

Package ‘BigMatrix’

September 22, 2012

Type Package

Title An extension of the bigmemory package with added safety, convenience, and a factor class.

Version 0.99.0

Author Peter M. Haverty

Maintainer Peter M. Haverty <phaverty@gene.com>

Description This package defines a “BigMatrix” ReferenceClass which adds safety and convenience features to the filebacked.big.matrix class from the bigmemory package. BigMatrix protects against segfaults by monitoring and gracefully restoring the connection to on-disk data and it also protects against accidental data modification with a filesystem-based permissions system. We provide utilities for using BigMatrix-derived classes as assayData matrices within the Biobase package’s eSet family of classes. BigMatrix provides some optimizations related to attaching to, and indexing into, file-backed matrices with dimnames. Additionally, the package provides a “BigMatrixFactor” class, a file-backed matrix with factor properties.

License Artistic-2.0

LazyLoad yes

ByteCompile TRUE

Depends R (>= 2.12)

Imports bigmemory, methods, Biobase

Suggests RUnit, BiocGenerics (>= 0.1.0)

biocViews Infrastructure, DataRepresentation

Collate ‘BigMatrix.R’ ‘eset.R’ ‘test_BigMatrix_package.R’ ‘zzz.R’

InstallableEverywhere yes

R topics documented:

BigMatrix-package	2
annotatedDataFrameFrom	2
attachAssayDataElements	3
BigMatrix	3
BigMatrix-class	4
BigMatrixFactor	6
BigMatrixFactor-class	7
updateAssayDataElementPaths	8

Index	10
--------------	-----------

BigMatrix-package	<i>An extension of the bigmemory package with added safety, convenience, and a factor class.</i>
-------------------	--

Description

This package defines a "BigMatrix" ReferenceClass which adds safety and convenience features to the filebacked.big.matrix class from the bigmemory package. BigMatrix protects against segfaults by monitoring, and gracefully restoring, the connection to on-disk data. It also protects against accidental data modification with a filesystem-based permissions system. We provide functionality for using BigMatrix-derived classes as assayData matrices within the Biobase package's eSet family of classes. BigMatrix provides some optimizations related to attaching to, and indexing into, file-backed matrices with dimnames. Additionally, the package provides a "BigMatrixFactor" class, a file-backed matrix with factor properties.

Details

BigMatrix stores the filesystem path to its on-disk data. When a big.matrix object is saved, e.g. with "save" the external pointer to the on-disk data becomes "nil". Any access to this nil pointer causes a segfault. A BigMatrix object will gracefully re-attach itself to its on-disk components any time use of the "nil" pointer is attempted.

See Also

BigMatrix-class BigMatrixFactor-class BigMatrix BigMatrixFactor filebacked.big.matrix ReferenceClasses

annotatedDataFrameFrom	<i>Use BigMatrix in eSet assayData</i>
------------------------	--

Description

The Biobase package's eSet objects hold one or more matrices within an environment stored in their assayData slot. Use of classes other than a base matrix require a method to create an annotated-DataFrame from that class. Users are unlikely to ever call this method.

 attachAssayDataElements

Attach on-disk matrices into assayData

Description

Biobase's eSet-derived objects (e.g. ExpressionSet, GenoSet) can hold one or more BigMatrix objects in their assayData slot. After re-loading an eSet from disk, these objects will each need to be re-attached to their on-disk component using their stored path information. This function checks each assayDataElement to see if it is an un-attached BigMatrix object, re-attaching if necessary. All other assayDataElements are left untouched. In later releases this function may also handle other on-disk types, like HDF5-based matrices.

Usage

```
attachAssayDataElements( aData )
```

Arguments

aData assayData object (environment or list style). Usually from the assayData slot of an eSet.

Details

Environment type assayData objects, even "lockedEnvironment" objects, will be updated in place (same pointer).

Value

assayData matching the supplied aData argument. Any BigMatrix objects will be attached. Re-assignment back original eSet only necessary if using a list type assayData, which is rare.

Examples

```
## Not run: assayData(eset) = attachAssayDataElements( assayData(eset) )
```

 BigMatrix

Create a new BigMatrix

Description

Create a new BigMatrix

Usage

```
BigMatrix(x = NULL, backingfile, nrow, ncol, dimnames,
  type = "double")
```

Arguments

x	NULL, matrix, or big.matrix. Optional data or big.matrix for new BigMatrix
backingfile	character, full path to the file that will contain the data matrix
nrow	integer, number of rows in the matrix we are about to create
ncol	integer, number of columns in the matrix we are about to create
dimnames	list, list(rownames,colnames), as for a typical matrix
type	character type of big.matrix (double, integer, char)

Value

BigMatrixFactor

Examples

```

dnames = dimnames=list(letters[1:3],LETTERS[1:3])
x = matrix(1:9,ncol=3,dimnames=dnames)
ds = BigMatrix(x,tempfile())
ds = BigMatrix(backingfile=tempfile(),nrow=3,ncol=3,dimnames=dnames)

```

BigMatrix-class

Class "BigMatrix"

Description

This package defines a "BigMatrix" ReferenceClass which adds safety and convenience features to the filebacked.big.matrix class from the bigmemory package. BigMatrix protects against segfaults by monitoring and gracefully restoring the connection to on-disk data and it also protects against accidental data modification with a filesystem-based permissions system. We provide utilities for using BigMatrix-derived classes as assayData matrices within the Biobase package's eSet family of classes. BigMatrix provides some optimizations related to attaching to, and indexing into, file-backed matrices with dimnames.

A big.matrix object is backed by two files, a data file (backingfile) and a text metadata file (descriptor file). BigMatrix stores the metadata in an additional binary "rds" file, for faster access. The location of this file is stored in the descpath field. The data in the metadata file is used to "attach" a big.matrix object its on-disk data. BigMatrix stores the big.matrix object and checks for a valid connection to the on-disk data each time it is necessary to access the big.matrix object. Re-attachment, using the attach method and the descpath field is performed as necessary. This avoids a segfault caused by accessing what could otherwise be a nil pointer. For example, A nil pointer occurs when a big.matrix data is saved and loaded using save() and load().

BigMatrix objects check the filesystem write permissions on the metadata file (the *.rds file) before modifying data in order to prevent undesired modification of any of the on-disk data. This feature will eventually be replaced with read-only or read/write access which is in development in the bigmemory package.

Reference classes are a relatively new and unused feature of R. They represent a new object model that is quite similar to OOP you might have seen in java or python. The ?ReferenceClasses help page is an excellent resource on this topic. As ReferenceClasses are an extension of R's S4 class system, S4-style methods can also be defined. In order to give the familiar matrix API, BigMatrix's ReferenceClass methods have a corresponding S4-style method. Please see the examples below.

Extends

All reference classes extend and inherit methods from "envRefClass".

Methods

S4-style methods supplement ReferenceClass methods to provide a standard matrix API.

[signature(x = "BigMatrix", i = "ANY", j = "ANY", drop = "ANY"): Access matrix data.

[<- signature(x = "BigMatrix", i = "ANY", j = "ANY"): Set matrix data.

[<- signature(x = "BigMatrix", i = "ANY", j = "ANY", NA = "ANY"): Set matrix data

annotatedDataFrameFrom signature(object = "BigMatrix"): Used in creation of a Biobase eSet with a BigMatrix as an assayDataElement.

coerce signature(from = "BigMatrix", to = "matrix"): Coerce BigMatrix to a base matrix.

dim signature(x = "BigMatrix"): Get dimensions of the matrix.

dimnames signature(x = "BigMatrix"): Get dimension names of the matrix.

dimnames<- signature(x = "BigMatrix", value = "ANY"): Set dimension names

length signature(x = "BigMatrix"): Get length of the vector used as a matrix (nrow * ncol).

ncol signature(x = "BigMatrix"): Get the number of columns in the matrix.

nrow signature(x = "BigMatrix"): Get the number of rows in the matrix.

Fields

bigmat: Object of class activeBindingFunction Public access to the big.matrix object.

datapath: Object of class activeBindingFunction Full file path the big.matrix backing file.

descpath: Object of class character Full file path the big.matrix metadata file.

.bm: Object of class big.matrix Private storage of the big.matrix object.

Class-Based Methods

save(): Saves metadata about big.matrix in .bm field to file at location given by descpath field.

colnames signature(x = "BigMatrix"): Get column names of the matrix.

rownames signature(x = "BigMatrix"): Get row names of the matrix.

dimnames signature(x = "BigMatrix"): Get dimension names of the matrix.

dim signature(x = "BigMatrix"): Get the dimensions of the matrix.

nrow signature(x = "BigMatrix"): Get the number of rows in the matrix.

ncol signature(x = "BigMatrix"): Get the number of columns in the matrix.

length signature(x = "BigMatrix"): Get the number of cells in the matrix.

setValues(i, j, value): Setter for on-disk data.

getValues(i, j, drop): Getter for on-disk data.

attach(): Uses path from descpath field to connect object to on-disk data.

Author(s)

Peter M. Haverty

See Also

See Also [ReferenceClasses](#), [BigMatrix](#), [BigMatrixFactor](#), [makeActiveBinding](#), [filebacked.big.matrix](#)

Examples

```
showClass("BigMatrix")
back.file = tempfile()
x = matrix(1:9,ncol=3,dimnames=list(letters[1:3],LETTERS[1:3]))
ds = BigMatrix(x,back.file)
ds[1,3]
ds[3,3] = 5
as(ds,"matrix")
ds[,]
rownames(ds)
ds$rownames()
colnames(ds)
ds$colnames()
dimnames(ds)
ds$dimnames()
dim(ds)
ds$dim()
nrow(ds)
ds$nrow()
ncol(ds)
ds$ncol()
```

BigMatrixFactor

Create a new BigMatrixFactor

Description

Create a new BigMatrixFactor

Usage

```
BigMatrixFactor(x = NULL, backingfile, nrow, ncol,
  dimnames, levels)
```

Arguments

x	NULL, matrix, or big.matrix. Optional data or big.matrix for new BigMatrix
backingfile	character, full path to the file that will contain the data matrix
nrow	integer, number of rows in the matrix we are about to create
ncol	integer, number of columns in the matrix we are about to create
dimnames	list, list(rownames,colnames), as for a typical matrix
levels	character, as for a typical factor

Value

BigMatrixFactor

Examples

```

dnames = dimnames=list(letters[1:3],LETTERS[1:3])
x = matrix( sample( 1:3, 9, replace=TRUE), ncol=3, dimnames=dnames)
ds = BigMatrixFactor(x,tempfile(),levels=c("AA","AB","BB"))
ds = BigMatrixFactor(backingfile=tempfile(),nrow=3,ncol=3,dimnames=dnames,levels=c("AA","AB","BB"))

```

BigMatrixFactor-class *Class* "BigMatrixFactor"

Description

The `BigMatrixFactor` class extends the `BigMatrix` class to add factor functionality. Data are stored as 32-bit or 8-bit integers on-disk depending on the number of levels. Reference Class methods are also provided in S4 style to give the familiar matrix and factor API (please see the examples below). Subsetting returns a integer matrix with the S3 class "factor" and an appropriate levels attribute.

Extends

Class "[BigMatrix](#)", directly.

All reference classes extend and inherit methods from "[envRefClass](#)".

Methods

S4-style methods supplement ReferenceClass methods to provide a standard factor API.

`signature(from = "BigMatrixFactor", to = "factor")`: ...

coerce `signature(from = "BigMatrixFactor", to = "matrix")`: ...

nlevels `signature(x = "BigMatrixFactor")`: Query the number of levels, as in a typical factor.

levels `signature(x = "BigMatrixFactor")`: Access the levels, as in a typical factor.

levels<- `signature(x = "BigMatrixFactor")`: Set the levels, as in a typical factor. Disallowed for now.

Fields

bigmat: Object of class `activeBindingFunction` Public access to the `big.matrix` object.

datapath: Object of class `activeBindingFunction` Full file path to the `big.matrix` backing file.

descpath: Object of class `character` Full file path to the `big.matrix` metadata file.

.bm: Object of class `big.matrix` Private storage of the `big.matrix` object.

levels: Object of class `character` Levels as in a typical factor.

Class-Based Methods

nlevels(): Check number of levels, as in a typical factor.

getValues(i, j, drop): Getter for on-disk data. Returns a factor with the appropriate number of dimensions.

setValues(i, j, value): Setter for on-disk data. Data converted to integer, matching to levels for character input.

The following methods are inherited (from the corresponding class): `setValues ("BigMatrix")`, `getValues ("BigMatrix")`, `attach ("BigMatrix")`, `save ("BigMatrix")`

Author(s)

Peter M. Haverty

See Also

See Also [ReferenceClasses](#), [BigMatrix](#), [factor](#), [BigMatrix](#), [makeActiveBinding](#), [filebacked.big.matrix](#)

Examples

```
showClass("BigMatrixFactor")
back.file = tempfile()
x = matrix( sample( 1:3, 9, replace=TRUE), ncol=3, dimnames=list(letters[1:3],LETTERS[1:3]))
ds = BigMatrixFactor(x,back.file,levels=c("AA","AB","BB"))
ds[1,3]
ds[3,3] = "AB"
as(ds,"matrix")
as(ds,"factor")
ds[,]
nlevels(ds)
ds$nlevels()
levels(ds)
ds$levels
```

updateAssayDataElementPaths

Update directory for BigMatrix assayDataElement to new location

Description

Update directory for BigMatrix assayDataElement to new location. Assumes files have already been moved on the filesystem. Assumes names of description and data files are the same.

Usage

```
updateAssayDataElementPaths(ds, dir)
```

Arguments

ds	eSet
dir	character, path to directory holding all BigMatrix files

Details

If you have renamed specific backing files or descriptor files, you will want to update the datapath and descpath fields of the relevant BigMatrix objects.

Value

eSet

Author(s)

Peter M. Haverty <phaverty@gene.com>

Examples

```
## Not run: eset = updateAssayDataElementPathss(eset, "/new/path/to/bigmat/dir")
```

Index

*Topic classes

- BigMatrix-class, 4
- BigMatrixFactor-class, 7
- [,BigMatrix,ANY,ANY,ANY-method (BigMatrix-class), 4
- [<-,BigMatrix,ANY,ANY,ANY-method (BigMatrix-class), 4
- [<-,BigMatrix,ANY,ANY-method (BigMatrix-class), 4
- annotatedDataFrameFrom, 2
- annotatedDataFrameFrom,BigMatrix-method (BigMatrix-class), 4
- attachAssayDataElements, 3
- BigMatrix, 3, 6–8
- BigMatrix-class, 4
- BigMatrix-package, 2
- BigMatrixFactor, 6, 6
- BigMatrixFactor-class, 7
- coerce,BigMatrix,matrix-method (BigMatrix-class), 4
- coerce,BigMatrixFactor,factor-method (BigMatrixFactor-class), 7
- coerce,BigMatrixFactor,matrix-method (BigMatrixFactor-class), 7
- dim,BigMatrix-method (BigMatrix-class), 4
- dimnames,BigMatrix-method (BigMatrix-class), 4
- dimnames<-,BigMatrix,ANY-method (BigMatrix-class), 4
- envRefClass, 5, 7
- factor, 8
- filebacked.big.matrix, 6, 8
- length,BigMatrix-method (BigMatrix-class), 4
- levels,BigMatrixFactor-method (BigMatrixFactor-class), 7
- levels<-,BigMatrixFactor-method (BigMatrixFactor-class), 7
- makeActiveBinding, 6, 8
- ncol,BigMatrix-method (BigMatrix-class), 4
- nlevels,BigMatrixFactor-method (BigMatrixFactor-class), 7
- nrow,BigMatrix-method (BigMatrix-class), 4
- ReferenceClasses, 6, 8
- updateAssayDataElementPaths, 8