



Arthur Oliveira Reis, Jonathan Coutinho Alves Santos
e Leonardo Castelo Branco Melo

CESAR School

Definição de Minimundo e Modelagem de Banco de Dados

Recife
2019

Equipe E

Projeto Mapeamento de Comportamento Digital em Tempo Real

MODELAGEM DE BANCO DE DADOS E DEFINIÇÃO DE MINIMUNDO

Este documento refere-se ao processo de modelagem de banco de dados e definição do minimundo do Projeto IV, desenvolvido pela Equipe E, do quarto período (turma 2018.1).

Recife

2019

Sumário Executivo

Este documento tem por finalidade descrever o processo de desenvolvimento do banco de dados do Projeto de Mapeamento do Comportamento Digital em Tempo Real, referente ao desafio apresentado pela Neurotech, desde a definição do minimundo. Aqui, encontra-se uma breve introdução sobre o projeto, descrição do objetivo da aplicação utilizando o modelo de dados criado, descrição textual de alto nível sobre o minimundo do projeto. Além, o documento aprofunda-se tanto na identificação e descrição das entidades envolvidas no projeto (juntamente com seus atributos), como nos relacionamentos entre as entidades do sistema (incluindo cardinalidades e possíveis atributos). A partir da leitura destes, entende-se o Modelo Entidade-Relacionamento (MER) apresentado e, conseqüentemente, o modelo normalizado.

SUMÁRIO

Introdução	4
Objetivo da Aplicação	4

Descrição dos Dados (Minimundo)	5
Entidades	6
Relacionamentos	7
Modelo Entidade-Relacionamento	8
Modelo Relacional Normalizado	9

Introdução

O projeto surgiu a partir de uma parceria entre a CESAR School e a Neurotech, que assumiram os papéis de mediadora e cliente, respectivamente. A Neurotech analisa dados com aprendizagem de máquina, assim como utiliza análise preditiva para identificar padrões para e prever comportamentos e sugerir as melhores decisões para seus clientes. Com isto em mente,

o tema acerca do projeto refere-se ao processo de coleta de dados, disponíveis em fontes públicas e privadas, para cruzar com as variáveis e facilitar decisões no contexto de concessão de crédito.

Objetivo da Aplicação

A Neurolake, plataforma de serviço MLaaS (Machine Learning as a Service), que oferece ferramentas de aprendizagem de máquina, visualização de dados e análise preditiva, é utilizada exaustivamente pela Neurotech para coletar o máximo de dados possíveis de fontes de dados diversas. Dados estes, que permitem mapear o comportamento digital de perfis de usuários, cuja finalidade fortalecer a plataforma de conhecimento da empresa, a modo de oferecer melhores tomadas de decisão e, no contexto de concessão de crédito, assertividade na concessão do limite adequado para cada cliente.

De qualquer forma, uma vez coletados, os dados demoram a ser atualizados - quanto maior o tempo entre a atualização dos dados, menos assertiva a predição é. Como consequência, caso um usuário solicite crédito e os dados referentes à eles não estejam atualizados, o mesmo pode ter seu pedido negado ou aprovado, trazendo um risco maior a quem está concedendo crédito. O mapeamento do comportamento digital, em tempo real, permite à Neurotech avaliar os desafios na concessão de crédito de forma mais assertiva e justa. Com isto, a aplicação tem por objetivo avaliar o seguinte desafio:

“Como podemos mapear o comportamento digital em tempo real, respeitando a Lei Geral de Proteção de Dados (LGPD)?”

Descrição dos Dados (Minimundo)

Uma pessoa, para solicitar crédito em uma instituição credora, deve apresentar as seguintes informações: nome, data de nascimento, sexo, identidade de gênero, RG, CPF, pelo menos um telefone e pelo menos um email.

A instituição credora deve ter, além do seu código único de identificação, CNPJ e pelo menos uma agência que, por sua vez, tem um código identificador da agência, telefone e endereço (composto pelo CEP — região, estado, cidade, bairro e rua —, número e complemento).

Uma vez que tanto uma pessoa, como uma instituição credora, tenham atendido seus respectivos critérios, uma pessoa pode solicitar. Ao solicitar o crédito, a instituição credora fará uma análise do comportamento digital do solicitador, através do mapeamento dos sites visitados, assim como das contas associadas à ele. Cada site tem um link único, informações sobre a última visita de cada pessoa no site, e uma categoria, e cada conta associada tem um email, nome, link para conta e uma categoria qual a conta se encaixa (rede social, fórum, comunidade, etc). Lembre-se que, ao longo da sua vida, um indivíduo pode não criar contas ou visitar sites, assim como pode visitar um ou mais sites e criar uma ou mais contas em um ou mais lugares.

Uma vez que a análise do comportamento digital foi realizada, se a instituição credora conceder crédito à pessoa, a data da concessão é mantida. Além disso, o crédito concedido tem um limite, tipo de crédito (cartão, seguro, etc), está associado à instituição credora através do código da mesma, valor do crédito que foi pago até o momento, e cpf da pessoa que recebeu o crédito. É possível que uma pessoa solicite crédito de tipos diversos em uma mesma instituição, ou um mesmo tipo de crédito em instituições diferentes. Ao mesmo passo, um mesmo tipo de crédito pode ser concedido para diversas pessoas e diversas instituições podem ter os mesmos tipos de crédito. Caso a instituição não conceda o crédito, nada é registrado.

Entidades

As entidades aqui representadas consistem de um conjunto de informações sobre o tema do projeto. Cada entidade tem também um conjunto de atributos, que representam informações que referenciam à entidade. Seguem, abaixo, as entidades e atributos que foram utilizadas para modelagem de banco de dados do projeto:

Endereço: entidade fraca que representa endereço físico de uma pessoa, ou agência. Os atributos que referenciam esta entidade são: CEP (atributo discriminador), apartamento (número do edifício, casa, etc), complemento (número do apartamento, torre, etc), rua e descrição (casa, trabalho, lazer).

Pessoa: entidade forte que representa usuário que pede crédito ao banco. Os atributos que referenciam esta entidade são: nome da pessoa, sexo (masculino ou feminino), identidade de gênero, CPF (chave primária), RG, data de nascimento, emails (atributo multivalorado, visto

que uma pessoa pode ter diversos emails), endereço (entidade fraca, descrita acima) e telefones (atributo multivalorado, visto que uma pessoa pode ter diversos números de telefone).

Banco: entidade forte que representa a instituição credora. Os atributos que referenciam esta entidade são: código do banco (chave primária), nome da instituição e CNPJ.

Crédito: entidade fraca que representa o crédito concedido por uma instituição financeira. Os atributos que referenciam esta entidade são: CPF e código do banco (chaves estrangeiras que compõem a chave primária composta), limite (valor do crédito concedido), tipo (como se categoriza o crédito concedido) e valor pago (pelo cliente, até o momento atual).

Agencia: entidade fraca que representa a agência bancária de uma instituição financeira. Os atributos que referenciam esta entidade são: código da agência e código do banco (chaves estrangeiras que compõem a chave primária composta) e telefone da agência.

Site: entidade forte que representa uma fonte de dados digital. Os atributos que referenciam esta entidade são: última visita (data), categoria a qual esta fonte de dados pertence (e-commerce, stream de vídeos, jogos), e link da fonte de dados (atributo discriminador).

Contas Associadas: entidade fraca que representa as contas, digitais, que estão associadas à uma pessoa. Os atributos que referenciam esta entidade são: nome (da pessoa à qual está associada), link (endereço digital da conta), categoria a qual a conta está associada (rede social, fórum, etc) e e-mail (chave primária).

Relacionamentos

É possível, uma vez que definidas, entender como se dá o relacionamento entre as entidades. Seguem, abaixo, os relacionamentos que estabelecidos para modelagem de banco de dados do projeto:

Concede Crédito: um Banco podem conceder crédito para várias pessoas assim como um pessoa pode ter o crédito concedido por vários bancos, o crédito deve somente existir caso o banco tenha concedido a pessoa, além de conter um atributo com a data em que o crédito foi concedido (n:n:n).

Tem Agência: um banco pode ter várias agências, porém uma agência necessariamente deve pertencer a somente um banco (n:1).

Tem Contas Associadas: um relacionamento fraco de várias contas associadas que existem somente se pertencerem a uma pessoa (n:1).

Visita: uma pessoa pode visitar vários sites, assim como um site pode ser visitado por várias pessoas (n:n).

Tem Endereço (Pessoa): uma pessoa pode ter vários endereços assim como um endereço deve pertencer a pelo menos uma pessoa (n:n).

Tem Endereço (Agência): uma agência pode ter somente um endereço, assim como um endereço pode pertencer somente a uma agência (1:1).

Modelo Entidade-Relacionamento

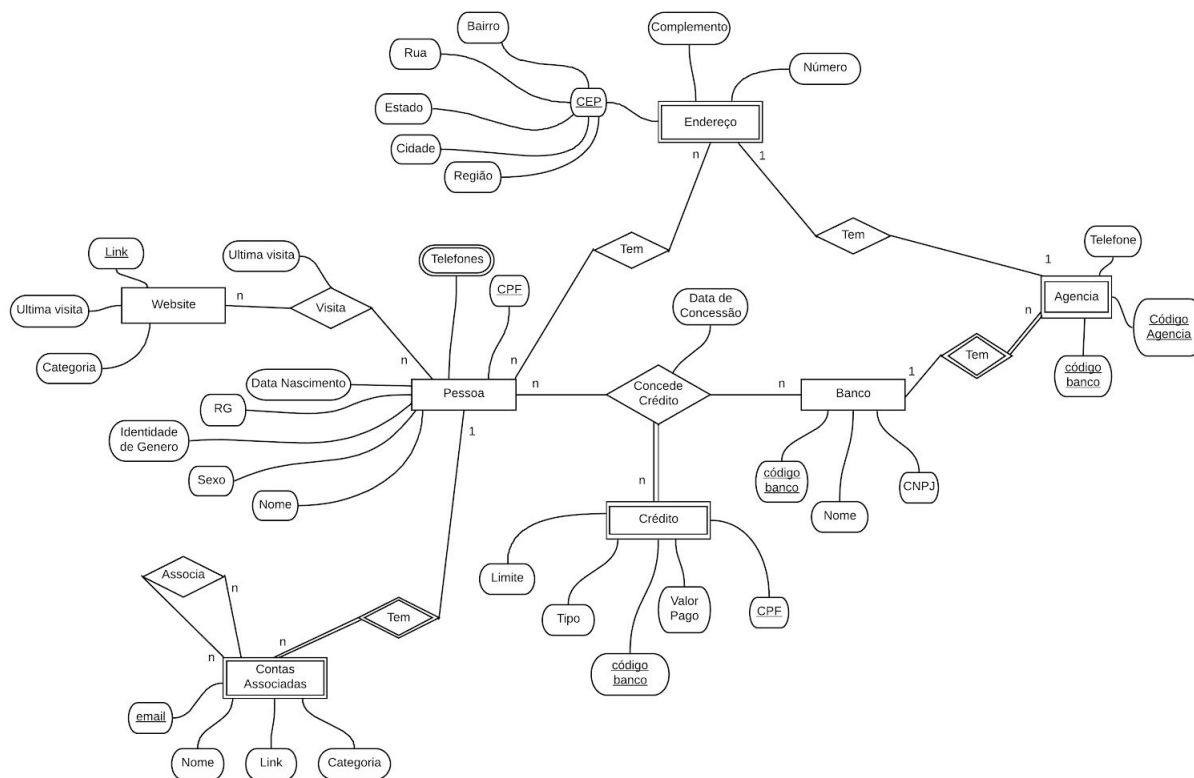


Figura 01. Modelo Entidade-Relacionamento não normalizado.

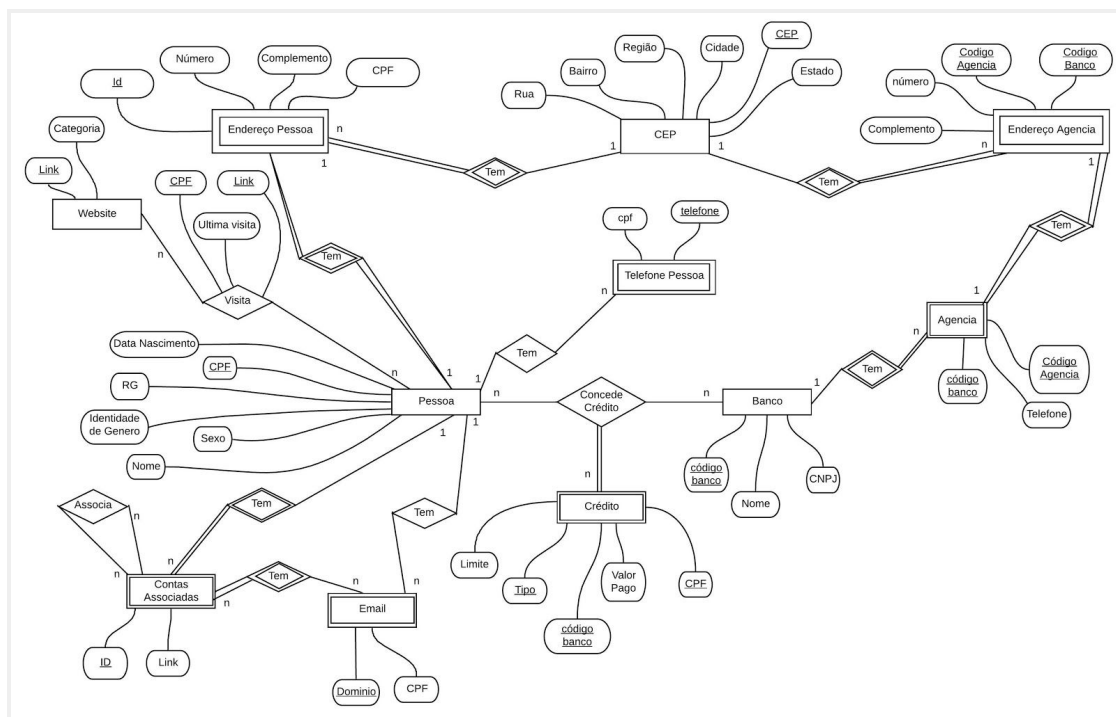


Figura 02. Modelo Entidade-Relacionamento normalizado (3FN).

Modelo Relacional Normalizado

A fim de minimizar, ou até eliminar, redundância no banco de dados, diminuir espaço de armazenamento e otimizar as atualizações das tabelas, o modelo entidade-relacionamento apresentado na **figura 02** foi normalizado para Terceira Forma Normal.

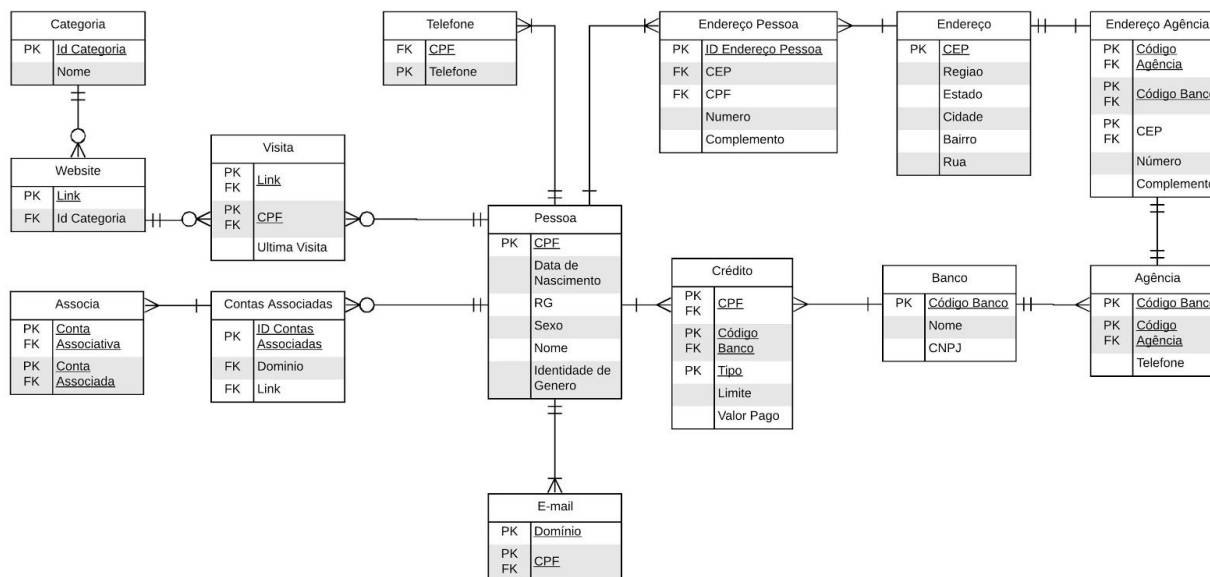


Figura 03. Modelo Relacional *normalizado (3FN)*.

Seguem, abaixo, as etapas para normalização do modelo:

Primeira Forma Normal (1FN): todos atributos do modelo são atômicos, ou indivisíveis. Os grupos repetidos, assim como atributos valorados e compostos, foram eliminados a partir da criação de uma nova entidade, cuja chave primária composta é formada pela com a chave primária do do grupo repetido e da entidade anterior.

Segunda Forma Normal (2FN): além de estar na Primeira Forma Normal, não contém dependências funcionais parciais - quando uma coluna depende apenas de parte de uma chave primária. Para isso, foram removidas das entidades todos estes atributos e novas entidades foram criadas para este subconjunto de atributos que foram removidos. Além, é preciso criar uma relação deles como chave primária.

Terceira Forma Normal (3FN): além de estar na Segunda Forma Normal, não contém dependências funcionais transitivas - quando uma coluna, além de depender da chave primária da tabela, depende também de uma ou mais outras colunas da relação (colunas não-chave não devem depender de outras colunas não-chave). Para isso, foram removidos da relação todos os atributos que são funcionalmente dependentes de outros atributos não chave.