

Dados Públicos Enem 2013

Ricardo Souza Rachaus¹, Juarez A. Sampaio Filho²

¹ Departamento de Ciência da Computação
Universidade de Brasília(UnB) – Brasília, DF – Brasil

juarez.asf@gmail.com, ricardorachaus@gmail.com

Resumo. *Dados públicos passaram a ser de grande interesse das pessoas com o advento da tecnologia, assim, muitos os buscam com diversos objetivos, desde estudos científicos até análises estatísticas. Com o Enem não seria diferente, sendo o maior exame do país, trabalhar sobre esses dados pode atrair o interesse de diversas áreas, e com seus dados públicos, torna-se uma ótima fonte de dados para pesquisas e trabalhos, como será visto neste texto.*

1. Introdução

No início, o Enem era apenas uma prova avaliativa do Ensino Médio, porém, hoje, é um dos principais instrumentos de seleção para ingresso em instituições de Ensino Superior. Assim, seus dados podem ter informações de grande utilidade para pesquisas.

Os microdados do Enem são o conjunto de dados obtidos após sua aplicação de um referido ano, contendo cidades, escolas, notas, dados de alunos e outros, sendo que nenhum destes dados implica na identificação direta de nenhum dos participantes.

Neste trabalho, será utilizado os dados do Enem referentes ao ano de 2013, armazenando estes em um Banco de Dados Relacional, mais especificadamente usando o SGBD PostgreSQL, onde as informações são armazenadas em tabelas e essas tabelas se relacionam. Com isso, busca-se obter informações interessantes sobre esses dados.

2. Diagrama Entidade Relacionamento

O Diagrama Entidade Relacionamento busca apresentar as principais entidades e seus relacionamentos através dos dados.

Esse diagrama serve como base para toda a construção do banco de dados, além de ser o primeiro momento em que os dados que serão armazenados são pensados de forma a satisfazer os requisitos do problema.

Aqui segue o modelo referente aos dados do Enem.

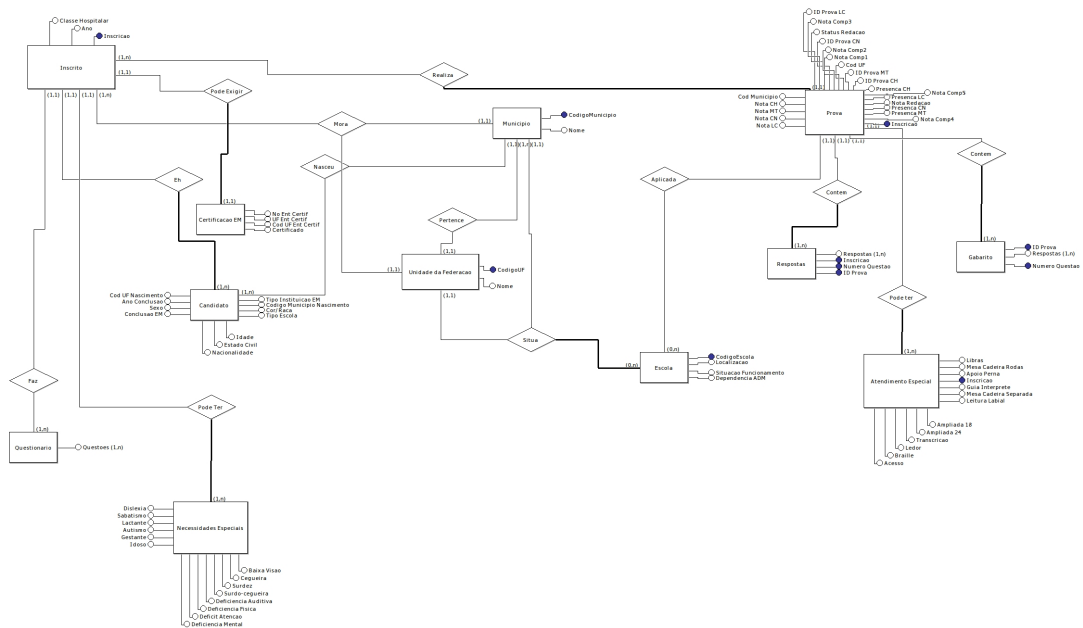


Figura 1. Diagrama Entidade Relacionamento

3. Modelo Relacional

O Modelo Relacional é a construção de tabelas, onde os relacionamentos entre as entidades são dados por atributos ou outras tabelas. O modelo gerado a partir do Diagrama Entidade Relacionamento acima é este:

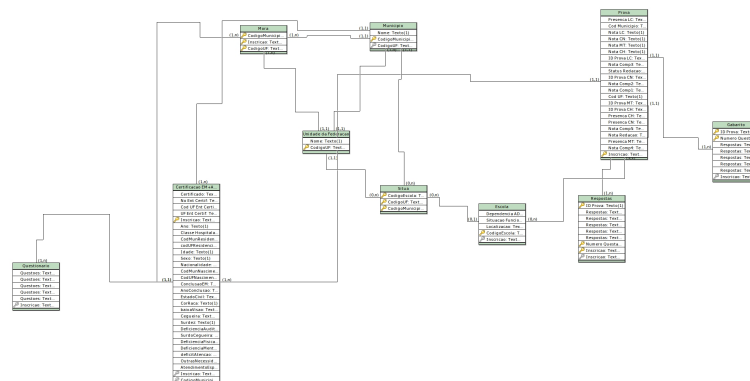


Figura 2. Modelo Relacional

4. Formas Normais

As formas normais servem para a criação de um banco de dados com dados consistentes e que seja eficiente no acesso ao banco. Essas formas são divididas em 3: 1FN (primeira forma normal), 2FN (segunda forma normal) e 3FN (terceira forma normal).

A 1FN, primeira forma normal, já é feita automaticamente ao ter o Diagrama Entidade Relacionamento transformado no Modelo Relacional, pois a teoria Relacional

já garante não haver atributos não-atômicos. Ou seja, não há campos multivalorados nas colunas que constituem as tabelas do Modelo Relacional.

A 2FN, segunda forma normal, refere-se ao fato de que cada tabela, todos os atributos não-chave não servem para identificar a entidade em questão, ou seja, apenas os atributos que constituem a chave servem para identificar uma entidade na tabela. Nesse caso, cada tabela do Modelo Relacional apresentado segue isso. Observando que para uma tabela estar na segunda forma normal, precisa estar também na primeira forma normal.

A 3FN, terceira forma normal, serve para retirar de uma tabela todos os atributos não dependentes da chave dessa tabela, ou seja, cada coluna da tabela, seu valor dependente diretamente da chave. O modelo apresentado também encontra-se na 3FN.

Obs.: Para uma tabela estar na terceira forma normal, é preciso estar tanto na primeira forma normal como na segunda forma normal.

5. Criação do Banco

A criação do banco foi dada através da ferramenta PostgreSQL. Nesse caso, a partir de um servidor criado no SGBD, criou-se um Database para armazenar os dados do Enem 2013.

Para a criação das tabelas, foi utilizado o arquivo "tabelas.sql" contido na pasta "sql", onde este arquivo contém todo o código SQL para a criação do banco, ou seja, contém todos os "CREATE TABLE" para criar cada uma das tabelas que se encontra no modelo.

6. ETL - Extract, Transform, Load

A extração, transformação e insreção dos dados a partir dos MICRODADOS do Enem 2013 se deu a partir de uma série de *scripts Python* onde as informações foram extraídos, e assim, era gerado um arquivo ".sql" com o código para realizar as inserções no banco de dados.

Para a extração, o *script Python* lia todos os dados e fazia uma separação deles considerando as tabelas onde seriam inseridos. A transformação se deu através do resultado da extração, onde a separação dos dados em arquivos ".sql" transformou o dado bruto, o original, em dado que para ser inserido no banco. Já o carregamento é realizado através da execução do *script SQL* no PostgreSQL.

7. Visualização dos Dados

8. View

9. Procedure

10. Trigger

11. Conclusão

12. Bibliografia

INSTITUTO NACIONAL DE ESTUDOS E PESQUISAS EDUCACIONAIS ANÍSIO TEIXEIRA. Microdados do Enem 2013. Brasília: Inep, 2015. Disponível em: <http://portal.inep.gov.br/basica-levantamentos-acessar>. Acesso em: 30 mai. 2015.