

Exemplo de banco de dados

Elias T Krainski

Curso de Especialização em
Data Science & Big Data
Universidade Federal do Paraná

28 Abril 2018



Bancos de dados

SGDBs, https://pt.wikipedia.org/wiki/Sistema_de_gerenciamiento_de_banco_de_dados

Sistema de Gerenciamento de Banco de Dados - SGBD, do inglês *Data Base Management System - DBMS*

- ▶ *softwares* de gerenciamento de um banco de dados
- ▶ Relacionais: interface via *Application Programming Interface - API* ou drivers
 - ▶ executam comandos na linguagem *Structured Query Language - SQL*

Exemplos de SGBDs, https://pt.wikipedia.org/wiki/Sistema_de_gerenciamento_de_banco_de_dados

- ▶ PostgreSQL
- ▶ Firebird
- ▶ HSQLDB
- ▶ IBM DB2
- ▶ IBM Informix
- ▶ mSQL
- ▶ MySQL
- ▶ MariaDB
- ▶ Oracle
- ▶ SQL-Server
- ▶ TinySQL
- ▶ ZODB
- ▶ JADE
- ▶ Sybase
- ▶ Microsoft Access (SGBD para alguns, mas é um SGBDR)
- ▶ Microsoft Visual Foxpro
- ▶ MongoDB

Exemplo

- ▶ Criação do banco de dados descrita no arquivo README.txt no diretório **menagerie**

Usando os dados em R

- Connecta o banco de dados

```
library(DBI)  
con <- dbConnect(RMariaDB::MariaDB(), user='dsbd01',  
                  password='1234567', dbname='menagerie')
```

- Lista tabelas do banco de dados

```
dbListTables(con)  
## [1] "event" "pet"
```

Carrega uma tabela

```
d1q <- dbSendQuery(con, 'select * from pet')
```

```
d1 <- dbFetch(d1q)
```

```
d1
```

##	pet_id	name	owner	species	sex	birth	death
## 1	1	Fluffy	Harold	cat	f	1993-02-04	<NA>
## 2	2	Claws	Gwen	cat	m	1994-03-17	<NA>
## 3	3	Buffy	Harold	dog	f	1989-05-13	<NA>
## 4	4	Fang	Benny	dog	m	1990-08-27	<NA>
## 5	5	Bowser	Diane	dog	m	1979-08-31	1995-07-29
## 6	6	Chirpy	Gwen	bird	f	1998-09-11	<NA>
## 7	7	Whistler	Gwen	bird	<NA>	1997-12-09	<NA>
## 8	8	Slim	Benny	snake	m	1996-04-29	<NA>
## 9	9	Puffball	Diane	hamster	f	1999-03-30	<NA>

```
dbClearResult(d1q)
```

Seleção

```
d1fq <- dbSendQuery(con, 'select * from pet where sex="f"')
```

```
d1f <- dbFetch(d1fq)
```

```
d1f
```

##	pet_id	name	owner	species	sex	birth	death
## 1	1	Fluffy	Harold	cat	f	1993-02-04	<NA>
## 2	3	Buffy	Harold	dog	f	1989-05-13	<NA>
## 3	6	Chirpy	Gwen	bird	f	1998-09-11	<NA>
## 4	9	Puffball	Diane	hamster	f	1999-03-30	<NA>

```
dbClearResult(d1fq)
```


Seleciona algumas colunas e algumas linhas

Variáveis *name*, *species*, *sex* da tabela **pet** quando a variável *birth* é maior que 1995/01/01

```
d2q <- dbSendQuery(  
con, "select name,species,sex from pet where birth>'1995-01-01'"  
d2 <- dbFetch(d2q)  
d2  
  
##      name species  sex  
## 1  Chirpy    bird   f  
## 2 Whistler   bird <NA>  
## 3    Slim   snake   m  
## 4 Puffball hamster   f  
  
dbClearResult(d2q)
```

Merge

```
d3q <- dbSendQuery(  
  con, paste('select * from pet left join',  
             'event on pet.pet_id=event.pet_id'))  
(d3 <- dbFetch(d3q))
```

##	pet_id	name	owner	species	sex	birth	death	event
## 1	1	Fluffy	Harold	cat	f	1993-02-04	<NA>	
## 2	2	Claws	Gwen	cat	m	1994-03-17	<NA>	
## 3	3	Buffy	Harold	dog	f	1989-05-13	<NA>	
## 4	3	Buffy	Harold	dog	f	1989-05-13	<NA>	
## 5	4	Fang	Benny	dog	m	1990-08-27	<NA>	
## 6	4	Fang	Benny	dog	m	1990-08-27	<NA>	
## 7	5	Bowser	Diane	dog	m	1979-08-31	1995-07-29	
## 8	6	Chirpy	Gwen	bird	f	1998-09-11	<NA>	
## 9	7	Whistler	Gwen	bird	<NA>	1997-12-09	<NA>	
## 10	8	Slim	Benny	snake	m	1996-04-29	<NA>	
## 11	9	Puffball	Diane	hamster	f	1999-03-30	<NA>	
##	pet_id	date	type	remark				
## 1	1	1995-05-15	litter	4 kittens, 3 female, 1 male				
## 2	2	1998-03-17	birthday	Gave him a new flea collar				
## 3	3	1993-06-23	litter	5 puppies, 2 female, 3 male				
## 4	3	1994-06-19	litter	3 puppies, 3 female				
## 5	4	1991-10-12	kennel	<NA>				
## 6	4	1998-08-28	birthday	Gave him a new chew toy				

Selecciona columnas e merge

```
d4q <- dbSendQuery(  
  con, paste('select name,sex,species,date,type',  
             'from pet left join',  
             'event on pet.pet_id=event.pet_id'))
```

```
d4 <- dbFetch(d4q)
```

```
d4
```

##	name	sex	species	date	type
## 1	Fluffy	f	cat	1995-05-15	litter
## 2	Claws	m	cat	1998-03-17	birthday
## 3	Buffy	f	dog	1993-06-23	litter
## 4	Buffy	f	dog	1994-06-19	litter
## 5	Fang	m	dog	1991-10-12	kennel
## 6	Fang	m	dog	1998-08-28	birthday
## 7	Bowser	m	dog	1991-10-12	kennel
## 8	Chirpy	f	bird	1999-03-21	vet
## 9	Whistler	<NA>	bird	1998-12-09	birthday
## 10	Slim	m	snake	1997-08-03	vet
## 11	Puffball	f	hamster	<NA>	<NA>

```
dbClearResult(d4q)
```

Inner join

```
d5q <- dbSendQuery(  
  con, paste('select name,sex,owner,date,type',  
             'from pet inner join',  
             'event on pet.pet_id=event.pet_id'))  
(d5 <- dbFetch(d5q))
```

##	name	sex	owner	date	type
## 1	Fluffy	f	Harold	1995-05-15	litter
## 2	Buffy	f	Harold	1993-06-23	litter
## 3	Buffy	f	Harold	1994-06-19	litter
## 4	Chirpy	f	Gwen	1999-03-21	vet
## 5	Slim	m	Benny	1997-08-03	vet
## 6	Bowser	m	Diane	1991-10-12	kennel
## 7	Fang	m	Benny	1991-10-12	kennel
## 8	Fang	m	Benny	1998-08-28	birthday
## 9	Claws	m	Gwen	1998-03-17	birthday
## 10	Whistler	<NA>	Gwen	1998-12-09	birthday

```
dbClearResult(d5q)
```

Merge e restricao (apenas sex='f')

```
d6q <- dbSendQuery(  
  con, paste('select name,sex,owner,date,type',  
             'from pet inner join',  
             'event on pet.pet_id=event.pet_id',  
             'where sex="f"'))
```

```
d6 <- dbFetch(d6q)
```

```
d6
```

##	name	sex	owner	date	type
## 1	Fluffy	f	Harold	1995-05-15	litter
## 2	Buffy	f	Harold	1993-06-23	litter
## 3	Buffy	f	Harold	1994-06-19	litter
## 4	Chirpy	f	Gwen	1999-03-21	vet

```
dbClearResult(d6q)
```

```
dbDisconnect(con)
```

Outras opções

<https://db.rstudio.com/> **dplyr**, **DBI**, **odbc**, **keyring** and **pool**

- ▶ **dplyr**: dados em memória ou em banco de dados
 - ▶ usa **DBI**
 - ▶ RMySQL/MariaDB, RPostgreSQL, RSQLite, odbc, bigrquery

Exemplos diversos usando pool

<https://shiny.rstudio.com/articles/overview.html> ha exemplos em cluster Amazon

```
library(pool)
library(dplyr)
scon <- dbPool(
  RMySQL::MySQL(),
  dbname = "shinydemo",
  host = "shiny-demo.csa7qlmguqrf.us-east-1.rds.amazonaws.com",
  username = "guest",
  password = "guest"
)
```

Consulta com pool

```
scon %>% tbl("City") %>% head(5)
```

```
## # Source:   lazy query [?? x 5]
```

```
## # Database: mysql 5.5.5-10.0.17-MariaDB
```

```
## # [guest@shiny-demo.csa7qlmguqrf.us-east-1.rds.amazonaws.]
```

```
##      ID Name      CountryCode District      Population
```

```
##   <dbl> <chr>      <chr>      <chr>      <dbl>
```

```
## 1      1 Kabul      AFG      Kabul      1780000
```

```
## 2      2 Qandahar    AFG      Qandahar    237500
```

```
## 3      3 Herat      AFG      Herat      186800
```

```
## 4      4 Mazar-e-Sharif AFG      Balkh      127800
```

```
## 5      5 Amsterdam    NLD      Noord-Holland 731200
```