * ext/psych/lib/psych/visitors/to_ruby.rb: raise an exception when deserializing an alias that does not exist.
- * test/psych/test_merge_keys.rb: corresponding test.
- * ext/psych/lib/psych.rb (load, parse): stop parsing or loading after the first document has been parsed.
- * test/psych/test_stream.rb: pertinent tests.
- * ext/psych/lib/psych.rb (parse_stream, load_stream): if a block is given, documents will be yielded to the block as they are parsed. [ruby-core:42404] [Bug #5978]
- * ext/psych/lib/psych/handlers/document_stream.rb: add a handler that yields documents as they are parsed
- * test/psych/test_stream.rb: corresponding tests.
- * ext/psych/lib/psych/core_ext.rb: only extend Kernel if IRB is loaded in order to stop method pollution.
- * ext/psych/lib/psych.rb: default open YAML files with utf8 external encoding. [ruby-core:42967]
- * test/psych/test_tainted.rb: ditto
- * ext/psych/parser.c: prevent a memory leak by protecting calls to handler callbacks.
- * test/psych/test_parser.rb: test to demonstrate leak.
- * ext/psych/parser.c: set parser encoding based on the YAML input rather than user configuration.
- * test/psych/test_encoding.rb: corresponding tests.
- * test/psych/test_parser.rb: ditto
- * test/psych/test_tainted.rb: ditto
- * ext/psych/parser.c: removed external encoding setter, allow parser to be reused.
- * ext/psych/lib/psych/parser.rb: added external encoding setter.
- * test/psych/test_parser.rb: test parser reuse
- * ext/psych/lib/psych/visitors/to_ruby.rb: Added support for loading subclasses of String with ivars
- * ext/psych/lib/psych/visitors/yaml_tree.rb: Added support for dumping subclasses of String with ivars
- * test/psych/test_string.rb: corresponding tests
- * ext/psych/lib/psych/visitors/to_ruby.rb: Added ability to load array subclasses with ivars.
- * ext/psych/lib/psych/visitors/yaml_tree.rb: Added ability to dump array subclasses with ivars.
- * test/psych/test_array.rb: corresponding tests
- * ext/psych/emitter.c: fixing clang warnings. Thanks Joey!
- * ext/psych/lib/psych/visitors/to_ruby.rb: BigDecimals can be restored from YAML.
- * ext/psych/lib/psych/visitors/yaml_tree.rb: BigDecimals can be dumped to YAML.
- * test/psych/test_numeric.rb: tests for BigDecimal serialization
- * ext/psych/lib/psych/scalar_scanner.rb: Strings that look like dates should be treated as strings and not dates.
- * test/psych/test_scalar_scanner.rb: corresponding tests.
- * ext/psych/lib/psych.rb (module Psych): parse and load methods take

```

    an optional file name that is used when raising Psych::SyntaxError
    exceptions
* ext/psych/lib/psych/syntax_error.rb (module Psych): allow nil file
names and handle nil file names in the exception message
* test/psych/test_exception.rb (module Psych): Tests for changes.

* ext/psych/parser.c (parse): parse method can take an option file
name for use in exception messages.
* test/psych/test_parser.rb: corresponding tests.

* ext/psych/lib/psych.rb: remove autoload from psych
* ext/psych/lib/psych/json.rb: ditto

* ext/psych/lib/psych/tree_builder.rb: dump complex numbers,
rationals, etc with reference ids.
* ext/psych/lib/psych/visitors/yaml_tree.rb: ditto
* ext/psych/lib/psych/visitors/to_ruby.rb: loading complex numbers,
rationals, etc with reference ids.
* test/psych/test_object_references.rb: corresponding tests

* ext/psych/lib/psych/scalar_scanner.rb: make sure strings that look
like base 60 numbers are serialized as quoted strings.
* test/psych/test_string.rb: test for change.

* ext/psych/parser.c: remove unused variable.

* ext/psych/lib/psych/syntax_error.rb: Add file, line, offset, and
message attributes during parse failure.
* ext/psych/parser.c: Update parser to raise exception with correct
values.
* test/psych/test_exception.rb: corresponding tests.

* ext/psych/parser.c (parse): Use context_mark for indicating error
line and column.

* ext/psych/lib/psych/scalar_scanner.rb: use normal begin / rescue
since postfix rescue cannot receive the exception class. Thanks
nagachika!

```

git-svn-id: svn+ssh://ci.ruby-lang.org/ruby/branches/ruby_1_9_3@35165 b2dd03c8-39d4-4d8f-98ff-823fe69b080e

Revision 7d984d76 - 03/29/2012 01:25 AM - tenderlovmaking (Aaron Patterson)

merge revision(s)

32578,33401,33403,33404,33531,33655,33679,33809,33900,33965,34067,34069,34087,34328,34330,34527,34772,34783,34839,34914,34953,34954,35153: [Backport #6212]

```

* ext/psych/lib/psych.rb: updating version to match gem
* ext/psych/psych.gemspec: ditto
* ext/psych/lib/psych/visitors/to_ruby.rb: fixing deprecation warning

* ext/psych/lib/psych.rb: define a new BadAlias error class.

* ext/psych/lib/psych/visitors/to_ruby.rb: raise an exception when
deserializing an alias that does not exist.

* test/psych/test_merge_keys.rb: corresponding test.

* ext/psych/lib/psych.rb (load, parse): stop parsing or loading after
the first document has been parsed.

* test/psych/test_stream.rb: pertinent tests.

* ext/psych/lib/psych.rb (parse_stream, load_stream): if a block is
given, documents will be yielded to the block as they are parsed.
[ruby-core:42404] [Bug #5978]

* ext/psych/lib/psych/handlers/document_stream.rb: add a handler that
yields documents as they are parsed

* test/psych/test_stream.rb: corresponding tests.

* ext/psych/lib/psych/core_ext.rb: only extend Kernel if IRB is loaded
in order to stop method pollution.

```

- * ext/psych/lib/psych.rb: default open YAML files with utf8 external encoding. [ruby-core:42967]
- * test/psych/test_tainted.rb: ditto

- * ext/psych/parser.c: prevent a memory leak by protecting calls to handler callbacks.
- * test/psych/test_parser.rb: test to demonstrate leak.

- * ext/psych/parser.c: set parser encoding based on the YAML input rather than user configuration.
- * test/psych/test_encoding.rb: corresponding tests.
- * test/psych/test_parser.rb: ditto
- * test/psych/test_tainted.rb: ditto

- * ext/psych/parser.c: removed external encoding setter, allow parser to be reused.
- * ext/psych/lib/psych/parser.rb: added external encoding setter.
- * test/psych/test_parser.rb: test parser reuse

- * ext/psych/lib/psych/visitors/to_ruby.rb: Added support for loading subclasses of String with ivars
- * ext/psych/lib/psych/visitors/yaml_tree.rb: Added support for dumping subclasses of String with ivars
- * test/psych/test_string.rb: corresponding tests

- * ext/psych/lib/psych/visitors/to_ruby.rb: Added ability to load array subclasses with ivars.
- * ext/psych/lib/psych/visitors/yaml_tree.rb: Added ability to dump array subclasses with ivars.
- * test/psych/test_array.rb: corresponding tests

- * ext/psych/emitter.c: fixing clang warnings. Thanks Joey!

- * ext/psych/lib/psych/visitors/to_ruby.rb: BigDecimals can be restored from YAML.
- * ext/psych/lib/psych/visitors/yaml_tree.rb: BigDecimals can be dumped to YAML.
- * test/psych/test_numeric.rb: tests for BigDecimal serialization

- * ext/psych/lib/psych/scalar_scanner.rb: Strings that look like dates should be treated as strings and not dates.

- * test/psych/test_scalar_scanner.rb: corresponding tests.

- * ext/psych/lib/psych.rb (module Psych): parse and load methods take an optional file name that is used when raising Psych::SyntaxError exceptions
- * ext/psych/lib/psych/syntax_error.rb (module Psych): allow nil file names and handle nil file names in the exception message
- * test/psych/test_exception.rb (module Psych): Tests for changes.

- * ext/psych/parser.c (parse): parse method can take an option file name for use in exception messages.
- * test/psych/test_parser.rb: corresponding tests.

- * ext/psych/lib/psych.rb: remove autoload from psych
- * ext/psych/lib/psych/json.rb: ditto

- * ext/psych/lib/psych/tree_builder.rb: dump complex numbers, rationals, etc with reference ids.
- * ext/psych/lib/psych/visitors/yaml_tree.rb: ditto
- * ext/psych/lib/psych/visitors/to_ruby.rb: loading complex numbers, rationals, etc with reference ids.
- * test/psych/test_object_references.rb: corresponding tests

- * ext/psych/lib/psych/scalar_scanner.rb: make sure strings that look like base 60 numbers are serialized as quoted strings.
- * test/psych/test_string.rb: test for change.

- * ext/psych/parser.c: remove unused variable.

- * ext/psych/lib/psych/syntax_error.rb: Add file, line, offset, and message attributes during parse failure.
- * ext/psych/parser.c: Update parser to raise exception with correct values.

```
* test/psych/test_exception.rb: corresponding tests.

* ext/psych/parser.c (parse): Use context_mark for indicating error
  line and column.

* ext/psych/lib/psych/scalar_scanner.rb: use normal begin / rescue
  since postfix rescue cannot receive the exception class. Thanks
  nagachika!
```

git-svn-id: svn+ssh://ci.ruby-lang.org/ruby/branches/ruby_1_9_3@35165 b2dd03c8-39d4-4d8f-98ff-823fe69b080e

History

#1 - 02/09/2012 02:23 AM - Anonymous

On Wed, Feb 08, 2012 at 01:47:31AM +0900, Thomas Sawyer wrote:

Issue [#5978](#) has been reported by Thomas Sawyer.

Bug [#5978](#): YAML.load_stream should process documents as they are read
<https://bugs.ruby-lang.org/issues/5978>

Author: Thomas Sawyer

Status: Open

Priority: Normal

Assignee:

Category:

Target version: 2.0.0

ruby -v: ruby 1.9.3p0 (2011-10-30 revision 33570) [x86_64-linux]

Psych say YAML.load_documents is deprecated and say to use YAML.load_stream instead.

Looking at the implementation for load_stream(), looks to me as if it waits for all documents in the stream to load before anything can be done with it.

```
# File 'lib/psych.rb', line 221
```

```
def self.load_stream yaml
  parse_stream(yaml).children.map { |child| child.to_ruby }
end
```

I don't think this should be the case. Ideally load_stream() would take a block, and if an IO object is given, read a document, yield it and then read the next document, and so on.

I imagine an Enumerator might be applicable to this as well.

I'd rather not change load_stream, but I want this functionality as well. What about something like this:

```
YAML::Reader.new(io).each do |doc|
  ...
end
```

Deserialized documents will be yielded as read. Does that seem acceptable? I'm hesitant to make it enumerable though because if we're truly doing stream processing, you couldn't iterate on the same object twice (imagine reading YAML from a socket or something).

--

Aaron Patterson

<http://tenderlovmaking.com/>

#2 - 02/09/2012 03:51 AM - trans (Thomas Sawyer)

Yea, that would suffice. It would still be nice to have a more intuitive/convenient class method though.

What about a new method, process_stream or each_document, or something like that, to wrap that code? Oh wait... why not just keep load_documents method for this and that way it will remain backward compatible with Syck API?

#3 - 02/09/2012 11:23 AM - Anonymous

On Thu, Feb 09, 2012 at 03:51:53AM +0900, Thomas Sawyer wrote:

Issue [#5978](#) has been updated by Thomas Sawyer.

Yea, that would suffice. It would still be nice to have a more intuitive/convenient class method though.

What about a new method, `process_stream` or `each_document`, or something like that, to wrap that code? Oh wait... why not just keep `load_documents` method for this and that way it will remain backward compatible with Syck API?

Honestly, I think you're right about the `load_stream` method. I'll just make it take a block and act the same as `load_documents`.

--
Aaron Patterson
<http://tenderlovmaking.com/>

#4 - 02/10/2012 12:53 AM - trans (Thomas Sawyer)

Cool.

#5 - 03/07/2012 07:48 AM - tenderlovmaking (Aaron Patterson)

- Assignee set to tenderlovmaking (Aaron Patterson)

#6 - 03/09/2012 06:21 AM - tenderlovmaking (Aaron Patterson)

- Status changed from Open to Closed

- % Done changed from 0 to 100

This issue was solved with changeset r34953.
Thomas, thank you for reporting this issue.
Your contribution to Ruby is greatly appreciated.
May Ruby be with you.

-
- `ext/psych/lib/psych.rb` (`parse_stream`, `load_stream`): if a block is given, documents will be yielded to the block as they are parsed.
[\[ruby-core:42404\]](#) [Bug [#5978](#)]
 - `ext/psych/lib/psych/handlers/document_stream.rb`: add a handler that yields documents as they are parsed
 - `test/psych/test_stream.rb`: corresponding tests.

Files			
noname	500 Bytes	02/09/2012	Anonymous
noname	500 Bytes	02/09/2012	Anonymous