

**UNIVERSIDADE DE ITAÚNA**

DAVI VENTURA CARDOSO PERDIGÃO  
EDMILSON LINO CORDEIRO  
ERIC HENRIQUE DE CASTRO CHAVES

**DATA MODELING (MODELAGEM DE DADOS)**

Trabalho Prático 2 - Ciência dos Dados

**ITAÚNA  
2022**

# SUMÁRIO

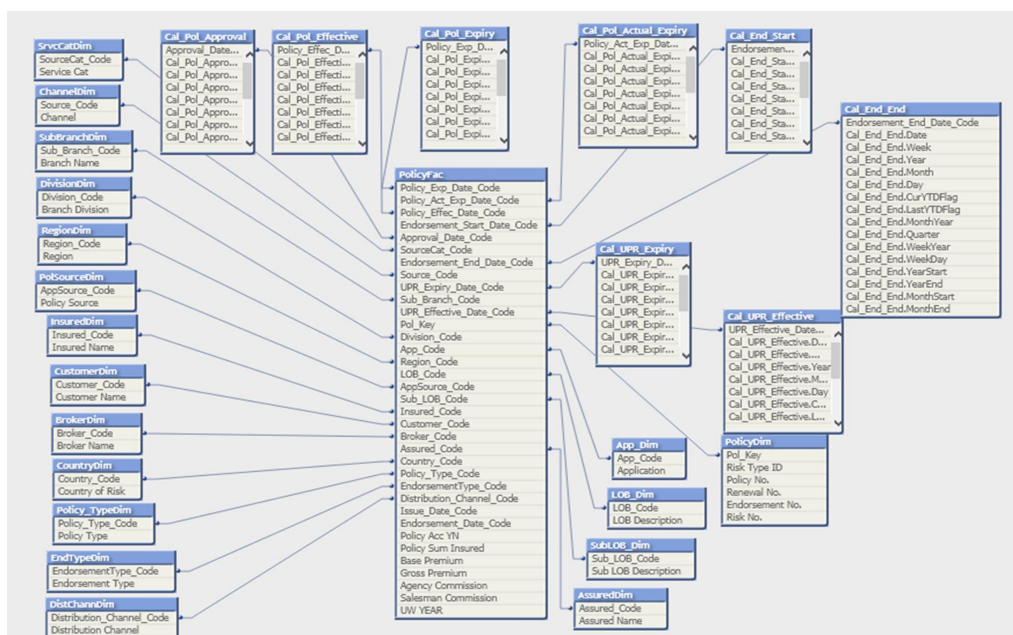
1. INTRODUÇÃO.....	2
2. POR QUE A MODELAGEM DE DADOS É FEITA?.....	3
3. PROCESSO DE MODELAGEM DE DADOS.....	4
4. TIPOS DE MODELAGEM DE DADOS.....	5, 6
5. TÉCNICAS DE MODELAGEM DE DADOS.....	6, 7, 8, 9 ,10
6. BENEFÍCIOS E DESAFIOS.....	10, 11, 12, 13, 14
7. CONCLUSÃO.....	14
8. REFERÊNCIAS BIBLIOGRÁFICAS.....	15

# 1. INTRODUÇÃO

A modelagem de dados é o processo de criação de um diagrama simplificado de um sistema de software e dos elementos de dados que ele contém, usando texto e símbolos para representar os dados e como eles fluem. Os modelos de dados fornecem um plano para projetar um novo banco de dados ou reengenharia de um aplicativo legado. Em geral, a modelagem de dados propõe uma organização a usar seus dados de forma eficaz para atender às necessidades de informações do negócio.

Os modelos de dados são construídos em torno das necessidades de negócios. As regras e os requisitos são definidos antecipadamente por meio do feedback das partes interessadas do negócio para que possam ser incorporados ao projeto de um novo sistema ou adaptados na iteração de um existente. Alternativamente, os modelos de dados podem ser criados por meio de esforços de engenharia reversa que os extraem de sistemas existentes.

Idealmente, os modelos de dados são documentos vivos que evoluem junto com as mudanças nas necessidades de negócios. Eles desempenham um papel importante no suporte aos processos de negócios e no planejamento da arquitetura e estratégia de TI. Os modelos de dados podem ser compartilhados com fornecedores, parceiros e/ou colegas do setor.



Fonte: Fórum KNIME (2020)

## 2. POR QUE A MODELAGEM DE DADOS É FEITA?

A modelagem de dados é uma disciplina central de gerenciamento de dados. Ao fornecer uma representação visual de conjuntos de dados e seu contexto de negócios, ele ajuda a identificar as necessidades de informações para diferentes processos de negócios. Em seguida, especifica as características dos elementos de dados que serão incluídos nos aplicativos e nas estruturas de banco de dados ou sistema de arquivos usados para processar, armazenar e gerenciar os dados.

A modelagem de dados também pode ajudar a estabelecer definições de dados comuns e padrões de dados internos, geralmente relacionados a programas de governança de dados . Além disso, desempenha um papel importante nos processos de arquitetura de dados que documentam ativos de dados, mapeiam como os dados se movem pelos sistemas de TI e criam uma estrutura conceitual de gerenciamento de dados. Os modelos de dados são um componente chave da arquitetura de dados, juntamente com diagramas de fluxo de dados , projetos arquitetônicos, um vocabulário de dados unificado e outros artefatos.

Tradicionalmente, os modelos de dados são construídos por modeladores de dados, arquitetos de dados e outros profissionais de gerenciamento de dados com informações de analistas de negócios, executivos e usuários. Mas a modelagem de dados também é agora uma habilidade importante para cientistas e analistas de dados envolvidos no desenvolvimento de aplicativos de inteligência de negócios e ciência de dados mais complexos e análises avançadas .

Ao modelar seus dados, você documenta quais dados você tem, como os usa e quais são seus requisitos em relação ao uso, proteção e governança. Através da modelagem de dados, sua organização:

- Cria uma estrutura para colaboração entre sua equipe de TI e suas equipes de negócios.
- Expõe oportunidades para melhorar os processos de negócios definindo as necessidades e os usos dos dados.
- Economiza tempo e dinheiro em investimentos em TI e processos por meio de um planejamento adequado com antecedência.
- Reduz erros (e entrada de dados redundante propensa a erros), enquanto melhora a integridade dos dados.
- Aumenta a velocidade e o desempenho da recuperação e análise de dados ao planejar a capacidade e o crescimento.

Portanto, não é apenas o que você obtém com a modelagem de dados, mas também como você a obtém. O próprio processo oferece benefícios significativos.

### 3. PROCESSO DE MODELAGEM DE DADOS

Como disciplina, a modelagem de dados convida as partes interessadas a avaliar o processamento e o armazenamento de dados em detalhes minuciosos. As técnicas de modelagem de dados têm convenções diferentes que determinam quais símbolos são usados para representar os dados, como os modelos são dispostos e como os requisitos de negócios são transmitidos. Todas as abordagens fornecem fluxos de trabalho formalizados que incluem uma sequência de tarefas a serem executadas de maneira iterativa.

Peter Aiken, consultor de gerenciamento de dados e professor associado de sistemas de informação da Virginia Commonwealth University, listou as seis etapas a seguir para projetar um modelo de dados durante um webinar Dataversity de 2019:

- Identifique as entidades de negócios que são representadas no conjunto de dados.
- Identifique as principais propriedades de cada entidade para diferenciá-las.
- Crie um modelo de rascunho de entidade-relacionamento para mostrar como as entidades estão conectadas.
- Identifique os atributos de dados que precisam ser incorporados ao modelo.
- Mapeie os atributos para entidades para ilustrar o significado comercial dos dados.
- Finalize o modelo de dados e valide sua precisão.

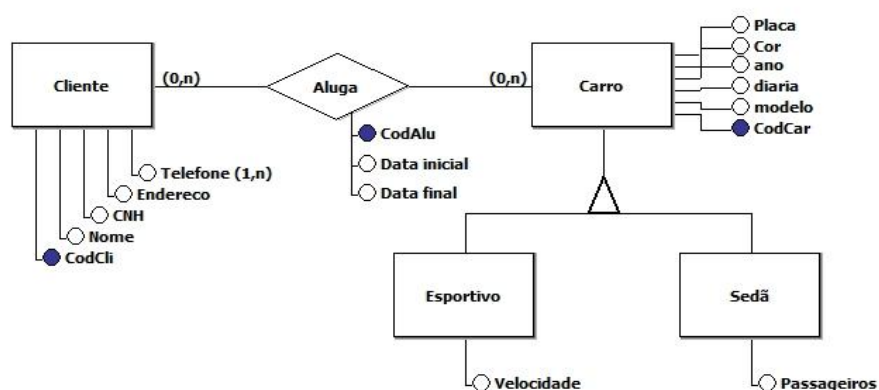
Mesmo depois disso, o processo geralmente não é concluído: os modelos de dados geralmente precisam ser atualizados e revisados à medida que os ativos de dados de uma organização e as necessidades de negócios mudam.

## 4. TIPOS DE MODELAGEM DE DADOS

A modelagem de dados pode ser dividida em três categorias: **Modelo Conceitual**, o **Modelo Lógico** e o **Modelo Físico** de Dados e que variam de acordo com o seu grau de abstração. O processo começa com um modelo conceitual, progride para um modelo lógico e termina com um modelo físico.

**Modelo Conceitual de Dados:** oferece uma grande visão do que o sistema irá conter, como será organizado, e que regras de negócio estão envolvidas. São geralmente criados como parte do processo de recolha dos requisitos iniciais do projeto. Tipicamente, incluem classes de entidades (definindo os tipos de coisas que são importantes para a empresa representar no modelo de dados), as suas características e restrições, as relações entre elas e os requisitos relevantes de segurança e integridade dos dados. Porém, nesse nível, ainda não existem detalhes relacionados à implementação.

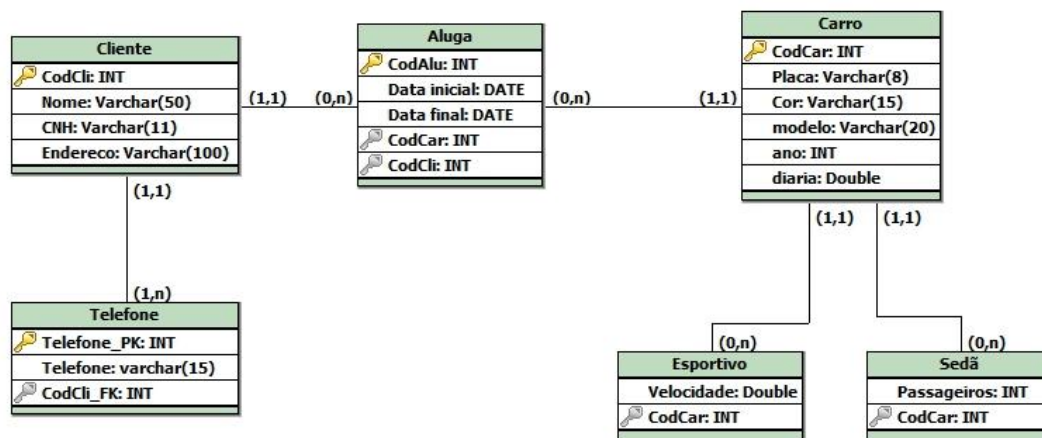
**Modelo Lógico de Dados:** é menos abstrato que o modelo conceitual e fornece detalhes sobre os conceitos e relações no domínio em consideração. Indicam atributos de dados, tais como tipos de dados e comprimentos correspondentes, e mostram as relações entre entidades, porém não especificam quaisquer requisitos técnicos do sistema. A partir deste ponto, o banco possui conceitos que o usuário, mesmo leigo, é capaz de compreender e, apesar de não necessitar de um SGDB, está muito próximo de um modelo físico.



Fonte: iStack (2020)

**Modelo Físico de Dados:** Fornecem um esquema de como os dados serão fisicamente armazenados dentro de uma base de dados. Como tal, eles são os menos abstratos de todos. Oferecem um desenho finalizado que pode ser implementado como uma base de dados relacional, incluindo tabelas associativas que ilustram as relações entre entidades, bem como as chaves primárias e as chaves estrangeiras que serão utilizadas para manter essas relações. Os modelos

de dados físicos podem incluir propriedades específicas do sistema de gestão de bases de dados (SGBD), incluindo a afinação do desempenho.



Fonte: iStack(2020)

## 5. TÉCNICAS DE MODELAGEM DE DADOS

A modelagem de dados surgiu na década de 1960, quando os bancos de dados se tornaram mais amplamente utilizados em mainframes e depois em minicomputadores. Permitiu que as organizações trouxessem consistência, repetibilidade e desenvolvimento disciplinado ao processamento e gerenciamento de dados. Ainda é assim, mas as técnicas usadas para criar modelos de dados evoluíram junto com o desenvolvimento de novos tipos de bancos de dados e sistemas computacionais. Essas são algumas das abordagens de modelagem de dados mais usadas ao longo dos anos:

### 1. Modelagem de dados hierárquicos

Os modelos de dados hierárquicos organizam os dados em um arranjo em forma de árvore de registros pai e filho. Um registro filho pode ter apenas um pai, tornando este um método de modelagem um para muitos. A abordagem hierárquica originada em bancos de dados de mainframe – o Information Management System (IMS) da IBM é o exemplo mais conhecido. Embora os modelos de dados hierárquicos tenham sido substituídos principalmente pelos relacionais a partir da década de 1980, o IMS ainda está disponível e é usado por muitas organizações. Um método hierárquico semelhante também é usado hoje em XML, formalmente conhecido como Extensible Markup Language.

## **2. Modelagem de dados de rede**

Essa também era uma opção popular de modelagem de dados em bancos de dados de mainframe que não é tão usada agora. Os modelos de dados de rede expandiram os hierárquicos, permitindo que os registros filho sejam conectados a vários registros pai. A Conferência sobre Linguagens de Sistemas de Dados, um grupo de padrões técnicos agora extinto comumente chamado de CODASYL, adotou uma especificação de modelo de dados de rede em 1969. Por causa disso, a técnica de rede é frequentemente chamada de modelo CODASYL.

## **3. Modelagem de dados relacionais**

O modelo de dados relacional foi criado como uma alternativa mais flexível aos modelos hierárquicos e de rede. Descrito pela primeira vez em um artigo técnico de 1970 do pesquisador da IBM Edgar F. Codd, o modelo relacional mapeia os relacionamentos entre elementos de dados armazenados em diferentes tabelas que contêm conjuntos de linhas e colunas. A modelagem relacional preparou o terreno para o desenvolvimento de bancos de dados relacionais, e seu uso generalizado tornou a técnica de modelagem de dados dominante em meados da década de 1990.

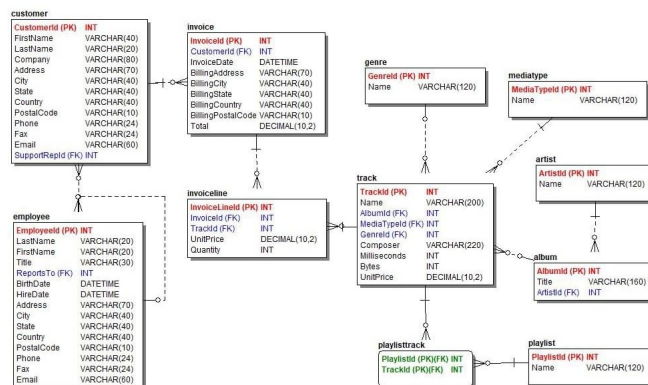
## **4. Modelagem de dados de entidade-relacionamento**

Uma variação do modelo relacional que também pode ser utilizado com outros tipos de banco de dados, os modelos entidade-relacionamento (ER) mapeiam visualmente as entidades, seus atributos e os relacionamentos entre as diferentes entidades. Por exemplo, os atributos de uma entidade de dados de funcionários podem incluir sobrenome, nome, anos de serviço e outros dados relevantes. Os modelos ER fornecem uma abordagem eficiente para captura de dados e processos de atualização, tornando-os particularmente adequados para aplicativos de processamento de transações.

A seguir podemos ver um exemplo de modelo de dados entidade-relacionamento criado a partir do banco de dados de exemplo AdventureWorks da Microsoft.



## Sample entity-relationship data model



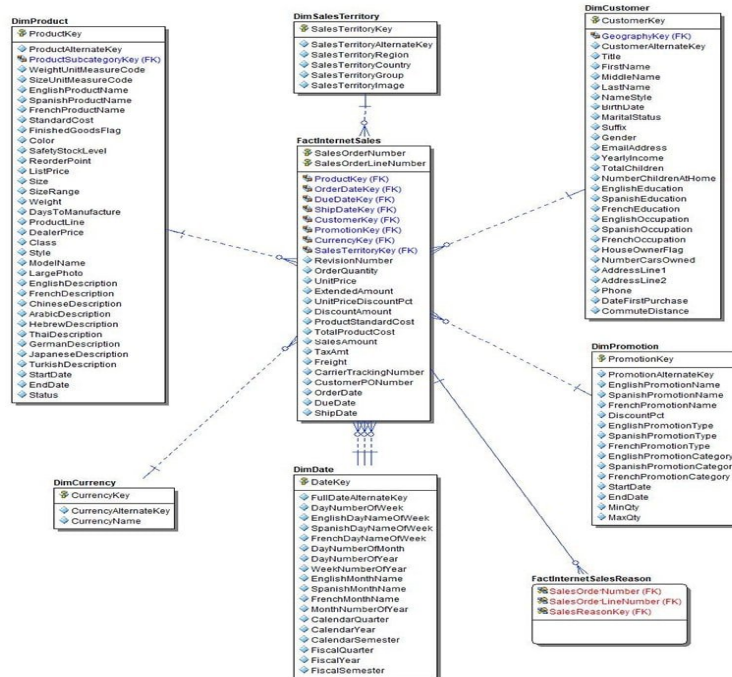
Fonte: AdventureWorks - Microsoft (2020).

## 5. Modelagem de dados dimensionais

Os modelos de dados dimensionais são usados principalmente em data warehouses e data marts que oferecem suporte a aplicativos de inteligência de negócios. Eles consistem em tabelas de fatos que contêm dados sobre transações ou outros eventos e tabelas de dimensões que listam os atributos das entidades nas tabelas de fatos. Por exemplo, uma tabela de fatos pode detalhar as compras de produtos por clientes, enquanto as tabelas de dimensões conectadas contêm dados sobre os produtos e clientes. Tipos notáveis de modelos dimensionais são esquemas em estrela, que conectam uma tabela de fatos a diferentes tabelas de dimensão e esquemas de floco de neve, que incluem vários níveis de tabelas de dimensão.

A seguir temos um exemplo de modelo de dados dimensional que foi criado a partir do banco de dados de exemplo AdventureWorks da Microsoft.

# Sample dimensional data model



Fonte: AdventureWorks - Microsoft (2020).

## 6. Modelagem de dados orientada a objetos

À medida que a programação orientada a objetos avançou na década de 1990 e os fornecedores de software desenvolveram bancos de dados de objetos, a modelagem de dados orientada a objetos também surgiu. A abordagem orientada a objetos é semelhante ao método ER na forma como representa dados, atributos e relacionamentos, mas abstrai entidades em objetos. Diferentes objetos que possuem os mesmos atributos e comportamentos podem ser agrupados em classes, e novas classes podem herdar os atributos e comportamentos das existentes. Mas os bancos de dados de objetos continuam sendo uma tecnologia de nicho para aplicativos específicos, o que limitou o uso da modelagem orientada a objetos.

## 7. Modelagem de dados gráficos

O modelo de dados gráfico é uma ramificação mais moderna dos modelos de rede e hierárquicos. Normalmente emparelhado com bancos de dados de gráficos, geralmente é usado para descrever conjuntos de dados que contêm relacionamentos complexos. Por exemplo, a modelagem de dados gráficos é uma abordagem popular em redes sociais, mecanismos de recomendação e aplicativos de detecção de fraudes. Os modelos de dados de grafos de propriedades são um tipo comum -- neles, nós que representam entidades de dados e documentam suas propriedades são conectados por relacionamentos, também conhecidos como arestas ou links, que definem como os diferentes nós estão relacionados uns aos outros.

## 6. BENEFÍCIOS E DESAFIOS

A modelagem de dados torna mais fácil para desenvolvedores, arquitetos de dados, analistas de negócios e outras partes interessadas visualizar e entender as relações entre os dados em um banco de dados ou data warehouse. Além disso, pode:

- Reduzir erros no desenvolvimento de software e banco de dados.
- Aumentar a consistência na documentação e no design do sistema em toda a empresa.
- Melhorar o desempenho do aplicativo e do banco de dados.
- Facilitar o mapeamento de dados em toda a organização.
- Melhorar a comunicação entre desenvolvedores e equipes de inteligência de negócios.
- Facilitar e acelerar o processo de projeto de banco de dados nos níveis conceitual, lógico e físico.

No entanto, a modelagem de dados é um processo complicado que pode ser difícil de fazer com sucesso. Estes são alguns dos desafios comuns que são encontrados nesse processo:

- A falta de comprometimento organizacional e de adesão dos negócios. Isso significa que as equipes de gerenciamento de dados devem garantir o suporte executivo antecipadamente.
- Falta de compreensão por parte dos usuários de negócios.
- Complexidade de modelagem e fluência de escopo. É importante definir prioridades e manter um escopo de projeto alcançável.
- Requisitos de negócios indefinidos ou pouco claros (mais perguntas e menos ação)

- Escolha da ferramenta utilizada para a modelagem de dados

### ***Ferramentas de modelagem de dados:***

Para escolher uma ferramenta ideal para o seu contexto, faça a si mesmo essas perguntas ao avaliar as ferramentas por seu potencial de modelagem e análise de dados:

#### **1. É intuitivo?**

O pessoal técnico que implementa o modelo pode ser capaz de lidar com qualquer ferramenta que você jogue neles, mas seus estrategistas de negócios e usuários diários de análise - e sua empresa como um todo - não obterão o valor ideal da ferramenta se não for fácil usar. Procure uma experiência de usuário intuitiva e direta.

#### **2. Desempenho**

Velocidade e eficiência. O melhor modelo de dados planejado não é realmente o melhor se não puder funcionar sob o estresse das condições do mundo real – que esperamos envolver o crescimento dos negócios e volumes crescentes de dados, recuperação e análise.

#### **3. Manutenção**

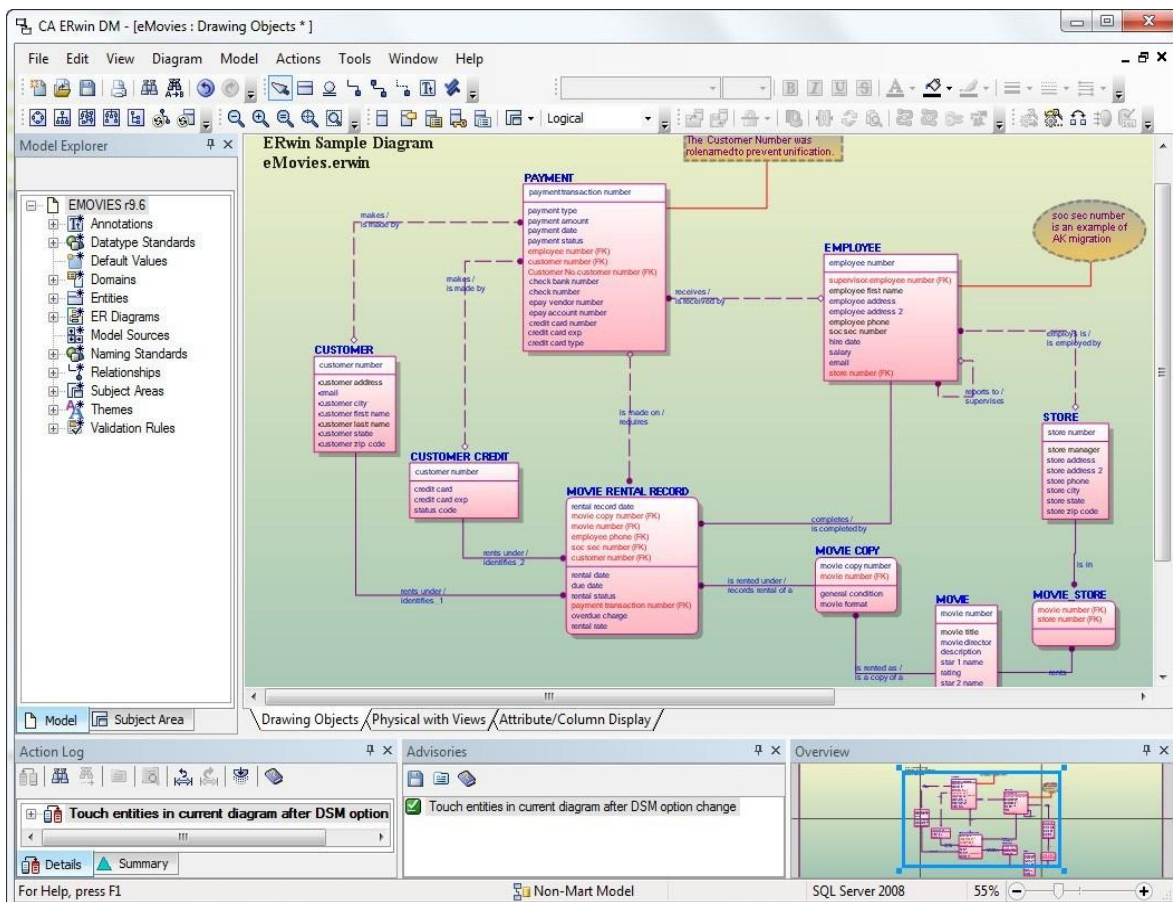
Se todas as alterações em seu modelo de negócios exigirem alterações complicadas em seu modelo de dados, sua empresa não obterá o melhor do modelo ou da análise associada. Procure uma ferramenta que facilite a manutenção e as atualizações, para que sua empresa possa dinamizar conforme necessário e ainda ter acesso aos dados mais atualizados.

#### **4. Meus dados estarão seguros?**

As regulamentações governamentais exigem que você proteja os dados de seus clientes, mas a viabilidade de seus negócios exige a proteção de todos os seus dados como o ativo valioso que são. Você deve certificar-se de que as ferramentas escolhidas tenham fortes medidas de segurança incorporadas, incluindo controles para conceder acesso a quem precisa e bloquear quem não precisa.

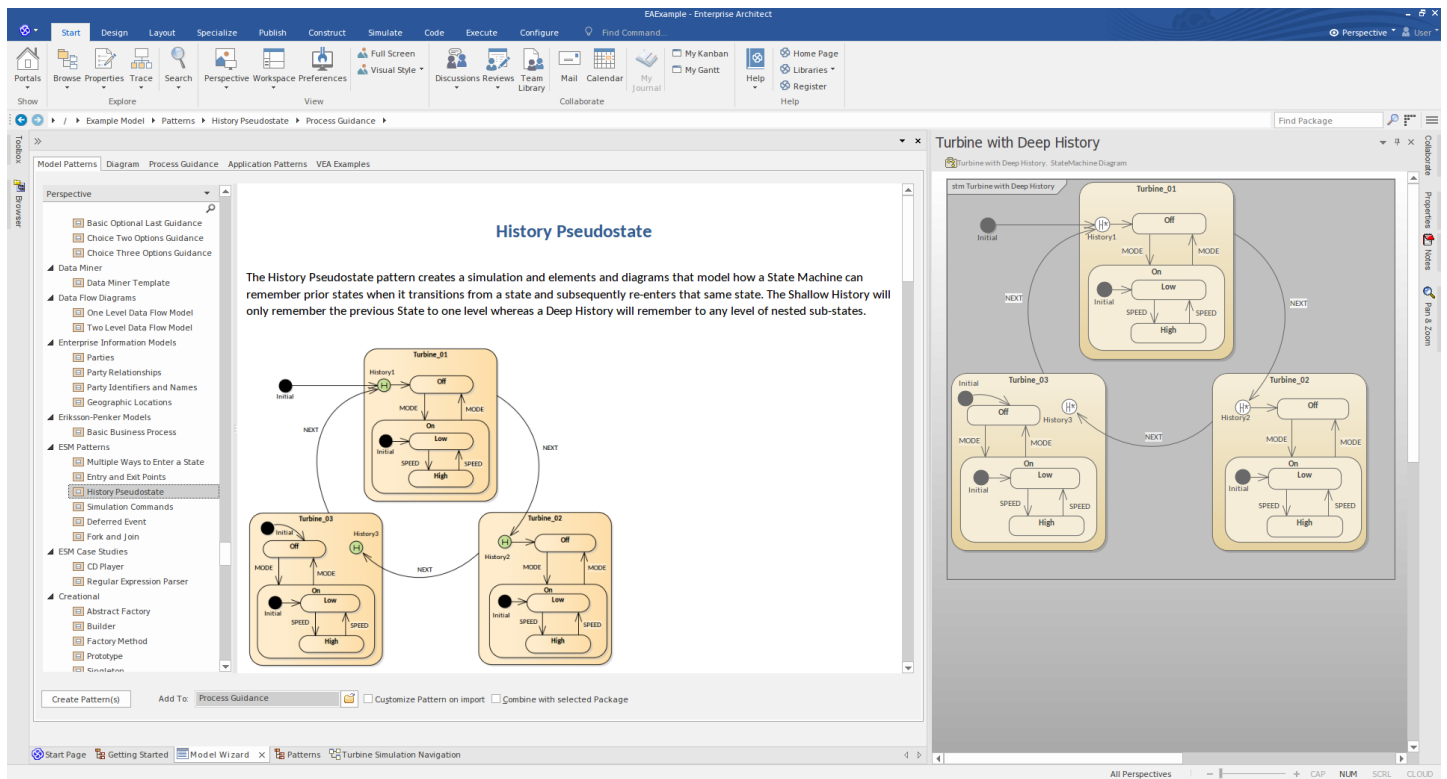
Inúmeras soluções de engenharia de software assistida por computador (CASE) comerciais e de código aberto são amplamente utilizadas hoje, incluindo várias ferramentas de modelagem, diagramação e visualização de dados. Aqui estão alguns exemplos:

- O **erwin Data Modeler** é uma ferramenta de modelagem de dados baseada na linguagem de modelagem de dados Integration DEFinition for Information Modeling (IDEF1X) que agora suporta outras metodologias de notação, incluindo uma abordagem dimensional.



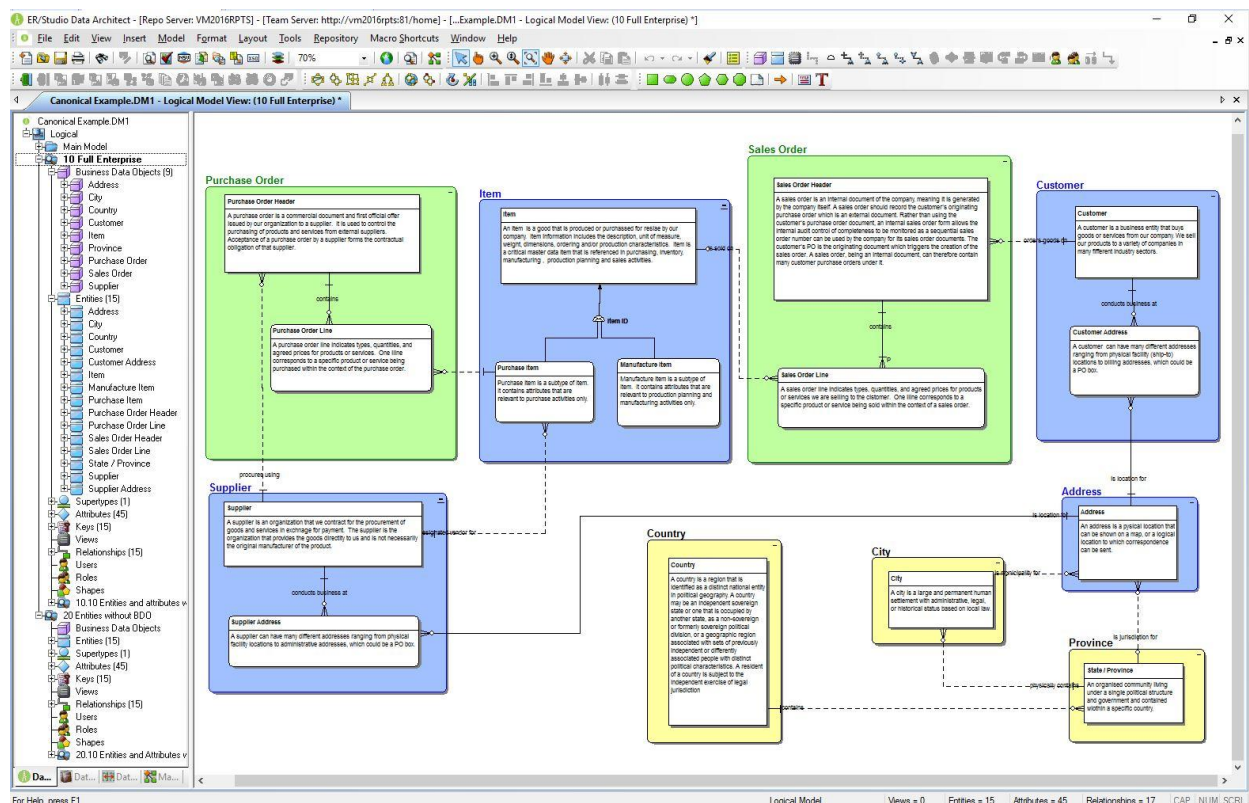
Fonte: GetWinPcSoft (2022)

- O **Enterprise Architect** é uma ferramenta visual de modelagem e design que suporta a modelagem de sistemas e arquiteturas de informações corporativas, bem como aplicativos de software e bancos de dados. É baseado em linguagens e padrões orientados a objetos.



Fonte: Sparx Systems (2022)

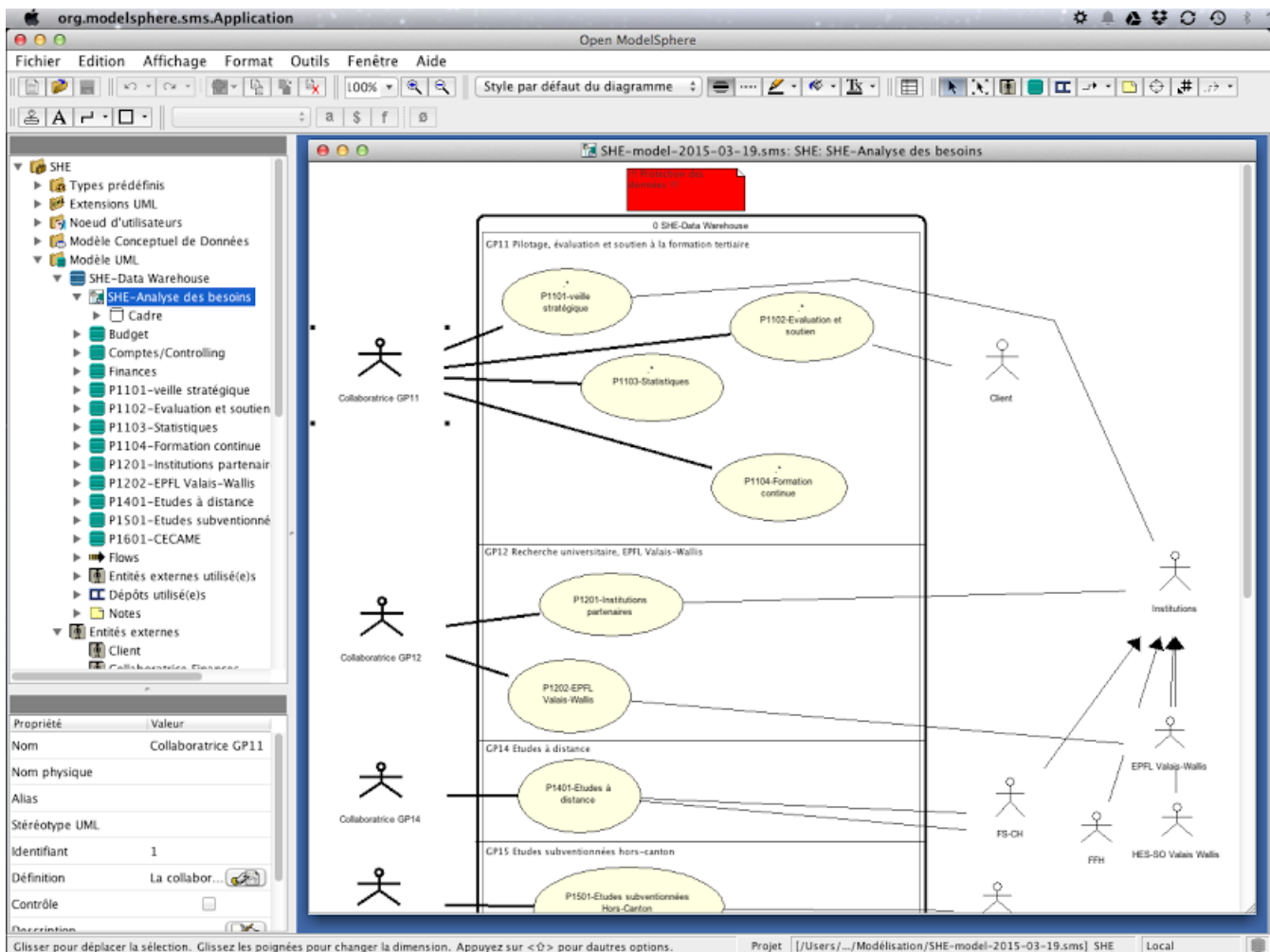
- O **ER/Studio** é um software de design de banco de dados compatível com vários dos sistemas de gerenciamento de banco de dados mais populares da atualidade. Ele suporta modelagem de dados relacional e dimensional.



Fonte: Idera (2022)



- Existem também as ferramentas gratuitas de modelagem de dados, que incluem soluções de código aberto, como o **Open ModelSphere**.



Fonte: BERTHOUSOZ, Cathy (2021)

## 7. CONCLUSÃO

A modelagem de dados desempenha um papel vital no armazenamento dos dados de acordo com os requisitos. Como as organizações lidam com grandes quantidades de dados, elas precisam ser capazes de organizar e dar sentido aos dados e ser capazes de comunicá-los a outras pessoas. São construídos durante a fase de design e análise de um projeto para garantir que esses requisitos do aplicativo sejam atendidos.

Em suma, os modelos de dados garantem consistência nas convenções de nomenclatura, valores padrão, semântica, segurança, garantindo a qualidade dos dados.

## 8. REFERÊNCIAS BIBLIOGRÁFICAS

- Datameer. **“What is Data Modeling and How Do You Do It”**. 2022. Disponível em: <<https://www.datameer.com/what-is-data-modeling-and-how-do-you-do-it/>>. Acesso em: 29/04/2022.
- GAUR, Chandan. **“Data Modelling - Understanding Tools and Techniques Involved”**. 2020. Disponível em: <<https://www.xenonstack.com/insights/data-modelling>>. Acesso em: 29/04/2022.
- IBM Cloud Education. **“Data Modeling”**. 2020. Disponível em: <<https://www.ibm.com/cloud/learn/data-modeling#toc-types-of-d-Fne2rmPT>>. Acesso em: 23/04/2022.
- Microsoft. **“What is Data Modeling”**. 2022. Disponível em: <<https://powerbi.microsoft.com/en-my/what-is-data-modeling/#:~:text=Data%20modelling%20is%20the%20process,between%20those%20bits%20of%20data>>. Acesso em: 29/04/2022.
- Simplilearn. **“What is Data Modelling? Overview, Basic Concepts, and Types in Detail”**. 2022. Disponível em: <<https://www.simplilearn.com/what-is-data-modeling-article>>. Acesso em: 23/04/2022.
- STEDMAN, Craig. **“Data Modeling”**. 2021. Disponível em: <<https://www.techtarget.com/searchdatamanagement/definition/data-modeling>>. Acesso em: 02/05/2022.