Package 'RHive'

October 26, 2011

Type Package

Title R and Hive

Version 0.1-0

Description RHive is an R extension facilitating distributed computing via HIVE query.
It provides an easy to use HQL like SQL and R objects and functions in HQL.

Author NexR

Maintainer NexR <rhive@nexr.com>

License Apache License (== 2.0)

Depends R (>= 2.13.0), rJava (>= 0.9-0), Rserve (>= 0.6-0)

Enhances rJava

SystemRequirements
Hadoop core >= 0.20.3 (http://hadoop.apache.org/core/),Hive >= 0.8.0 (http://hive.apache.org/)

Repository CRAN

R topics documented:

`rhive` *rhive: R and Hive*

Description

RHive is an R extension facilitating distributed computing via HIVE query. It allows easy usage of HQL in R, and allows easy usage of R objects and R functions in Hive.

Details

The rhive package supplies functions to interact with Hive from within R. There are functions for exporting and connecting as well as querying hive.

See Also

hiveConnect hiveQuery hiveExport hiveExportAll hiveClose

`rhive-connect` *Manage connection to Hive using functions in Package 'RHive'*

Usage

```
hiveConnect(host,port=10000)
hiveClose(hiveclient)
```

Arguments

`host`	hive-server address to connect hive.
`port`	hive-server listen port.
`hiveclient`	a client for hive.

Examples

```
## try to connect hive server
hivecon <- hiveConnect("localhost")

## close connection
hiveClose(hivecon)
```

rhive-export *Export R function to Hive using functions in Package 'RHive'*

Usage

```
hiveExport(exportname,hosts="hivethost",port=6311)
hiveExportAll(exportname,hosts="localhost",port=6311)
```

Arguments

hiveclient	a client for hive.
hosts	list of Rserve's address
query	hive query.
exportname	function name to be exported.

Examples

```
## try to connect hive server
hivecon <- hiveConnect("localhost")

## execute HQL(hive query)
hiveQuery(hivecon, "show tables")


## define R function
coff <- 5.2
scoring <- function(sal) {
coff * sal
}

## export R objects (scoring and coff) to Hive
hiveExportAll('scoring')

## execute HQL using exported R objects
## name of UDF is 'R'
hiveQuery(hivecon,"select R('scoring',sal,0.0) from emp")

## define R aggregation function
## define iterate operator
hsum <- function(prev,sal) {
c(prev[1] + sal[1])
}
## define partial aggregation operator
hsum.partial <- function(agg_sal) {
agg_sal
}
## define merge operator
hsum.merge <- function(prev, agg_sal) {
c(prev[1] + agg_sal[1])
}
## define final aggregation operator
hsum.terminate <- function(agg_sal) {
agg_sal
```

```
}

hiveExportAll('hsum')

## name of UDAF is 'RA'
hiveQuery(hivecon,"select RA('hsum',sal) from emp group by empno")

## close connection
hiveClose(hivecon)
```

`rhive-query` *Execute HQL(Hive Query) in R using functions in Package 'RHive'*

Usage

```
hiveQuery(hiveclient,query,fetchsize = 40, limit = -1)
```

Arguments

`hiveclient`	a client for hive.
`query`	hive query.
`fetchsize`	fetch size for result.
`limit`	total result size.

Examples

```
## try to connect hive server
hivecon <- hiveConnect("localhost")

## execute hive query
hiveQuery(hivecon, "show tables")

## close connection
hiveClose(hivecon)
```

Index