



Universidade de São Paulo
Instituto de Ciências Matemáticas e de Computação
Departamento de Ciências de Computação

Exercícios sobre Álgebra Relacional - Explorando a Base de Cidades

Prof. Dr. Caetano Traina Júnior
Entrega dia 30 de outubro de 2017

1 Descrição das Relações

Este trabalho é feito sobre quatro relações obtidas da *web*, e portanto se caracterizam como uma base de dados diferente do que é o usual em bases de dados corporativas, porque os dados não estão necessariamente consistentes entre as diversas relações. No entanto, os problemas que a falta de um planejamento unificado integrando as diversas relações, ao mesmo tempo em que propiciam desafios interessantes para o treinamento, também refletem uma realidade cada vez mais comum na informática: a necessidade de trabalhar e integrar dados provenientes de diversas fontes obtidas de distintos provedores/coletores de dados. As relações passaram por um pré-processamento para limpeza de dados e foram criadas as tabelas descritas a seguir.

1.1 Tabela Cidades

A tabela *Cidades* foi obtida do *site* do IBGE, e corresponde a todos os municípios oficialmente reconhecidos no Brasil em 2010. Existem muitas informações sobre cada município disponibilizadas pelo IBGE. Para quem estiver interessado, o arquivo “*Brasil-Censo2010.zip*” contém 17 tabelas sobre aspectos variados medidos para o Censo 2010, e o arquivo *base_MUNIC_ods.2013.zip* contém informações políticas e administrativas das cidades. O arquivo *Tab1.xls* foi extraído do censo 2010 e armazenado em Posgres como a tabela *Cidades*, cuja definição em SQL e os dados de carga (extraídos de *Tab.xls*) estão em *CreateInsertTab_Cidades.sql*. As demais tabelas não foram extraídas. Para extraí-las, basta abrir cada uma numa planilha e gerar os comandos “*insert into*” correspondentes (como feito para a tabela *Tab.xls*) e fazer a junção pelo atributo *Codigo*.

A tabela *Cidades* informa a população de cada cidade, com as porcentagens de quem vive em zona rural ou urbana, e porcentagens de homens e de mulheres, nos seguintes atributos:

```
Codigo decimal(7) -- Código univoco definido pelo IBGE para cada cidade;
Estado char(2)
Municipio char(70)
População decimal(10) -- População total
pctUrbana decimal(7,4) -- Porcentagem da população urbana
pctRural decimal(7,4) -- Porcentagem da população rural
pctHomem decimal(7,4) -- Porcentagem da população masculina
PctMulher decimal(7,4) -- Porcentagem da população feminina
pctHM decimal(7,4) -- Proporção entre homens e mulheres
```

Cada cidade pode ser identificada por seu Código entre as diversas tabelas do IBGE (por exemplo no arquivo *Tab1.xls*), ou pelos atributos {*Estado*, *Município*}. Atenção: no processo de limpeza, alguns valores em *Município* foram modificados, portanto podem estar diferentes dos mesmos municípios nas demais tabelas não limpas (dos arquivos *Tab1.xls* e *base_MUNIC_ods.2013.zip*), mas o atributo *Codigo* não foi modificado.

1.2 Tabela BRCitiesCoord

A tabela *BRCitiesCoord* indica a coordenada geográfica (em *latitude*, *longitude* de cada cidade brasileira oficialmente reconhecidas em 2004. Ela está definida e tem os dados de carga no arquivo

CreateInsertTab-BRCities-Coord. Note-se que novos municípios surgiram a partir de então, e portanto não estão nessa tabela. Seus atributos são:

```
Municipio varchar(70)
UF char(4)
Lat decimal(5,2)
Long decimal(5,2)
```

Os atributos {Município, UF} correspondem aos atributos {Estado, Município} da tabela de Cidades.

1.3 Tabela Aeroportos

A tabela **Aeroportos** contém uma lista de aeroportos brasileiros e internacionais, com ênfase nos aeroportos brasileiros. Estão presentes apenas os aeroportos internacionais que têm conexão internacional, e todos os aeroportos brasileiros com sigla ICAO. A tabela **Aeroportos** informa a Sigla ICAO, e a localização do aeroporto pelo município, estado, País e continente.

```
Sigla char(5) -- Sigla ICAO do aeroporto.
Descricao varchar(100) -- Descrição sucinta, tal como o nome do aeroporto quando
houver.
Municipio varchar(70)
UF char(4)-- O Estado, quando o aeroporto é Brasileiro.
País char(25)
Continente char(16)
```

Os aeroportos brasileiros têm a indicação de seu estado, de maneira que os atributos {Município, UF} correspondem aos atributos {Estado, Município} da tabela de Cidades. Os aeroportos internacionais não têm informação sobre o estado correspondente. Os comandos de definição de carga estão no arquivo no arquivo **CreateInsertTab-Aerodromos.sql**.

1.4 Tabela Estacoes

A tabela **Estacoes** foi obtida do *site* do INPE-CPTEC, e corresponde às estações meteorológicas registradas em todo o mundo. Como quase todo aeroporto tem junto uma estação meteorológica, as estações têm um designador internacional (WMO) e se a estação estiver em um aeroporto, ela tem também o código ICAO correspondente. A tabela **Estacoes** armazena o País, os códigos WMO e ICAO (quando aplicável), um nome para a **estação** (que geralmente está associado à cidade ou região onde ela está), e as coordenadas **latitude**, **Longitude** e **altitude** de cada estação. Seus atributos são:

```
País char(2)
WMO char(5)
ICAO char(4)
Estacao varchar(70) -- Nome simbólico
Latitude decimal(7,4)
Longitude decimal(7,4)
Altitude decimal(7,2) -- em metros
```

Os comandos de definição de carga estão no arquivo no arquivo **CreateInsertTab-Estacoes.sql**.

2 Definição do Trabalho

Este trabalho deve ser feito em grupo de três alunos, e entregue dia 30 de outubro (segunda-feira). Identifique todos os participantes pelo seu nome, número USP e e-mail.

A entrega do trabalho deve ser feita pelo tidia-4, em um arquivo em formato .zip contendo para cada exercício:

1. O comando em SQL, em um arquivo em formato texto (.sql);
2. Um arquivo em formato pdf, contendo
 - Comentário que explique como o comando funciona (documentação do comando); e
 - Uma listagem do resultado da execução do comando em Postgres. Quando o resultado tiver até 20 tuplas, listar todas elas. Quando o resultado tiver mais de 20 tuplas, listar apenas de 10 a 20 das tuplas mais significativas, que mostrem tuplas de maior interesse para mostrar a corretude do comando executado.

Portanto, o arquivo .zip deve conter 20 arquivos, Exec01.sql a Exec10.sql e Exec01.pdf a Exec10.pdf

3 Consultas em SQL

Exercício 1) Crie as Restrições de Integridade Fundamentais que existem nessa base de dados: Unicidade da Chave, Integridade da Entidade e Integridade de Referencial.

Exercício 2) Quantas cidades brasileiras se chamam 'São Carlos'? Quantas têm 'São Carlos' no nome? Quais são elas, e em que estado está cada uma?

Exercício 3) Quais cidades brasileiras têm maior percentagem de população feminina? e masculina? e Rural? e Urbana?

Note-se que pode haver empate, e haver mais de uma cidade com a mesma porcentagem de algo.

Exercício 4) Quais cidades brasileiras têm a maior porcentagem de população feminina em cada estado? Existe algum estado em que mais de uma cidade empata nesse quesito?

Exercício 5) Liste as cidades do mundo com mais de um Aeroporto, indicando quantos aeroportos cada cidade tem. Ordene a resposta pela quantidade de aeroportos, começando pelas cidades com mais aeroportos, e dentro de cada quantidade, ordene pelo Estado da cidade, com os aeroportos estrangeiros precedendo os nacionais.

Exercício 6) Liste todas as cidades brasileiras incluindo as coordenadas dos aeroportos que elas eventualmente tenham. Gere o resultado ordenado pelo nome do município.

Exercício 7) Como as tabelas **BRCitiesCoord** e **Cidades** foram obtidas de fontes não relacionadas e em épocas diferentes, elas podem ter cidades diferentes, ou com nomes diferentes, ou mesmo que não existiam no período. Mostre quais são as cidades em **BRCitiesCoord** e em **Cidades** que não têm correspondentes na outra tabela.

Exercício 8) Quantos aeroportos estão na tabela de aeroportos mas não na tabela de estações, identificados pela sua sigla ICAO? Indique o total por país.

Exercício 9) Listar os aeroportos do Estado de São Paulo com suas respectivas coordenadas, quando disponíveis.

Exercício 10) Qual é o aeroporto mais elevado do estado de São Paulo?

[Ultima atualização desta lista: 25 de setembro de 2017]