Package: HDF5DataFrame (via r-universe)

June 4, 2026

Version 0.99.1

Date 2025-04-23

Title HDF5-backed DataFrame objects and methods

Description Implements bindings for h5 files that are compatible with Bioconductor S4 data structures, namely the DataFrame and DelayedArray. This allows HDF5-backed data to be easily used as data frames with arbitrary sets of columns.

License MIT + file LICENSE

Depends R (>= 4.5.0), DelayedArray, S4Vectors

Imports stats, methods, BiocGenerics, h5mread

Suggests testthat, knitr, HDF5Array, rhdf5

biocViews DataRepresentation, Infrastructure, Software

VignetteBuilder knitr

RoxygenNote 7.3.3

Encoding UTF-8

URL https://github.com/BIMSBbioinfo/HDF5DataFrame

BugReports https://github.com/BIMSBbioinfo/HDF5DataFrame/issues

Config/pak/sysreqs libssl-dev zlib1g-dev

Repository https://artur-man.r-universe.dev

Date/Publication 2025-09-20 13:51:25 UTC

RemoteUrl https://github.com/Artur-man/HDF5DataFrame

RemoteRef HEAD

RemoteSha 4d62ea45c480d31acbf6d16018858c112f9edb8d

Contents

HDF5ColumnSeed *HDF5ColumnSeed*

Description

Represent a column of a HDF5-based data frame as a 1-dimensional DelayedArray. This allows us to use HDF5-backed data inside DataFrame without loading them into memory.

Usage

```
HDF5ColumnSeed(path, name, column, type = NULL, length = NULL)

HDF5ColumnVector(x, ...)

## S4 method for signature 'HDF5ColumnSeed'
DelayedArray(seed)

## S4 method for signature 'HDF5ColumnSeed'
dim(x)

## S4 method for signature 'HDF5ColumnSeed'
type(x)

## S4 method for signature 'HDF5ColumnSeed'
path(object)

## S4 method for signature 'HDF5ColumnSeed'
extract_array(x, index)
```

Arguments

path	String containing a path to a HDF5-based data frame.
name	String containing the HDF5 group of the h5 file.
column	String containing the name of the column inside the file.
type	String specifying the type of the data. If NULL, this is determined by inspecting the file. Users may specify this to avoid a look-up, or to coerce the output into a different type.
length	Integer containing the number of rows. If NULL, this is determined by inspecting the file. This should only be supplied for efficiency purposes, to avoid a file look-up on construction.

x	Either a string containing the path to an HDF5-based data frame file (to be used as path), or an existing HDF5ColumnSeed object.
...	Further arguments to be passed to the HDF5ColumnSeed constructor.
seed, object	A HDF5ColumnSeed object
index	An unnamed list of integer vectors, one per dimension in x. See extract_array

Value

For HDF5ColumnSeed, a HDF5ColumnSeed is returned, obviously.

For HDF5ColumnVector, a HDF5ColumnVector is returned.

Author(s)

Artür Manukyan

Examples

```
# libraries
library(rhdf5)
library(HDF5Array)
library(HDF5DataFrame)

# h5
output_h5ad <- tempfile(fileext = ".h5")
h5createFile(output_h5ad)
h5createGroup(output_h5ad, group = "metadata")

# data
data("chickwts")
metadata <- chickwts

# set metadata
meta.data_list <- list()
for(i in 1:ncol(metadata)){
  cur_column <- as.vector(subset(metadata,
                                 select = colnames(metadata)[i]))[[1]]
  if(is.character(cur_column) || is.factor(cur_column))
    cur_column <- as.character(cur_column)
  cur_column <- as.array(cur_column)
  meta.data_list[[colnames(metadata)[i]]] <-
    writeHDF5Array(cur_column,
                   output_h5ad,
                   name = paste0("metadata", "/",
                                 colnames(metadata)[i]),
                   with.dimnames = FALSE)
}

# define hd5columnseed
columnseed <- HDF5ColumnSeed(path = path(meta.data_list[[1]]),
                             name = "metadata",
                             column = colnames(metadata)[i],
```

```
                                type = type(meta.data_list[[1]]))

# methods
dim(columnseed)
path(columnseed)
type(columnseed)
```

HDF5ColumnSeed-class *HDF5ColumnSeed Class*

Description

The HDF5ColumnSeed class for HDF5ColumnVector.

Arguments

path	The path (as a single string or H5File object) to the HDF5 file (.h5 or .h5ad) where the dataset is located.
name	The name of the dataset in the HDF5 file.
column	the names of the columns, see HDF5ColumnVector
length	the length of the HDF5Array.

HDF5ColumnVector-class
HDF5ColumnVector Class

Description

The HDF5ColumnVector class for each column of a HDF5DataFrame class

Arguments

seed	An HDF5ColumnSeed object

HDF5DataFrame *HDF5-backed DataFrame*

Description

Create a HDF5-backed DataFrame, where the data are kept on disk until requested.

Usage

```
HDF5DataFrame(x, name, columns = NULL, nrows = NULL)

## S4 method for signature 'HDF5DataFrame'
nrow(x)

## S4 method for signature 'HDF5DataFrame'
length(x)

## S4 method for signature 'HDF5DataFrame'
path(object)

## S4 method for signature 'HDF5DataFrame'
rownames(x)

## S4 method for signature 'HDF5DataFrame'
names(x)

## S4 replacement method for signature 'HDF5DataFrame'
rownames(x) <- value

## S4 replacement method for signature 'HDF5DataFrame,ANY'
names(x) <- value

## S4 method for signature 'HDF5DataFrame,ANY,ANY'
x[[i, j, ...]]

## S4 replacement method for signature 'HDF5DataFrame,ANY,ANY,ANY'
x[[i, j, ...]] <- value

## S4 method for signature 'HDF5DataFrame'
cbind(..., deparse.level = 1)

## S4 method for signature 'HDF5DataFrame'
as.data.frame(x, row.names = NULL, optional = FALSE, ...)
```

Arguments

x, object	A set of HDF5Arrays that are the columns of the HDF5DataFrame object.

name	String containing the HDF5 group of the h5 file.
columns	Character vector containing the names of columns in a HDF5-based data frame. If NULL, this is determined from path.
nrows	Integer scalar specifying the number of rows in a HDF5-based data frame. If NULL, this is determined from path.
value	rownames, names or new columns for HDF5DataFrame object
i	Depends on the usage
j	Depends on the usage
...	arguments passed to other methods
deparse.level	See ?base::cbind for a description of description of these arguments.
row.names, optional	See ?base::as.data.frame for a description of these arguments.

Value

A HDF5DataFrame object where each column is a HDF5ColumnVector.

HDF5DataFrame object

number of rows of HDF5DataFrame object

length of HDF5DataFrame object

path to hdf5 file of HDF5DataFrame object

rownames of HDF5DataFrame object

names of columns of HDF5DataFrame object

HDF5DataFrame object

data.frame object

Author(s)

Artür Manukyan

Examples

```
# libraries
library(rhdf5)
library(HDF5Array)
library(HDF5DataFrame)

# h5
output_h5ad <- tempfile(fileext = ".h5")
h5createFile(output_h5ad)
h5createGroup(output_h5ad, group = "metadata")

# data
data("chickwts")
metadata <- chickwts
```

```
# set metadata
meta.data_list <- list()
for(i in 1:ncol(metadata)){
  cur_column <- as.vector(subset(metadata,
                                 select = colnames(metadata)[i]))[[1]]
  if(is.character(cur_column) || is.factor(cur_column))
    cur_column <- as.character(cur_column)
  cur_column <- as.array(cur_column)
  meta.data_list[[colnames(metadata)[i]]] <-
    writeHDF5Array(cur_column,
                   output_h5ad,
                   name = paste0("metadata", "/",
                                 colnames(metadata)[i]),
                   with.dimnames = FALSE)
}
metadata_large <-
  HDF5DataFrame(meta.data_list,
                name = "metadata",
                columns = names(meta.data_list))

# coerce to data.frame
metadata_large <- as.data.frame(metadata_large)

# cbind
metadata_large <- cbind(metadata_large, metadata)
```

`HDF5DataFrame-class` *HDF5DataFrame Class*

Description

The HDF5DataFrame class is a DataFrame subclass for representing datasets with arbitrary collections of columns stored in HDF5.

Arguments

`path`	The path (as a single string or H5File object) to the HDF5 file (.h5 or .h5ad) where the dataset is located.
`name`	The name of the group in the HDF5 file.
`columns`	the names of the columns, see HDF5ColumnVector
`nrows`	the number of rows of the DataFrame.

subsetting-utils *subsetting-utils*

Description

Low-level utility functions and classes to support subsetting of vector-like objects. They are not intended to be used directly. See extractROWS.

Usage

```
## S4 method for signature 'HDF5DataFrame,ANY'
extractROWS(x, i)

## S4 method for signature 'HDF5DataFrame'
extractCOLS(x, i)

## S4 method for signature 'HDF5DataFrame'
replaceROWS(x, i, value)

## S4 method for signature 'HDF5DataFrame'
replaceCOLS(x, i, value)
```

Arguments

x	HDF5DataFrame object
i	row/column index or name
value	vector to be replaced

Value

HDF5DataFrame object

HDF5DataFrame object

Index