

Package ‘RBigQuery’

March 3, 2011

Version 0.5.2

Date 2011-1-21

Title BigQuery for R

Author David Xiao <dxiao@mit.edu>

Maintainer David Xiao <dxiao@mit.edu>

Depends R (>= 2.11.0), methods, DBI (> 0.2-2), rjson, RCurl, plyr

Description R interface for Google BigQuery

Collate dbObjectId.R RBigQuerySupport.R RBigQuery.R

LazyLoad yes

License GPL (>= 2)

R topics documented:

RBigQuery-package	2
BigQuery	3
BQConnection-class	4
bqDescribeTable	5
BQDriver-class	6
BQObject-class	7
BQResult-class	7
dbConnect-methods	8
dbDriver-methods	9
dbGetInfo-methods	10
dbGetQuery-methods	11
dbListConnections-methods	12
dbListFields-methods	12
dbReadTable-methods	13
dbSendQuery-methods	13
fetch-methods	14
summary-methods	15

RBigQuery-package *BigQuery for R*

Description

A DBI extension for querying and parsing results from Google's BigQuery database.

Overview

1. Connect and authenticate with Google:

```
conn <- dbConnect(BigQuery(), username="bquser@gmail.com",  
                  password="pass")
```

2. Get table fields:

```
info <- dbListFields(conn, "bigquery/samples/shakespeare")
```

3. Send a query and if successful get results as data.frame:

```
result <- dbSendQuery(conn, "SELECT word FROM  
[bigquery/samples/shakespeare] WHERE word CONTAINS 'raisin'")  
if (result@success)  
  fetch(result)  
else  
  dbGetInfo(result, "error")
```

4. Get information about connection, results, etc.:

```
summary(BigQuery(), verbose=TRUE)  
summary(conn, verbose=TRUE)  
summary(result, verbose=TRUE)  
dbGetException(conn)
```

Note

This package does not yet support data import or table deletion.

Author(s)

David Xiao <dxiao@csail.mit.edu>

See Also

[dbDriver](#) [dbConnect](#) [dbDisconnect](#) [dbSendQuery](#) [dbGetQuery](#) [fetch](#) [dbClearResult](#)
[summary](#) [dbGetInfo](#) [dbColumnInfo](#) [dbGetException](#) [dbGetStatement](#)

`BigQuery`*Creates a BigQuery driver.*

Description

Creates a BigQuery driver, or if the driver is already created and the `force.reload` flag is not set, returns the current BigQuery driver. (drivers are singletons)

Usage

```
BigQuery(max.con = 1, fetch.default.rec = 10000, force.reload = FALSE)
```

Arguments

`max.con` Maximum number of connections BigQuery can drive at once.

`fetch.default.rec`
 Default number of result rows to fetch at a time

`force.reload` Force the driver to restart itself.

Details

the current arguments to `BigQuery()` are not yet set to do anything. Don't expect them to do *_anything_* to the way RBigQuery operates.

Value

Returns a driver object, which can then be used to create a connection.

See Also

[DBIDriver-class](#)

Examples

```
## Not run:
  driver <- BigQuery()
  ## Alternatively, use
  driver <- dbDriver("BigQuery")

## End(Not run)
```

BQConnection-class *Class "BQConnection"*

Description

Holds auth information to connect to BigQuery.

Objects from the Class

Objects can be created by calls of the form `dbConnect(BigQuery(), "username@gmail.com", "password")`

Extends

Class "[DBIConnection](#)", directly. Class "[BQObject](#)", directly. Class "[DBIOObject](#)", by class "DBIConnection", distance 2. Class "[dbObjectId](#)", by class "BQObject", distance 2.

Methods

```
dbConnect signature(drv = "BQConnection", username="character", password="character",
  ...
dbDisconnect signature(conn = "BQConnection"): ...
dbGetException signature(conn = "BQConnection"): ...
dbGetInfo signature(dbObj = "BQConnection"): ...
dbSendQuery signature(conn = "BQConnection", statement = "character",
  verbose="logical", status="logical"): ...
dbGetQuery signature(conn = "BQConnection", statement = "character"):
  ...
dbListFields signature(conn = "BQConnection", name = "character"): ...
dbListResults signature(conn = "BQConnection"): ...
dbReadTable signature(conn = "BQConnection", name = "character"): ...
summary signature(object = "BQConnection"): ...
```

Examples

```
## Not run:
conn <- dbConnect(drv, username="username", password="password")
result <- dbGetQueyr(conn, "SELECT word FROM [bigquery/samples/shakespeare]
  WHERE word CONTAINS 'raisin'")

## End(Not run)
```

bqDescribeTable	<i>Get table metadata from remote server.</i>
-----------------	---

Description

Equivalent to the describe command in bq, asks for metadata regarding a BigQuery table - primarily column name and type

Usage

```
bqDescribeTable(con, table, what = "", verbose = FALSE, ...)
```

Arguments

con	Connection to BigQuery to use.
table	The name of the table to describe
what	Optional: A specific tag whose value to return
verbose	Print extra debugging output
...	Currently unimplemented

Value

kind	Type of query - should be bigquery#table
name	name of the table
fields	List of fields in the table and their types. Note: this can produce very unhelpful output - for column names, using bqListFields is recommended.

See Also

[BQConnection-class](#)

Examples

```
## Not run:  
  metadata <- bqDescribeTable(conn, "bigquery/samples/shakespeare")  
  
## End(Not run)
```

BQDriver-class	<i>Class "BQDriver"</i>
----------------	-------------------------

Description

Singleton holding some configuration options and open connections. For RBigQuery, not extremely important other than to conform to DBI.

Objects from the Class

Objects can be created by calls of the form `BigQuery()` or `dbDriver("BigQuery")`

Extends

Class "[DBIDriver](#)", directly. Class "[BQObject](#)", directly. Class "[DBIOObject](#)", by class "[DBIDriver](#)", distance 2. Class "[dbObjectId](#)", by class "[BQObject](#)", distance 2.

Methods

```
dbConnect  signature(drv = "BQDriver", username = "character", password="character")
            ...
dbGetInfo  signature(dbObj = "BQDriver", what="character"):...
dbListConnections  signature(drv = "BQDriver"): ...
dbListFields  signature(drv = "BQDriver", table="character"): ...
summary  signature(object = "BQDriver", verbose="logical"): ...
```

See Also

[BQObject-class](#) [DBIDriver-class](#) [BQConnection-class](#) [dbConnect-methods](#)
[dbGetInfo-methods](#) [dbListConnections](#) [summary](#)

Examples

```
## Not run:
drv <- dbDriver("BigQuery")
conn <- dbConnect(drv, "username", "password")

## End(Not run)
```

BQObject-class	Class "BQObject"
----------------	------------------

Description

Parent class to other BQ objects Driver, Connection, and Result.

Objects from the Class

A virtual Class: No objects may be created from it.

Slots

Id: Unique object id for be BQObject

Extends

Class "[DBIObject](#)", directly. Class "[dbObjectId](#)", directly.

Methods

No methods defined with class "BQObject" in the signature.

See Also

[DBIObject-class](#) [BQDriver-class](#) [BQConnection-class](#) [BQResult-class](#)

BQResult-class	Class "BQResult"
----------------	------------------

Description

Incapsulates information regarding a specific query, including query statement, result, and result metadata.

Objects from the Class

Results can be created by querying the BigQuery database, either with `dbSendQuery(conn, ...)` or `dbGetQuery(conn, ...)`

Extends

Class "[DBIResult](#)", directly. Class "[BQObject](#)", directly. Class "[DBIObject](#)", by class "[DBIResult](#)", distance 2. Class "[dbObjectId](#)", by class "[BQObject](#)", distance 2.

Methods

```

dbClearResult signature(res = "BQResult"): ...
dbColumnInfo signature(res = "BQResult"): ...
dbGetInfo signature(dbObj = "BQResult", what="character"): ...
dbGetStatement signature(res = "BQResult"): ...
fetch signature(res = "BQResult"): ...
summary signature(object = "BQResult", what="character"): ...

```

Examples

```

## Not run:
result <- dbSendQuery(conn, "SELECT word FROM [bigquery/samples/shakespeare]
                             WHERE word CONTAINS 'raisin'")

summary(result)
data <- fetch(result)

## or to get data directly, use
data <- dbGetQuery(conn, "SELECT word FROM [bigquery/samples/shakespeare]
                          WHERE word CONTAINS 'raisin'")

## End (Not run)

```

dbConnect-methods *Connect to the BigQuery server*

Description

Authenticates with the BigQuery server and creates a BQConnection

Arguments

username	Your google username: e.g. john.doe@gmail.com or just john.doe
password	Your google password.

Value

A BQConnection object which can then be used to query BigQuery.

Methods

```

signature(drv = "BQDriver", username="character", password="character")
  Connects to and authenticates with the Google servers, then creates a BQConnection object
  which can then be used to query the BigQuery servers.

signature(drv = "BQDriver", username="character", auth.token="character")
  Takes the authentication token provided as an argument and creates a BQConnection object
  with can then be used to query to BigQuery servers. N.B.: RBigQuery does not actually verify
  to see if the token is correct - you'll have to query BigQuery yourself to find out.

```


`signature(drv = "BQConnection")` Should clone the provided BQConnection object, not implemented for now.

See Also

[BQConnection-class](#)

Examples

```
## Not run:
  driver <- BigQuery()
  conn <- dbConnect(driver, username="username", password="password")
  data <- dbGetQuery(conn, "insert query here")

## End (Not run)
```

dbDriver-methods *constructs the DBI driver*

Description

initializes the appropriate driver from the driver string given.

Arguments

`drvName` should be the character name of the driver to initiate. in this package, "Big-Query"

Value

a DBIDriver object for the specified database. In BigQuery's case, a BQDriver. Note that BQDriver is a singleton; only one object of its class should be instantiated at a time.

Methods

`signature(drvName = "character"), ...` calls the initialization function appropriate to driver name specified. Syntactically equal to `drvName(...)`

See Also

[BQDriver-class](#) [BigQuery](#)

dbGetInfo-methods *Information about a BQObject*

Description

Get information about a BQObject, either general list or a specific element of that list.

Arguments

dbObj	BQObject to get information about
what	Optional: character name of specific element of list to return.

Methods

signature(dbObj = "BQDriver") Returns:

- drvName** Name of Package
- managerID** ID number of object
- fetch.default.rec** default number of rows per fetch (unused)
- max.con** maximum number of connections allows (unused)
- prc.con** number of processed connections (unused)
- connections** list of active connections under driver
- clientVersion** version of package

signature(dbObj = "BQConnection") Returns:

- driver** Driver the connection is using
- username** username of google account
- password** length of google account password (as opposed to password itself)
- auth.token** authorization token used when querying BigQuery
- last.result** last BQResult returned by connection

signature(dbObj = "BQResult") Returns:

- connection** Connection the result came from
- statement** statement used in query
- success** logical indicating success of query
- fields** result table metadata
- result** if success is TRUE, the data returned by query
- error** if success is FALSE, the error returned by query

See Also

[summary BQDriver-class](#) [BQConnection-class](#) [BQResult-class](#)

Examples

```
## Not run:
  dbGetInfo(driver) #returns list

  dbGetInfo(driver, what="drvName") #returns name of package

## End (Not run)
```

dbGetQuery-methods *query and return*

Description

Executes a query, parses the result. Returns data.frame if successful, FALSE if not.

Arguments

conn	BQConnection object to send query through
statement	character string of statement to execute
...	other arguments are passed to dbSendQuery call

Methods

```
signature(conn = "BQConnection", statement = "character"), ...
```

See Also

[dbSendQuery-methods](#) [fetch BQConnection-class](#)

Examples

```
## Not run:
  data <- dbGetQuery(conn, "SELECT word FROM [bigquery/samples/shakespeare]
                           WHERE word CONTAINS 'raisin'")

## End (Not run)
```

```
dbListConnections-methods
```

List open connections

Description

Lists all open and query-able connections that are associated with a provided driver.

Synonomous to `dbGetInfo(driver, what="connections")`

Methods

`signature(drv = "BQDriver")` Returns a list of BQConnections

See Also

[BQDriver-class](#) [BQConnection-class](#) [dbGetInfo-methods](#)

```
dbListFields-methods
```

Lists table column names

Description

Returns a list of the column names for a given table. Uses `bqDescribeTable` to get metadata

Methods

`signature(conn = "BQConnection", name = "character")` name is the character name of the table, with without square brackets.

See Also

[bqDescribeTable](#) [BQConnection-class](#)

Examples

```
## Not run:
  dbListFields(conn, "bigquery/samples/shakespeare")

## End(Not run)
```

 dbReadTable-methods

Query all columns of a table

Description

Queries over all columns of a table.

Arguments

name	name of the table to query - do not enclose in square brackets
addend	(Optional) argument for adding additional clauses to the query string.

Methods

signature(conn = "BQConnection", name = "character"), ... returns a data.frame of the results of the query "SELECT * FROM [name] <addend>".

See Also

[dbListFields-methods](#) [dbGetQuery-methods](#) [dbSendQuery-methods](#) [BQConnection-class](#)

Examples

```
## Not run:
dbReadTable(conn, "bigquery/samples/shakespeare")

dbReadTable(conn, "bigquery/samples/shakespeare", "WHERE word CONTAINS 'th'")

## End (Not run)
```

 dbSendQuery-methods

Query the database

Description

Sends a query to the BigQuery database and returns a BQResult object describing it.

Arguments

statement	character string of query. See BigQuery query reference for syntax
status	(Optional) flag to print status messages about the state of execution - recommended
verbose	(Optional) flag to print all debugging information during the query - do not set unless you know what you're doing

Methods

signature(conn = "BQConnection", statement = "character"), ... Returns a BQResult. Use fetch() to retrieve full results.

See Also

[BQConnection-class](#) [BQResult-class](#) [dbGetQuery-methods](#) [fetch-methods](#) [dbReadTable-methods](#) [dbGetInfo-methods](#)

Examples

```
## Not run:
conn <- dbConnect(driver, "username", "password")
result <- dbSendQuery(conn, "SELECT word FROM [bigquery/samples/shakespeare]
                             WHERE word CONTAINS 'raisin'", status=TRUE)
data <- fetch(result)

## End(Not run)
```

fetch-methods

Fetch data out of an BQResult object

Description

Retrive data from a BQResult in the form of a data.frame

Methods

signature(res = "BQResult", n = "missing") Retrieves as many rows of the data as possible.

signature(res = "BQResult", n = "numeric") Retrieves the first n rows of data.

See Also

[BQResult-class](#) [dbSendQuery-methods](#)

Examples

```
## Not run:
result <- dbSendQuery(conn, "SELECT word FROM [bigquery/samples/shakespeare]
                             WHERE word CONTAINS 'raisin'")
data <- fetch(result)

## End(Not run)
```

summary-methods	<i>pretty-print object information</i>
-----------------	--

Description

prints a human-readable string describing the object and its data

Arguments

`verbose` (Optional) flag which will print additional information about the object

Methods

`signature(object = "BQDriver")` Will give driver name, records per fetch, max allowed connections, number of processed connections, number of open connections. If `verbose` is set, the connection ids, and DBI and BQ versions.

`signature(object = "BQConnection")` Will give driver, username, and object id of last result returned from connection. If `verbose` is set, will give the length of password, the stored authorization token, and the BQ client version.

`signature(object = "BQResult")` Will give connection, statement queried, a flag indicating success or failure, and if success, the field names returned, if failure, the error message. If `verbose` is set, it will return the data as well.

See Also

[BQDriver-class](#) [BQConnection-class](#) [BQResult-class](#)

Examples

```
## Not run:
  summary(driver)
  summary(conn, TRUE)
  summary(result, verbose=TRUE)

## End(Not run)
```

Index

*Topic **BigQuery**

BigQuery, [3](#)

*Topic **classes**

BQConnection-class, [4](#)

BQDriver-class, [6](#)

BQObject-class, [7](#)

BQResult-class, [7](#)

*Topic **methods**

dbConnect-methods, [8](#)

dbDriver-methods, [9](#)

dbGetInfo-methods, [10](#)

dbGetQuery-methods, [11](#)

dbListConnections-methods, [12](#)

dbListFields-methods, [12](#)

dbReadTable-methods, [13](#)

dbSendQuery-methods, [13](#)

fetch-methods, [14](#)

summary-methods, [15](#)

*Topic **package**

RBigQuery-package, [1](#)

*Topic **table**

bqDescribeTable, [5](#)

BigQuery, [3](#), [9](#)

BQConnection-class, [5–7](#), [9–15](#)

BQConnection-class, [4](#)

bqDescribeTable, [5](#), [12](#)

BQDriver-class, [7](#), [9](#), [10](#), [12](#), [15](#)

BQDriver-class, [6](#)

BQObject, [4](#), [6](#), [7](#)

BQObject-class, [6](#)

BQObject-class, [7](#)

BQResult-class, [7](#), [10](#), [14](#), [15](#)

BQResult-class, [7](#)

dbClearResult, [2](#)

dbClearResult, BQResult-method
(BQResult-class), [7](#)

dbColumnInfo, [2](#)

dbColumnInfo, BQResult-method
(BQResult-class), [7](#)

dbConnect, [2](#)

dbConnect, BQConnection-method
(dbConnect-methods), [8](#)

dbConnect, BQDriver-method
(dbConnect-methods), [8](#)

dbConnect-methods, [6](#)

dbConnect-methods, [8](#)

dbDisconnect, [2](#)

dbDisconnect, BQConnection-method
(BQDriver-class), [6](#)

dbDriver, [2](#)

dbDriver, character-method
(dbDriver-methods), [9](#)

dbDriver-methods, [9](#)

dbGetException, [2](#)

dbGetException, BQConnection-method
(BQDriver-class), [6](#)

dbGetInfo, [2](#)

dbGetInfo, BQConnection-method
(dbGetInfo-methods), [10](#)

dbGetInfo, BQDriver-method
(dbGetInfo-methods), [10](#)

dbGetInfo, BQResult-method
(dbGetInfo-methods), [10](#)

dbGetInfo-methods, [6](#), [12](#), [14](#)

dbGetInfo-methods, [10](#)

dbGetQuery, [2](#)

dbGetQuery, BQConnection, character-method
(dbGetQuery-methods), [11](#)

dbGetQuery-methods, [13](#), [14](#)

dbGetQuery-methods, [11](#)

dbGetStatement, [2](#)

dbGetStatement, BQResult-method
(BQResult-class), [7](#)

DBIConnection, [4](#)

DBIDriver, [6](#)

DBIDriver-class, [3](#), [6](#)

DBIObject, 4, 6, 7
DBIObject-class, 7
DBIResult, 7
dbListConnections, 6
dbListConnections, BQDriver-method
 (*dbListConnections-methods*),
 12
dbListConnections-methods, 12
dbListFields, BQConnection, character-method
 (*dbListFields-methods*), 12
dbListFields, BQConnection-method
 (*BQDriver-class*), 6
dbListFields-methods, 13
dbListFields-methods, 12
dbListResults, BQConnection-method
 (*BQDriver-class*), 6
dbObjectId, 4, 6, 7
dbReadTable, BQConnection, character-method
 (*dbReadTable-methods*), 13
dbReadTable-methods, 14
dbReadTable-methods, 13
dbSendQuery, 2
dbSendQuery, BQConnection, character-method
 (*dbSendQuery-methods*), 13
dbSendQuery-methods, 11, 13, 14
dbSendQuery-methods, 13

fetch, 2, 11
fetch, BQResult, missing-method
 (*fetch-methods*), 14
fetch, BQResult, numeric-method
 (*fetch-methods*), 14
fetch-methods, 14
fetch-methods, 14

RBigQuery (*RBigQuery-package*), 1
RBigQuery-package, 1

summary, 2, 6, 10
summary, BQConnection-method
 (*summary-methods*), 15
summary, BQDriver-method
 (*summary-methods*), 15
summary, BQResult-method
 (*summary-methods*), 15
summary-methods, 15