Package 'basilisk.utils'

November 20, 2024

Version 1.19.0

Date 2024-09-10

Title Basilisk Installation Utilities

Imports utils, methods, tools, dir.expiry

Suggests reticulate, knitr, rmarkdown, BiocStyle, testthat, basilisk

biocViews Infrastructure

Description Implements utilities for installation of the basilisk package, primarily for creation of the underlying Conda instance. This allows us to avoid re-writing the same R code in both the configure script (for centrally administered R installations) and in the lazy installation mechanism (for distributed package binaries). It is highly unlikely that developers - or, heaven forbid, end-users! - will need to interact with this package directly; they should be using the basilisk package instead.

License GPL-3

RoxygenNote 7.2.3

VignetteBuilder knitr

git_url https://git.bioconductor.org/packages/basilisk.utils

git_branch devel

git_last_commit f596ce0

git_last_commit_date 2024-10-29

Repository Bioconductor 3.21

Date/Publication 2024-11-20

Author Aaron Lun [aut, cre, cph]

Maintainer Aaron Lun <infinite.monkeys.with.keyboards@gmail.com>

Contents

`activateEnvironment` *Activate a Conda environment*

Description

Mimic the (de)activation of a Conda environment by modifying environment variables in the current R process.

Usage

```
activateEnvironment(envpath = NULL, full.activation = NA, loc = getCondaDir())

deactivateEnvironment(listing)
```

Arguments

`envpath`	String containing the path to the Conda environment to activate. If `NULL`, the base Conda instance at `getCondaDir()` is activated.
`full.activation`	Logical scalar indicating whether the conda `activate` command should be run on the environment. Default is `NA`, which means `TRUE` on Windows and `FALSE` anywhere else.
`loc`	String containing the path to the root of a conda instance.
`listing`	Named list of strings containing name:value pairs for environment variables, typically the output of `activateEnvironment`.

Details

Conda environments generally need to be activated with the conda `activate` command to function properly. This is especially relevant on Windows where the `"PATH"` variable needs to be modified for the DLL search. When performing full activation, the `activateEnvironment` function mimics the effect conda `activate` by modifying environment variables in the current R session. This can be reversed by `deactivateEnvironment` once the Conda environment is no longer in use.

The `activateEnvironment` function will also unset a few bothersome environment variables:

- `"PYTHONPATH"`: to avoid compromising the version guarantees if **reticulate**'s `import` is allowed to search other locations beyond the specified Conda environment.
- `"PYTHONNOUSERSITE"`: similarly, to avoid searching the user's site libraries.
- `"RETICULATE_PYTHON"`: this would otherwise override any choice of Python, even after explicit specification via **reticulate**'s `use_Condaenv`!
- `"RETICULATE_PYTHON_ENV"`: for similar reasons.

Value

`activateEnvironment` will modify environment variables to mimic activation of the Conda environment. It returns a named list of the previous values of all variables modified in this manner. (`NA` values indicate that the corresponding variable was not previously set.)

`deactivateEnvironment` restores the environment variables to their pre-activation state. It returns `NULL` invisibly.

Author(s)

Aaron Lun

Examples

```
# We can't actually run activateEnvironment() here, as it
# either relies on basilisk already being installed or
# it has a hard-coded path to the basilisk system dir.
print("dummy test to pass BiocCheck")
```

`cleanConda` *Run* conda `clean`

Description

Clean the Conda installation to remove unused packages and tarballs.

Usage

```
cleanConda(loc = getCondaDir())
```

Arguments

loc	String containing the path to the root of a conda instance.

Details

This should only be performed to save disk space for system installations, as the cached extras will never be used by any other **basilisk** package.

In contrast, for an externally cached conda, the extras may be re-used by other **basilisk** clients. So it's usually worth keeping them around, though this function can be called directly by the user if the cache gets too big.

Value

An integer value indicating whether the cleaning was successful.

Author(s)

Aaron Lun

Examples

```
# We can't actually run cleanConda() here, as it
# either relies on basilisk already being installed or
# it has a hard-coded path to the basilisk system dir.
print("dummy test to pass BiocCheck")
```

---

clearExternalDir *Clear the external installation directory*

Description

Clear the external installation directory by removing all or obsolete conda instances/environments. This can be used to free up some space if the expiry mechanism is not fast enough at deleting unused environments.

Usage

```
clearExternalDir(
  path = getExternalDir(),
  package = NULL,
  obsolete.only = FALSE
)
```

Arguments

path	String containing a path to the external directory containing the current conda installation and environments.
package	String containing the name of a client R package. If provided, all environments will be removed for this package.
obsolete.only	Logical scalar indicating whether to only remove environments for obsolete package versions.

Value

If package=NULL and obsolete.only=FALSE, all of the conda instances (and associated environments) in the external directory are destroyed. If obsolete.only=TRUE, the conda instances and environments associated with **basilisk** versions older than the current version are destroyed.

If package is supplied and obsolete.only=FALSE, all conda environments for the specified client package are destroyed. If obsolete.only=FALSE, only the environments for older versions of the client package are destroyed.

Author(s)

Aaron Lun

See Also

getExternalDir, which determines the location of the external directory.

installConda, for the motivation behind this function.

Examples

```
# We can't actually run clearExternalDir() here, as it
# relies on basilisk already being installed.
print("dummy test to pass BiocCheck")
```

---

destroyOldVersions *Destroy old versions?*

Description

Should we destroy old installations of Conda from previous versions of **basilisk** (or old environment installations, for **basilisk** client packages)?

Usage

```
destroyOldVersions()
```

Details

The default value is TRUE, in order to save some hard drive space. This can be changed by setting BASILISK_NO_DESTROY environment variable to "1".

Value

Logical scalar providing an answer to the above.

Author(s)

Aaron Lun

See Also

installConda, where this function is used.

clearObsoleteDir, which may be triggered by this function.

---

dir.create2 *Safe directory construction*

Description

Create a directory with an error message if it does not succeed.

Usage

```
dir.create2(path, recursive = TRUE, ...)
```

Arguments

path, recursive, ...
: Further arguments to pass to dir.create.

Details

This is primarily necessary to avoid incomprehensible errors when a directory cannot be created, usually due to insufficient permissions. We set recursive=TRUE by default for convenience.

Note that the presence of an existing directory at path will cause this function to fail. This is usually desirable in the context of **basilisk.utils** as stale directories should be unlink2ed beforehand.

Value

Either path is created or an error is raised. NULL is invisibly returned.

See Also

unlink2, for a similarly safe deletion function.

Examples

```
out <- tempfile()
dir.create2(out)
```

`getBinaries` *Get binary paths*

Description

Get binary paths

Usage

```
getCondaBinary(loc)

getPythonBinary(loc)
```

Arguments

`loc`	String containing the path to the root of a conda instance or environment.

Details

This code is largely copied from **reticulate**, and is only present here as they do not export these utilities for general consumption.

Value

String containing the path to the conda or Python executable inside `loc`. If `loc` is not supplied, the relative path from the root of the environment is returned.

Author(s)

Aaron Lun

Examples

```
getCondaBinary()

getPythonBinary()
```

getCondaDir *Get the* **basilisk** *Conda directory*

Description

Find the installation directory for the **basilisk**-managed Conda instance.

Usage

```
getCondaDir(installed = TRUE)
```

Arguments

installed	Logical scalar indicating whether **basilisk** is already installed.

Details

By default, conda is installed to a location specified by `getExternalDir`. This ensures that R package build systems do not attempt to generate binaries that include the conda instance; such binaries are not relocatable due to the presence of hard-coded paths, resulting in run-time failures.

If the `BASILISK_EXTERNAL_CONDA` environment variable is set to a path to an existing conda instance, the function will return it directly without modification. This allows users to use their own conda instances with **basilisk** but, in turn, they are responsible for managing it.

If the `BASILISK_USE_SYSTEM_DIR` environment variable is set to `"1"`, the function will return a path to a location inside the `basilisk` system installation directory. This is the ideal approach when installing from source as any conda and **basilisk** re-installations are synchronized. It also ensures that any R process that can load **basilisk** will also have permissions to access the conda instance, which makes life easier for sysadmins of clusters or other shared resources.

We suggest always calling this function after an `installConda` call, which guarantees the presence of the conda installation directory (or dies trying). Setting `installed=FALSE` should only happen inside the **basilisk** `configure` script.

Value

String containing the path to the conda instance.

Author(s)

Aaron Lun

Examples

```
# Setting the environment variable to run this example:
# all other modes rely on installation of basilisk.
old <- Sys.getenv("BASILISK_USE_SYSTEM_DIR")
Sys.setenv(BASILISK_USE_SYSTEM_DIR=1)
```

```
getCondaDir(installed=FALSE)

Sys.setenv(BASILISK_USE_SYSTEM_DIR=old)
```

---

getExternalDir *Get an external conda directory*

Description

Define an external location for installing the conda instance and **basilisk** environments.

Usage

```
getExternalDir()
```

Details

The default path contains the version number so that re-installation of **basilisk** will install a new instance of Conda. (This assumes that **basilisk** and **basilisk.utils** have synchronized version bumps.) See `installConda` for more details on how old versions of Conda are managed in this external directory.

If the `BASILISK_EXTERNAL_DIR` environment variable is set to some location, this will be used instead as the installation directory. Setting this variable is occasionally necessary if the default path returned by `R_user_dir` has spaces; or on Windows, if the 260 character limit is exceeded after combining the default path with deeply nested conda paths.

We assume that the user has read-write access to the external directory. Write access is necessary to generate new environments and to handle locking in `lockExternalDir`.

Value

String containing a path to an appropriate external folder. The last component of the path will always be the **basilisk** version number.

Author(s)

Aaron Lun

See Also

`getCondaDir`, to obtain the Conda installation directory.

Examples

```
# We can't actually run getExternalDir() here, as it
# either relies on basilisk already being installed.
print("dummy test to pass BiocCheck")
```

getFallbackREnv *Get the path to the fallback R environment*

Description

Get the path to the conda environment containing an internal R installation. This is used as a last resort fallback for **reticulate** when there are shared library version conflicts with the current R installation.

Usage

```
getFallbackREnv()
```

Details

If the environment does not yet exist and `useSystemDir` is `FALSE`, it is created on the fly. Otherwise, if `useSystemDir` is `TRUE`, an error is thrown; this is because the fallback environment should have been created by `installConda`, unless `noFallbackR` is set to `TRUE`.

Value

String containing the path to the conda environment with a fallback version of R installed.

Author(s)

Aaron Lun

Examples

```
# We can't actually run getFallbackEnv() here, as it
# either relies on basilisk already being installed or
# it has a hard-coded path to the basilisk system dir.
print("dummy test to pass BiocCheck")
```

getSystemDir *Get the system installation directory*

Description

Get the system installation directory for a package. This is not entirely trivial as it may be called before the package is installed.

Usage

```
getSystemDir(pkgname, installed)
```

Arguments

pkgname	String containing the package name.
installed	Logical scalar specifying whether the package is likely to be installed yet.

Value

String containing the path to the (likely, if installed=FALSE) installation directory for pkgname.

Author(s)

Aaron Lun

Examples

```
getSystemDir("basilisk", installed=FALSE)
```

---

installConda *Install (Mini)conda*

Description

Install conda - specifically Miniconda, though historically we used Anaconda - to an appropriate destination path, skipping the installation if said path already exists.

Usage

```
installConda(installed = TRUE)
```

Arguments

installed	Logical scalar indicating whether **basilisk** is already installed. Should only be set to FALSE in **basilisk** configure scripts.

Details

This function was originally created from code in https://github.com/hafen/rminiconda, also borrowing code from **reticulate**'s install_miniconda for correct Windows installation. It downloads and runs a Miniconda installer to create a dedicated Conda instance that is managed by **basilisk**, separate from other instances that might be available on the system.

The installer itself is cached to avoid re-downloading it when, e.g., re-installing **basilisk** across separate R sessions. Users can obtain/delete the cached installer by looking at the contents of the parent directory of getExternalDir. This caching behavior is disabled for system installations (see useSystemDir), which touch nothing except the system directories; in such cases, only repeated installation attempts in the same R session will re-use the same installer.

Currently, we use version 4.12.0 of the Miniconda3 installer, which also comes with Python 3.9. Users can change this by setting the BASILISK_MINICONDA_VERSION environment variable, e.g., to

"py38_4.11.0". Any change should be done with a great deal of caution, typically due to some system-specific problem with a particular Miniconda version. If it must be done, users should try to stick to the same Python version.

Value

A conda instance is created at the location specified by getCondaDir. Nothing is performed if a complete instance already exists at that location. A logical scalar is returned indicating whether a new instance was created.

Destruction of old instances

Whenever installConda is re-run and BASILISK_USE_SYSTEM_DIR is not set, any old conda instances and their associated **basilisk** environments are deleted from the external installation directory. This avoids duplication of large conda instances after their obselescence. Client packages are expected to recreate their environments in the latest conda instance.

Users can disable this destruction by setting the BASILISK_NO_DESTROY environment variable to "1". This may be necessary on rare occasions when running multiple R instances on the same Bioconductor release. Note that setting this variable is not required for R instances using different Bioconductor releases; the destruction process is smart enough to only remove conda instances generated from the same release.

Skipping the fallback R

When BASILISK_USE_SYSTEM_DIR is set, installConda will automatically create a conda environment with its own copy of R. This is used as the "last resort fallback" for running **reticulate** code in the presence of shared library incompatibilities with the main R installation. If users know that no incompatibilities exist in their application, they can set the BASILISK_NO_FALLBACK_R variable to "1". This will instruct installConda to skip the creation of the fallback environment, saving some time and disk space.

Author(s)

Aaron Lun

Examples

```
# We can't actually run installConda() here, as it
# either relies on basilisk already being installed or
# it has a hard-coded path to the basilisk system dir.
print("dummy test to pass BiocCheck")
```

isWindows *Find the operating system or architecture.*

Description

Indicate whether we are on Windows or MacOSX. For MacOSX and Linux, we can also determine if we are on an x86-64 or Arm-based architecture.

Usage

```
isWindows()

isMacOSX()

isMacOSXArm()

isLinux()

isLinuxAarch64()
```

Value

Logical scalar indicating whether we are on the specified OS and/or architecture.

Author(s)

Aaron Lun

Examples

```
isWindows()
isMacOSX()
isLinux()
```

lockExternalDir *Lock external directory*

Description

Lock the external Conda installation directory so that multiple processes cannot try to install at the same time.

Usage

```
lockExternalDir(path = getExternalDir(), ...)

unlockExternalDir(lock.info, ...)
```

Arguments

path	String containing the path to the external directory.
...	For lockExternalDir, further arguments to pass to lockDirectory such as exclusive. For unlockExternalDir, further arguments to pass to unlockDirectory such as clear.
lock.info	A lock object generated by lockDirectory.

Details

This will apply a lock to the (possibly user-specified) external Conda installation directory, so that a user trying to run parallel **basilisk** processes will not have race conditions during lazy Conda installation. We use **dir.expiry** to manage the locking process for us, with the following strategy:

- If a system installation is being performed, we do not perform any locking. Rather, the R package manager will lock the entire R installation directory for us.
- If the external directory is not yet present, we establish an exclusive lock. We then proceed to the creation of said directory and installation of Conda.
- If an external installation directory is already present, we establish a shared lock. This will wait for any exclusive lock to expire (and thus any currently running installation to finish). No waiting is required if there are no existing exclusive locks.

Note that locking is only required during installation of Conda (or its environments), not during actual use. Once an installation/environment is created, we assume that it is read-only for all processes. Technically, this might not be true if one were to install a new version of **basilisk** halfway through an R session, which would prompt installConda to wipe out the old Conda installations; but one cannot in general guarantee the behavior of running R sessions when package versions change anyway, so we won't bother to protect against that.

Value

lockExternalDir will return a lock object from lockDirectory.

unlockExternalDir will unlock the file and return NULL invisibly.

Author(s)

Aaron Lun

See Also

installConda, for an example of how to implement this locking approach.

Examples

```
loc <- lockExternalDir()
unlockExternalDir(loc)
```

noFallbackR *Don't install a fallback version of R*

Description

By default, installConda will install a conda-managed R in a new environment, to provide a fallback for GLIBCXX-mismatch errors in C++-dependent Python packages like **scipy**. This cost can be avoided if the version of the C++ standard library used by R is known to be compatible with client environments.

Usage

```
noFallbackR()
```

Details

The default value is FALSE in order to provide a fallback when clients observe GLIBCXX errors. This can be changed by setting the BASILISK_NO_FALLBACK_R environment variable to "1".

Value

Logical scalar providing an answer to the above.

Author(s)

Aaron Lun

See Also

installConda, where this function is used.

setCondaPackageDir *Set or unset the Conda package directory*

Description

Set or unset the directory used to store the cached Conda packages, e.g., tarballs and such. This should be a non-temporary location as other packages may link to its contents.

Usage

```
setCondaPackageDir(loc)
```

Arguments

loc	A string containing a path to the desired directory (that should already exist). Alternatively NA, in which case any existing setting is removed.

Value

The previous value of CONDA_PKGS_DIRS, invisibly.

Author(s)

Aaron Lun

Examples

```
# Setting it to something new:
out <- setCondaPackageDir(tempdir())

# Setting it back
setCondaPackageDir(out)
```

---

setVariable *Set an environment variable*

Description

Set an environment variable safely, unsetting it if the supplied value is NA.

Usage

```
setVariable(name, value)
```

Arguments

name	String containing the name of an environment variable.
value	String containing the value of an environment variable. This can be NA to unset the variable.

Value

String containing the value of the variable before running this function; or NA, if the variable was not set.

Author(s)

Aaron Lun

`unlink2` *Safe file deletion*

Description

Delete files or directories with an error message if it does not succeed.

Usage

```
unlink2(x, recursive = TRUE, force = TRUE, ...)
```

Arguments

`x`, `recursive`, `force`, `...`
: Further arguments to pass to `unlink`.

Details

This is primarily necessary to avoid incomprehensible errors when a directory containing a stale environment or installation is not successfully deleted. We set `recursive=TRUE` by default for convenience; we also set `force=TRUE` by default to avoid difficulties due to rogue permissions.

Value

Either all `x` are successfully deleted or an error is raised. `NULL` is invisibly returned.

See Also

`dir.create2`, for a similarly safe directory creation function.

Examples

```
out <- tempfile()
unlink2(out) # no error from deleting non-existent file.

write(file=out, "whee")
unlink2(out)
```

`useSystemDir` *Use the R system directory?*

Description

Should we use the R system directory for installing **basilisk**'s Conda instance (or client environments)?

Usage

```
useSystemDir()
```

Details

The default value is `FALSE` to avoid problems with position-dependent code in packaged binaries. This can be changed by setting `BASILISK_USE_SYSTEM_DIR` environment variable to `"1"`.

Value

Logical scalar providing an answer to the above.

Author(s)

Aaron Lun

See Also

`installConda` and `getCondaDir`, where this function is used.

Index