Oracle queries in R

Arni Magnusson

Hafro, 9 Nov 2010

Outline

1. Background - history, interfaces, problems
2. ROracle - connection
3. ROracleUI - sql, tables, desc, to_char

Outline

SQL history

1974	SEQUEL	Chamberlin and Boyce at IBM, based on Codd (1970) paper
1987	SQL-86	ANSI and ISO (based on IBM)
1989	SQL-89	CREATE
1992	SQL-92	LEFT/RIGHT JOIN, CASE, AS, IS NULL, string/date funcs
1999	SQL:1999	LOB (large object), user types/funcs, schemas, regexp
2003	SQL:2003	XML features
2006	SQL:2006	XQuery (W3C) support
2008	SQL:2008	TRUNCATE, INSTEAD OF

Oracle history

1979	2	version 1 never existed, queries and JOIN
1983	3	implemented in C, COMMIT and ROLLBACK
1984	4	read-consistency
1985	5	distributed computing
1989	6	PL/SQL
1992	7	referential integrity
1997	8	object orientation
1999	8i	Java virtual machine, available for Linux (i for internet)
2001	9i	XML support, RAC distributed computing
2003	10g	grid distributed computing (g for grid)
2007	11g	PIVOT and UNPIVOT

ROracle and ROracleUI history

ROracle

2001	0.3-1	James and Luciani at Bell Labs
2002	0.3-3	
	0.4-0	DBI compliant, almost all functions renamed
2003	0.5-0	
	0.5-3	ported to Windows 2000 (now obsolete)
2004	0.5-4	
	0.5-5	
2006	0.5-7	
	0.5-8	
2007	0.5-9	
2010		ported to Windows (locally by Gunnar Orvarsson at Hafro)

ROracleUI

2010	1.0-0	Arni Magnusson at Hafro
	1.1-3	Windows support
	1.2-0	new arg ‘stringsAsFactors’ in sql()

Interfaces

sqlplus

interactive session, also within Emacs

sql++

Hafro Perl script, was often used from S-Plus and R

xSql.pl

Hafro Perl script, less used but based on newer Perl packages

sqldeveloper

GUI application

R

import directly as data frame, without intermediate text file

Three problems

1. Unreliable – queries can return fewer lines than intended, because of certain data types
2. Dates are difficult to convert from Icelandic format to something that statistical software can analyze
3. Overview of tables and columns

Outline

ROracle

Query

```
query <- "SELECT sysdate FROM dual"  # character
```

Import

```
drv <- dbDriver("Oracle")            # OraDriver
con <- dbConnect(drv)                # OraConnection
res <- dbSendQuery(con, query)       # OraResult
out <- fetch(res, n=-1)              # data.frame
```

Clean up

```
dbClearResult(res)
suppressWarnings(dbUnloadDriver(dbDriver("Oracle")))
```

Outline

sql

Query

```
query <- "SELECT sysdate FROM dual"  # or "file.sql"
```

Import

```
out <- sql(query)
```

Default options

`tolower`	`COLNAMES` → `colnames`	`TRUE`
`dots`	`col_names` → `col.names`	`TRUE`
`posix`	try converting dates to `POSIXct`	`TRUE`
`stringsAsFactors`	convert string columns to factors	`FALSE`

tables, views, desc

Describe table (or view)

```
desc("fiskar.stodvar")
```

List tables

```
tables(owner="fiskar")
```

```
tables(table="%tegund%")
```

List views

```
views(owner="fiskar")
```

to_char

Import date as Icelandic strings

```
x <- sql("SELECT username,created FROM all_users")
```

Import date as POSIX using to_char in R

```
y <- sql(paste("SELECT username,",
               to_char(created),
               "FROM all_users"))
```

Import date as POSIX using to_char in Oracle

```
z <- sql("SELECT username,
          to_char(created,'YYYY-MM-DD HH24:MI:SS')
          AS created FROM all_users")
```