Overview

This R package provides basic connectivity to HBASE, using the [Thrift server](#). R programmers can browse, read, write, and modify tables stored in HBASE. The following functions are part of this package

- Table Maninpulation
  hb.new.table, hb.delete.table, hb.describe.table, hb.set.table.mode, hb.regions.table
- Read/Write
  hb.insert, hb.get, hb.delete, hb.insert.data.frame, hb.get.data.frame, hb.scan
- Utility
  hb.list.tables
- Initialization
  hb.defaults, hb.init

Prerequisites

- Installing the package requires that you first [install and build Thrift](#). Once you have the libraries built, be sure they are in a path where the R client can find them (i.e. /usr/lib). **This package was built and tested using Thrift 0.8**

  Here is an example for building the libraries on CentOS:

  1. Install all Thrift pre-requisites: http://wiki.apache.org/thrift/GettingCentOS5Packages
  2. Build Thrfit according to instructions: http://wiki.apache.org/thrift/ThriftInstallation
  3. Update PKG_CONFIG_PATH: `export PKG_CONFIG_PATH=$PKG_CONFIG_PATH:/usr/local/lib/pkgconfig/`
  4. Verifiy pkg-config path is correct: `pkg-config --cflags thrift`, returns: `-I/usr/local/include/thrift`
  5. Copy Thrift library `sudo cp /usr/local/lib/libthrift.so.0 /usr/lib/`

- The Thrift server by default starts on port 9090.

  ```
  [hbase-root]/bin/hbase thrift start
  ```

  If you are running on rhbase on a different hostname:port you will have to change how the package is initialized

  ```
  hb.init(host=127.0.0.1, port=9090)
  ```

- By default the rhbase uses "native" R serialization (serialize/unserialize) to read and write data from hbase. You can switch this to "raw" (i.e treat everything as a string) serialization by specifying "serialization="raw"" during the initialization of the package

  ```
  hb.init(serialize="raw")
  ```

  See the sample " /rhbase/pkg/inst/samples/StringSerializer.r" for details