Kepler - Bug #4578

tracking bug for changes to .kepler

11/24/2009 01:24 PM - Derik Barseghian

Status:	Resolved	**Start date:**	11/24/2009
Priority:	Immediate	**Due date:**	
Assignee:	Derik Barseghian	**% Done:**	0%
Category:	core	**Estimated time:**	0.00 hour
Target version:	2.0.0	**Spent time:**	0.00 hour
Bugzilla-Id:	4578		

Description

Thanks everyone for your input to the "proposed changes to .kepler" email. The threads are here:
http://mercury.nceas.ucsb.edu/kepler/pipermail/kepler-dev/2009-November/016568.html
http://mercury.nceas.ucsb.edu/kepler/pipermail/kepler-dev/2009-November/016577.html

I don't have an exact plan yet, but this bug will track progress. I think it's clear we want to at least divide cache into temporary and persistent data areas. Things get more complicated when you ask how and when to separate by "version" of kepler.

Related issues:

Blocks Kepler - Bug #4785: move cache object from 1.0 to 2.0	**New**	**02/10/2010**
Blocks Kepler - Bug #4702: KAR to module conversion utility	**Resolved**	**01/27/2010**

History

#1 - 12/02/2009 05:37 PM - Derik Barseghian

.kepler now has cache and persistent areas like:

.kepler/cache/gamma.msi.ucsb.edu.OpenAuth.1278/modules/actors (etc)
.kepler/persistent/gamma.msi.ucsb.edu.OpenAuth.1278/modules/actors (etc)
.kepler/persistent/gamma.msi.ucsb.edu.OpenAuth.1278/modules/actors/configuration/

The idea is that if a developer writes something into a persistent folder (like the configuration folder shown above) they're assuming the responsiblity of dealing with that data during version upgrades. Otherwise they should probably write their temporary items into the appropriate module's cache dir.

Today in a meeting we decided it would probably be better to move persistent out of .kepler, and put it somewhere more apparent to the user. For example if a user's migrating to a different machine, it will be more apparent they should take their ~/myKepler dir with. By default this might be ~/myKepler/persistent. We also decided it would be good to put the local repository library folders (e.g. Actors, Directors, R) into ~/myKepler too. (And "myKepler" was just a suggestion).

I'm planning to try to move separate the provenance DB back out into it's own DB inside the provenance persistent folder, and move the kepler DB into the appropriate cache module folder.

#2 - 12/15/2009 03:25 PM - Aaron Aaron

I have removed the use of the AuthNamespace from the .kepler directory structure.

The cache directory is now
~/.kepler/cache

and the persistent directory is
~/KeplerData

#3 - 02/05/2010 01:58 PM - Derik Barseghian

We've now got KeplerData with module dirs for persistent data, and .kepler for temporary data.

Chad recently added a versionMarker.txt to .kepler. I'm thinking we should probably put similar files in all the modules dirs in KeplerData so that during the install of a future post 2.0 Kepler we can easily id and upgrade data as needed. Sound ok?

I think Chad did some work to upgrade existing 1.0 .kepler to 2.0, but not everything can be safely upgraded. So we have a problem: .kepler can no longer be deleted at will, since the user may still have 1.0 data that they wish to use in with 1.0. Since kepler 1.0 is hardcoded to look for .kepler, we can't move the data into KeplerData. Maybe a solution is to create .kepler2 for use by 2.0 and beyond. Then we know .kepler2 can always safely be deleted, but .kepler should be left alone.

We no longer use the AuthorizedNamespace as part of the directory structures in KeplerData. This file resides in KeplerData/core/. So this file serves as a unique identifier for a KeplerData dir, not for a Kepler instance. Multiple instances of Kepler running on the same machine connect to the same databases and KeplerData dir, and use this same namespace. Did we decided we're OK with this?

#4 - 02/05/2010 02:17 PM - Chad Berkley

the upgrade work I did happens in two parts:

First, if DotKeplerManager (DKM) sees the directory signature of a 1.0 .kepler directory, it renames the kepler 2.0 cache directory to 'cache-2.0.0'. It then sets kepler 2.0 to use that cache dir. That dir can be removed at anytime.

The second thing it does is to add a versionMarker file. The version marker serves two purposes. It saves the version number of the latest version of kepler to use the directory, and it also writes any changes to the DKM variables that result from an upgrade. I did this just to save cycles in determining and redetermining if the directory names should be changed. If the version marker is removed, it will be recreated.

I decided to do this instead of making a .kepler2 directory just because I thought it might be annoying to have more than one directory around. It seems like we will also end up with .kepler3 and .kepler4, etc in the future, which is not desirable (IMHO). It is still totally possible (and easy) (if anyone has a good reason to do it that way) to use .keplerX instead of the way I did it.

#5 - 02/05/2010 02:27 PM - Derik Barseghian

I'm ok with that -- I was erroneously thinking the build system deleted all of .kepler with ant clean-cache.(In reply to comment #4)

> the upgrade work I did happens in two parts:
>
> First, if DotKeplerManager (DKM) sees the directory signature of a 1.0 .kepler
> directory, it renames the kepler 2.0 cache directory to 'cache-2.0.0'. It then
> sets kepler 2.0 to use that cache dir. That dir can be removed at anytime.
>
> The second thing it does is to add a versionMarker file. The version marker
> serves two purposes. It saves the version number of the latest version of
> kepler to use the directory, and it also writes any changes to the DKM
> variables that result from an upgrade. I did this just to save cycles in
> determining and redetermining if the directory names should be changed. If the
> version marker is removed, it will be recreated.
>
> I decided to do this instead of making a .kepler2 directory just because I
> thought it might be annoying to have more than one directory around. It seems
> like we will also end up with .kepler3 and .kepler4, etc in the future, which
> is not desirable (IMHO). It is still totally possible (and easy) (if anyone
> has a good reason to do it that way) to use .keplerX instead of the way I did
> it.

#6 - 02/05/2010 02:29 PM - Chad Berkley

I'm actually not sure that it doesn't. I'll have to check that out.

#7 - 02/05/2010 02:53 PM - Aaron Aaron

Yeah the AuthNamespace as Kepler Instance Identifier fell through when we decided to not write anything to the root of the installation. Now it uniquely identifies the KeplerData directory...

I think renaming the .kepler directory may be a good way to go (.kep, .kepler_cache, .kepcache, .keepler, .kepler_transient, or something). Since it is now truly a transient directory we would not need to update it any more in the future (i.e. no need for .kepler3, .kepler4, etc.)

#8 - 02/05/2010 03:24 PM - Christopher Brooks

I suggest creating a version specific directory inside .kepler instead of renaming it.

Something like:

.kepler/v/2.0/

The reasons:
I have enough dot files in my home directory as it is.
The /v/ means version, which means all version specific files would be
down a level. This is good in case there are other files added to
.kepler/. Pushing things down a level helps keep things organized.

Just some thoughts.

#9 - 02/10/2010 03:38 PM - Derik Barseghian

After talking w/ Chad and Matt the plan is to mostly leave things as they are. We just need to add version info to some of the files we're serializing in KeplerData, and we also decided to change 2.0 to **always** use .kepler/cache-2.0.0/. I think if we complete bug#4785 and bug#4702, it will then be safe for 2.0 to delete all of .kepler at will, provided the conversions have run. Otherwise we should change the build system to not delete all of .kepler during ant clean-all.

Matt reminded me modules are currently supposed to keep track of their own versioning. So devs, if you think at some point we'll want to know the version of serialized data in KeplerData/modules/[module-name]/, you probably want to be writing that into the files, into a text file, or storing the data in a versioned directory, etc. I've looked through my wrp ~/KeplerData, and the only things I see as candidates for needing version info are the coreDB, core/configuration.xml and provenanceDB. I'll look into these.

#10 - 02/11/2010 12:25 PM - Derik Barseghian

changed to always use cache-2.0.0 in r23059.

#11 - 02/17/2010 02:06 PM - ben leinfelder

'ant clean-cache' seems to not be deleting anything at the moment. shouldn't it?

#12 - 02/17/2010 02:12 PM - Derik Barseghian

(In reply to comment #11)

> 'ant clean-cache' seems to not be deleting anything at the moment. shouldn't
> it?

Whoops, it should. I can look into that.

#13 - 02/17/2010 02:27 PM - Derik Barseghian

ant clean-cache deletes .kepler/cache-2.0.0 in r23085

#14 - 02/17/2010 02:35 PM - Aaron Aaron

The kar directories that used to be in the cache are now in the KeplerData directory, so they are not deleted/rebuilt after an "ant clean-cache". We probably need a new build task for "ant clean-persistent". See bug 4723

#15 - 02/17/2010 02:43 PM - Chad Berkley

I would argue against having a task to clean the persistent directory. It's supposed to be persistent and not have to be "cleaned". If there is a case where it needs to be removed, doing it manually should suffice, or even better, write a task to just change what is there and upgrade it.

#16 - 02/23/2010 09:46 AM - Chad Berkley

Since 4785 has been closed, I don't think we're going to make any more changes to .kepler, so I'm closing this tracking bug.

#17 - 03/27/2013 02:27 PM - Redmine Admin

Original Bugzilla ID was 4578