Package: spark.sas7bdat (via r-universe)

August 31, 2024

Type Package

Title Read in 'SAS' Data ('.sas7bdat' Files) into 'Apache Spark'

Description Read in 'SAS' Data ('.sas7bdat' Files) into 'Apache Spark' from R. 'Apache Spark' is an open source cluster computing framework available at <http://spark.apache.org>. This R package uses the 'spark-sas7bdat' 'Spark' package (<https://spark-packages.org/package/saurfang/spark-sas7bdat>) to import and process 'SAS' data in parallel using 'Spark'. Hereby allowing to execute 'dplyr' statements in parallel on top of 'SAS' data.

Maintainer Jan Wijffels <jwijffels@bnosac.be>

License GPL-3

Version 1.4

URL https://github.com/bnosac/spark.sas7bdat

VignetteBuilder knitr

Imports sparklyr (>= 0.3)

Suggests knitr, rmarkdown

RoxygenNote 7.1.1

Repository https://bnosac.r-universe.dev

RemoteUrl https://github.com/bnosac/spark.sas7bdat

RemoteRef HEAD

RemoteSha 5c38ff6959fe158a50168cb8717a85233a6b5602

Contents

spark.sas7bdat-package *Read in SAS datasets (.sas7bdat files) into Spark*

Description

'spark.sas7bdat' uses the spark-sas7bdat Spark package to process SAS datasets in parallel using Spark. Hereby allowing to execute dplyr statements on top of SAS datasets.

spark_read_sas *Read in SAS datasets in .sas7bdat format into Spark by using the spark-sas7bdat Spark package.*

Description

Read in SAS datasets in .sas7bdat format into Spark by using the spark-sas7bdat Spark package.

Usage

```
spark_read_sas(sc, path, table)
```

Arguments

sc	Connection to Spark local instance or remote cluster. See the example
path	full path to the SAS file either on HDFS (hdfs://), S3 (s3n://), as well as the local file system (file://). Mark that files on the local file system need to be specified using the full path.
table	character string with the name of the Spark table where the SAS dataset will be put into

Value

an object of class tbl_spark, which is a reference to a Spark DataFrame based on which dplyr functions can be executed. See https://github.com/sparklyr/sparklyr

References

https://spark-packages.org/package/saurfang/spark-sas7bdat, https://github.com/saurfang/spark-sas7bdat, https://github.com/sparklyr/sparklyr

See Also

spark_connect, sdf_register

Examples

```
## Not run:
## If you haven't got a Spark cluster, you can install Spark locally like this
library(sparklyr)
spark_install(version = "2.0.1")

## Define the SAS .sas7bdat file, connect to the Spark cluster to read + process the data
myfile <- system.file("extdata", "iris.sas7bdat", package = "spark.sas7bdat")
myfile

library(spark.sas7bdat)
sc <- spark_connect(master = "local")
x <- spark_read_sas(sc, path = myfile, table = "sas_example")
x

library(dplyr)
x %>% group_by(Species) %>%
  summarise(count = n(), length = mean(Sepal_Length), width = mean(Sepal_Width))

## End(Not run)
```

Index