

# Modelagem e consultas de dados relativos a taxas de rendimento escolar na educação básica brasileira

Estevan Gladstone<sup>1</sup>, João Luis Guio<sup>1</sup>, Matheus Andrade<sup>1</sup>, Tiago Montalvão<sup>1</sup>

<sup>1</sup>Departamento de Ciência da Computação  
Universidade Federal do Rio de Janeiro (UFRJ)

**Abstract.** *This paper describes the process of modeling a database found in a government website, since the creation of an ER model, going through the logical model up to the creation of tables in a SQL physical system through the MySQL Database Management System. A set of queries is presented together with the Web application developed to access the database.*

**Resumo.** *Este artigo descreve o processo de modelagem de uma base de dados encontrada em site do governo, desde a criação de um modelo ER, passando pelo modelo lógico e por fim a criação de tabelas em um sistema físico SQL, através do Sistema de Gerenciamento de Banco de Dados MySQL. Um conjunto de consultas é apresentado, juntamente com a descrição da aplicação Web desenvolvida para acessar o banco de dados.*

## Introdução

Este artigo começa descrevendo brevemente a base de dados apresentada em <http://dados.gov.br/dataset/taxas-de-rendimento-escolar-na-educacao-basica>. Esta base apresenta dados indicativos do Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (Inep) sobre diferentes tipos de taxas escolares da educação básica brasileira: aprovação, reprovação e abandono. Estas taxas são separadas por escolas, por ano e por ano escolar na escola.

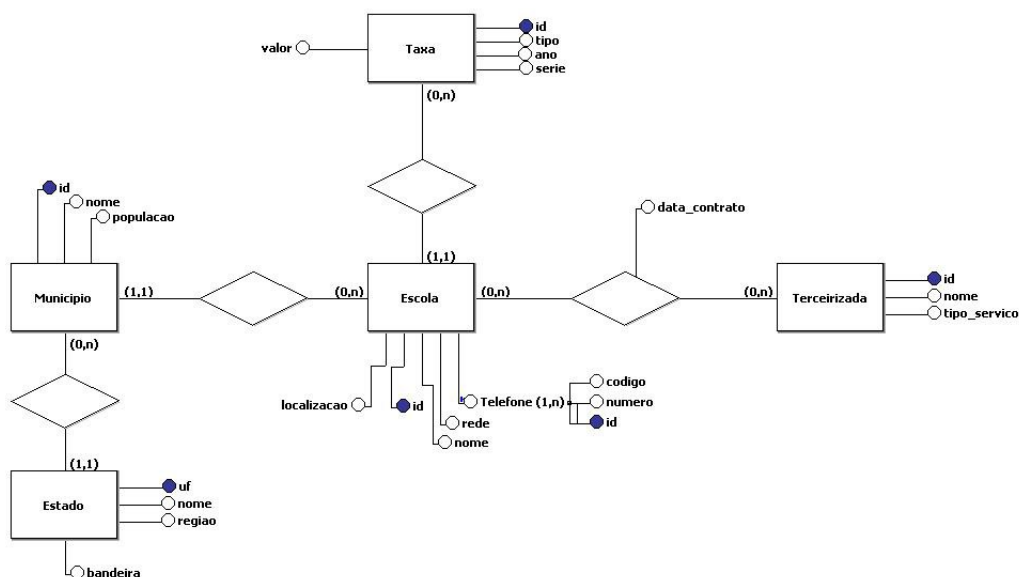
A fim de deixar a modelagem mais rica em relações, o nosso grupo criou mais uma entidade, representando empresas terceirizadas que prestam serviços às escolas.

As modelagens das entidades no modelo ER e no lógico, e a subsequente tradução para modelo físico, foram realizadas com o uso do software brModelo.

## Modelagem ER

A modelagem Entidade-Relacionamento (Fig.1) leva em consideração os objetos mais importantes a serem modelados e os transforma em entidades. Sendo assim, temos 5 entidades:

- Escola
- Municipio
- Estado
- Taxa
- Terceirizada



**Figura 1. Modelagem ER da base de dados**

Todos eles possuem um atributo identificador e pelo menos mais dois outros atributos, além da entidade Escola possuir um atributo (Telefone) multivalorado.

As cardinalidades das relações presentes são 1:N e N:N.

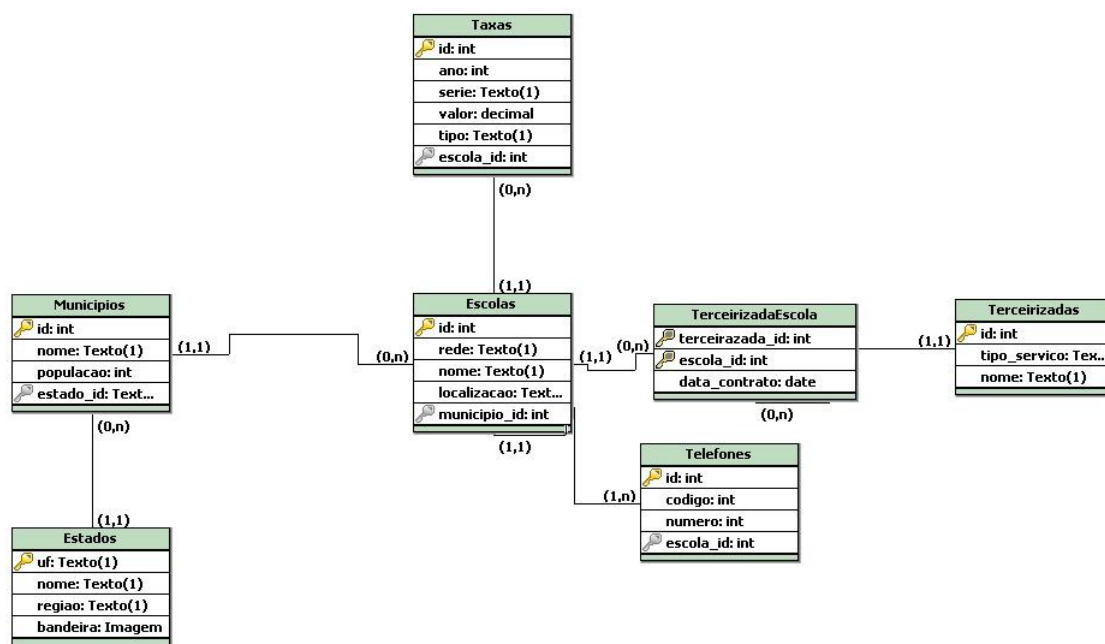
A relação N:N entre Escola e Terceirizada é uma relação que possui um atributo, representando a data do contrato dos serviços prestados pela empresa terceirizada.

## Modelo Lógico

A transformação de modelo ER para lógico (Fig.2) dá-se de maneira semi-automática pelo software brModelo. Algumas configurações manuais são necessárias, assim como alguns ajustes no modelo gerado:

- **Atributo multivalorado Telefone:** foi criada uma tabela separada para tal, com chave estrangeira escola\_id para a respectiva escola. Havia a possibilidade de incluir os atributos de Telefone na tabela Escola, mas isto causaria grandes redundâncias.
- **Chave primária da tabela TerceirizadaEscola:** esta é uma tabela relacional, que guarda informações de todas as relações entre uma empresa e uma escola. Ela contém chaves estrangeiras referenciando as chaves primárias das respectivas tabelas do relacionamento. Estas chaves devem constituir a chave primária da nova tabela, o que não acontecia inicialmente no modelo gerado.

Podem ser informados os tipos de cada atributo de cada tabela na hora da criação do modelo lógico no brModelo. Isto facilita a futura tradução em modelo físico. É válido mencionar que a tabela Estado possui um campo que será uma imagem, e assim é do tipo BLOB (Binary Large Object), que representa no caso uma imagem, mas poderia representar arquivos de mídia, como áudio ou vídeo, ou até mesmo grandes documentos.



**Figura 2. Modelo lógico da base de dados**

## Análise da forma normal

Os dados nas planilhas da base de dados da qual extraímos os dados encontrava-se em nenhuma forma normal, sendo assim classificada como pertencente à 0FN.

A modelagem apresentada encontra-se pelo menos na 3FN, pois não é encontrada dependência funcional nem parcial nem transitiva de chave. Isto foi garantido na hora de fazer a modelagem ER e a respectiva tradução para o modelo lógico. Um exemplo de situação que não permitiria a modelagem de estar na 2FN seria se as informações de escolas, municípios e estados estivessem todos na mesma tabela. Estado depende da escola, mas também depende de município que, por sua vez, depende do atributo escola.

## Modelo físico

A tradução para o modelo físico também foi realizado pelo software brModelo. Isto é feito de maneira automática, mas uma pequena modificação (inserção do caractere ';' após cada comando SQL) também foi necessária.

Após a criação de um arquivo com os comandos SQL para a criação das tabelas, este foi executado em um banco de dados, através do SGBD MySQL, e as tabelas foram criadas, ainda vazias.

## Inserção de valores no banco de dados

A extração de dados das planilhas e a respectiva inserção no banco foram feitas em três passos:

1. Tradução do arquivo XLS das planilhas para o formato CSV

2. Criação de scripts (em PHP e em Lua) que manipulam os dados dos arquivos .CSV e geram um outro .CSV, agora pronto para ser inserido no BD
3. Inserção no BD, dos arquivos .CSV já tratados, através do phpMyAdmin, configurado em um servidor local.

## Consultas

A seguir, apresentamos uma lista de consultas, com sua descrição (o que envolve e seu enunciado):

– Id, nome, população e id do estado de todas os municípios do estado do Rio de Janeiro (envolve apenas seleção e projeção)

```
SELECT id , nome , populacao , estado_id FROM Municipio
WHERE estado_id = 'RJ';
```

– Obtém o nome, localização, rede, nome do municipio e UF de todas as Escolas cuja UF é MG ou PR (envolve junção de apenas duas relações)

```
SELECT Escola.nome , localizacao , rede , Municipio.nome
AS municipio , Municipio.estado_id AS uf
FROM Escola INNER JOIN Municipio ON Escola .
municipio_id=Municipio.id
WHERE Municipio.estado_id = 'MG' OR Municipio .
estado_id = 'PR';
```

Obtém todos os telefones da escola de nome CEFET CELSO SUCKOW DA FONSECA (envolve junção de apenas duas relações)

```
SELECT codigo , numero
FROM Escola INNER JOIN Telefone ON Telefone.escola_id=
Escola.id
WHERE nome = 'CEFET_CELSO_SUCKOW_DA_FONSECA';
```

Obtém o nome, rede, localização e uf de todas as Escola da região Sudeste (envolve junção de três relações)

```
SELECT Escola.nome , rede , localizacao , uf
FROM Escola INNER JOIN Municipio ON Escola .
municipio_id=Municipio.id INNER JOIN Estado ON
Estado.uf=Municipio.estado_id
WHERE Estado.regiao = 'Sudeste';
```

Obtém a média das taxas de aprovação do terceiro ano do ensino medio por estado (envolve junção de três relações e um agrupamento)

```
SELECT avg(Taxa.valor) AS mediaAprovacao , uf , bandeira
FROM Escola INNER JOIN Taxa ON Escola.id=Taxa .
escola_id INNER JOIN Municipio ON Escola .
municipio_id=Municipio.id INNER JOIN Estado ON
Estado.uf=Municipio.estado_id
```

```
WHERE Taxa.serie = 12
GROUP BY Estado.uf;
```

As escolas que tem reprovacão maior que 50% ou abandono maior que 50% em qualquer ano (consulta envolvendo operações sobre conjuntos)

```
SELECT distinct Escola.nome
FROM Escola left JOIN Taxa on Escola.id = Taxa.
    escola_id
WHERE Taxa.tipo = 'Reprovacão' and Taxa.valor >= 50
union
SELECT distinct Escola.nome
FROM escola left JOIN taxa on Escola.id = Taxa.
    escola_id
WHERE Taxa.tipo = 'Abandono' and Taxa.valor >= 50
```

Obtém o número de Escola cadastradas no sistema por estado (consulta envolvendo função de agregação)

```
SELECT count(distinct Escola.id), Estado.nome, uf,
    regiao, bandeira
FROM Escola INNER JOIN Municipio ON Escola.
    municipio_id=Municipio.id INNER JOIN Estado ON
    Estado.uf=Municipio.estado_id
GROUP BY Estado.uf
```

Número de Escola cadastradas que tem taxa de aprovacão acima de 50% do terceiro ano do ensino medio por região (consulta envolvendo função de agregação)

```
SELECT count(distinct Escola.id), regiao
FROM Escola INNER JOIN Municipio ON Escola.
    municipio_id=Municipio.id INNER JOIN Estado ON
    Estado.uf=Municipio.estado_id LEFT JOIN Taxa ON
    Escola.id=Taxa.escola_id
WHERE Taxa.valor > 50.0 AND Taxa.serie = 12 AND Taxa.
    tipo = 'Aprovacão'
GROUP BY regiao;
```

Estados que possui média de reprovacao das escolas maior ou igual a 30% (consulta envolvendo subconsultas aninhadas)

```
SELECT name FROM
(
    SELECT Estado.nome as name, avg(taxa.valor) as
        media
    FROM estado JOIN municipio on estado.id =
        municipio.estado_id left JOIN escola on
        municipio.id = escola.municipio_id left JOIN
        taxa on escola.id = taxa.escola_id
```

```

        group by Estado.UF
    ) as t
WHERE t.media >= 30

```

Escolas que possuem mais de 2 terceirizadas contratadas (consulta envolvendo sub-consultas aninhadas)

```

SELECT name FROM
(
    SELECT Escola.nome as name, count(terceirizada.id)
        as c
    FROM TerceirizadaEscola JOIN Terceirizada on
        Terceirizada.id = TerceirizadaEscola.
        terceirizada_id JOIN Escola on Escola.id =
        TerceirizadaEscola.escola_id
    group by Escola.id
) as t
WHERE c > 2

```

A escola com maior taxa de reprovação do terceiro ano do ensino medio na cidade do rio de janeiro agrupadas por rede (consulta do tipo relatório)

```

SELECT max(Taxa.valor), Escola.nome, rede, localizacao
FROM Escola INNER JOIN Taxa ON Escola.id=Taxa.
    escola_id
WHERE Taxa.serie = 12 AND Taxa.tipo = 'Reprovação'
GROUP BY Escola.rede;

```

A escola com menor taxa de aprovação do terceiro ano do ensino medio na cidade do rio de janeiro agrupadas por rede (consulta do tipo relatório)

```

SELECT min(Taxa.valor), Escola.nome, rede, localizacao
FROM Escola INNER JOIN Taxa ON Escola.id=Taxa.
    escola_id
WHERE Taxa.serie = 12 AND Taxa.tipo = 'Aprovação'
GROUP BY Escola.rede;

```

## Participações

Todos os membros participaram indiretamente de todas as partes do trabalho. A seguir, segue mais detalhadamente, por ordem de participação, o que cada membro do grupo fez:

- Estevan Gladstone
  - Criação de consultas
  - Criação de scripts de manipulação para importação de dados para o banco de dados
  - Ajudou na aplicação web
  - Modelagem das entidades e tabelas
- João Guio

- Criação de scripts de manipulação para importação de dados para o banco de dados
  - Modelagem das entidades e tabelas
  - Criação de consultas
  - Geração de dados abstratos falsos para as tabelas inventadas
- Matheus Andrade
  - Desenvolvimento da aplicação web
  - Modelagem das entidades e tabelas
- Tiago Montalvão
  - Escrita do relatório
  - Exportação de planilhas para arquivos CSV
  - Criação de consultas
  - Modelagem das entidades e tabelas