**Projeto de Banco de Dados**

# 1. Introdução

Durante muito tempo criou-se a ideia de que projetar bancos de dados era uma disciplina com identidade própria, uma atividade específica e até certo ponto isolada no ciclo de vida de um sistema, que tinha existência própria e podia ser realizada a partir de primitivas e conceitos exclusivos da técnica de modelagem de dados.

Com a evolução das tecnologias de desenvolvimento de aplicações, novas linguagens e principalmente com o advento da orientação a objetos não mais restrita ao ambiente de programação, mas extensiva às metodologias de análise de sistemas, o trabalho de projetar as bases de dados que serão utilizadas por um sistema em desenvolvimento assume, nos dias de hoje, características que objetivam mixar um projeto orientado a objetos com as necessidades de esse mesmo sistema interagir com um banco de dados relacional, constituído por um conjunto de tabelas, que equivale à denominada camada de persistência de dados.

Essa necessidade de mixagem é real pela absoluta ausência de projetos comerciais que utilizem bancos de dados orientados a objetos que sejam confiáveis a grandes massas de dados, à não popularização desses produtos e aos grandes investimentos já realizados em softwares de Sistemas Gerenciadores de Bancos de Dados Relacionais existentes no mercado nacional e mundial.

É indiscutível a vantagem obtida em um projeto orientado a objetos; logo, surge a necessidade de uma nova técnica de projetar banco de dados, que não é a formatação pura de classes de dados, mas uma interação alta entre o ambiente de análise orientada a objetos e a nossa sempre presente modelagem de dados, que é estritamente necessária à administração de dados da organização de TI. A utilização de ferramentas para a camada de persistência, tais como o Hybernate, acaba levando o desenvolvedor a deixar de lado a preocupação com as estruturas do banco de dados, assim como a preocupação com a qualidade das queries realizadas em uma aplicação.

Não existem técnicas nem ferramentas que possibilitem tanto ao administrador de dados (DA) quanto ao administrador de bancos de dados (DBA) realizar suas funções sobre classes de dados, pois esses mesmos bancos de dados relacionais atuam e têm todas as suas funcionalidades sobre tabelas relacionais de dados, as quais são hoje de domínio maior dos usuários de aplicações.

Há muito tempo escrevemos sobre modelagem de dados e o processo continua existindo como sempre existiu, porém o que desejamos neste livro é apresentar essas técnicas integradas com a análise orientada a objetos, de forma a permitir que quem trabalha tanto com OO como com análise estruturada tenha domínio de projetos de bancos de dados eficientes a uma aplicação, seja qual for o ambiente de desenvolvimento em que esteja. Busca-se ainda permitir que um projeto totalmente desenvolvido em OO seja facilmente inserido em um ambiente de banco de dados relacional, com o mínimo de traumas e mantida a coerência com os objetos de negócios de uma organização.

# 2. O Projeto de Banco de Dados

Todo projeto de sistema de banco de dados necessita de um foco, um centro nervoso. A modelagem de um banco de dados por meio da abordagem Entidade-Relacionamento representa esse ponto central no projeto conceitual de um sistema.

O objetivo da modelagem de dados é transmitir e apresentar uma representação única, não redundante e resumida, dos dados de uma aplicação. Em projetos conceituais de aplicações em banco de dados o modelo Entidade-Relacionamento é o mais largamente utilizado para representação e entendimento dos dados que compõem a essência de um sistema.

O projeto de banco de dados para sistemas de aplicação, hoje em dia, não é mais uma tarefa a ser realizada somente por profissionais especializados em bancos de dados da área de desenvolvimento de sistemas, mas também por analistas de sistemas independentemente de experiência anterior em um determinado produto de banco de dados, pela utilização de técnicas estruturadas como a modelagem conceitual de dados Entidade-Relacionamento.

O projeto de um sistema de informações é atividade complexa que inclui planejamento, especificações e desenvolvimento de vários componentes.

O que se propõe é situar a sequência dessas atividades em uma ordem que possa resultar ganhos de produtividade e confiabilidade dos sistemas desenvolvidos, eliminando sensivelmente as falhas de sistema após sua implantação.

Desafortunadamente as metodologias de projeto de banco de dados, para sistemas de aplicação, apesar de já serem muito populares entre a comunidade técnica, não são utilizadas corretamente.

Em várias organizações os profissionais utilizam-se de pequenas técnicas pessoais, ou ainda pior, de uma inexistência completa de metodologia para esses projetos e com distorções na utilização das técnicas, sendo esta uma das maiores causas de falhas nos sistemas de informação desenvolvidos, mesmo com a existência de modelos de dados.

A utilização de abordagem correta de metodologia orientada a modelagem de banco de dados envolve a estruturação nos três níveis de visão de dados vistos anteriormente, ou seja, três etapas na execução de um projeto: conceitual, lógico e físico.

Isola-se desta forma a realidade a ser retratada em dados de suas implicações lógicas e físicas, determinando o momento para adequação do modelo construído à ponderação desses fatores.

É evidente e óbvio que a realidade dos negócios de uma empresa é simplesmente diferente da realidade de outra empresa, mesmo que os ambientes sejam similares. Existem particularidades que só dizem respeito ao funcionamento de um ambiente específico.

Devido a essa não similaridade completa entre ambientes de mesma natureza, será sempre necessária a criação de um modelo específico para cada nova realidade observada.

Fica bem claro, então, a necessidade de ter uma metodologia estruturada que permita construir vários modelos, resultando o chamado metamodelo, ou seja, uma técnica específica de ampla abrangência e flexibilidade.

O metamodelo a ser utilizado deve ter a característica de poder modelar qualquer realidade, ter uma forma de trabalho bastante simples e características gráficas que sejam bem simples de construir e entender.

O nosso metamodelo é o modelo Entidade-Relacionamento (ER).

Como já destacamos, o modelo de dados é um plano para construir o banco de dados. Para ser efetivo, deve ser simples o bastante para comunicar ao usuário final a estrutura de dados requerida pelo banco de dados e detalhada o suficiente para criar a estrutura física de banco de dados.

O modelo Entidade-Relacionamento (ER) é o método mais comum para a construção de modelos de dados para bancos de dados relacionais.

# 3. Modelagem de Dados

Modelagem de dados é o estudo das informações existentes em um contexto sob observação para a construção de um modelo de representação e entendimento de tal contexto. A modelagem de dados minera as informações que representam um contexto, estruturando-as em um conjunto que denominamos modelo lógico de dados.

Uma das principais características da abordagem de modelagem de banco de dados é que ela fornece níveis de abstração de dados que omitem do usuário final detalhes sobre o armazenamento dos dados. Não existe a preocupação com um banco de dados tecnologicamente falando.

O modelo de dados é um conjunto de conceitos que podem ser utilizados para descrever as estruturas lógicas e físicas de um banco de dados.

Já um modelo de classes de objetos não se detém somente às informações e dados, sendo mais amplo no momento em que integra as informações e os seus métodos de acesso, recuperação, processamento e outros em um único objeto.

Logo, temos de encontrar e separar os dados dos seus processos em um determinado momento, para que possamos efetivamente obter domínio do negócio para o qual a aplicação está sendo desenvolvida. Os objetivos reais da aplicação continuam sendo mais facilmente compreendidos por meio de um modelo de dados, e não de um modelo de objetos.

Normalmente as aplicações têm sido desenvolvidas com AOO (Análise Orientada a Objetos) mesclada com a tradicional modelagem de dados, com a utilização de paterns, ou melhor, padrões de modelos de dados, que não são obtidos tão simplesmente quando da pura criação de diagramas de classes. Desenvolvemos com isso um processo que pode ser chamado two fase commit entre diagrama de classes e modelo de dados Entidade-Relacionamento.

Para que isso possa ser realizado, é preciso entender e dominar as técnicas de modelagem de dados e depois, finalmente, aprender a lidar com a sua equivalência em um modelo de classes de objetos.

O objetivo deste curso é explorar e ensinar da melhor forma possível como modelar dados para ambientes de bancos de dados relacionais, complementando com a implementação desses bancos seja em um ambiente tradicional, seja em um ambiente orientado a objetos, tal como uma aplicação desenvolvida em Java, com base em experiências por nós realizadas.

Estrutura é a forma como esses dados estão agrupados, os relacionamentos entre eles e as restrições que podem recair sobre eles, o que não é totalmente explícito em um diagrama de classes da análise orientada a objetos.

Historicamente, os primeiros modelos de bancos de dados datam da década de 1960. Desde então, a pesquisa científica na área procura evoluir no sentido de definir, encontrar modelos que representem da melhor maneira possível os dados de uma realidade, ou seja, que organizem os dados de uma forma mais próxima da maneira como são vistos e manipulados pelas pessoas no mundo real.

A ideia é definir abstrações que facilitem a compreensão da organização dos dados pelo usuário, tornando cada vez mais transparente e independente o conhecimento da organização física, independente de precisar utilizar o conhecimento técnico de um conjunto de técnicas orientadas a objetos para ter esse entendimento.

O objetivo de um modelo de dados é ter certeza de que todos os objetos de dados existentes em determinado contexto e requeridos pela aplicação e pelo banco de dados estão completamente representados e com precisão. Pelo fato de os dados modelados usarem notações facilmente compreendidas e em um idioma natural, eles podem ser revisados e verificados pelos usuários finais.

O modelo de dados também deve ser detalhado o bastante para ser usado pelo implementador (DBA) do banco de dados como uma espécie de fotocópia para construir o banco de dados físico.

Será utilizada toda a informação que está no modelo de dados lógico para definir as tabelas de um banco de dados relacional, chaves primárias e estrangeiras, procedimentos armazenados (stored procedures) e gatilhos (triggers).

Um banco de dados mal projetado requer mais tempo e retrabalho a longo prazo.

Sem planejamento e análise cuidadosa você pode criar um banco de dados que omita alguns dados exigidos, ou inconsistente em relação ao contexto de informações que ele deve refletir. Os resultados incorretos ou incompatíveis em uma aplicação sistêmica impossibilitam inclusive acomodar as mudanças de exigências dos usuários ao longo do tempo, ou implicam que o banco de dados tenha constante manutenção em suas estruturas e a aplicação fique extremamente dependente do próprio banco de dados.

Para ser efetivo, um modelo de dados deve ser simples o bastante para comunicar ao usuário final a estrutura de dados requerida pelo banco de dados e bastante detalhado para o administrador de banco de dados usar para criação da estrutura física correspondente em um SGBD.

O modelo de Entidade-Relacionamento (ER) é o método mais comum para construção de modelos de dados para bancos de dados de relacionais.

O mais comum em um modelo de dados é uma pequena coleção de mecanismos de abstração ditos primitivos, que são classificação, agregação e generalização.

Abstrações ajudam o analista a entender, classificar e modelar uma realidade, alguma coisa do mundo real que está em observação.

# 4. O que é Abstração?

Uma abstração ou a capacidade de abstração é um processo, com certeza mental, que usamos quando selecionamos várias características e propriedades de um conjunto de objetos ou fatos, e excluímos outras que não são relevantes em um contexto. Em outras palavras, aplicamos abstração quando nos concentramos nas propriedades de um conjunto de objetos ou coisas que consideramos essenciais, e desprezamos outras que não consideramos importantes, sempre dentro de um determinado contexto sob observação ou análise.

O analista, durante a modelagem conceitual dos dados, deve se concentrar na obser- vação dos objetos relevantes que existem em uma realidade ou contexto, com a finalidade de construir um modelo de compreensão e conceitos existentes nessa realidade. Esse primeiro modelo denominamos de minimundo, sem pensar no momento em automatizar ou processar a informação dessa realidade.

Abstração, em síntese, nada mais é do que a visão, sem conceituações técnicas, que obtemos na mente de uma realidade qualquer do mundo real.

Por exemplo, quando olhamos para uma bicicleta, temos o resultado de um processo de abstração em que excluímos alguns detalhes, como os pedais, os freios, os mecanismos de tração e inclusive as possíveis diferenças entre várias bicicletas, e definimos o objeto como o que conceituamos e denominamos de "bicicleta". Normalmente associamos um nome a cada abstração. No caso da figura, o conceito bicicleta é uma representação dessa prática de abstração. Definimos um conceito para um objeto existente em contexto sob observação. Denominamos essa capacidade humana de modelo conceitual.

Em realidade o que criamos é um modelo de conceitos de objetos em um primeiro instante que denominamos de modelo conceitual.

Ao coletar e selecionar os objetos relevantes de um contexto, o analista deve identificar os elementos geradores de cada objeto, as leis que os regem no contexto, bem como as operações que incidem sobre os objetos básicos (dados).

Nesse momento, os documentos que registram esses objetos ou elementos de negócio só devem ser utilizados como apoio ao entendimento, e não como base para o desenvolvimento do sistema de informações, ou seja, não devemos ter a preocupação de simular o ambiente atual, seja ele manual ou automatizado.

A preocupação que o analista deve ter é retratar as necessidades de informação que as pessoas (que agem sobre um contexto) precisam para alcançar os objetivos em seu ambiente de negócios.

Para registrarmos as necessidades de informação de uma realidade e sempre dentro do contexto desta realidade, precisamos fazer uso de um modelo de dados, ou seja, algo que nos mostre as informações existentes e como elas estão relacionadas (regras de negócio).

Com base nesse modelo criado, os analistas podem interagir com os usuários validando seus objetivos e metas, permitindo o detalhamento do que denominamos base de dados para construção de um sistema de informações aderente às necessidades dos usuários finais.

Em seguida apresentamos uma descrição dos principais macroelementos construídos com a aplicação de técnicas de abstração.

## 4.1. Minimundo

Porção da realidade, captada pelo analista, que a gestão de negócios de uma organização tem interesse em observar, controlar e administrar. A complexidade existente no momento de analisar um minimundo pode levar o analista a subdividi-lo em partes menores, às quais damos o nome de visão de processo de negócio.

## 4.2. Banco de Dados

Um banco de dados pode ser definido como um conjunto de dados devidamente relacionados. Podemos compreender como dados os objetos conhecidos que podem ser armazenados e que possuem um significado implícito, porém o significado do termo banco de dados é mais restrito que simplesmente a definição dada anteriormente. Um banco de dados possui as seguintes propriedades:

É uma coleção lógica coerente de dados com um significado inerente; uma disposição desordenada dos dados não pode ser referenciada como banco de dados.

Ele é projetado, construído e preenchido com valores de dados para um propósito específico; um banco de dados possui um conjunto predefinido de usuários e de aplicações.

Ele representa algum aspecto do mundo real, o qual é chamado de minimundo; qualquer alteração efetuada no minimundo é automaticamente refletida no banco de dados.

## 4.3. Modelo Conceitual

Representa, descreve a realidade do ambiente do problema, constituindo-se em uma visão global dos principais dados e seus relacionamentos (estruturas de informação), completamente independente dos aspectos de sua implementação tecnológica.

Quando falamos em modelo conceitual, estamos nos referindo àquela que deve ser sempre a primeira etapa de um projeto de banco de dados.

O objetivo do modelo conceitual é descrever de forma simples e facilmente compreendida pelo usuário final as informações de um contexto de negócios, as quais devem ser armazenadas em um banco de dados.

É uma descrição de alto nível (macrodefinição), mas que tem a preocupação de captar e retratar toda a realidade de uma organização, processo de negócio, setor, repartição, depar-tamento etc.

É importante destacar que o modelo conceitual não retrata nem é vinculado aos aspectos ligados à abordagem do banco de dados que será utilizado, tampouco se preocupa com as formas de acesso ou estruturas físicas implementadas por um Sistema Gerenciador de Banco de Dados específico. Sem modelo conceitual não temos uma visão clara das regras do negócio e acabamos por criar aplicações sem entender para que elas foram criadas.

O resultado de um modelo conceitual é um esquema gráfico que representa a realidade das informações existentes em determinado contexto de negócios, assim como as estruturas de dados em que estão organizadas essas informações.

O modelo conceitual nunca deve ser construído com considerações sobre processos de negócio, com preocupações de acesso aos dados, não devendo existir nenhuma preocupação com o modo como serão realizadas as operações de consulta e manutenção dos dados nele apresentados. O foco deve ser dirigido sempre ao entendimento e à representação de uma realidade, de um contexto.

## 4.4. Modelo Lógico

Ele somente tem início após a criação do modelo conceitual, pois agora vamos considerar uma das abordagens possíveis da tecnologia Sistemas Gerenciadores de Banco de Dados (relacional, hierárquica, rede ou orientada a objetos) para estruturação e estabelecimento da lógica dos relacionamentos existentes entre os dados definidos no modelo conceitual.

A elaboração direta e um modelo lógico de dados, independentemente de já sabermos a abordagem para banco de dados, para a qual estamos realizando um projeto, levam à vinculação tecnológica de nosso raciocínio, perturbando a interpretação pura e simples de um contexto. Sempre que analisamos um contexto sob a óptica tecnológica, temos a tendência de sermos técnicos demais, distorcer a realidade, conduzindo-a às restrições da tecnologia empregada, o que sempre, e já estatisticamente comprovado, leva a erros de interpretação da realidade, criando assim modelos de dados que não possuem aderência ao minimundo descrito.

O modelo lógico descreve em formato as estruturas que estarão no banco de dados de acordo com as possibilidades permitidas pela sua abordagem, mas sem considerar, ainda, nenhuma característica específica de um Sistema Gerenciador de Banco de Dados (SGBD). Isso resulta um esquema lógico de dados sob a óptica de uma das abordagens citadas, pelo emprego de uma técnica de modelagem de dados orientada às restrições de cada abordagem.

## 4.5. Modelo Físico

O modelo físico será construído a partir do modelo lógico e descreve as estruturas físicas de armazenamento de dados, tais como:

Tipo e tamanho de campos;

Índices;

Domínio de preenchimento desses campos;

Nomenclaturas;

Exigência de conteúdo;

Gatilhos etc.

Elas são projetadas de acordo com os requisitos de processamento e uso mais econômico dos recursos computacionais. Esse modelo detalha o estudo dos métodos de acesso do SGBD para criação dos índices necessários para cada informação colocada nos modelos conceitual e lógico.

Esta é a etapa final do projeto de banco de dados, na qual será utilizada a linguagem de definição de dados do SGBD (DDL) para a realização da sua montagem no dicionário de dados. Em ambiente de banco de dados relacional denominamos script de criação do banco de dados o conjunto de comandos em SQL (DDL) que será executado no Sistema Gerenciador de Banco de Dados para a criação de banco de dados correspondente ao modelo físico.

# 5. O Modelo Entidade-Relacionamento (ER)

Ele foi proposto originalmente por Peter, em 1976 (CHEN, 1976), como um modo de unificar as visões de um banco de dados relacional, e teve como base a teoria relacional criada por E. F. Codd (1970).

Simplesmente declarado como modelo de ER, ele é conceitual e vê o mundo real como entidades e relacionamentos. Um componente básico do modelo é o diagrama de Entidade-Relacionamento, que visualmente representa objetos de dados.

Ao longo dos anos, vários estudiosos (Theorey, Fry, James Martin, entre outros) evoluíram e expandiram esse metamodelo.

Segundo Chen, a visão de uma determinada realidade baseia-se no relacionamento entre as entidades, as quais retratam os fatos que governam essa mesma realidade, e cada um (entidade ou relacionamento) pode possuir atributos (qualificadores descritivos dessa realidade).

Grande parte das extensões ao metamodelo baseia-se em alguns mecanismos de abstração: classificação, generalização e agregação.

O conceito de abstração, como já destacamos, permite ao analista separar da realidade em estudo as partes que são realmente relevantes para o desenvolvimento do sistema de informações e excluir da modelagem todos os aspectos que não exercem influência sobre o ambiente a ser modelado.

Os conceitos do modelo ER destinam-se prioritariamente ao projeto de banco de dados, entretanto eles podem ser utilizados para o entendimento de um determinado negócio, por exemplo, na modelagem de processos de negócios (modelo do negócio), bem como auxiliar o desenvolvimento de estruturas de dados que possam ser implementadas fora de um ambiente de banco de dados, utilizando-se uma linguagem de programação (Cobol, C, Pascal etc.).

O modelo ER é ainda a base para o desenvolvimento de sistemas orientados a objetos.

Somente o domínio das técnicas de modelagem não é suficiente para produzir bons modelos. O essencial na modelagem de dados é o total entendimento dos conceitos pertencentes à realidade em questão.

Ao modelar dados, ou qualquer outra coisa, o objeto observado, o ambiente sob análise será sempre o ponto de partida. Ele serve como referência para a construção de todo o modelo. Por este motivo a fase de levantamento de dados, investigação e análise dos dados propriamente dita é tão importante.

O processo de modelagem consiste em cinco aspectos importantes:

Observação: entrevistas, reuniões, questionários, análise de documentos aliados ao conhecimento e experiência prévios da área de negócios ou seu perfeito entendimento e compreensão.

Entendimento dos conceitos: núcleo do processo de modelagem. Essa fase destina-se a identificar, conceituar, entender e assimilar o objeto observado.

Representação dos objetos: aplicação de técnicas de modelagem de dados Entidade-Relacionamento.

Verificação de fidelidade e carências: detectar falhas e anomalias, identificando respectivas causas que podem residir em conceitos malformados, pontos de vista equivocados, falha na concepção ou aplicação errada da técnica de representação.

Validações: nessa fase busca-se a aprovação formal do modelo. Para que esse objetivo seja alcançado, é necessária a participação ativa do usuário final, bem como a visão de outros profissionais da área técnica (administrador de dados, analista de sistemas, administrador de banco de dados etc., de acordo com o caso). Esse processo deve ser rigoroso e crítico, tanto quanto possível.

As técnicas de modelagem de dados devem ser disseminadas à área usuária para que a validação não se transforme em um monólogo de analistas de sistemas e deixe de cumprir o seu verdadeiro objetivo (detectar falhas e anomalias), porém nunca tente fazer do usuário final um expert no assunto modelagem de dados.

# 6. A Implementação de Banco de Dados

Em síntese é a criação efetiva em um Sistema Gerenciador de Banco de Dados de uma estrutura de dados correspondente ao modelo físico derivado do modelo construído por meio das técnicas de modelagem Entidade-Relacionamento.

Muitas vezes ouvimos afirmações como: projeta-se o modelo Entidade-Relacionamento, mas o que se implementa é um conjunto de tabelas que muitas vezes não têm nada a ver com o modelo projetado.

Em suma isso até ocorre, mas constitui um ERRO da administração de dados em conflito com a administração de banco de dados, pois o modelo de tabelas, por assim dizer, implementado deve obrigatoriamente corresponder ao modelo lógico derivado do modelo conceitual projetado.

Em empresas são muitas as situações cujo modelo de dados é desprezado em decorrência de os projetistas não possuírem abrangência de aspectos físicos e operacionais de um Sistema Gerenciador de Banco de Dados no tocante à forma de recuperação dos dados nele inseridos, quanto à utilização de índices ou à disponibilidade de área de armazenamento e visão da população que vai ter um banco de dados. Isso faz com que o projetista de BD seja meramente conceitual, saltando-se desta forma o modelo lógico ou o conceitual, tratando o processo de modelagem em somente dois níveis:

Modelo conceitual e físico, ou Modelo lógico e físico.

Isso provoca buracos de análise bem sensíveis que conduzem a uma implementação equivocada e de alto índice de manutenção das aplicações e do banco de dados.

Sejam quais forem os processos utilizados, completos ou parciais, devemos validar esse modelo antes de implementá-lo efetivamente, pela aplicação de conceitos e comandos de álgebra relacional na busca de recuperação das informações que nele estarão.

É fundamental para o projetista de banco de dados que ele possua capacitação na navegação do banco de dados proposto e implementado e, principalmente, compreensão do modelo, conhecendo seus caminhos de navegação. Apresentamos a álgebra relacional em capítulo específico adiante.

Quando falamos de álgebra relacional, destacamos o fator de importância que ela assume em relação ao projeto de banco de dados, destacando a seguinte afirmativa:

* De que adianta projetar banco de dados se não sei como recuperar suas informações?
* De que adianta saber comandos SQL se não sei como é a estrutura interna de um comando para recuperar dados de um conjunto de tabelas?

Foi comprovado, ao longo do tempo, que o domínio e o conhecimento de álgebra relacional dotam o analista de sistemas da capacidade de resolver qualquer tipo de consulta necessária à navegação e recuperação de situações, da mais simples à mais complexa exigida em uma aplicação de banco de dados.