

# Projeto Final BD

## 1 Introdução

Quando há requisito de eficiência nas organizações, tanto com fins lucrativos ou não, é importante que haja conhecimento sobre os dados que transitam por tal organização. Tais dados devem ser trabalhados de modo a se adquirir informações úteis para um bom planejamento, do gerencial ao estratégico.

É possível armazenar tais dados em memória simples de um computador, sendo esta um HD, e realizar todas as operações sobre tais dados; mas há o caso de que, cada vez mais, a massa de dados vem aumentando. Tal aumento exige um gerenciamento adequado, promovido pelos *Sistemas Gerenciadores de Banco de Dados (SGBDs)*.

Primeiramente, é preciso compreender o que seria um *banco de dados* (BD): pode ser visto como o equivalente eletrônico de um armário de arquivamento, uma vez que é a coleção de dados persistentes utilizadas pelos sistemas de aplicação de uma empresa. Dessa forma, um SGBD é responsável pelo gerenciamento dessas coleções, garantindo integridade, segurança e outras características.

Os dados podem ser extraídos de qualquer fonte, mas, para este trabalho, será utilizada a *INDA (Infraestrutura Nacional de Dados Abertos)*; visto que são dados que estão à disposição livremente a todos, mas apenas alguns a acessam. Dessa forma, ao se trabalhar com tais dados e gerar informações consideráveis estaríamos exercendo a nossa cidadania, buscando discrepâncias entre os dados e apontando algumas irregularidades.

Como escopo, foi escolhido os dados sobre **Diárias e passagens** (Tais dados estão disponível *on-line*: <http://www.portaltransparencia.gov.br/>) limitados ao período: *jan/2015* a *jun/2015*, visando responder os seguintes questionamentos:

1. Qual o gasto total em cada mês?
2. Quais os órgãos que mais gastaram?
3. Quais os programas que mais gastaram?
4. Quais os servidores que mais gastaram?

5. Quais as funções que mais gastaram?

## 2 Diagrama Entidade Relacionamento (DER)

Um DER constitui uma forma de representação gráfica para os conceitos atribuídos ao *Modelo Entidade Relacionamento* (MER), sendo este um modelo de dados conceitual de auto nível, visto que está centrado na percepção dos usuários sobre os dados, não importando a maneira na qual os dados serão armazenados. Dessa forma, entendi o DER como representação visual sobre os dados que serão tratados no BD.

Como os dados fornecidos vem em forma de planilhas, foi preciso realizar uma leitura e uma interpretação dos mesmos, tentando identificar o que seriam as entidades e os relacionamentos para o nosso BD. Dessa forma, obteve-se a Figura 1 como resultado do estudo realizado.

## 3 Modelo Relacional (MR)

Um MR representa os dados num BD como uma coleção de relações, denominadas de tabelas. Essa coleção será implantada no SGBD, ou seja, o MR representa a construção física do BD. Por isso é desenvolvido a partir do DER e, no escopo deste trabalho, tem-se a Figura 2 como resultado.

## 4 Avaliação das formas normais

A normalização é importante para identificar um bom projeto relacional. Um bom MER e sua consequente conversão para um MR, praticamente, deixa o esquema relacional *normalizado*.

Ao se tratar de normalização, consideram-se três formas normais, onde: na primeira (1FN) há a caracterização de um valor de uma coluna de uma tabela é indivisível; na segunda (2FN), se estiver na 1FN e todo atributo do complemento de uma chave candidata é totalmente funcionalmente dependente daquela chave; por fim, na terceira, se estiver na 2FN e todos os atributos não-chave forem dependentes não-transitivos da chave primária.

Dessa forma, foram selecionadas três tabelas: Pessoa, Ação e Função; onde:

- **Pessoa:**

<u>Cpf</u>	Nome	Cod_superior	Cod_subordinado	Cod_unidade_gestora
------------	------	--------------	-----------------	---------------------

- 1FN: Todos os valores das colunas da tabela são indivisíveis;
- 2FN: Chave candidata: Cpf

*Cpf* → *Nome*

*Cpf* → *Cod\_superior*

*Cpf* → *Cod\_subordinado*

*Cpf* → *Cod\_unidade\_gestora*

- 3FN: Não há nenhuma transitividade nas colunas, uma vez que todos são definidos unicamente somente pelo Cpf.

- **Ação:**

<u>Código</u>	Nome	Linguagem_citada
---------------	------	------------------

- 1FN: Todos os valores das colunas da tabela são indivisíveis;
- 2FN: Chave candidata: Código

*Código* → *Nome*

*Código* → *Linguagem\_citada*

- 3FN: Não há nenhuma transitividade nas colunas, uma vez que todos são definidos unicamente somente pelo Código.

- **Função:**

<u>Código</u>	Nome
---------------	------

- 1FN: Todos os valores das colunas da tabela são indivisíveis;
- 2FN: Chave candidata: Código

*Código* → *Nome*

- 3FN: Não há nenhuma transitividade nas colunas, uma vez que todos são definidos unicamente somente pelo Código.

## 5 Criação do BD

## 6 Processo de ETL

## 7 Persistência e Visualização

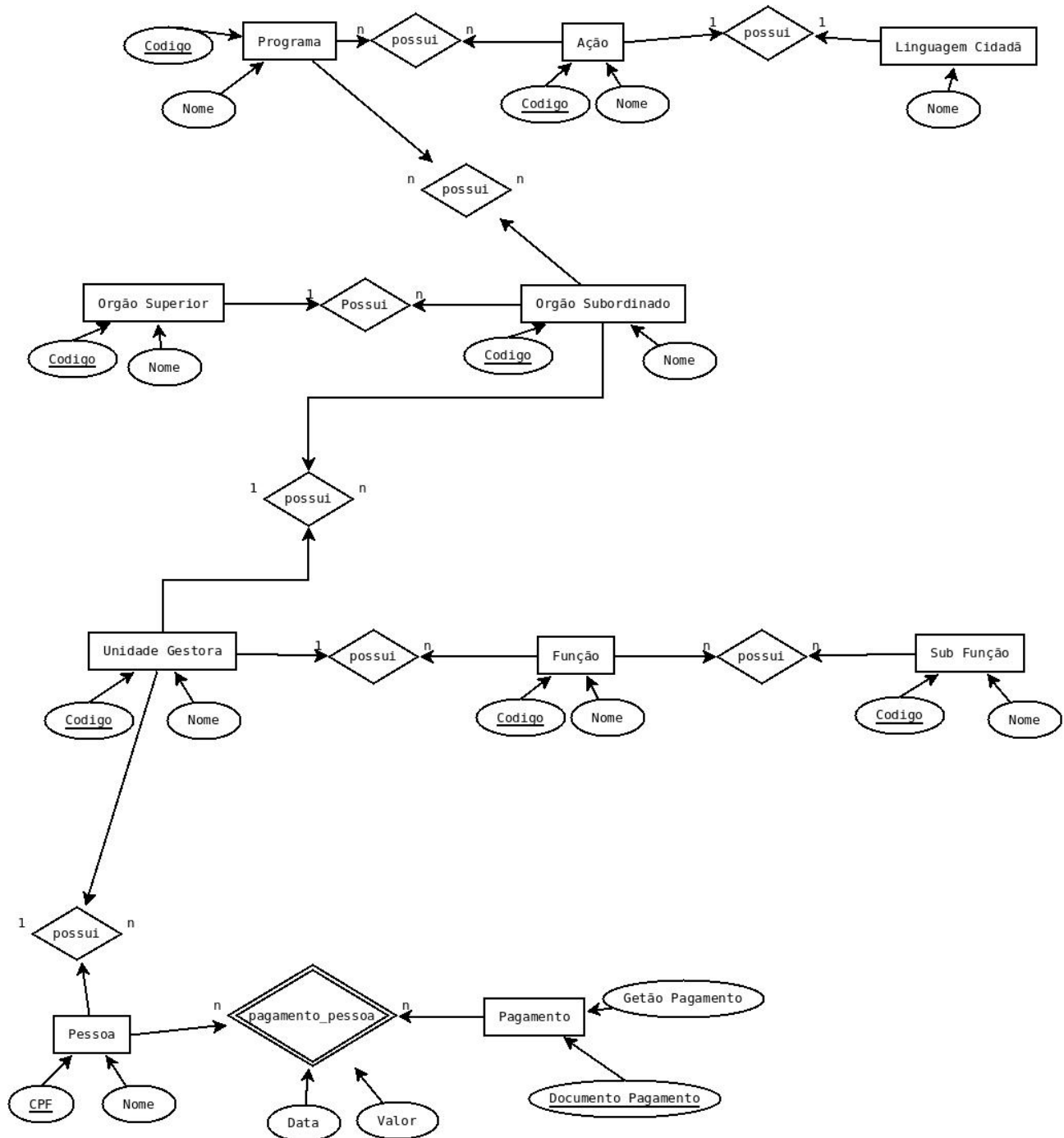


Figura 1: DER gerado a partir das tabelas e nossa interpretação.

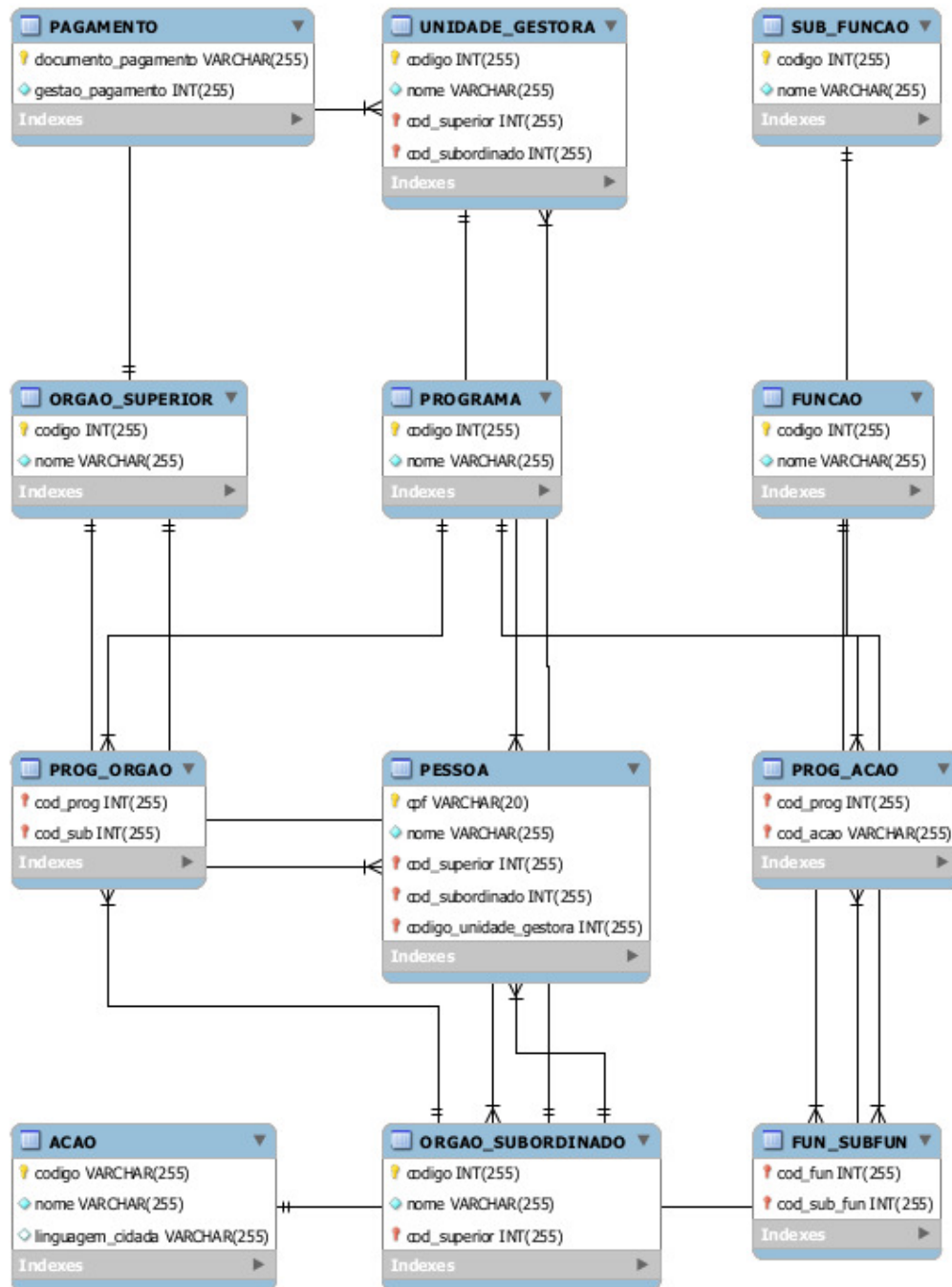


Figura 2: MR gerado a partir do DER.