Overview

This R package provides basic connectivity to the Hadoop Distributed File System. R programmers can browse, read, write, and modify files stored in HDFS. The following functions are part of this package

- File Manipulations
  hdfs.copy, hdfs.move, hdfs.rename, hdfs.delete, hdfs.rm, hdfs.del, hdfs.chown, hdfs.put, hdfs.get
- File Read/Write
  hdfs.file, hdfs.write, hdfs.close, hdfs.flush, hdfs.read, hdfs.seek, hdfs.tell, hdfs.line.reader, hdfs.read.text.file
- Directory
  hdfs.dircreate, hdfs.mkdir
- Utility
  hdfs.ls, hdfs.list.files, hdfs.file.info, hdfs.exists
- Initialization
  hdfs.init, hdfs.defaults

Prerequisites

- This package has a dependency on rJava
- Access to HDFS via this R package is dependent upon the `HADOOP_HOME` and `HADOOP_CONF` environment variables. Be sure that these are properly set. If these variables are not properly set, the package will be accessing the local file system instead of HDFS

Examples:

```
HADOOP_HOME=/usr/lib/hadoop
HADOOP_CONF=/etc/hadoop/conf
```

R Objects

R objects can be serialized to HDFS via the function: `hdfs.write`. An example is shown below:

```
model <- lm(...)
modelfilename <- "my_smart_unique_name"
modelfile <- hdfs.file(modelfilename, "w")
hdfs.write(model, modelfile)
hdfs.close(modelfile)
```

R objects can be deserialized to HDFS via the function: `hdfs.read`. An example is shown below:

```
modelfile = hdfs.file(modelfilename, "r")
m <- hdfs.read(modelfile)
model <- unserialize(m)
hdfs.close(modelfile)
```