News * 7/17/2012 rhdfs (1.0.4) change to handle different classpaths in the init function * 7/13/2012 rmr version 1.3 with vectorized API — see [[Changelog]]. * 5/18/2012 rhdfs (1.0.3) bug fix in function hdfs.file * 4/14/2012 rhbase (1.0.4) and rhdfs (1.0.2) minor bug fixes and some cleanup for R CMD check * 3/30/2012 rmr 1.2.2 fixes from.dfs for some obscure platforms and prepares for apache 1.0.2 compatibility (more flexible w.r.t. hadoop layout) * 3/13/2012 New version of rhbase (1.0.3) that supports both "native" and "raw" serialization * 2/27/2012 rmr version 1.2 with binary formats and other goodies available — see [[Changelog]]. * 2/11/2012 New version of rhbase (1.0.2) that installs with thrift 0.8 or greater. * 2/1/2012 Binary format now default in dev, passes all normal checks. Please test. * 24/1/2012 - Merged branch binary-io into dev.Please note some non-backward compatible changes in the API intended to strike a compromise between flexibility and ease of use in the IO department. * 12/7/2011 - Version 1.1 of the package rmr is available. See the [[Changelog]] for details. * 9/29/2011 - Version 1.0.1 available - fixes some minor defects with R CMD check tests on the packages

About

RHadoop is a collection of three R packages that allow users to manage and analyze data with Hadoop. The packages have been implemented and tested in Cloudera's distribution of Hadoop (CDH3). and R 2.15.0. THe packages have also been tested with Revolution R 4.3 and 5.0

RHadoop consists of the following packages:

- [[rmr]] - functions providing Hadoop MapReduce functionality in R
- [[rhdfs]] - functions providing file management of the HDFS from within R
- [[rhbase]] - functions providing database management for the HBase distributed database from within R

More information about RHadoop

- Overview of RHadoop, from the Revolution Analytics blog.
- Slides and Replay of 30-minute presentation about RHadoop, "Leveraging R in Hadoop Environments".
- [[Downloads]]
- [[Tutorial]]
- [[Contribute to the RHadoop project]]
- Live from the net

Contact: rhadoop@revolutionanalytics.com

Questions: Please participate in our discussion group. For private questions, please use the above email address.