



Digital
College

FORMAÇÃO EM

DATA ANALYTICS

UNIDADE 1:

BANCO DE DADOS

MÓDULO 2:

> BANCO DE DADOS DIMENSIONAIS >

Sumário

Introdução	03
1.Data Warehouse	05
1.1. Quais são os principais benefícios em usar Data Warehouse nas empresas?	08
1.2.Entenda as principais desvantagens do Data Warehouse	09
1.3.Como o Data Warehouse é usado em Business Intelligence?	10
1.4. Qual é a diferença entre Data Warehouse e Database?	11
1.5.Data Warehouse: caso real	12
1.6. O que esperar do Data Warehouse?	13
2. Arquitetura Data Warehouse	14
2.1. Modelo Multidimensional	18
3.1. Objetivo	22
3.2. Desenvolvimento	23
3.2.1. Fatores críticos de sucesso	23
3.2.2. Definição da técnica para desenvolvimento de sistemas a ser utilizada	24
3.2.3. Visualizar as necessidades do usuário	24
3.2.4. Nova visão da informação para a tomada de decisões	25
3.2.5. Análise do negócio a ser modelado	25
3.2.6. Análise do ambiente de origem	26
3.2.7. Modelagem dimensional dos dados	27
3.2.8. Aspectos da implementação física	28
3.2.9. Aspectos da visualização das informações	31
4. Estudo de Caso	32
4.1. Definindo o modelo de dados dimensional	34
Anexo I - Criação da tabela	36

Introdução

De acordo com Heinrichs (2003), para competir no mercado global de hoje, as empresas precisam deter mais conhecimento do que antigamente e, ainda, para obter sucesso, elas precisam saber mais sobre seus clientes, mercados, tecnologias e processos, e precisam ter essas informações antes de seus concorrentes.

Para uma tomada de decisão, é necessário ter em mãos informações estratégicas e, tendo em vista esse princípio, as empresas começaram a extrair dados dos seus sistemas operacionais e a armazená-los, separando-os dos dados operacionais. Daí surgiram os Sistemas de Apoio à Decisão. Esses sistemas são soluções computacionais desenvolvidas para apoiar a tomada de decisões complexas durante a resolução de problemas. Ferramentas clássicas de Sistemas de Apoio à Decisão compreendem componentes para gerenciamento de banco de dados sofisticados com poderosas funções de modelagem e projetos de interface com o usuário, as quais permitem trabalhar interativamente com questões, relatórios e funções gráficas.

Segundo Bidgoli (1989), as decisões são classificadas em:

- Decisão estruturada: apresenta procedimento operacional padrão, bem definido e claramente projetado. Este tipo de decisão conta com sistemas de informação relativamente fáceis de definir, programáveis, baseados em lógica clássica, fatos e resultados bem definidos, horizonte de tempo pequeno, rotinas repetitivas e voltadas para baixos níveis da organização;

- Decisão semi-estruturada: não é totalmente bem definida, porém inclui aspectos de estruturação; pode, em grande parte, contar com apoio dos sistemas de informação;
- Decisão não-estruturada: não apresenta qualquer padrão de procedimento operacional, não se repete. No tocante aos sistemas de informação, estes podem apenas apoiar o decisor, o qual precisa contar fortemente com sua intuição, experiência etc. É difícil de formalizar, envolve heurística, tentativa e erro, senso comum em adição à lógica, horizonte de tempo longo, raramente replica decisões prévias e voltadas para os níveis intermediários e alta gerência.

Um sistema de apoio à decisão possui como arquitetura básica: dados, modelo e usuário. De acordo com Westmacott (2001), os componentes básicos são descritos na **Figura 1**.

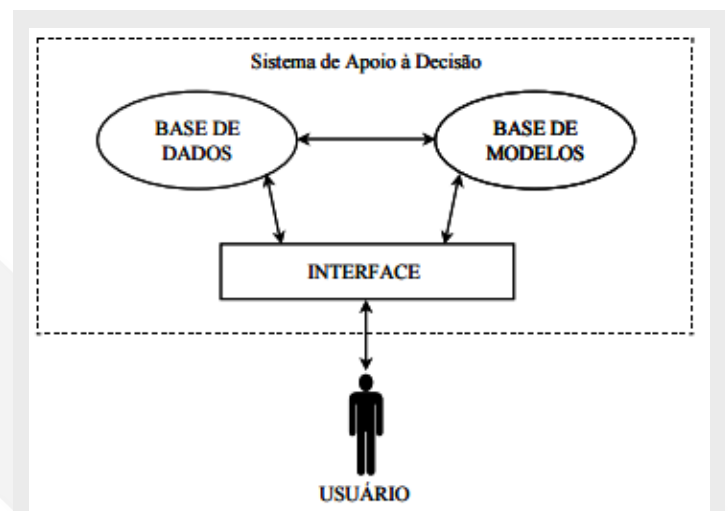


Figura 1 – Estrutura Genérica de um SAD
Fonte: WESTMACOTT (2001)

- Interface com usuário: é a única parte do sistema com a qual o usuário terá contato, exercendo uma importante função do sistema;
- Base de Dados: é utilizada nas operações de gerenciamento de dados (armazenamento, atualização, recuperação e processamento). Nela estão contidos todos os dados e informações que irão alimentar o modelo;
- Base de Modelos: atualmente existe uma grande diversidade de modelos e técnicas de modelagem, a definição destes depende sobretudo das necessidades do usuário, dos objetivos do sistema e, ainda, dos recursos financeiros e do tempo disponível.

O processo de tomada de decisão pode ser traduzido como uma sequência de tarefas que envolvem uma grande quantidade de informação com relacionamentos complexos entre si, os Sistemas de Apoio à Decisão têm ênfase na simulação e exploração de dados, com o objetivo de dar suporte às decisões de simulação e exploração de dados. De acordo com Alter (1980), os Sistemas de Apoio à Decisão são classificados em:

- Sistemas de análise de informações: provêm acesso a uma série de dados orientados à decisão e a pequenos modelos para prover informação gerencial, possibilitando a análise por meio do uso de dados internos. Podem, ainda, a partir de dados do passado, gerar previsões para períodos futuros;
- Modelos de contas: calculam a consequência de ações planejadas sobre a base de definições de contas. Em geral, geram estimativas baseadas em variações das entradas nas fórmulas das contas;

- Modelos de representação: incluem modelos de simulação que estimam a consequência de ações sobre a base de modelos, tais como probabilidades de ocorrências;
- Modelos de otimização: oferecem linhas de ação para uma solução ótima, considerando as restrições necessárias;
- Modelos de sugestão: consideram uma sugestão específica para uma decisão, substituindo procedimentos menos eficientes.

Essas categorias possuem dois agrupamentos, dos quais os três primeiros itens pertencem aos sistemas orientados a dados e os demais itens pertencem aos sistemas orientados a modelos. Na medida em que os usuários começarem a aceitar os conceitos de um Sistema de Apoio à Decisão e a utilizar todos os recursos de sua organização para estimular o aprendizado acerca dos problemas e suas soluções, os Sistemas de Apoio à Decisão se tornarão mais efetivos e frequentes.



1.Data Warehouse

Como você utiliza o **Data Warehouse** hoje? Se essa tem sido uma lacuna na sua empresa, está em tempo de recuperar as oportunidades perdidas. O termo surgiu na década de 1980, quando o cientista da computação (INMON, 2007), tido como o pai do conceito, desenvolveu os primeiros processos operacionais em sistemas de suporte à decisão (DSSs). Desde então, muita coisa mudou. Hoje, eles são um ponto de apoio fundamental na formação de estratégias de negócios.

Podemos definir o **Data Warehouse** como um grande armazém de dados voltado para dar o suporte necessário nas decisões de usuários finais, geralmente gerentes e analistas de negócios. Os dados que compõem o **Data Warehouse** são derivados em sua grande maioria dos diversos bancos de dados operacionais e estão armazenados em locais diferentes dos referidos bancos.

Um **Data Warehouse** consiste em um banco de dados de diversas fontes, normalmente utilizado como base para análises avançadas. Na prática, para entender o que é um **Data Warehouse** e como ele funciona, é preciso antes enxergá-lo como uma solução voltada para empresas.

Ele pode ser compreendido como um suporte para orientar gestores de negócios de todas as áreas em seus processos decisórios. Ele pode servir como suporte desde empresas que atuam no varejo até instituições financeiras e de ensino que precisem de recursos que garantam segurança e transparência em suas operações. É o caso das companhias que se valem de sistemas **Online Transaction Processing**, ou Processamento de Transações em Tempo Real (OLTP).

O **Data Warehouse** permite que sejam feitas consultas e análises eficazes, transformando em dados esparsos informações antes inacessíveis ou subaproveitadas ao reunir informações dispersas nos diversos bancos de dados operacionais de origem que poderiam estar em plataformas distintas. Essas informações podem ser convertidas em estratégias para o negócio.

Um dos grandes benefícios proporcionados pelo **Data Warehouse** é a diminuição do tempo que os gerentes levam para obter as informações necessárias aos seus processos decisórios com a eliminação de tarefas operacionais como pesquisa e identificação dos dados necessários. O **Data Warehouse** contém somente os dados necessários aos gerentes para realizarem as referidas tarefas.

O objetivo do **Data Warehouse** deve ser o de satisfazer as necessidades de análises de informações dos seus usuários, como, por exemplo, monitorar e comparar dados de transações atuais com as passadas e prever tendências futuras, permitindo tomar as medidas cabíveis ainda a tempo de eventuais correções ou aperfeiçoamentos. Utilizar os bancos de dados dos aplicativos operacionais para realizar uma análise ou uma consulta é muito mais trabalhoso, o que ainda pode interferir na performance do aplicativo. Outros problemas também podem ocorrer com esta prática, como enfrentar problemas oriundos da redundância de dados dispersos pelos diversos aplicativos, ou encontrar dados que não estão no formato adequado para se realizar as análises necessárias.

“A grande vantagem de um **Data Warehouse** é permitir a tomada de decisões baseadas em fatos”. O **Data Warehouse** deve se tornar uma ferramenta imprescindível para os gerentes conseguirem administrar seus negócios nas próximas décadas. Como o ambiente de negócios está se tornando cada vez mais dinâmico, é extremamente necessário que as regras de negócios sejam incorporadas às aplicações, que as estruturas dos sistemas se ajustem aos negócios, e que o tempo de resposta dos sistemas seja cada vez menor.

Um **Data Warehouse** se caracteriza por ser um sistema ativo de prospecção e tratamento de dados para atender a finalidades específicas. É diferente, portanto, dos **Data Lakes**, que são repositórios de dados não estruturados de baixo custo e sem uma aplicação em especial.

Entre as suas principais características, destacamos:

- Em um DW, são compilados dados relacionais de sistemas transacionais, aplicativos voltados a negócios e bancos de dados operacionais;
- Os dados precisam ser de qualidade e estarem organizados;
- Permite consultas mais ágeis, graças à tecnologia de armazenamento local;
- Pode gerar relatórios em lote, conforme o conceito de Business Intelligence (BI);
- Os usuários finais geralmente são cientistas de dados, analistas de negócios ou desenvolvedores de dados;

- A arquitetura elementar de um armazenamento de dados (DW) tem como base as diferentes fontes de dados on-line ou em rede;
- A partir delas, é implementada uma solução chamada “área de datastage”, na qual as informações são coletadas e filtradas – e também onde redundâncias são eliminadas;
- Essa área é interligada a um data mart, cuja função é realizar uma nova filtragem de dados para enviá-los às ferramentas utilizadas pelo usuário final.

A crescente utilização do **Data Warehouse** por parte das empresas se deve à necessidade de um maior domínio das informações estratégicas para garantir uma ação mais rápida e eficaz, assegurando, assim, uma maior competitividade no mercado. Dentre os principais fatores que contribuíram para essa maior demanda da tecnologia de *Data Warehousing*, podemos destacar as mudanças organizacionais e estruturais nos negócios e abertura de mercado e a globalização da economia.

Para a adoção da tecnologia de **Data Warehousing** em uma organização, a empresa necessita ter, pelo menos:

- Várias plataformas de *hardware* e de *software*;
- Constantes alterações nos sistemas transacionais corporativos;
- Dificuldade acentuada na recuperação de dados históricos em períodos superiores ao ano atual de operações;
- Existência de sistemas “pacotes” de fornecedores diferentes;

- Falta de padronização e integração dos dados existentes nos diversos sistemas;
- Carência de documentação e segurança no armazenamento de dados;
- Dificuldade de aplicação de sistemas devido a dependências múltiplas de sistemas corporativos.

Segundo Inmon (2007), o **Data Warehouse** possui as seguintes características:

- Baseado em assuntos – armazena informações agrupadas por assunto de acordo com a necessidade da empresa, fornecendo dessa forma informações estratégicas para o negócio. Os projetistas de **Data Warehouse** devem focar apenas nos dados relevantes para as tomadas de decisão que sejam de interesse do negócio;
- Integrado – durante o projeto do modelo de dados, deve-se ter o cuidado de evitar dados que provoquem ambiguidade. As convenções de nomes, os valores de variáveis, tais como sexo masculino e feminino, e outros atributos físicos de dados como *data types* são formalmente unificados e integrados;
- Não é volátil – um princípio do **Data Warehouse** é que depois que o dado é inserido jamais deve ser alterado; ele possui somente duas operações básicas: a carga (inicial e incremental) e a consulta em modo de leitura;
- Variável em relação ao tempo – as informações contidas no **Data Warehouse** são como um *snapshot*, um conjunto estático de registros, acumulando diversos dados sobre diversos períodos de tempo, fornecendo dessa maneira subsídios para análises do negócio em tempos diferentes.

1.1. Quais são os principais benefícios em usar Data Warehouse nas empresas?

Agora que você entendeu o que é o Data Warehouse e conheceu os seus tipos, vamos apontar as principais vantagens em contar com um armazenamento de dados nas empresas. Veja quais são:

Agilidade nas consultas: sistemas de data warehouse não são apenas capazes de armazenar dados, mas também são uma solução completa para companhias que lidam frequentemente com a informação.

Maior capacidade de processamento de dados: com a expansão da cloud computing, a capacidade de armazenamento e processamento dos sistemas de data warehouse vem aumentando bastante.

Acesso a dados históricos: quando é necessário ter uma referência histórica para efetuar uma operação on-line, os armazenamentos de dados se revelam ainda mais valiosos, já que trabalham com sistemas OLTP.

Centralização de dados: outra importante vantagem é que eles operam a partir de dados centralizados e compilados em um único repositório.



1.2. Entenda as principais desvantagens do Data Warehouse

Estruturar um Data Warehouse tem seus pontos de atenção. Conheça alguns desafios que podem surgir no dia a dia:

- Dificuldade em integrá-lo com sistemas e softwares legados;
- Problemas no controle de acesso aos dados;
- Complicações ao estruturar dados e ao agregar valor a eles;
- Sua estruturação pode ser trabalhosa demais;
- Rápida obsolescência;
- Dificuldade em estabelecer regras para as distintas fases de operação;
- Imprevisibilidade em relação aos problemas.

1.3. Como o Data Warehouse é usado em Business Intelligence?

Na prática, um Data Warehouse serve como a base na qual middlewares em ambientes de Business Intelligence fornecem aos usuários finais diferentes resultados. Ou seja, é a partir desse sistema que ferramentas de BI são operacionalizáveis. Logo, em processos de **Business Intelligence**, os DW funcionam como a espinha dorsal do armazenamento de dados. Isso porque a inteligência de negócios depende de consultas complexas e da comparação de vários conjuntos de dados para balizar desde decisões diárias a mudanças mais radicais ou que contemplem toda a empresa. Para facilitar isso, o BI se estrutura em três atividades abrangentes: organização de dados, armazenamento de dados e análise de dados.

A transformação de dados geralmente é viabilizada por tecnologias de extração, transformação e carregamento (ETL), enquanto a análise é feita usando ferramentas de inteligência de negócios.

1.4. Qual é a diferença entre Data Warehouse e Database?

Por tudo que vimos até aqui, podemos dizer que o **Data Warehouse** é um sistema de informação que armazena dados históricos e relacionais de fontes únicas ou múltiplas.

Ele é projetado para analisar, relatar e integrar dados de transações de diferentes fontes. O DW facilita o trabalho de análise e formação de relatórios de uma companhia e é também a fonte principal para orientar no processo de tomada de decisão e previsão.

Já o Database é uma coleção de dados relacionados que representam alguns aspectos do mundo real, sendo projetado para a gravação de tais elementos. Sendo assim, podemos apontar para algumas diferenças entre esses dois recursos:

- O *Database* é projetado para registrar dados, enquanto o *data warehouse* é projetado para analisá-los;
- O *Database* é uma coleção de dados orientada para aplicações, enquanto o armazenamento de dados é a coleção de dados orientada para o assunto;
- O primeiro usa o *Online Transaction Processing* (OLTP), enquanto o DW usa o *Online Analytical Processing* (OLAP);
- O *Database* é projetado utilizando técnicas de modelagem *Entity Relationship Diagram* (ERD), enquanto o armazém de dados usa técnicas de modelagem de dados para projetar.

1.5.Data Warehouse: caso real

Um caso real de empresa que utilizou **Data Warehouse** para melhorar o tratamento dos seus dados é o da Universidade Cornell, nos Estados Unidos. A instituição usava o Cognos Data Manager para transformar e mesclar dados em um DW da Oracle. Em um certo momento, a IBM decidiu encerrar o suporte para o produto. O motivo foi assim descrito pelo então gerente de DW da universidade, Jeff Christen:

“Infelizmente, tínhamos milhões de linhas de código escritas no Data Manager, então, tivemos que procurar por uma substituição.”

Ele viu nesse acontecimento uma oportunidade para adicionar novas funcionalidades para que seu **Data Warehouse** funcionasse com mais eficiência. O gestor começou a procurar ferramentas de ETL a fim de adicionar as otimizações desejadas. Para isso, ele se concentrou em áreas-chave ao avaliar os fornecedores: documentação, custos de licenciamento, melhoria do desempenho e capacidade de trabalhar dentro das limitações de pessoal existentes.

A saída encontrada foi o WhereScape, uma solução em automação de dados que, segundo Christen, “é uma ferramenta robusta, mas também intuitiva o suficiente para ser dominada em poucas semanas”.

1.6. O que esperar do Data Warehouse?

Cada vez mais integrados a soluções e recursos de BI, machine learning e inteligência artificial, a tendência para o futuro dos **Data Warehouses** é de se tornarem mais intuitivos.

É o que se espera a partir do novo conceito de **Data Warehouse** 2.0, no qual a arquitetura mais avançada trata os dados como em um ciclo de vida. Outra tendência muito forte é o uso cada vez mais intenso da *cloud computing*.

Afinal, as empresas estão mudando para tecnologias de armazenamento de dados em nuvem por motivos de desempenho, segurança, agilidade e simplificação operacional. Para o futuro, os DWs tendem a ser também verdadeiros ecossistemas de análise completos.

Isso porque processos e projetos analíticos dependem de dados de diversos tipos (dados transacionais, dados de eventos e dados de referência) que vêm de sistemas e bancos de dados corporativos, bem como de fontes de *big data*.

Sendo assim, daqui por diante, os dados existentes nos DWs deverão se integrar ao ecossistema de análise, trabalhando em conjunto com um *data lake* para fornecer toda a gama de dados necessários para que possam ser analisados.

2. Arquitetura Data Warehouse

A escolha da arquitetura do **Data Warehouse** é uma decisão gerencial do projeto e está relacionada com a capacidade da infraestrutura disponível. A decisão da arquitetura pode trazer enormes impactos ao projeto, uma decisão errada pode acarretar o tempo de execução do projeto, com isso causando a insatisfação aos envolvidos.

Existem 10 fatores que afetam a decisão relativa à escolha da arquitetura, descritos a seguir:

- Interdependência de informações entre as unidades da organização;
- Demanda de informações da alta administração;
- Nível de urgência de um Data Warehouse;
- Natureza das tarefas do usuário final;
- Limitação de recursos;
- Visão estratégica do Data Warehouse antes da implementação;
- Compatibilidade com os sistemas existentes;
- Capacidade interna da equipe de Tecnologia da Informação (TI);
- Questões de natureza técnica;
- Fatores sociais / políticos.

Os componentes de um **Data Warehouse** são: extração, armazenamento e apresentação. A arquitetura de um **Data Warehouse** determina como se dá a organização de seus componentes. A arquitetura demonstrada na Figura 2 compreende a camada dos dados operacionais que serão acessados pela camada de extração de dados, em seguida os dados serão armazenados no **Data Warehouse**. A camada de acesso à informação possibilita a extração das informações do **Data Warehouse** utilizando um conjunto de ferramentas.

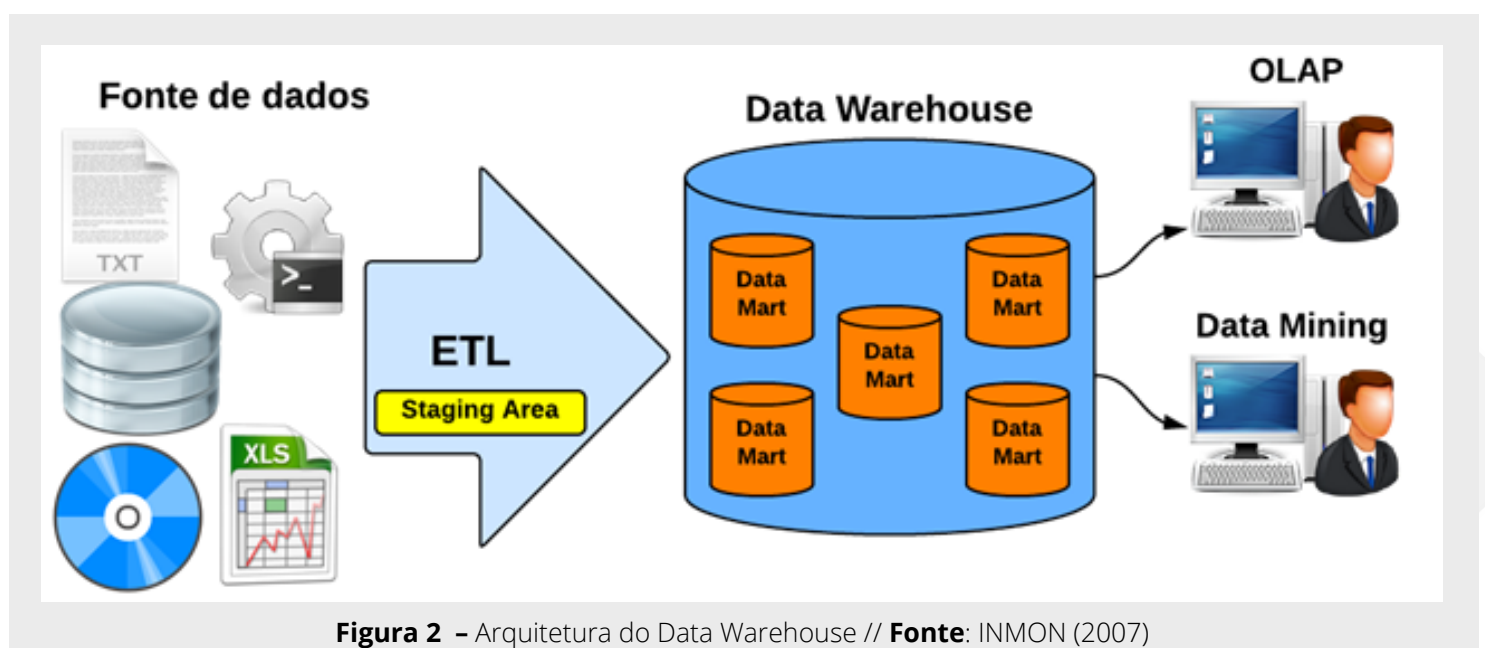


Figura 2 – Arquitetura do Data Warehouse // **Fonte:** INMON (2007)

Na Figura 3, é demonstrada a implementação top-down que foi introduzida por Inmon (1997) e iniciou-se com a extração, transformação e integração das informações dos sistemas operativos e dados externos. Em seguida, os dados e metadados são enviados a um **Data Warehouse** centralizado que reúne todos os dados relativos à organização. Nessa implementação os dados irão ficar disponíveis nos Data Marts, que são um pequeno conjunto de um **Data Warehouse** organizado por assunto. Para Machado (2013), as vantagens da implementação top-down são:

- Herança de arquitetura: todos os Data Marts originados de um **Data Warehouse** utilizam a arquitetura e os dados desse **Data Warehouse**, permitindo uma fácil manutenção;
- Visão de empreendimento: o **Data Warehouse** concentra todos os negócios da empresa, sendo possível extrair dele níveis menores de informação;
- Repositório de metadados centralizado e simples: o **Data Warehouse** provê um repositório de metadados central para o sistema. Essa centralização permite manutenções mais simples do que aquelas realizadas em múltiplos repositórios;
- Controle e centralização de regras: a arquitetura top-down garante a existência de um único conjunto de aplicações para extração, limpeza e integração dos dados, além de processos centralizados de manutenção e monitoração.

As desvantagens da implementação *top-down* segundo Machado (2013) são:

- Implementação muito longa: em média, 15 ou mais meses para que a primeira área de assunto entre em produção, dificultando a garantia de apoio político e orçamentário;
- Alta taxa de risco: não existem garantias para o investimento nesse tipo de ambiente;
- Heranças de cruzamentos funcionais: é necessária uma equipe de desenvolvedores e usuários finais altamente capacitada para avaliar as informações e consultas que garantam à empresa habilidades para sobreviver;
- Expectativas relacionadas ao ambiente: a demora do projeto e a falta de retorno podem induzir a expectativa dos usuários.

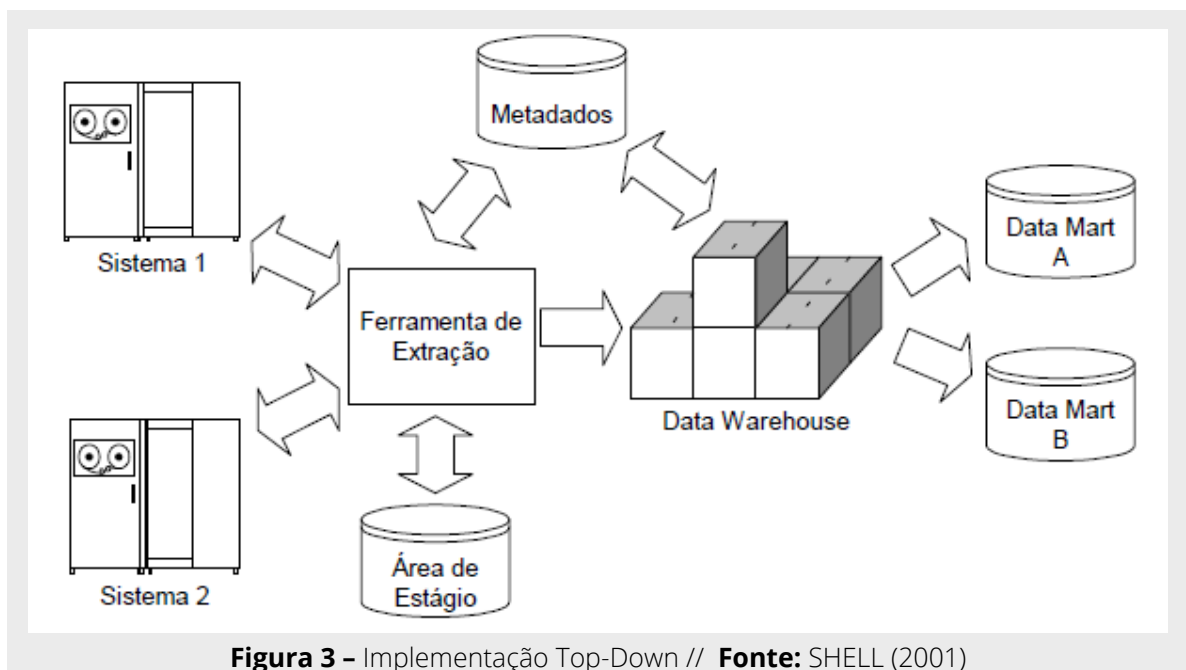


Figura 3 – Implementação Top-Down // Fonte: SHELL (2001)

A implementação *bottom-up* permite que o planejamento e o desenho do *Data Mart* possa ser realizado antes da criação do *Data Warehouse*. Essa implementação caracteriza-se pelo armazenamento e extração a partir da criação incremental de vários *Data Marts* independentes, com metadados e área de extração individualizadas e, no conjunto, formando as fontes de dados do *Data Warehouse*. A área de extração é composta por áreas de estágio e mecanismos de extração distintos para cada *Data Mart*. A arquitetura é ilustrada na Figura 4. As vantagens da implementação *bottom-up* são:

- Implementação rápida: a construção dos *Data Marts* é altamente direcionada, permitindo um rápido desenvolvimento;
- Retorno rápido: a arquitetura baseada em *Data Marts* com incremento demonstra rapidamente seu valor;
- Manutenção do enfoque da equipe: a elaboração de *Data Marts* incrementais permite que os principais negócios sejam criados inicialmente;
- Herança incremental: os *Data Marts* incrementais obrigam a entrega de recursos passo a passo.

As desvantagens da implementação *bottom-up* são:

- Perigo de legado: essas soluções podem não considerar a arquitetura global, transformando-se assim em legados.
- Desafio de possuir a visão de empreendimento: é necessário que se mantenha a visão do negócio como um todo.
- Administrar e coordenar múltiplas equipes e iniciativas: desenvolvimento de múltiplos *Data Marts* em paralelo.
- A maldição do sucesso: os usuários com seus *Data Marts* solicitam novas atualizações cada vez mais, isso traz constantes desafios para a equipe de desenvolvimento.

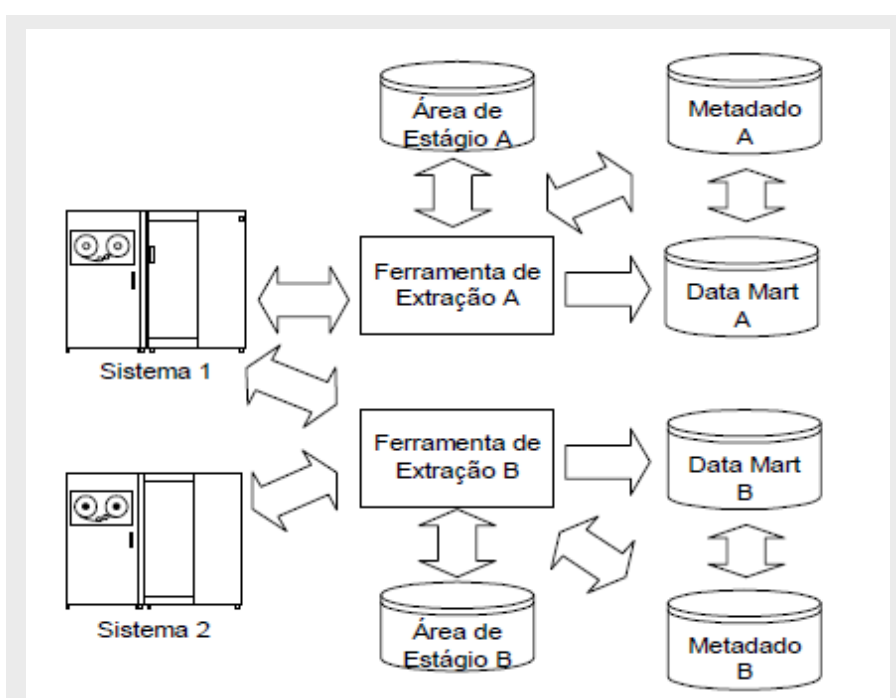
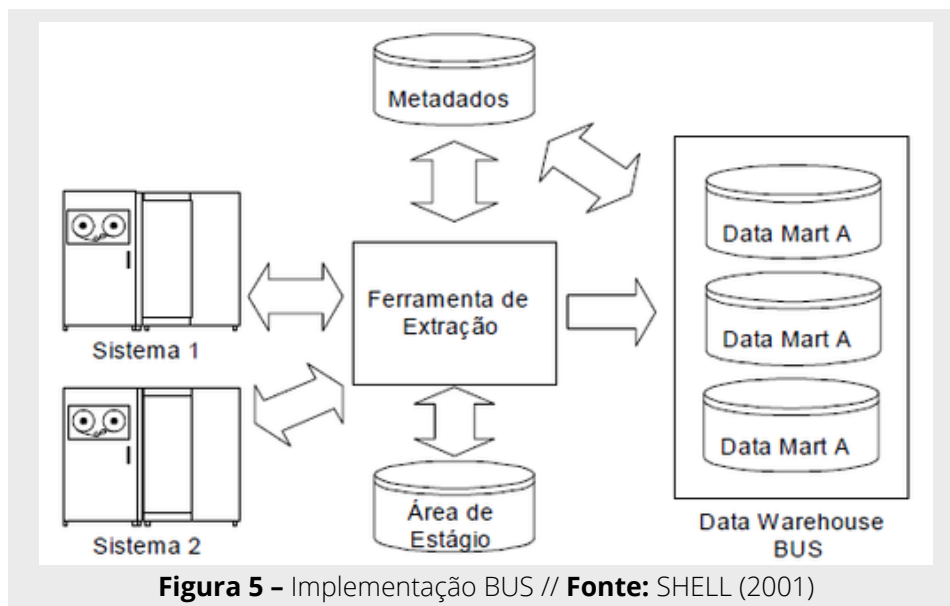


Figura 4 – Implementação *Bottom-Up* // Fonte: SHELL (2001)

A implementação BUS foi introduzida por Kimball (1998) e tem o propósito de integrar as implementações Top-Down e Bottom-Up. Nessa implementação efetua-se a criação do modelo de **Data Warehouse** em sua forma macro, em seguida os Data Marts são planejados e integrados através dos metadados e de tabelas de fatos e dimensões padronizadas. Antes de iniciar a construção do **Data Warehouse**, são definidos os Data Marts a serem construídos e as dimensões e fatos comuns. Cada Data Mart é construído respeitando a pré-estruturação dos fatos e dimensões comuns. O **Data Warehouse** é composto pela união dos Data Marts, sendo coordenado pelos metadados. A área de extração é composta por uma única área de estágio e por um único mecanismo de extração. Algumas variações desta arquitetura são discutidas em Firestone (2000), sendo que as mesmas variações apresentam áreas de extração independentes e alternativas ao esquema de metadados originalmente proposto.



A granularidade representa o nível de detalhamento dos dados contidos no **Data Warehouse** e a definição do nível de granularidade constitui um dos passos mais importantes da construção de um **Data Warehouse**. Quanto menor o nível de detalhes, mais alto é o nível de granularidade. A importância do nível de granularidade em um **Data Warehouse** determina o volume de dados armazenados e o tipo de consulta que pode ser atendido e deve ser balanceado de acordo com o detalhamento requerido nas consultas. Machado (2013) defende que em quase todas as situações, um menor nível de granularidade para suas dimensões deve ser usado, para atender o maior número de consultas possíveis, inclusive as não esperadas. Ele também argumenta que o menor nível de granularidade torna o **Data Warehouse** muito mais resistente a novos elementos de dados. No entanto, essa adoção de uma granularidade muito baixa, além de aumentar o volume de dados armazenados no **Data Warehouse**, pode também elevar a complexidade das consultas finais se uma parte das consultas pretendidas envolve operações mais complexas do que simples agregações. A escolha inadequada pode comprometer todo o restante do projeto.

Se tomarmos como base o modelo estrela, a definição da granularidade corresponde ao menor grão que será armazenado na tabela de fatos e a granularidade da tabela de dimensões não poderá ser menor do que a de fatos, no entanto, se essa for maior, não irá acarretar qualquer contradição lógica, mas poderá causar a perda de informações úteis para o usuário.

2.1. Modelo Multidimensional

A modelagem multidimensional é uma técnica de concepção e visualização de um modelo de dados de forma intuitiva. É utilizada especialmente para sumarizar e reestruturar dados e apresentá-los em visões que suportem a análise dos valores desses dados.

A modelagem multidimensional tem sido amplamente adotada como a técnica de representação de um **Data Warehouse**, a sua simplicidade é a chave fundamental que possibilita aos utilizadores o entendimento fácil dos dados apresentados.

A modelagem multidimensional é uma das mais importantes diferenças entre o modelo operacional e um **Data Warehouse**, os modelos entidade-relacionamentos estão sendo utilizados na maioria dos ambientes operacionais. Já os **Data Warehouse** utilizam a modelagem multidimensional, necessitando, assim, de análises de dados multidimensionais. De acordo com Peixinho (2012), o modelo multidimensional possui três elementos básicos:

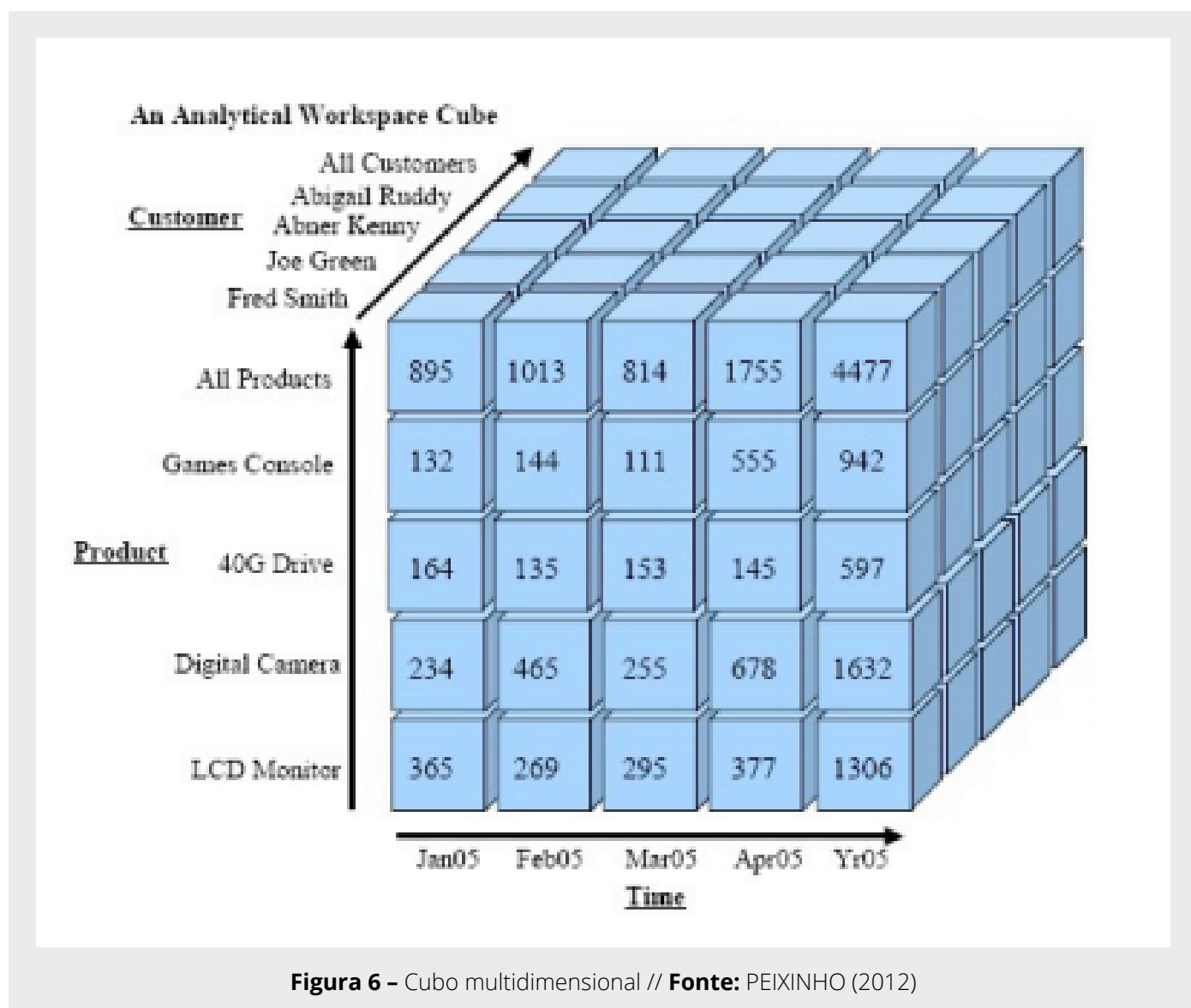
- **Fato:** um fato é uma coleção de itens de dados, cada fato representa um item, um evento de negócio de uma empresa. É representado pelos valores numéricos e implementado pelas tabelas denominadas Tabelas Fatos;
- **Dimensões:** são os elementos que participam de um fato, as dimensões determinam o contexto de um assunto de negócios;
- **Medidas:** são os atributos numéricos que representam um fato, a performance de um indicador de negócios relativo às dimensões que participam desse fato.

Modelo Entidade Relacionamento x Modelo Multidimensional	
Modelo Entidade Relacionamento	Modelo Multidimensional
Mais complexo	Estrutura mais fácil e intuitiva
Anos 70	Anterior a Entidade Relacionamento e recriada posteriormente
Tabelas representam conjunto de entidades e relacionamentos	Tabelas representam fatos e dimensões
Tabelas resultantes normalizadas	Tabelas Fatos normalizadas, Tabelas Dimensões não podem ser normalizadas

Tabelas acessadas por todos filtros	Tabelas de dimensões são pontos de entrada de acesso
Maior dificuldade de consulta e leitura por usuário não especializado	Leitura e consulta mais fáceis para usuários não especializados

Tabela 1 - Comparativo Entidade Relacionamento x Multidimensional // **Fonte:** Elaborado pelo autor

Uma medida está localizada como atributo de um fato e é determinada pelas combinações das dimensões desse fato. A modelagem multidimensional tem como ideia principal a representação dos dados do negócio como um cubo de dados, no qual as células do cubo contêm as medidas, e os lados são definidos pelas dimensões.



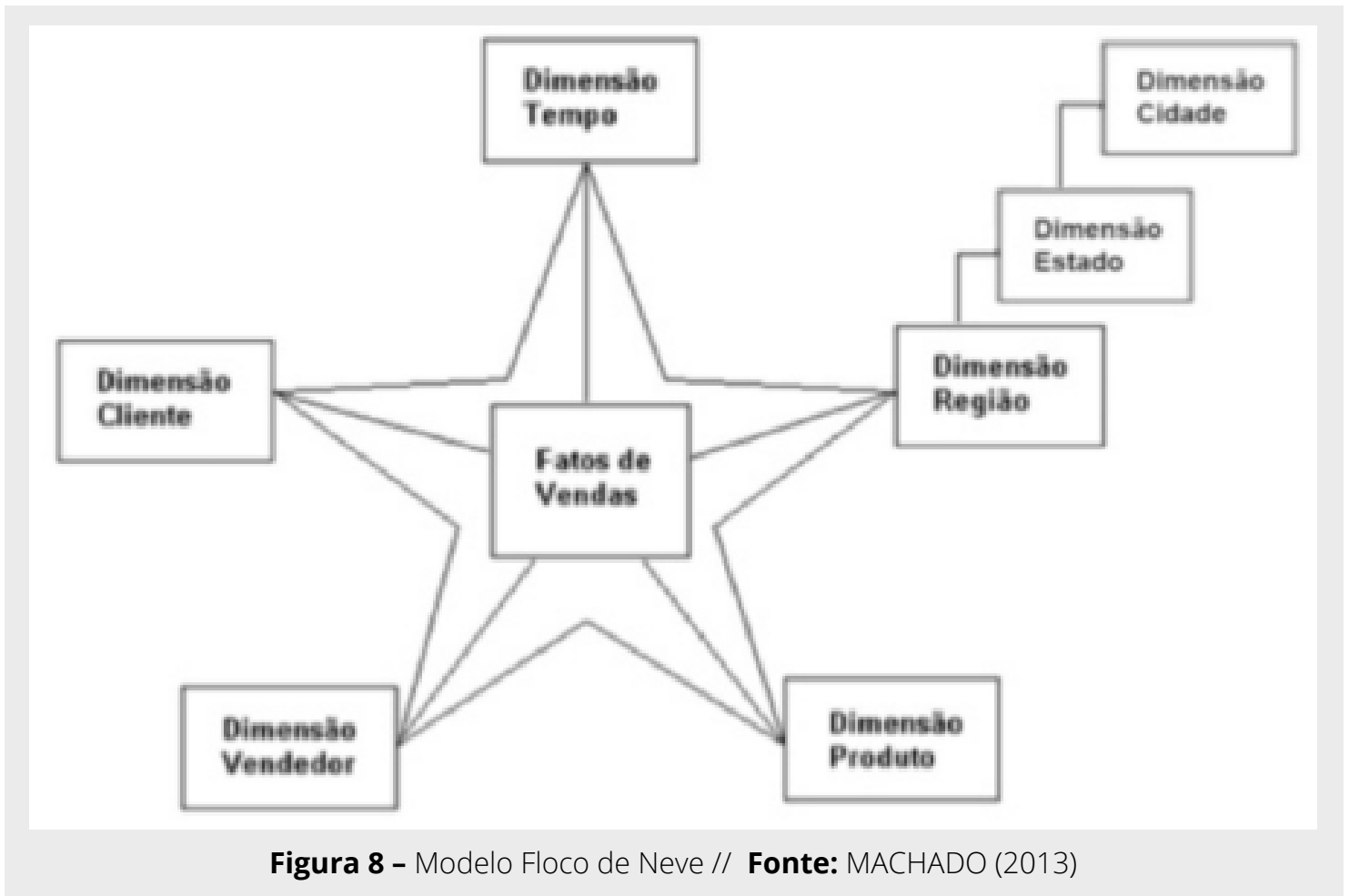
Conforme demonstrado na Figura 6, é possível observar que o cubo agrupa as dimensões e que o cubo completo pode ser dividido para que possa ser possível uma visualização de um determinado fato, sendo assim, a forma como os dados são armazenados é de vital importância para uma boa performance.

A modelagem multidimensional possui dois tipos de modelos: o modelo estrela (*star schema*) e o modelo floco de neve (*snow flake*). Cada um com aplicabilidade diferente a depender da especificidade do problema. As dimensões do modelo estrela são desnormalizadas, ao contrário do floco de neve, que parcialmente possui normalização.



Figura 7 – Modelo Estrela // Fonte: MACHADO (2013)

O modelo estrela é a estrutura básica de um modelo multidimensional, este modelo é constituído por uma tabela de fatos cercada por tabelas de dimensões, conforme a Figura 7. As dimensões servem como ponto de entrada simetricamente iguais para a tabela de fatos, essa simetria permite um padrão previsível do modelo que resiste a mudanças não previstas no comportamento do usuário.



O modelo floco de neve exposto na Figura 8 é uma variação do modelo estrela, no qual todas as tabelas de dimensões podem ser normalizadas, reduzindo dessa maneira a redundância dos dados, mas aumentando a complexidade do modelo e, conseqüentemente, a compreensão por parte dos usuários, dificultando também a implementação de ferramentas de visualização dos dados.

3. Passo a Passo para Modelagem Dimensional

3.1. Objetivo

O enfoque principal é propor um roteiro que especifique as etapas necessárias no desenvolvimento de um projeto de **Data Warehouse** visando prover o conhecimento mínimo necessário àqueles profissionais que não têm referência alguma sobre esse assunto.

3.2. Desenvolvimento

As etapas para a construção do roteiro estão em uma ordem sequencial, algumas delas podem ser realizadas paralelamente a outras ou a ordem pode ser alterada, de acordo com a necessidade; isso será definido pela equipe que está elaborando o projeto. Sugere-se essa sequência, pois facilita o entendimento do objetivo do projeto e descreve etapas interligadas e fundamentais para o sucesso dele.

Abaixo, o roteiro é apresentado em uma forma esquemática e resumida.

3.2.1. Fatores críticos de sucesso

Primeiramente, é de extrema importância elencar os principais itens que determinarão o sucesso do projeto. A seguir estão alguns itens importantes:

- É fundamental planejar o projeto;
- É importante ter o orçamento aprovado para desenvolver o projeto de acordo com o plano de investimento da empresa;
- É importante que a empresa esteja disposta a modernizar-se para garantir o crescimento e a competitividade;
- É importante ter o apoio e o empenho dos usuários responsáveis e suas respectivas gerências por intermédio de atitudes participativas e cooperativas;
- É fundamental ressaltar a importância das informações para a empresa e, em especial, o compartilhamento delas, evitando duplicidade de dados e informações proprietárias;
- É necessário negociar um acordo com a alta administração referente à proposta de implementação do projeto;
- É importante primeiro analisar as áreas de atuação mais lucrativas da empresa;
- É importante analisar a contenção de custos;
- É imprescindível que as informações sejam precisas, consistentes e rápidas de se obter;
- É importante ter a capacidade de adaptação às necessidades de negócio;
- É necessário apresentar resultados entre as etapas do desenvolvimento do projeto.

Após estarem definidos os fatores críticos de sucesso do projeto, deve-se iniciar a análise de qual técnica de desenvolvimento de sistemas mais se integra à construção de um projeto de **Data Warehouse**.

3.2.2. Definição da técnica para desenvolvimento de sistemas a ser utilizada

É de extrema importância utilizar uma técnica para desenvolvimento de sistemas para auxiliar no projeto do **Data Warehouse**. Indica-se a técnica da Engenharia da Informação em virtude de os dados serem o enfoque principal desta técnica, e em um **Data Warehouse** o objetivo principal é a transformação dos dados em informações. Ao utilizar a técnica da Engenharia da Informação, torna-se mais fácil a forma de visualização dos relacionamentos entre os requisitos básicos que estruturam o negócio objeto de análise. Como produto principal desta técnica, temos o Modelo de Entidades e Relacionamentos (MER). Um MER é válido e confiável quando consegue responder às perguntas dos processos que serão por ele atendidos. Portanto, é um modelo obrigatório em qualquer desenvolvimento de sistemas, até porque é requisito básico para a implementação física das bases de dados. Tendo definida e assimilada a técnica de desenvolvimento de sistemas a ser utilizada no projeto, é importante entender quais são as necessidades do usuário, diferenciando-as entre o ambiente analítico e o operacional.

3.2.3. Visualizar as necessidades do usuário

O usuário é o fator chave para o sucesso do projeto. Sua participação efetiva durante todo o projeto é o requisito mais importante, pois ele é quem domina a área a ser analisada e irá dirimir as dúvidas do analista da informação que implementará o projeto. Há, ainda, a necessidade de entender a diferença entre os sistemas de processamento operacional e analítico.

Sistemas de processamento operacional são sistemas que suportam as operações do dia a dia da empresa. São sistemas de processamento on-line atualizados ao longo do dia.

Sistemas de processamento analítico disponibilizam informações usadas para analisar um problema ou uma situação. É feito por meio de comparações ou da análise de padrões ou tendências. As informações refletem um instante específico no tempo.

Tendo definidas e assimiladas quais são as necessidades do usuário, é importante entender a mudança de enfoque em um projeto de Data Warehouse no que se refere às informações para a tomada de decisões. Durante este processo, é fundamental a participação do usuário e o pleno entendimento dele nesse novo processo de análise da informação.

3.2.4. Nova visão da informação para a tomada de decisões

O que buscamos quando estamos modelando sistemas de informação é o entendimento operacional do negócio que está sendo modelado, sem pensar na aplicação em si, sem pensar na tecnologia envolvida, mas com a visão relacional de dados. Quando vamos executar uma modelagem multidimensional, estamos desenhando soluções de gestão de negócios, buscando entender como os executivos e gestores de uma organização avaliam as informações para adoção de estratégias e avaliação de desempenho. A análise multidimensional requer um modelo de dados que permita que os dados sejam facilmente visualizados de acordo com diversas perspectivas.

O passo seguinte é como analisar a área do negócio a ser modelado.

3.2.5. Análise do negócio a ser modelado

Primeiramente, é preciso definir os objetivos do projeto, que podem ser resumidamente descritos nos dois itens abaixo:

1. Disponibilizar as bases de dados do ambiente operacional (OLTP) para um ambiente analítico (OLAP);
2. Permitir aos usuários a extração dos dados por meio de análise exploratória no ambiente OLAP.

Definir o escopo do projeto, observando:

- Caracterizar o problema: o dado deve ser representado como informação para que a empresa possa tomar decisões;
- Definir quais processos são mais críticos para o negócio e quais precisam de decisões sobre ele;
- Assimilar o conhecimento e a definição da área de atuação (negócio) do cliente específico e uma tomada de decisão. É importante representar este conhecimento em um modelo de entidades e relacionamentos, pois será muito útil na modelagem dimensional;
- Definir o grupo gestor da informação que irá participar durante todo o processo de desenvolvimento do projeto. O grupo gestor pode ser considerado como os usuários que dominam a área de negócio e que tomam decisões, juntamente com o pessoal de informática.

É extremamente importante a definição do escopo do projeto, pois é a base dele. No escopo deve ser descrito se a implementação será em um **Data Warehouse** (base corporativa que atenderá toda a organização) ou em um ou vários "data marts" (base de informação por linha de negócio que contém um subconjunto dos dados corporativos da organização). Recomenda-se iniciar com "data marts" integrando-os a um **Data Warehouse** ao longo do tempo. Isso se justifica pelos seguintes aspectos:

- O custo é bem inferior a implantar um *Data Warehouse* de toda a empresa;
- O tempo de implementação é reduzido;
- É mais fácil assimilar a atuação de uma área de negócio específica a entender todo o processo da empresa.

Optando-se por implementar em "*data marts*", deve-se ter o cuidado de não esquecer a integração entre os eles, caso contrário o projeto é inviabilizado.

Tendo definida a área de negócio da empresa a ser implementada no projeto, deve-se analisar onde estão as informações que alimentarão o "*data warehouse*" / "*data mart(s)*".

3.2.6. Análise do ambiente de origem

Partindo das necessidades de informação por área de negócio já estarem definidas, nesta análise deve-se agir da seguinte maneira:

- Analisar o ambiente computacional da empresa como, por exemplo: "mainframe" IBM, ambiente DOS, ambiente "windows", ambiente UNIX etc;
- Constatar se as informações necessárias estão armazenadas nos aplicativos operacionais ou, caso contrário, manter os aplicativos operacionais para que obtenham os dados que faltam. É preciso também verificar a viabilidade deste tipo de manutenção para gerar a informação que falta no sistema analítico;
- Definir como integrar as várias fontes de informação, se houver;

Um fator importante é como analisar as fontes de informação do ambiente de origem como: tipos de bancos de dados (relacionais, em rede, hierárquicos), arquivos sequenciais, planilhas etc. Também é importante verificar os modelos de dados, dicionários de dados dos sistemas do ambiente operacional, relatórios gerenciais dos aplicativos, entre outros. Estas informações auxiliam muito na construção do modelo dimensional.

A próxima etapa do roteiro é a construção da modelagem dimensional.

3.2.7. Modelagem dimensional dos dados

Sugere-se utilizar, como técnica principal para a modelagem de dados dimensional, o "star schema" que é a técnica mais indicada para projetos de *Data Warehouse*. A modelagem dimensional compõe-se de Tabelas Dimensões e Fatos:

Dimensões:

Representam as possíveis formas de visualizar os dados. São as entradas para as consultas (tempo, região, cliente etc). A base para entendimento de qualquer negócio é responder às quatro perguntas fundamentais abaixo:

1. QUANDO? (Período de tempo a que se refere a análise)
2. O QUÊ? (O principal objeto de análise)
3. ONDE? (Localização física ou geográfica para análise)
4. QUEM? (Um objeto específico e detalhado para análise: opcional)

Fatos:

É a tabela central que interliga as dimensões e tem os indicadores de análise ou métricas (quantidade, valores etc). A Tabela Fatos deverá possuir, no mínimo, as três primeiras dimensões. A pergunta "QUEM?" refere-se a um nível de análise mais detalhado.

Agregações:

Fator importante a analisar durante a modelagem já que o principal objetivo de uma agregação é prever um aumento de performance no acesso aos dados e reduzir o volume de dados na tabela.

Resumo das etapas importantes para a modelagem dimensional:

1. Obter a necessidade executiva: qual o negócio objeto de gestão;
2. Entender quais são os indicadores de valor do negócio, definindo as métricas de controle;
3. Descrever o negócio em um modelo de dados conceitual;
4. Se houver modelos de relatórios do ambiente operacional: simular;
5. Definir o que descreve o negócio: dimensões;
6. Definir a organização da dimensão tempo: qual a menor unidade de tempo;
7. Trabalhar com o conceito de hierarquia nas dimensões de consulta. Por exemplo, uma hierarquia na dimensão tempo: Dia -> Mês -> Trimestre -> Semestre -> Ano;
8. Definir a Tabela Fatos;
9. Montar o "star schema".

A modelagem dimensional é aplicada para visualizarmos os dados referentes às necessidades do usuário para obtenção da informação em um sistema analítico. O modelo de dados dimensional independe de em qual banco de dados será implementado (ROLAP ou MOLAP).

Muito importante é documentar o modelo de dados em uma ferramenta CASE na qual se definem as tabelas, atributos, domínios, enfim, todas as informações necessárias para garantir a plena documentação do modelo e facilitar a geração dos esquemas físicos no banco de dados, principalmente em bancos de dados relacionais. Outro fator importante são os metadados que podem ser entendidos como verdadeiros documentadores dos dados que compõem um *Data Warehouse*. Estas informações são armazenadas em um repositório que tem como objetivo principal documentar e administrar os metadados. Os metadados englobam o *Data Warehouse* e mantêm informações sobre a estrutura dos dados, as fontes de dados, a transformação dos dados, as rotinas de extração de dados, o modelo de dados, os relacionamentos e demais informações pertinentes a dar significado ao dado.

3.2.8. Aspectos da implementação física

Um *Data Warehouse* exige grande capacidade de armazenamento e processamento dos dados, pois armazena dados analíticos destinados às necessidades de tomada de decisão. Estes dados podem ser armazenados tanto em um banco de dados relacional (ROLAP) quanto em um multidimensional (MOLAP). A diferença básica é que em uma estrutura ROLAP deve-se criar vários índices atrelados às tabelas fatos e dimensões para um acesso mais rápido e eficiente ao banco de dados, enquanto que, em um banco multidimensional, somente precisa-se informar quais são as dimensões e os fatos, pois o próprio banco encarrega-se de gerar os cubos.

Levantamento de volumes de dados:

É necessário analisar a necessidade de sumarizar os dados (agregações). Um fator importante é que as agregações reduzem o volume de dados, considerando que um sistema para tomada de decisões geralmente não analisa o dado em nível detalhado, pois esta informação já se encontra no ambiente operacional (OLTP). O volume de dados a ser carregado e a previsão de crescimento é fator importante ao decidir qual tipo de banco de dados será utilizado (MOLAP ou ROLAP). Um banco MOLAP não suporta eficientemente um volume muito grande de dados. Ao analisar o volume de dados na tabela fatos, há de considerar o crescimento no número de linhas na tabela, número de dimensões associadas e qual o nível de detalhe desejado. Muitas vezes, é importante analisar a necessidade de separar a tabela fatos em outras tabelas menores ou em agregações em nível mais alto.

Periodicidade de Carga:

As rotinas de extração dos dados do ambiente operacional devem ser executadas de acordo com um esquema de processamento pré-definido. A periodicidade de execução pode ser diária (inviável quando a extração necessita de muitas fontes externas), semanal, quinzenal, mensal etc, sempre levando em consideração o volume de processamento.

Tempo de Armazenagem dos Dados:

De acordo com o volume de dados carregados de forma incremental no Data Warehouse, é importante estimar por quanto tempo estes dados deverão estar disponíveis na base de dados. O principal objetivo do Data Warehouse é manter o histórico dos dados por um período de tempo necessário para a análise. Tudo depende do tipo de negócio do cliente. Geralmente a carga inicial do Data Warehouse corresponde a um volume histórico de 3 anos (no mínimo) para dar um respaldo necessário ao processo analítico. De acordo com o crescimento do banco de dados, é necessário gerenciar o expurgo das informações não mais utilizadas em um determinado período de tempo. Caso o usuário queira manter as informações por um período de tempo longo, deve-se deixar claro que o custo de manutenção de "hardware" e "software" aumentará.

Controle de "Backups":

O administrador do *Data Warehouse* deve controlar como o "backup" das bases de dados será efetuado (periodicidade, volume, dispositivos de armazenamento etc). É importante verificar se há necessidade de copiar todo o banco de dados (visto que, em projetos muito grandes, o custo desta atividade é alto e torna-se inviável) ou apenas um determinado período de tempo. Obviamente o "backup" é importante e deve ser analisado. Vale a pena verificar se as informações originárias dos sistemas transacionais para carga do *Data Warehouse* são facilmente recuperáveis pelas rotinas de extração, aproveitando o "backup" das bases operacionais (bancos de dados, arquivos sequenciais, etc) em caso de problemas. Neste caso pode-se utilizar como "backup" os próprios arquivos oriundos do sistema transacional com os movimentos (diários, semanais, mensais etc) carregados no *Data Warehouse*. Tudo depende de como gerenciar este processo e analisar por quanto tempo pode ficar indisponível em caso de recuperação do banco de dados e estimar quanto tempo levará a recarga a partir dos "backups" operacionais.

Análise de Performance:

Indica-se a técnica "star schema" pelos principais motivos na implementação física em um servidor ROLAP:

- As tabelas fatos contêm a maioria das linhas de dados;
- As tabelas dimensões tendem a ter um número menor de linhas;
- Os "joins" (junções, relacionamentos) em um "star schema" tipicamente envolvem uma tabela fatos grande e uma ou mais tabelas dimensões pequenas, o que torna a operação de consulta mais rápida;
- Sempre consultar as tabelas fatos pelas entradas, que são as dimensões;
- Esquema altamente desnormalizado para melhorar performance.

A desnormalização provoca redundância e duplicação de dados, o que privilegia a performance em detrimento do consumo de espaço.

Uma análise importante refere-se aos índices a serem criados nas tabelas fatos e dimensões quando da implementação em uma estrutura ROLAP. Os índices são extremamente importantes, pois permitem o acesso às informações de uma maneira rápida, visto que o processo de tomada de decisão pode envolver consultas complexas que necessitem acessar um grande número de registros. Geralmente, criam-se índices das chaves primárias nas tabelas dimensões e fatos mais outros atributos das dimensões necessários às consultas e também à combinação de dimensões na tabela fatos. Ao criarmos índices em bancos de dados relacionais, devemos considerar a granularidade das informações, a seletividade e alguns outros fatores pertinentes à estrutura do banco, pois os bancos relacionais analisam seu próprio plano de acesso antes de decidir acessar à estrutura de índices ou se é mais eficiente dar um *scan table* (pesquisa em toda a tabela). Geralmente, definem-se índices *clustered* para cada chave primária. Este tipo de suporte é fornecido pelo administrador do banco de dados (DBA).

Ao utilizarmos um servidor MOLAP (com um banco multidimensional), não é necessário criar índices, visto que a própria estrutura deste tipo de banco de dados cria os cubos com todas as combinações possíveis das dimensões. Devemos informar ao banco de dados quais são as dimensões e quais são os fatos. Esta tecnologia multidimensional tem alta performance, porém ocupa um alto espaço de armazenamento, dependendo do volume de dados da tabela fatos. A tendência dos bancos de dados é reunir funcionalidades ROLAP e MOLAP.

Segurança:

Além de utilizar o esquema de segurança de "logins" do próprio banco de dados (chaves autorizadas e o nível de acesso de cada uma), é importante que o administrador do Data Warehouse tenha um controle do perfil de usuários com autorização de acesso. Muitas ferramentas OLAP possuem um controle de segurança próprio. Outra alternativa é o desenvolvimento de um aplicativo específico de controle de usuários autorizados, antes de executarem uma consulta ao banco de dados onde reside o "data warehouse"/"data mart".

3.2.9. Aspectos da visualização das informações

Queries Simples:

São comandos SQL simples desenvolvidos pelo usuário, fazendo o acesso diretamente ao SGBDR e retornando às linhas acessadas em uma forma tabular. Este usuário deve possuir um perfil técnico avançado na criação de consultas SQL.

Stored Procedures (SP):

Uma "stored procedure" fica armazenada no banco de dados relacional e o acesso é mais rápido. Recomenda-se desenvolver "stored procedures" quando as consultas são pré-definidas pelo usuário.

Ferramentas OLAP:

As ferramentas OLAP de visualização dos dados transformam números enfadonhos em excitantes apresentações visuais e são ferramentas facilmente operadas pelo usuário final.

Aplicativos de Consulta:

Muitas vezes o usuário solicita "queries" pré-definidas que podem ser armazenadas no banco de dados na forma de "stored procedures" ou gravadas de acordo com o tipo de ferramenta OLAP. A viabilidade de construir um aplicativo de consulta está diretamente relacionada às necessidades de informação do alto escalão da empresa (presidência, diretoria, gerências). Obviamente estes funcionários não irão desenvolver "queries", e sim desejam somente clicar em um ícone para obter a informação. Nestes casos, um aplicativo de consulta que integra as "queries" pré-definidas deve ser implementado, prevendo a integração de novas consultas que são executadas frequentemente.

Para consultas esporádicas, não se justifica implementá-las em um aplicativo, pois o usuário pode desenvolvê-las de acordo com sua necessidade. Depende da análise do usuário se quer ou não implementar determinada consulta em um aplicativo.

4. Estudo de Caso

Para esse estudo de caso, vamos utilizar o Dataset “ds_salarios”. Nele constam diversos cargos dentro da Tecnologia da Informação, mais especificamente na área de Analytics. Segue abaixo o dicionário de dados:

COLUNA	TIPO	OBSERVAÇÃO
id	integer	Chave primária
ano	integer	Ano da contratação
nivel_experiencia	character varying	Nível de experiência
tipo_trabalho	character varying	Tipo de trabalho
titulo	character varying	Nome do cargo
salario	numeric(18,2)	Valor do salário na moeda corrente
moeda_corrente	character varying	Moeda corrente
salario_dolar	numeric(18,2)	Valor do salário em Dólar
residencia_empregado	character varying	Localização do país de residência do empregado
percentual_remoto	character varying	Percentual de trabalho remoto
localizacao_empresa	character varying	Localização do país da empresa
tamanho_empresa	character varying	Porte da empresa

A seguir, tem-se uma amostra dos dados. Em anexo a essa apostila, temos o script para criação da tabela e a inserção das informações no Anexo I.

id	ano	nivel_experiencia	tipo_trabalho	titulo	salario	moeda_corrente	salario_dolar	residencia_empregado	percentual_remoto	localizacao_empresa	tamanho_empresa
0	2020	MI	FT	Data Scientist	70000	EUR	79833	DE	0	DE	L
1	2020	SE	FT	Machine Learning Scientist	260000	USD	260000	JP	0	JP	S
2	2020	SE	FT	Big Data Engineer	85000	GBP	109024	GB	50	GB	M
3	2020	MI	FT	Product Data Analyst	20000	USD	20000	HN	0	HN	S
4	2020	SE	FT	Machine Learning Engineer	150000	USD	150000	US	50	US	L
5	2020	EN	FT	Data Analyst	72000	USD	72000	US	100	US	L
6	2020	SE	FT	Lead Data Scientist	190000	USD	190000	US	100	US	S
7	2020	MI	FT	Data Scientist	11000000	HUF	35735	HU	50	HU	L
8	2020	MI	FT	Business Data Analyst	135000	USD	135000	US	100	US	L
9	2020	SE	FT	Lead Data Engineer	125000	USD	125000	NZ	50	NZ	S
10	2020	EN	FT	Data Scientist	45000	EUR	51321	FR	0	FR	S

Tabela 1 – Dataset

4.1. Definindo o modelo de dados dimensional

O primeiro passo é separar os atributos categóricos do Dataset. São eles:

- ano;
- nivel_experiencia;
- tipo_trabalho;
- titulo;
- moeda_corrente;
- residencia_empregado;
- localizacao_empresa;
- tamanho_empresa.

Para o atributo “ano”, iremos criar a dimensão tempo, e para os atributos “localização_empresa” e “residencia_empregado” iremos criar a dimensão “pais”. Os demais atributos irão se transformar em dimensões levando seus respectivos nomes. O modelo de dados utilizado será o “star schema”, conforme imagem abaixo.

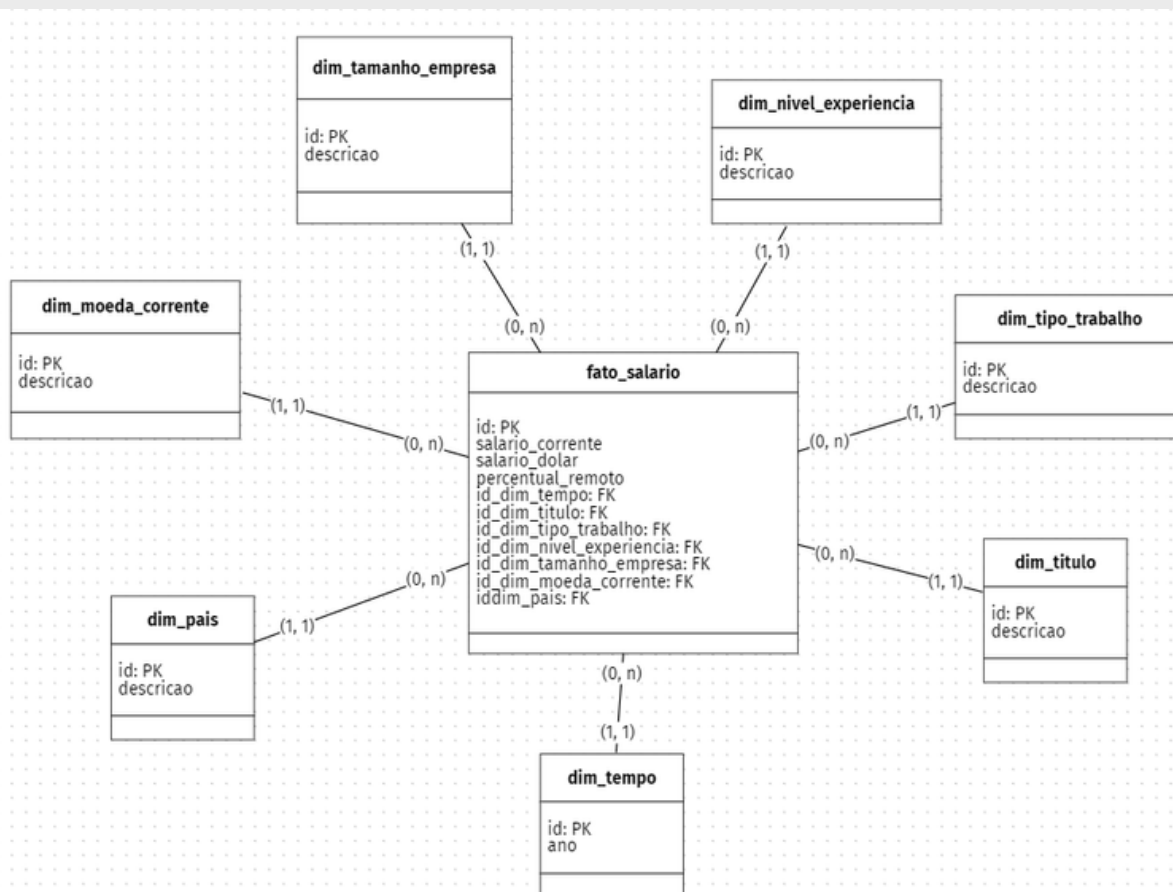


Figura 9 – Modelo Dimensional de Salários

Com o modelo de dados dimensional criado, o próximo passo será desenvolver uma solução de ETL para a carga desse novo modelo. Os dados serão extraídos do Dataset de origem e serão transportados para o modelo dimensional.



Referências bibliográficas

ALVES, William Pereira. **Banco de Dados [BV:MB]**. 1ª ed. São Paulo: Érica, 2014.

BALIEIRO, R. **Banco de Dados [BV:RE]**. 1ª ed. Rio de Janeiro: SESES, 2015.

ELMASRI, R.; NAVATHE, Shamkant B. **S. Sistemas de Banco de Dados [BV:PE]**. 7. ed. São Paulo: Pearson, 2018.

FONSECA, Cleber Costa da. **Implementação de banco de dados. Banco de Dados [BV:RE]**. 1ª ed. Rio de Janeiro: SESES, 2016.

HEUSER, C. **Projeto de Banco de Dados [BV:MB]**. 6. ed. Porto Alegre: ARTMED, 2009.

MACHADO, Felipe N. R. **Banco de Dados - Projeto e Implementação [BV:MB]**. 3. ed. São Paulo: Érica, 2014.

MACHADO, Felipe N. R. **Tecnologia e Projetos de Data Warehouse**. 6. ed. São Paulo: Érica, 2013.

NETO, Geraldo H. **MODELAGEM DE DADOS [BV:RE]**. 1ª ed. Rio de Janeiro: SESES, 2015.

PUGA, Sandra; FRANÇA, Edson; GOYA, Milton. **Banco de Dados: implementação em SQL, PL/SQL e Oracle 11g [BV:PE]**. 1ª ed. São Paulo: Pearson, 2013.

RAMAKRISHNAN, R. **Sistemas de gerenciamento de banco de dados [BE:MB]**. 3. ed. Porto Alegre: McGraw-Hill, 2008.



ANEXO I

Criação da tabela.

```
CREATE TABLE IF NOT EXISTS public.salarios
(
    id serial,
    ano integer,
    nivel_experiencia character varying,
    tipo_trabalho character varying,
    titulo character varying,
    salario numeric(18,2),
    moeda_corrente character varying,
    salario_dolar numeric(18,2),
    residencia_empregado character varying,
    percentual_remoto integer,
    localizacao_empresa character varying,
    tamanho_empresa character varying,
    CONSTRAINT salarios_pkey PRIMARY KEY (id)
)

TABLESPACE pg_default;

ALTER TABLE public.salarios
    OWNER to postgres;
```

Inserção de Dados.

```

INSERT INTO public.salarios Values(0,2020,'MI','FT','Data Scientist',70000,'EUR',79833,'DE','0','DE','L');
INSERT INTO public.salarios Values(1,2020,'SE','FT','Machine Learning Scientist',260000,'USD',260000,'JP','0','JP','S');
INSERT INTO public.salarios Values(2,2020,'SE','FT','Big Data Engineer',85000,'GBP',109024,'GB','50','GB','M');
INSERT INTO public.salarios Values(3,2020,'MI','FT','Product Data Analyst',20000,'USD',20000,'HN','0','HN','S');
INSERT INTO public.salarios Values(4,2020,'SE','FT','Machine Learning Engineer',150000,'USD',150000,'US','50','US','L');
INSERT INTO public.salarios Values(5,2020,'EN','FT','Data Analyst',72000,'USD',72000,'US','100','US','L');
INSERT INTO public.salarios Values(6,2020,'SE','FT','Lead Data Scientist',190000,'USD',190000,'US','100','US','S');
INSERT INTO public.salarios Values(7,2020,'MI','FT','Data Scientist',11000000,'HUF',35735,'HU','50','HU','L');
INSERT INTO public.salarios Values(8,2020,'MI','FT','Business Data Analyst',135000,'USD',135000,'US','100','US','L');
INSERT INTO public.salarios Values(9,2020,'SE','FT','Lead Data Engineer',125000,'USD',125000,'NZ','50','NZ','S');
INSERT INTO public.salarios Values(10,2020,'EN','FT','Data Scientist',45000,'EUR',51321,'FR','0','FR','S');
INSERT INTO public.salarios Values(11,2020,'MI','FT','Data Scientist',3000000,'INR',40481,'IN','0','IN','L');
INSERT INTO public.salarios Values(12,2020,'EN','FT','Data Scientist',35000,'EUR',39916,'FR','0','FR','M');
INSERT INTO public.salarios Values(13,2020,'MI','FT','Lead Data Analyst',87000,'USD',87000,'US','100','US','L');
INSERT INTO public.salarios Values(14,2020,'MI','FT','Data Analyst',85000,'USD',85000,'US','100','US','L');
INSERT INTO public.salarios Values(15,2020,'MI','FT','Data Analyst',8000,'USD',8000,'PK','50','PK','L');
INSERT INTO public.salarios Values(16,2020,'EN','FT','Data Engineer',4450000,'JPY',41689,'JP','100','JP','S');
INSERT INTO public.salarios Values(17,2020,'SE','FT','Big Data Engineer',100000,'EUR',114047,'PL','100','GB','S');
INSERT INTO public.salarios Values(18,2020,'EN','FT','Data Science Consultant',423000,'INR',5707,'IN','50','IN','M');
INSERT INTO public.salarios Values(19,2020,'MI','FT','Lead Data Engineer',56000,'USD',56000,'PT','100','US','M');
INSERT INTO public.salarios Values(20,2020,'MI','FT','Machine Learning Engineer',299000,'CNY',43331,'CN','0','CN','M');
INSERT INTO public.salarios Values(21,2020,'MI','FT','Product Data Analyst',450000,'INR',6072,'IN','100','IN','L');

```

```

INSERT INTO public.salarios Values(22,2020,'SE','FT','Data Engineer',42000,'EUR',47899,'GR','50','GR','L');
INSERT INTO public.salarios Values(23,2020,'MI','FT','BI Data Analyst',98000,'USD',98000,'US','0','US','M');
INSERT INTO public.salarios Values(24,2020,'MI','FT','Lead Data
Scientist',115000,'USD',115000,'AE','0','AE','L');
INSERT INTO public.salarios Values(25,2020,'EX','FT','Director of Data
Science',325000,'USD',325000,'US','100','US','L');
INSERT INTO public.salarios Values(26,2020,'EN','FT','Research
Scientist',42000,'USD',42000,'NL','50','NL','L');
INSERT INTO public.salarios Values(27,2020,'SE','FT','Data
Engineer',720000,'MXN',33511,'MX','0','MX','S');
INSERT INTO public.salarios Values(28,2020,'EN','CT','Business Data
Analyst',100000,'USD',100000,'US','100','US','L');
INSERT INTO public.salarios Values(29,2020,'SE','FT','Machine Learning
Manager',157000,'CAD',117104,'CA','50','CA','L');
INSERT INTO public.salarios Values(30,2020,'MI','FT','Data Engineering
Manager',51999,'EUR',59303,'DE','100','DE','S');
INSERT INTO public.salarios Values(31,2020,'EN','FT','Big Data
Engineer',70000,'USD',70000,'US','100','US','L');
INSERT INTO public.salarios Values(32,2020,'SE','FT','Data Scientist',60000,'EUR',68428,'GR','100','US','L');
INSERT INTO public.salarios Values(33,2020,'MI','FT','Research
Scientist',450000,'USD',450000,'US','0','US','M');
INSERT INTO public.salarios Values(34,2020,'MI','FT','Data Analyst',41000,'EUR',46759,'FR','50','FR','L');
INSERT INTO public.salarios Values(35,2020,'MI','FT','Data Engineer',65000,'EUR',74130,'AT','50','AT','L');
INSERT INTO public.salarios Values(36,2020,'MI','FT','Data Science
Consultant',103000,'USD',103000,'US','100','US','L');
INSERT INTO public.salarios Values(37,2020,'EN','FT','Machine Learning
Engineer',250000,'USD',250000,'US','50','US','L');
INSERT INTO public.salarios Values(38,2020,'EN','FT','Data Analyst',10000,'USD',10000,'NG','100','NG','S');
INSERT INTO public.salarios Values(39,2020,'EN','FT','Machine Learning
Engineer',138000,'USD',138000,'US','100','US','S');
INSERT INTO public.salarios Values(40,2020,'MI','FT','Data
Scientist',45760,'USD',45760,'PH','100','US','S');
INSERT INTO public.salarios Values(41,2020,'EX','FT','Data Engineering
Manager',70000,'EUR',79833,'ES','50','ES','L');
INSERT INTO public.salarios Values(42,2020,'MI','FT','Machine Learning Infrastructure
Engineer',44000,'EUR',50180,'PT','0','PT','M');
INSERT INTO public.salarios Values(43,2020,'MI','FT','Data
Engineer',106000,'USD',106000,'US','100','US','L');
INSERT INTO public.salarios Values(44,2020,'MI','FT','Data
Engineer',88000,'GBP',112872,'GB','50','GB','L');

```

```

INSERT INTO public.salarios Values(45,2020,'EN','PT','ML Engineer',14000,'EUR',15966,'DE','100','DE','S');
INSERT INTO public.salarios Values(46,2020,'MI','FT','Data
Scientist',60000,'GBP',76958,'GB','100','GB','S');
INSERT INTO public.salarios Values(47,2020,'SE','FT','Data
Engineer',188000,'USD',188000,'US','100','US','L');

INSERT INTO public.salarios Values(48,2020,'MI','FT','Data
Scientist',105000,'USD',105000,'US','100','US','L');
INSERT INTO public.salarios Values(49,2020,'MI','FT','Data Engineer',61500,'EUR',70139,'FR','50','FR','L');
INSERT INTO public.salarios Values(50,2020,'EN','FT','Data Analyst',450000,'INR',6072,'IN','0','IN','S');
INSERT INTO public.salarios Values(51,2020,'EN','FT','Data Analyst',91000,'USD',91000,'US','100','US','L');
INSERT INTO public.salarios Values(52,2020,'EN','FT','AI Scientist',300000,'DKK',45896,'DK','50','DK','S');
INSERT INTO public.salarios Values(53,2020,'EN','FT','Data
Engineer',48000,'EUR',54742,'PK','100','DE','L');
INSERT INTO public.salarios Values(54,2020,'SE','FL','Computer
Engineer',60000,'USD',60000,'RU','100','US','S');
INSERT INTO public.salarios Values(55,2020,'SE','FT','Principal
Scientist',130000,'EUR',148261,'DE','100','DE','M');
INSERT INTO public.salarios Values(56,2020,'MI','FT','Data Scientist',34000,'EUR',38776,'ES','100','ES','M');
INSERT INTO public.salarios Values(57,2020,'MI','FT','Data
Scientist',118000,'USD',118000,'US','100','US','M');
INSERT INTO public.salarios Values(58,2020,'SE','FT','Data
Scientist',120000,'USD',120000,'US','50','US','L');
INSERT INTO public.salarios Values(59,2020,'MI','FT','Data
Scientist',138350,'USD',138350,'US','100','US','M');
INSERT INTO public.salarios Values(60,2020,'MI','FT','Data
Engineer',110000,'USD',110000,'US','100','US','L');
INSERT INTO public.salarios Values(61,2020,'MI','FT','Data
Engineer',130800,'USD',130800,'ES','100','US','M');
INSERT INTO public.salarios Values(62,2020,'EN','PT','Data Scientist',19000,'EUR',21669,'IT','50','IT','S');
INSERT INTO public.salarios Values(63,2020,'SE','FT','Data
Scientist',412000,'USD',412000,'US','100','US','L');
INSERT INTO public.salarios Values(64,2020,'SE','FT','Machine
Engineer',40000,'EUR',45618,'HR','100','HR','S');
INSERT INTO public.salarios Values(65,2020,'EN','FT','Data Scientist',55000,'EUR',62726,'DE','50','DE','S');
INSERT INTO public.salarios Values(66,2020,'EN','FT','Data Scientist',43200,'EUR',49268,'DE','0','DE','S');
INSERT INTO public.salarios Values(67,2020,'SE','FT','Data
Manager',190200,'USD',190200,'US','100','US','M');
INSERT INTO public.salarios Values(68,2020,'EN','FT','Data
Scientist',105000,'USD',105000,'US','100','US','S');

```



```

INSERT INTO public.salarios Values(69,2020,'SE','FT','Data Scientist',80000,'EUR',91237,'AT','0','AT','S');
INSERT INTO public.salarios Values(70,2020,'MI','FT','Data Scientist',55000,'EUR',62726,'FR','50','LU','S');
INSERT INTO public.salarios Values(71,2020,'MI','FT','Data Scientist',37000,'EUR',42197,'FR','50','FR','S');
INSERT INTO public.salarios Values(72,2021,'EN','FT','Research Scientist',60000,'GBP',82528,'GB','50','GB','L');
INSERT INTO public.salarios Values(73,2021,'EX','FT','BI Data Analyst',150000,'USD',150000,'IN','100','US','L');
INSERT INTO public.salarios Values(74,2021,'EX','FT','Head of Data',235000,'USD',235000,'US','100','US','L');
INSERT INTO public.salarios Values(75,2021,'SE','FT','Data Scientist',45000,'EUR',53192,'FR','50','FR','L');
INSERT INTO public.salarios Values(76,2021,'MI','FT','BI Data Analyst',100000,'USD',100000,'US','100','US','M');
INSERT INTO public.salarios Values(77,2021,'MI','PT','3D Computer Vision Researcher',400000,'INR',5409,'IN','50','IN','M');
INSERT INTO public.salarios Values(78,2021,'MI','CT','ML Engineer',270000,'USD',270000,'US','100','US','L');
INSERT INTO public.salarios Values(79,2021,'EN','FT','Data Analyst',80000,'USD',80000,'US','100','US','M');
INSERT INTO public.salarios Values(80,2021,'SE','FT','Data Analytics Engineer',67000,'EUR',79197,'DE','100','DE','L');
INSERT INTO public.salarios Values(81,2021,'MI','FT','Data Engineer',140000,'USD',140000,'US','100','US','L');
INSERT INTO public.salarios Values(82,2021,'MI','FT','Applied Data Scientist',68000,'CAD',54238,'GB','50','CA','L');
INSERT INTO public.salarios Values(83,2021,'MI','FT','Machine Learning Engineer',40000,'EUR',47282,'ES','100','ES','S');
INSERT INTO public.salarios Values(84,2021,'EX','FT','Director of Data Science',130000,'EUR',153667,'IT','100','PL','L');
INSERT INTO public.salarios Values(85,2021,'MI','FT','Data Engineer',110000,'PLN',28476,'PL','100','PL','L');
INSERT INTO public.salarios Values(86,2021,'EN','FT','Data Analyst',50000,'EUR',59102,'FR','50','FR','M');
INSERT INTO public.salarios Values(87,2021,'MI','FT','Data Analytics Engineer',110000,'USD',110000,'US','100','US','L');
INSERT INTO public.salarios Values(88,2021,'SE','FT','Lead Data Analyst',170000,'USD',170000,'US','100','US','L');
INSERT INTO public.salarios Values(89,2021,'SE','FT','Data Analyst',80000,'USD',80000,'BG','100','US','S');
INSERT INTO public.salarios Values(90,2021,'SE','FT','Marketing Data Analyst',75000,'EUR',88654,'GR','100','DK','L');
INSERT INTO public.salarios Values(91,2021,'EN','FT','Data Science Consultant',65000,'EUR',76833,'DE','100','DE','S');
INSERT INTO public.salarios Values(92,2021,'MI','FT','Lead Data Analyst',1450000,'INR',19609,'IN','100','IN','L');

```

```

INSERT INTO public.salarios Values(93,2021,'SE','FT','Lead Engineer',276000,'USD',276000,'US','0','US','L');
INSERT INTO public.salarios Values(94,2021,'EN','FT','Data Scientist',2200000,'INR',29751,'IN','50','IN','L');
INSERT INTO public.salarios Values(95,2021,'MI','FT','Cloud Engineer',120000,'SGD',89294,'SG','50','SG','L');
INSERT INTO public.salarios Values(96,2021,'EN','PT','AI Scientist',12000,'USD',12000,'BR','100','US','S');
INSERT INTO public.salarios Values(97,2021,'MI','FT','Financial Analyst',450000,'USD',450000,'US','100','US','L');
INSERT INTO public.salarios Values(98,2021,'EN','FT','Computer Engineer',70000,'USD',70000,'US','100','US','M');
INSERT INTO public.salarios Values(99,2021,'MI','FT','Computer Engineer',81000,'EUR',95746,'DE','100','US','S');
INSERT INTO public.salarios Values(100,2021,'MI','FT','Data Analyst',75000,'USD',75000,'US','0','US','L');
INSERT INTO public.salarios Values(101,2021,'SE','FT','Data Engineer',150000,'USD',150000,'US','100','US','L');
INSERT INTO public.salarios Values(102,2021,'MI','FT','BI Analyst',11000000,'HUF',36259,'HU','50','US','L');
INSERT INTO public.salarios Values(103,2021,'MI','FT','Data Analyst',62000,'USD',62000,'US','0','US','L');
INSERT INTO public.salarios Values(104,2021,'MI','FT','Data Scientist',73000,'USD',73000,'US','0','US','L');
INSERT INTO public.salarios Values(105,2021,'MI','FT','Data Analyst',37456,'GBP',51519,'GB','50','GB','L');
INSERT INTO public.salarios Values(106,2021,'MI','FT','Research Scientist',235000,'CAD',187442,'CA','100','CA','L');
INSERT INTO public.salarios Values(107,2021,'SE','FT','Data Engineer',115000,'USD',115000,'US','100','US','S');
INSERT INTO public.salarios Values(108,2021,'SE','FT','Data Engineer',150000,'USD',150000,'US','100','US','M');
INSERT INTO public.salarios Values(109,2021,'EN','FT','Data Engineer',2250000,'INR',30428,'IN','100','IN','L');
INSERT INTO public.salarios Values(110,2021,'SE','FT','Machine Engineer',80000,'EUR',94564,'DE','50','DE','L');
INSERT INTO public.salarios Values(111,2021,'SE','FT','Director of Engineering',82500,'GBP',113476,'GB','100','GB','M');
INSERT INTO public.salarios Values(112,2021,'SE','FT','Lead Engineer',75000,'GBP',103160,'GB','100','GB','S');
INSERT INTO public.salarios Values(113,2021,'EN','PT','AI Scientist',12000,'USD',12000,'PK','100','US','M');
INSERT INTO public.salarios Values(114,2021,'MI','FT','Data Engineer',38400,'EUR',45391,'NL','100','NL','L');
INSERT INTO public.salarios Values(115,2021,'EN','FT','Machine Scientist',225000,'USD',225000,'US','100','US','L');
INSERT INTO public.salarios Values(116,2021,'MI','FT','Data Scientist',50000,'USD',50000,'NG','100','NG','L');

```

```

INSERT INTO public.salarios Values(117,2021,'MI','FT','Data Science
Engineer',34000,'EUR',40189,'GR','100','GR','M');
INSERT INTO public.salarios Values(118,2021,'EN','FT','Data
Analyst',90000,'USD',90000,'US','100','US','S');
INSERT INTO public.salarios Values(119,2021,'MI','FT','Data
Engineer',200000,'USD',200000,'US','100','US','L');
INSERT INTO public.salarios Values(120,2021,'MI','FT','Big Data
Engineer',60000,'USD',60000,'ES','50','RO','M');
INSERT INTO public.salarios Values(121,2021,'SE','FT','Principal Data
Engineer',200000,'USD',200000,'US','100','US','M');
INSERT INTO public.salarios Values(122,2021,'EN','FT','Data
Analyst',50000,'USD',50000,'US','100','US','M');
INSERT INTO public.salarios Values(123,2021,'EN','FT','Applied Data
Scientist',80000,'GBP',110037,'GB','0','GB','L');
INSERT INTO public.salarios Values(124,2021,'EN','PT','Data Analyst',8760,'EUR',10354,'ES','50','ES','M');
INSERT INTO public.salarios Values(125,2021,'MI','FT','Principal Data
Scientist',151000,'USD',151000,'US','100','US','L');
INSERT INTO public.salarios Values(126,2021,'SE','FT','Machine Learning
Scientist',120000,'USD',120000,'US','50','US','S');
INSERT INTO public.salarios Values(127,2021,'MI','FT','Data Scientist',700000,'INR',9466,'IN','0','IN','S');
INSERT INTO public.salarios Values(128,2021,'EN','FT','Machine Learning
Engineer',20000,'USD',20000,'IN','100','IN','S');
INSERT INTO public.salarios Values(129,2021,'SE','FT','Lead Data
Scientist',3000000,'INR',40570,'IN','50','IN','L');
INSERT INTO public.salarios Values(130,2021,'EN','FT','Machine Learning
Developer',100000,'USD',100000,'IQ','50','IQ','S');
INSERT INTO public.salarios Values(131,2021,'EN','FT','Data
Scientist',42000,'EUR',49646,'FR','50','FR','M');
INSERT INTO public.salarios Values(132,2021,'MI','FT','Applied Machine Learning
Scientist',38400,'USD',38400,'VN','100','US','M');
INSERT INTO public.salarios Values(133,2021,'SE','FT','Computer Vision
Engineer',24000,'USD',24000,'BR','100','BR','M');

INSERT INTO public.salarios Values(134,2021,'EN','FT','Data
Scientist',100000,'USD',100000,'US','0','US','S');
INSERT INTO public.salarios Values(135,2021,'MI','FT','Data
Analyst',90000,'USD',90000,'US','100','US','M');
INSERT INTO public.salarios Values(136,2021,'MI','FT','ML Engineer',7000000,'JPY',63711,'JP','50','JP','S');
INSERT INTO public.salarios Values(137,2021,'MI','FT','ML Engineer',8500000,'JPY',77364,'JP','50','JP','S');
INSERT INTO public.salarios Values(138,2021,'SE','FT','Principal Data
Scientist',220000,'USD',220000,'US','0','US','L');

```

```

INSERT          INTO          public.salarios          Values(139,2021,'EN','FT','Data
Scientist',80000,'USD',80000,'US','100','US','M');
INSERT          INTO          public.salarios          Values(140,2021,'MI','FT','Data
Analyst',135000,'USD',135000,'US','100','US','L');
INSERT          INTO          public.salarios          Values(141,2021,'SE','FT','Data          Science
Manager',240000,'USD',240000,'US','0','US','L');
INSERT          INTO          public.salarios          Values(142,2021,'SE','FT','Data          Engineering
Manager',150000,'USD',150000,'US','0','US','L');
INSERT          INTO          public.salarios          Values(143,2021,'MI','FT','Data
Scientist',82500,'USD',82500,'US','100','US','S');
INSERT          INTO          public.salarios          Values(144,2021,'MI','FT','Data
Engineer',100000,'USD',100000,'US','100','US','L');
INSERT          INTO          public.salarios          Values(145,2021,'SE','FT','Machine          Learning
Engineer',70000,'EUR',82744,'BE','50','BE','M');
INSERT          INTO          public.salarios          Values(146,2021,'MI','FT','Research
Scientist',53000,'EUR',62649,'FR','50','FR','M');
INSERT          INTO          public.salarios          Values(147,2021,'MI','FT','Data
Engineer',90000,'USD',90000,'US','100','US','L');
INSERT          INTO          public.salarios          Values(148,2021,'SE','FT','Data          Engineering
Manager',153000,'USD',153000,'US','100','US','L');
INSERT          INTO          public.salarios          Values(149,2021,'SE','FT','Cloud          Data
Engineer',160000,'USD',160000,'BR','100','US','S');
INSERT          INTO          public.salarios          Values(150,2021,'SE','FT','Director          of          Data
Science',168000,'USD',168000,'JP','0','JP','S');
INSERT          INTO          public.salarios          Values(151,2021,'MI','FT','Data
Scientist',150000,'USD',150000,'US','100','US','M');
INSERT          INTO          public.salarios          Values(152,2021,'MI','FT','Data
Scientist',95000,'CAD',75774,'CA','100','CA','L');
INSERT          INTO          public.salarios          Values(153,2021,'EN','FT','Data
Scientist',13400,'USD',13400,'UA','100','UA','L');
INSERT          INTO          public.salarios          Values(154,2021,'SE','FT','Data          Science
Manager',144000,'USD',144000,'US','100','US','L');
INSERT          INTO          public.salarios          Values(155,2021,'SE','FT','Data          Science
Engineer',159500,'CAD',127221,'CA','50','CA','L');
INSERT          INTO          public.salarios          Values(156,2021,'MI','FT','Data
Scientist',160000,'SGD',119059,'SG','100','IL','M');

```



```

INSERT INTO public.salarios Values(157,2021,'MI','FT','Applied Machine Learning
Scientist',423000,'USD',423000,'US','50','US','L');
INSERT INTO public.salarios Values(158,2021,'SE','FT','Data Analytics
Manager',120000,'USD',120000,'US','100','US','M');
INSERT INTO public.salarios Values(159,2021,'EN','FT','Machine Learning
Engineer',125000,'USD',125000,'US','100','US','S');
INSERT INTO public.salarios Values(160,2021,'EX','FT','Head of
Data',230000,'USD',230000,'RU','50','RU','L');
INSERT INTO public.salarios Values(161,2021,'EX','FT','Head of Data
Science',85000,'USD',85000,'RU','0','RU','M');
INSERT INTO public.salarios Values(162,2021,'MI','FT','Data
Engineer',24000,'EUR',28369,'MT','50','MT','L');
INSERT INTO public.salarios Values(163,2021,'EN','FT','Data Science
Consultant',54000,'EUR',63831,'DE','50','DE','L');
INSERT INTO public.salarios Values(164,2021,'EX','FT','Director of Data
Science',110000,'EUR',130026,'DE','50','DE','M');
INSERT INTO public.salarios Values(165,2021,'SE','FT','Data
Specialist',165000,'USD',165000,'US','100','US','L');
INSERT INTO public.salarios Values(166,2021,'EN','FT','Data
Engineer',80000,'USD',80000,'US','100','US','L');
INSERT INTO public.salarios Values(167,2021,'EX','FT','Director of Data
Science',250000,'USD',250000,'US','0','US','L');
INSERT INTO public.salarios Values(168,2021,'EN','FT','BI Data
Analyst',55000,'USD',55000,'US','50','US','S');
INSERT INTO public.salarios Values(169,2021,'MI','FT','Data
Architect',150000,'USD',150000,'US','100','US','L');
INSERT INTO public.salarios Values(170,2021,'MI','FT','Data
Architect',170000,'USD',170000,'US','100','US','L');
INSERT INTO public.salarios Values(171,2021,'MI','FT','Data
Engineer',60000,'GBP',82528,'GB','100','GB','L');
INSERT INTO public.salarios Values(172,2021,'EN','FT','Data
Analyst',60000,'USD',60000,'US','100','US','S');
INSERT INTO public.salarios Values(173,2021,'SE','FT','Principal Data
Scientist',235000,'USD',235000,'US','100','US','L');
INSERT INTO public.salarios Values(174,2021,'SE','FT','Research
Scientist',51400,'EUR',60757,'PT','50','PT','L');
INSERT INTO public.salarios Values(175,2021,'SE','FT','Data Engineering
Manager',174000,'USD',174000,'US','100','US','L');

```



```

INSERT INTO public.salarios Values(176,2021,'MI','FT','Data Scientist',58000,'MXN',2859,'MX','0','MX','S');
INSERT INTO public.salarios Values(177,2021,'MI','FT','Data Scientist',30400000,'CLP',40038,'CL','100','CL','L');
INSERT INTO public.salarios Values(178,2021,'EN','FT','Machine Learning Engineer',81000,'USD',81000,'US','50','US','S');
INSERT INTO public.salarios Values(179,2021,'MI','FT','Data Scientist',420000,'INR',5679,'IN','100','US','S');
INSERT INTO public.salarios Values(180,2021,'MI','FT','Big Data Engineer',1672000,'INR',22611,'IN','0','IN','L');
INSERT INTO public.salarios Values(181,2021,'MI','FT','Data Scientist',76760,'EUR',90734,'DE','50','DE','L');
INSERT INTO public.salarios Values(182,2021,'MI','FT','Data Engineer',22000,'EUR',26005,'RO','0','US','L');
INSERT INTO public.salarios Values(183,2021,'SE','FT','Finance Analyst',45000,'GBP',61896,'GB','50','GB','L');
INSERT INTO public.salarios Values(184,2021,'MI','FL','Machine Learning Scientist',12000,'USD',12000,'PK','50','PK','M');
INSERT INTO public.salarios Values(185,2021,'MI','FT','Data Engineer',4000,'USD',4000,'IR','100','IR','M');
INSERT INTO public.salarios Values(186,2021,'SE','FT','Data Analytics Engineer',50000,'USD',50000,'VN','100','GB','M');
INSERT INTO public.salarios Values(187,2021,'EX','FT','Data Science Consultant',59000,'EUR',69741,'FR','100','ES','S');
INSERT INTO public.salarios Values(188,2021,'SE','FT','Data Engineer',65000,'EUR',76833,'RO','50','GB','S');
INSERT INTO public.salarios Values(189,2021,'MI','FT','Machine Learning Engineer',74000,'USD',74000,'JP','50','JP','S');
INSERT INTO public.salarios Values(190,2021,'SE','FT','Data Science Manager',152000,'USD',152000,'US','100','FR','L');
INSERT INTO public.salarios Values(191,2021,'EN','FT','Machine Learning Engineer',21844,'USD',21844,'CO','50','CO','M');
INSERT INTO public.salarios Values(192,2021,'MI','FT','Big Data Engineer',18000,'USD',18000,'MD','0','MD','S');
INSERT INTO public.salarios Values(193,2021,'SE','FT','Data Science Manager',174000,'USD',174000,'US','100','US','L');
INSERT INTO public.salarios Values(194,2021,'SE','FT','Research Scientist',120500,'CAD',96113,'CA','50','CA','L');
INSERT INTO public.salarios Values(195,2021,'MI','FT','Data Scientist',147000,'USD',147000,'US','50','US','L');
INSERT INTO public.salarios Values(196,2021,'EN','FT','BI Data Analyst',9272,'USD',9272,'KE','100','KE','S');
INSERT INTO public.salarios Values(197,2021,'SE','FT','Machine Learning Engineer',1799997,'INR',24342,'IN','100','IN','L');

```

```

INSERT INTO public.salarios Values(198,2021,'SE','FT','Data Science
Manager',4000000,'INR',54094,'IN','50','US','L');
INSERT INTO public.salarios Values(199,2021,'EN','FT','Data Science
Consultant',90000,'USD',90000,'US','100','US','S');
INSERT INTO public.salarios Values(200,2021,'MI','FT','Data
Scientist',52000,'EUR',61467,'DE','50','AT','M');
INSERT INTO public.salarios Values(201,2021,'SE','FT','Machine Learning Infrastructure
Engineer',195000,'USD',195000,'US','100','US','M');
INSERT INTO public.salarios Values(202,2021,'MI','FT','Data
Scientist',32000,'EUR',37825,'ES','100','ES','L');
INSERT INTO public.salarios Values(203,2021,'SE','FT','Research
Scientist',50000,'USD',50000,'FR','100','US','S');
INSERT INTO public.salarios Values(204,2021,'MI','FT','Data
Scientist',160000,'USD',160000,'US','100','US','L');
INSERT INTO public.salarios Values(205,2021,'MI','FT','Data Scientist',69600,'BRL',12901,'BR','0','BR','S');
INSERT INTO public.salarios Values(206,2021,'SE','FT','Machine Learning
Engineer',200000,'USD',200000,'US','100','US','L');
INSERT INTO public.salarios Values(207,2021,'SE','FT','Data
Engineer',165000,'USD',165000,'US','0','US','M');
INSERT INTO public.salarios Values(208,2021,'MI','FL','Data Engineer',20000,'USD',20000,'IT','0','US','L');
INSERT INTO public.salarios Values(209,2021,'SE','FT','Data Analytics
Manager',120000,'USD',120000,'US','0','US','L');
INSERT INTO public.salarios Values(210,2021,'MI','FT','Machine Learning
Engineer',21000,'EUR',24823,'SI','50','SI','L');
INSERT INTO public.salarios Values(211,2021,'MI','FT','Research
Scientist',48000,'EUR',56738,'FR','50','FR','S');
INSERT INTO public.salarios Values(212,2021,'MI','FT','Data
Engineer',48000,'GBP',66022,'HK','50','GB','S');
INSERT INTO public.salarios Values(213,2021,'EN','FT','Big Data
Engineer',435000,'INR',5882,'IN','0','CH','L');
INSERT INTO public.salarios Values(214,2021,'EN','FT','Machine Learning
Engineer',21000,'EUR',24823,'DE','50','DE','M');
INSERT INTO public.salarios Values(215,2021,'SE','FT','Principal Data
Engineer',185000,'USD',185000,'US','100','US','L');
INSERT INTO public.salarios Values(216,2021,'EN','PT','Computer Vision
Engineer',180000,'DKK',28609,'DK','50','DK','S');
INSERT INTO public.salarios Values(217,2021,'MI','FT','Data Scientist',76760,'EUR',90734,'DE','50','DE','L');
INSERT INTO public.salarios Values(218,2021,'MI','FT','Machine Learning
Engineer',75000,'EUR',88654,'BE','100','BE','M');

```



```

INSERT INTO public.salarios Values(219,2021,'SE','FT','Data Analytics
Manager',140000,'USD',140000,'US','100','US','L');
INSERT INTO public.salarios Values(220,2021,'MI','FT','Machine Learning
Engineer',180000,'PLN',46597,'PL','100','PL','L');
INSERT INTO public.salarios Values(221,2021,'MI','FT','Data
Scientist',85000,'GBP',116914,'GB','50','GB','L');
INSERT INTO public.salarios Values(222,2021,'MI','FT','Data Scientist',2500000,'INR',33808,'IN','0','IN','M');
INSERT INTO public.salarios Values(223,2021,'MI','FT','Data
Scientist',40900,'GBP',56256,'GB','50','GB','L');
INSERT INTO public.salarios Values(224,2021,'SE','FT','Machine Learning
Scientist',225000,'USD',225000,'US','100','CA','L');

INSERT INTO public.salarios Values(225,2021,'EX','CT','Principal Data
Scientist',416000,'USD',416000,'US','100','US','S');
INSERT INTO public.salarios Values(226,2021,'SE','FT','Data
Scientist',110000,'CAD',87738,'CA','100','CA','S');
INSERT INTO public.salarios Values(227,2021,'MI','FT','Data Scientist',75000,'EUR',88654,'DE','50','DE','L');
INSERT INTO public.salarios Values(228,2021,'SE','FT','Data
Scientist',135000,'USD',135000,'US','0','US','L');
INSERT INTO public.salarios Values(229,2021,'SE','FT','Data
Analyst',90000,'CAD',71786,'CA','100','CA','M');
INSERT INTO public.salarios Values(230,2021,'EN','FT','Big Data
Engineer',1200000,'INR',16228,'IN','100','IN','L');
INSERT INTO public.salarios Values(231,2021,'SE','FT','ML
Engineer',256000,'USD',256000,'US','100','US','S');
INSERT INTO public.salarios Values(232,2021,'SE','FT','Director of Data
Engineering',200000,'USD',200000,'US','100','US','L');
INSERT INTO public.salarios Values(233,2021,'SE','FT','Data
Analyst',200000,'USD',200000,'US','100','US','L');
INSERT INTO public.salarios Values(234,2021,'MI','FT','Data
Architect',180000,'USD',180000,'US','100','US','L');
INSERT INTO public.salarios Values(235,2021,'MI','FT','Head of Data
Science',110000,'USD',110000,'US','0','US','S');
INSERT INTO public.salarios Values(236,2021,'MI','FT','Research
Scientist',80000,'CAD',63810,'CA','100','CA','M');
INSERT INTO public.salarios Values(237,2021,'MI','FT','Data
Scientist',39600,'EUR',46809,'ES','100','ES','M');
INSERT INTO public.salarios Values(238,2021,'EN','FT','Data Scientist',4000,'USD',4000,'VN','0','VN','M');
INSERT INTO public.salarios Values(239,2021,'EN','FT','Data
Engineer',1600000,'INR',21637,'IN','50','IN','M');

```

```

INSERT INTO public.salarios Values(240,2021,'SE','FT','Data
Scientist',130000,'CAD',103691,'CA','100','CA','L');
INSERT INTO public.salarios Values(241,2021,'MI','FT','Data Analyst',80000,'USD',80000,'US','100','US','L');
INSERT INTO public.salarios Values(242,2021,'MI','FT','Data
Engineer',110000,'USD',110000,'US','100','US','L');
INSERT INTO public.salarios Values(243,2021,'SE','FT','Data
Scientist',165000,'USD',165000,'US','100','US','L');
INSERT INTO public.salarios Values(244,2021,'EN','FT','AI Scientist',1335000,'INR',18053,'IN','100','AS','S');
INSERT INTO public.salarios Values(245,2021,'MI','FT','Data
Engineer',52500,'GBP',72212,'GB','50','GB','L');
INSERT INTO public.salarios Values(246,2021,'EN','FT','Data Scientist',31000,'EUR',36643,'FR','50','FR','L');
INSERT INTO public.salarios Values(247,2021,'MI','FT','Data Engineer',108000,'TRY',12103,'TR','0','TR','M');
INSERT INTO public.salarios Values(248,2021,'SE','FT','Data
Engineer',70000,'GBP',96282,'GB','50','GB','L');
INSERT INTO public.salarios Values(249,2021,'SE','FT','Principal Data
Analyst',170000,'USD',170000,'US','100','US','M');
INSERT INTO public.salarios Values(250,2021,'MI','FT','Data
Scientist',115000,'USD',115000,'US','50','US','L');
INSERT INTO public.salarios Values(251,2021,'EN','FT','Data
Scientist',90000,'USD',90000,'US','100','US','S');
INSERT INTO public.salarios Values(252,2021,'EX','FT','Principal Data
Engineer',600000,'USD',600000,'US','100','US','L');
INSERT INTO public.salarios Values(253,2021,'EN','FT','Data
Scientist',2100000,'INR',28399,'IN','100','IN','M');
INSERT INTO public.salarios Values(254,2021,'MI','FT','Data Analyst',93000,'USD',93000,'US','100','US','L');
INSERT INTO public.salarios Values(255,2021,'SE','FT','Big Data
Architect',125000,'CAD',99703,'CA','50','CA','M');
INSERT INTO public.salarios Values(256,2021,'MI','FT','Data
Engineer',200000,'USD',200000,'US','100','US','L');
INSERT INTO public.salarios Values(257,2021,'SE','FT','Principal Data
Scientist',147000,'EUR',173762,'DE','100','DE','M');
INSERT INTO public.salarios Values(258,2021,'SE','FT','Machine Learning
Engineer',185000,'USD',185000,'US','50','US','L');
INSERT INTO public.salarios Values(259,2021,'EX','FT','Director of Data
Science',120000,'EUR',141846,'DE','0','DE','L');
INSERT INTO public.salarios Values(260,2021,'MI','FT','Data
Scientist',130000,'USD',130000,'US','50','US','L');
INSERT INTO public.salarios Values(261,2021,'SE','FT','Data Analyst',54000,'EUR',63831,'DE','50','DE','L');

```



```

INSERT INTO public.salarios Values(262,2021,'MI','FT','Data
Scientist',1250000,'INR',16904,'IN','100','IN','S');
INSERT INTO public.salarios Values(263,2021,'SE','FT','Machine Learning
Engineer',4900000,'INR',66265,'IN','0','IN','L');
INSERT INTO public.salarios Values(264,2021,'MI','FT','Data
Scientist',21600,'EUR',25532,'RS','100','DE','S');
INSERT INTO public.salarios Values(265,2021,'SE','FT','Lead Data
Engineer',160000,'USD',160000,'PR','50','US','S');
INSERT INTO public.salarios Values(266,2021,'MI','FT','Data Engineer',93150,'USD',93150,'US','0','US','M');
INSERT INTO public.salarios Values(267,2021,'MI','FT','Data
Engineer',111775,'USD',111775,'US','0','US','M');
INSERT INTO public.salarios Values(268,2021,'MI','FT','Data
Engineer',250000,'TRY',28016,'TR','100','TR','M');
INSERT INTO public.salarios Values(269,2021,'EN','FT','Data
Engineer',55000,'EUR',65013,'DE','50','DE','M');
INSERT INTO public.salarios Values(270,2021,'EN','FT','Data
Engineer',72500,'USD',72500,'US','100','US','L');
INSERT INTO public.salarios Values(271,2021,'SE','FT','Computer Vision
Engineer',102000,'BRL',18907,'BR','0','BR','M');
INSERT INTO public.salarios Values(272,2021,'EN','FT','Data Science
Consultant',65000,'EUR',76833,'DE','0','DE','L');
INSERT INTO public.salarios Values(273,2021,'EN','FT','Machine Learning
Engineer',85000,'USD',85000,'NL','100','DE','S');
INSERT INTO public.salarios Values(274,2021,'SE','FT','Data Scientist',65720,'EUR',77684,'FR','50','FR','M');
INSERT INTO public.salarios Values(275,2021,'EN','FT','Data
Scientist',100000,'USD',100000,'US','100','US','M');
INSERT INTO public.salarios Values(276,2021,'EN','FT','Data
Scientist',58000,'USD',58000,'US','50','US','L');
INSERT INTO public.salarios Values(277,2021,'SE','FT','AI Scientist',55000,'USD',55000,'ES','100','ES','L');
INSERT INTO public.salarios Values(278,2021,'SE','FT','Data Scientist',180000,'TRY',20171,'TR','50','TR','L');
INSERT INTO public.salarios Values(279,2021,'EN','FT','Business Data
Analyst',50000,'EUR',59102,'LU','100','LU','L');
INSERT INTO public.salarios Values(280,2021,'MI','FT','Data
Engineer',112000,'USD',112000,'US','100','US','L');
INSERT INTO public.salarios Values(281,2021,'EN','FT','Research
Scientist',100000,'USD',100000,'JE','0','CN','L');
INSERT INTO public.salarios Values(282,2021,'MI','PT','Data
Engineer',59000,'EUR',69741,'NL','100','NL','L');
INSERT INTO public.salarios Values(283,2021,'SE','CT','Staff Data
Scientist',105000,'USD',105000,'US','100','US','M');

```

```

INSERT INTO public.salarios Values(284,2021,'MI','FT','Research
Scientist',69999,'USD',69999,'CZ','50','CZ','L');
INSERT INTO public.salarios Values(285,2021,'SE','FT','Data Science
Manager',7000000,'INR',94665,'IN','50','IN','L');
INSERT INTO public.salarios Values(286,2021,'SE','FT','Head of Data',87000,'EUR',102839,'SI','100','SI','L');
INSERT INTO public.salarios Values(287,2021,'MI','FT','Data
Scientist',109000,'USD',109000,'US','50','US','L');
INSERT INTO public.salarios Values(288,2021,'MI','FT','Machine Learning
Engineer',43200,'EUR',51064,'IT','50','IT','L');
INSERT INTO public.salarios Values(289,2022,'SE','FT','Data
Engineer',135000,'USD',135000,'US','100','US','M');
INSERT INTO public.salarios Values(290,2022,'SE','FT','Data
Analyst',155000,'USD',155000,'US','100','US','M');
INSERT INTO public.salarios Values(291,2022,'SE','FT','Data
Analyst',120600,'USD',120600,'US','100','US','M');
INSERT INTO public.salarios Values(292,2022,'MI','FT','Data
Scientist',130000,'USD',130000,'US','0','US','M');
INSERT INTO public.salarios Values(293,2022,'MI','FT','Data Scientist',90000,'USD',90000,'US','0','US','M');
INSERT INTO public.salarios Values(294,2022,'MI','FT','Data
Engineer',170000,'USD',170000,'US','100','US','M');
INSERT INTO public.salarios Values(295,2022,'MI','FT','Data
Engineer',150000,'USD',150000,'US','100','US','M');
INSERT INTO public.salarios Values(296,2022,'SE','FT','Data
Analyst',102100,'USD',102100,'US','100','US','M');
INSERT INTO public.salarios Values(297,2022,'SE','FT','Data
Analyst',84900,'USD',84900,'US','100','US','M');
INSERT INTO public.salarios Values(298,2022,'SE','FT','Data
Scientist',136620,'USD',136620,'US','100','US','M');
INSERT INTO public.salarios Values(299,2022,'SE','FT','Data
Scientist',99360,'USD',99360,'US','100','US','M');
INSERT INTO public.salarios Values(300,2022,'SE','FT','Data
Scientist',90000,'GBP',117789,'GB','0','GB','M');

INSERT INTO public.salarios Values(301,2022,'SE','FT','Data
Scientist',80000,'GBP',104702,'GB','0','GB','M');
INSERT INTO public.salarios Values(302,2022,'SE','FT','Data
Scientist',146000,'USD',146000,'US','100','US','M');
INSERT INTO public.salarios Values(303,2022,'SE','FT','Data
Scientist',123000,'USD',123000,'US','100','US','M');

```

```

INSERT INTO public.salarios Values(304,2022,'EN','FT','Data
Engineer',40000,'GBP',52351,'GB','100','GB','M');
INSERT INTO public.salarios Values(305,2022,'SE','FT','Data Analyst',99000,'USD',99000,'US','0','US','M');
INSERT INTO public.salarios Values(306,2022,'SE','FT','Data
Analyst',116000,'USD',116000,'US','0','US','M');
INSERT INTO public.salarios Values(307,2022,'MI','FT','Data
Analyst',106260,'USD',106260,'US','0','US','M');
INSERT INTO public.salarios Values(308,2022,'MI','FT','Data
Analyst',126500,'USD',126500,'US','0','US','M');
INSERT INTO public.salarios Values(309,2022,'EX','FT','Data
Engineer',242000,'USD',242000,'US','100','US','M');
INSERT INTO public.salarios Values(310,2022,'EX','FT','Data
Engineer',200000,'USD',200000,'US','100','US','M');
INSERT INTO public.salarios Values(311,2022,'MI','FT','Data Scientist',50000,'GBP',65438,'GB','0','GB','M');
INSERT INTO public.salarios Values(312,2022,'MI','FT','Data Scientist',30000,'GBP',39263,'GB','0','GB','M');
INSERT INTO public.salarios Values(313,2022,'MI','FT','Data
Engineer',60000,'GBP',78526,'GB','0','GB','M');
INSERT INTO public.salarios Values(314,2022,'MI','FT','Data
Engineer',40000,'GBP',52351,'GB','0','GB','M');
INSERT INTO public.salarios Values(315,2022,'SE','FT','Data
Scientist',165220,'USD',165220,'US','100','US','M');
INSERT INTO public.salarios Values(316,2022,'EN','FT','Data
Engineer',35000,'GBP',45807,'GB','100','GB','M');
INSERT INTO public.salarios Values(317,2022,'SE','FT','Data
Scientist',120160,'USD',120160,'US','100','US','M');
INSERT INTO public.salarios Values(318,2022,'SE','FT','Data
Analyst',90320,'USD',90320,'US','100','US','M');
INSERT INTO public.salarios Values(319,2022,'SE','FT','Data
Engineer',181940,'USD',181940,'US','0','US','M');
INSERT INTO public.salarios Values(320,2022,'SE','FT','Data
Engineer',132320,'USD',132320,'US','0','US','M');
INSERT INTO public.salarios Values(321,2022,'SE','FT','Data
Engineer',220110,'USD',220110,'US','0','US','M');
INSERT INTO public.salarios Values(322,2022,'SE','FT','Data
Engineer',160080,'USD',160080,'US','0','US','M');
INSERT INTO public.salarios Values(323,2022,'SE','FT','Data
Scientist',180000,'USD',180000,'US','0','US','L');
INSERT INTO public.salarios Values(324,2022,'SE','FT','Data
Scientist',120000,'USD',120000,'US','0','US','L');

```

INSERT INTO public.salarios	Values(325,2022,'SE','FT','Data
Analyst',124190,'USD',124190,'US','100','US','M');	
INSERT INTO public.salarios	Values(326,2022,'EX','FT','Data
Analyst',130000,'USD',130000,'US','100','US','M');	
INSERT INTO public.salarios	Values(327,2022,'EX','FT','Data
Analyst',110000,'USD',110000,'US','100','US','M');	
INSERT INTO public.salarios	Values(328,2022,'SE','FT','Data
Analyst',170000,'USD',170000,'US','100','US','M');	
INSERT INTO public.salarios	Values(329,2022,'MI','FT','Data
Analyst',115500,'USD',115500,'US','100','US','M');	
INSERT INTO public.salarios	Values(330,2022,'SE','FT','Data
Analyst',112900,'USD',112900,'US','100','US','M');	
INSERT INTO public.salarios	Values(331,2022,'SE','FT','Data
Analyst',90320,'USD',90320,'US','100','US','M');	
INSERT INTO public.salarios	Values(332,2022,'SE','FT','Data
Analyst',112900,'USD',112900,'US','100','US','M');	
INSERT INTO public.salarios	Values(333,2022,'SE','FT','Data
Analyst',90320,'USD',90320,'US','100','US','M');	
INSERT INTO public.salarios	Values(334,2022,'SE','FT','Data
Engineer',165400,'USD',165400,'US','100','US','M');	
INSERT INTO public.salarios	Values(335,2022,'SE','FT','Data
Engineer',132320,'USD',132320,'US','100','US','M');	
INSERT INTO public.salarios	Values(336,2022,'MI','FT','Data
Analyst',167000,'USD',167000,'US','100','US','M');	
INSERT INTO public.salarios	Values(337,2022,'SE','FT','Data
Engineer',243900,'USD',243900,'US','100','US','M');	
INSERT INTO public.salarios	Values(338,2022,'SE','FT','Data
Analyst',136600,'USD',136600,'US','100','US','M');	
INSERT INTO public.salarios	Values(339,2022,'SE','FT','Data
Analyst',109280,'USD',109280,'US','100','US','M');	
INSERT INTO public.salarios	Values(340,2022,'SE','FT','Data
Engineer',128875,'USD',128875,'US','100','US','M');	
INSERT INTO public.salarios	Values(341,2022,'SE','FT','Data
Engineer',93700,'USD',93700,'US','100','US','M');	


```

INSERT INTO public.salarios Values(342,2022,'EX','FT','Head of Data
Science',224000,'USD',224000,'US','100','US','M');
INSERT INTO public.salarios Values(343,2022,'EX','FT','Head of Data
Science',167875,'USD',167875,'US','100','US','M');
INSERT INTO public.salarios Values(344,2022,'EX','FT','Analytics
Engineer',175000,'USD',175000,'US','100','US','M');
INSERT INTO public.salarios Values(345,2022,'SE','FT','Data
Engineer',156600,'USD',156600,'US','100','US','M');
INSERT INTO public.salarios Values(346,2022,'SE','FT','Data
Engineer',108800,'USD',108800,'US','0','US','M');
INSERT INTO public.salarios Values(347,2022,'SE','FT','Data Scientist',95550,'USD',95550,'US','0','US','M');
INSERT INTO public.salarios Values(348,2022,'SE','FT','Data
Engineer',113000,'USD',113000,'US','0','US','L');
INSERT INTO public.salarios Values(349,2022,'SE','FT','Data
Analyst',135000,'USD',135000,'US','100','US','M');
INSERT INTO public.salarios Values(350,2022,'SE','FT','Data Science
Manager',161342,'USD',161342,'US','100','US','M');
INSERT INTO public.salarios Values(351,2022,'SE','FT','Data Science
Manager',137141,'USD',137141,'US','100','US','M');
INSERT INTO public.salarios Values(352,2022,'SE','FT','Data
Scientist',167000,'USD',167000,'US','100','US','M');
INSERT INTO public.salarios Values(353,2022,'SE','FT','Data
Scientist',123000,'USD',123000,'US','100','US','M');
INSERT INTO public.salarios Values(354,2022,'SE','FT','Data
Engineer',60000,'GBP',78526,'GB','0','GB','M');
INSERT INTO public.salarios Values(355,2022,'SE','FT','Data
Engineer',50000,'GBP',65438,'GB','0','GB','M');
INSERT INTO public.salarios Values(356,2022,'SE','FT','Data
Scientist',150000,'USD',150000,'US','0','US','M');
INSERT INTO public.salarios Values(357,2022,'SE','FT','Data
Scientist',211500,'USD',211500,'US','100','US','M');
INSERT INTO public.salarios Values(358,2022,'SE','FT','Data
Architect',192400,'USD',192400,'CA','100','CA','M');
INSERT INTO public.salarios Values(359,2022,'SE','FT','Data
Architect',90700,'USD',90700,'CA','100','CA','M');
INSERT INTO public.salarios Values(360,2022,'SE','FT','Data
Analyst',130000,'USD',130000,'CA','100','CA','M');
INSERT INTO public.salarios Values(361,2022,'SE','FT','Data
Analyst',61300,'USD',61300,'CA','100','CA','M');

```




```

INSERT INTO public.salarios Values(362,2022,'SE','FT','Data
Analyst',130000,'USD',130000,'CA','100','CA','M');
INSERT INTO public.salarios Values(363,2022,'SE','FT','Data
Analyst',61300,'USD',61300,'CA','100','CA','M');
INSERT INTO public.salarios Values(364,2022,'SE','FT','Data
Engineer',160000,'USD',160000,'US','0','US','L');
INSERT INTO public.salarios Values(365,2022,'SE','FT','Data
Scientist',138600,'USD',138600,'US','100','US','M');
INSERT INTO public.salarios Values(366,2022,'SE','FT','Data
Engineer',136000,'USD',136000,'US','0','US','M');
INSERT INTO public.salarios Values(367,2022,'MI','FT','Data Analyst',58000,'USD',58000,'US','0','US','S');
INSERT INTO public.salarios Values(368,2022,'EX','FT','Analytics
Engineer',135000,'USD',135000,'US','100','US','M');
INSERT INTO public.salarios Values(369,2022,'SE','FT','Data
Scientist',170000,'USD',170000,'US','100','US','M');
INSERT INTO public.salarios Values(370,2022,'SE','FT','Data
Scientist',123000,'USD',123000,'US','100','US','M');
INSERT INTO public.salarios Values(371,2022,'SE','FT','Machine Learning
Engineer',189650,'USD',189650,'US','0','US','M');
INSERT INTO public.salarios Values(372,2022,'SE','FT','Machine Learning
Engineer',164996,'USD',164996,'US','0','US','M');
INSERT INTO public.salarios Values(373,2022,'MI','FT','ETL
Developer',50000,'EUR',54957,'GR','0','GR','M');
INSERT INTO public.salarios Values(374,2022,'MI','FT','ETL
Developer',50000,'EUR',54957,'GR','0','GR','M');
INSERT INTO public.salarios Values(375,2022,'EX','FT','Lead Data
Engineer',150000,'CAD',118187,'CA','100','CA','S');
INSERT INTO public.salarios Values(376,2022,'SE','FT','Data
Analyst',132000,'USD',132000,'US','0','US','M');
INSERT INTO public.salarios Values(377,2022,'SE','FT','Data
Engineer',165400,'USD',165400,'US','100','US','M');
INSERT INTO public.salarios Values(378,2022,'SE','FT','Data
Architect',208775,'USD',208775,'US','100','US','M');

```



```

INSERT INTO public.salarios Values(379,2022,'SE','FT','Data
Architect',147800,'USD',147800,'US','100','US','M');
INSERT INTO public.salarios Values(380,2022,'SE','FT','Data
Engineer',136994,'USD',136994,'US','100','US','M');
INSERT INTO public.salarios Values(381,2022,'SE','FT','Data
Engineer',101570,'USD',101570,'US','100','US','M');
INSERT INTO public.salarios Values(382,2022,'SE','FT','Data
Analyst',128875,'USD',128875,'US','100','US','M');
INSERT INTO public.salarios Values(383,2022,'SE','FT','Data
Analyst',93700,'USD',93700,'US','100','US','M');
INSERT INTO public.salarios Values(384,2022,'EX','FT','Head of Machine
Learning',6000000,'INR',79039,'IN','50','IN','L');
INSERT INTO public.salarios Values(385,2022,'SE','FT','Data
Engineer',132320,'USD',132320,'US','100','US','M');
INSERT INTO public.salarios Values(386,2022,'EN','FT','Machine Learning
Engineer',28500,'GBP',37300,'GB','100','GB','L');
INSERT INTO public.salarios Values(387,2022,'SE','FT','Data
Analyst',164000,'USD',164000,'US','0','US','M');
INSERT INTO public.salarios Values(388,2022,'SE','FT','Data
Engineer',155000,'USD',155000,'US','100','US','M');
INSERT INTO public.salarios Values(389,2022,'MI','FT','Machine Learning
Engineer',95000,'GBP',124333,'GB','0','GB','M');
INSERT INTO public.salarios Values(390,2022,'MI','FT','Machine Learning
Engineer',75000,'GBP',98158,'GB','0','GB','M');
INSERT INTO public.salarios Values(391,2022,'MI','FT','AI Scientist',120000,'USD',120000,'US','0','US','M');
INSERT INTO public.salarios Values(392,2022,'SE','FT','Data
Analyst',112900,'USD',112900,'US','100','US','M');
INSERT INTO public.salarios Values(393,2022,'SE','FT','Data
Analyst',90320,'USD',90320,'US','100','US','M');
INSERT INTO public.salarios Values(394,2022,'SE','FT','Data Analytics
Manager',145000,'USD',145000,'US','100','US','M');
INSERT INTO public.salarios Values(395,2022,'SE','FT','Data Analytics
Manager',105400,'USD',105400,'US','100','US','M');
INSERT INTO public.salarios Values(396,2022,'MI','FT','Machine Learning
Engineer',80000,'EUR',87932,'FR','100','DE','M');
INSERT INTO public.salarios Values(397,2022,'MI','FT','Data
Engineer',90000,'GBP',117789,'GB','0','GB','M');
INSERT INTO public.salarios Values(398,2022,'SE','FT','Data
Scientist',215300,'USD',215300,'US','100','US','L');

```

```

INSERT INTO public.salarios Values(399,2022,'SE','FT','Data
Scientist',158200,'USD',158200,'US','100','US','L');
INSERT INTO public.salarios Values(400,2022,'SE','FT','Data
Engineer',209100,'USD',209100,'US','100','US','L');
INSERT INTO public.salarios Values(401,2022,'SE','FT','Data
Engineer',154600,'USD',154600,'US','100','US','L');
INSERT INTO public.salarios Values(402,2022,'SE','FT','Data
Analyst',115934,'USD',115934,'US','0','US','M');
INSERT INTO public.salarios Values(403,2022,'SE','FT','Data Analyst',81666,'USD',81666,'US','0','US','M');
INSERT INTO public.salarios Values(404,2022,'SE','FT','Data
Engineer',175000,'USD',175000,'US','100','US','M');
INSERT INTO public.salarios Values(405,2022,'MI','FT','Data
Engineer',75000,'GBP',98158,'GB','0','GB','M');
INSERT INTO public.salarios Values(406,2022,'MI','FT','Data Analyst',58000,'USD',58000,'US','0','US','S');
INSERT INTO public.salarios Values(407,2022,'SE','FT','Data
Engineer',183600,'USD',183600,'US','100','US','L');
INSERT INTO public.salarios Values(408,2022,'MI','FT','Data
Analyst',40000,'GBP',52351,'GB','100','GB','M');
INSERT INTO public.salarios Values(409,2022,'SE','FT','Data
Scientist',180000,'USD',180000,'US','100','US','M');
INSERT INTO public.salarios Values(410,2022,'MI','FT','Data Scientist',55000,'GBP',71982,'GB','0','GB','M');
INSERT INTO public.salarios Values(411,2022,'MI','FT','Data Scientist',35000,'GBP',45807,'GB','0','GB','M');
INSERT INTO public.salarios Values(412,2022,'MI','FT','Data
Engineer',60000,'EUR',65949,'GR','100','GR','M');
INSERT INTO public.salarios Values(413,2022,'MI','FT','Data
Engineer',45000,'EUR',49461,'GR','100','GR','M');
INSERT INTO public.salarios Values(414,2022,'MI','FT','Data
Engineer',60000,'GBP',78526,'GB','100','GB','M');
INSERT INTO public.salarios Values(415,2022,'MI','FT','Data
Engineer',45000,'GBP',58894,'GB','100','GB','M');
INSERT INTO public.salarios Values(416,2022,'SE','FT','Data
Scientist',260000,'USD',260000,'US','100','US','M');
INSERT INTO public.salarios Values(417,2022,'SE','FT','Data Science
Engineer',60000,'USD',60000,'AR','100','MX','L');
INSERT INTO public.salarios Values(418,2022,'MI','FT','Data Engineer',63900,'USD',63900,'US','0','US','M');
INSERT INTO public.salarios Values(419,2022,'MI','FT','Machine Learning
Scientist',160000,'USD',160000,'US','100','US','L');
INSERT INTO public.salarios Values(420,2022,'MI','FT','Machine Learning
Scientist',112300,'USD',112300,'US','100','US','L');

```

```

INSERT INTO public.salarios Values(421,2022,'MI','FT','Data Science
Manager',241000,'USD',241000,'US','100','US','M');
INSERT INTO public.salarios Values(422,2022,'MI','FT','Data Science
Manager',159000,'USD',159000,'US','100','US','M');
INSERT INTO public.salarios Values(423,2022,'SE','FT','Data
Scientist',180000,'USD',180000,'US','0','US','M');
INSERT INTO public.salarios Values(424,2022,'SE','FT','Data Scientist',80000,'USD',80000,'US','0','US','M');
INSERT INTO public.salarios Values(425,2022,'MI','FT','Data Engineer',82900,'USD',82900,'US','0','US','M');
INSERT INTO public.salarios Values(426,2022,'SE','FT','Data
Engineer',100800,'USD',100800,'US','100','US','L');
INSERT INTO public.salarios Values(427,2022,'MI','FT','Data
Engineer',45000,'EUR',49461,'ES','100','ES','M');
INSERT INTO public.salarios Values(428,2022,'SE','FT','Data
Scientist',140400,'USD',140400,'US','0','US','L');
INSERT INTO public.salarios Values(429,2022,'MI','FT','Data
Analyst',30000,'GBP',39263,'GB','100','GB','M');
INSERT INTO public.salarios Values(430,2022,'MI','FT','Data Analyst',40000,'EUR',43966,'ES','100','ES','M');
INSERT INTO public.salarios Values(431,2022,'MI','FT','Data Analyst',30000,'EUR',32974,'ES','100','ES','M');
INSERT INTO public.salarios Values(432,2022,'MI','FT','Data
Engineer',80000,'EUR',87932,'ES','100','ES','M');
INSERT INTO public.salarios Values(433,2022,'MI','FT','Data
Engineer',70000,'EUR',76940,'ES','100','ES','M');
INSERT INTO public.salarios Values(434,2022,'MI','FT','Data
Engineer',80000,'GBP',104702,'GB','100','GB','M');
INSERT INTO public.salarios Values(435,2022,'MI','FT','Data
Engineer',70000,'GBP',91614,'GB','100','GB','M');
INSERT INTO public.salarios Values(436,2022,'MI','FT','Data
Engineer',60000,'EUR',65949,'ES','100','ES','M');
INSERT INTO public.salarios Values(437,2022,'MI','FT','Data
Engineer',80000,'EUR',87932,'GR','100','GR','M');
INSERT INTO public.salarios Values(438,2022,'SE','FT','Machine Learning
Engineer',189650,'USD',189650,'US','0','US','M');
INSERT INTO public.salarios Values(439,2022,'SE','FT','Machine Learning
Engineer',164996,'USD',164996,'US','0','US','M');
INSERT INTO public.salarios Values(440,2022,'MI','FT','Data
Analyst',40000,'EUR',43966,'GR','100','GR','M');
INSERT INTO public.salarios Values(441,2022,'MI','FT','Data
Analyst',30000,'EUR',32974,'GR','100','GR','M');
INSERT INTO public.salarios Values(442,2022,'MI','FT','Data
Engineer',75000,'GBP',98158,'GB','100','GB','M');

```

```

INSERT INTO public.salarios Values(443,2022,'MI','FT','Data
Engineer',60000,'GBP',78526,'GB','100','GB','M');
INSERT INTO public.salarios Values(444,2022,'SE','FT','Data
Scientist',215300,'USD',215300,'US','0','US','L');
INSERT INTO public.salarios Values(445,2022,'MI','FT','Data
Engineer',70000,'EUR',76940,'GR','100','GR','M');
INSERT INTO public.salarios Values(446,2022,'SE','FT','Data
Engineer',209100,'USD',209100,'US','100','US','L');
INSERT INTO public.salarios Values(447,2022,'SE','FT','Data
Engineer',154600,'USD',154600,'US','100','US','L');
INSERT INTO public.salarios Values(448,2022,'SE','FT','Data
Engineer',180000,'USD',180000,'US','100','US','M');
INSERT INTO public.salarios Values(449,2022,'EN','FT','ML Engineer',20000,'EUR',21983,'PT','100','PT','L');
INSERT INTO public.salarios Values(450,2022,'SE','FT','Data
Engineer',80000,'USD',80000,'US','100','US','M');
INSERT INTO public.salarios Values(451,2022,'MI','FT','Machine Learning
Developer',100000,'CAD',78791,'CA','100','CA','M');
INSERT INTO public.salarios Values(452,2022,'EX','FT','Director of Data
Science',250000,'CAD',196979,'CA','50','CA','L');
INSERT INTO public.salarios Values(453,2022,'MI','FT','Machine Learning
Engineer',120000,'USD',120000,'US','100','US','S');
INSERT INTO public.salarios Values(454,2022,'EN','FT','Computer Vision
Engineer',125000,'USD',125000,'US','0','US','M');
INSERT INTO public.salarios Values(455,2022,'MI','FT','NLP
Engineer',240000,'CNY',37236,'US','50','US','L');
INSERT INTO public.salarios Values(456,2022,'SE','FT','Data
Engineer',105000,'USD',105000,'US','100','US','M');
INSERT INTO public.salarios Values(457,2022,'SE','FT','Lead Machine Learning
Engineer',80000,'EUR',87932,'DE','0','DE','M');
INSERT INTO public.salarios Values(458,2022,'MI','FT','Business Data
Analyst',1400000,'INR',18442,'IN','100','IN','M');
INSERT INTO public.salarios Values(459,2022,'MI','FT','Data
Scientist',2400000,'INR',31615,'IN','100','IN','L');
INSERT INTO public.salarios Values(460,2022,'MI','FT','Machine Learning Infrastructure
Engineer',53000,'EUR',58255,'PT','50','PT','L');
INSERT INTO public.salarios Values(461,2022,'EN','FT','Financial Data
Analyst',100000,'USD',100000,'US','50','US','L');

```

```

INSERT INTO public.salarios Values(462,2022,'MI','PT','Data
Engineer',50000,'EUR',54957,'DE','50','DE','L');
INSERT INTO public.salarios Values(463,2022,'EN','FT','Data
Scientist',1400000,'INR',18442,'IN','100','IN','M');
INSERT INTO public.salarios Values(464,2022,'SE','FT','Principal Data
Scientist',148000,'EUR',162674,'DE','100','DE','M');
INSERT INTO public.salarios Values(465,2022,'EN','FT','Data
Engineer',120000,'USD',120000,'US','100','US','M');
INSERT INTO public.salarios Values(466,2022,'SE','FT','Research
Scientist',144000,'USD',144000,'US','50','US','L');
INSERT INTO public.salarios Values(467,2022,'SE','FT','Data
Scientist',104890,'USD',104890,'US','100','US','M');
INSERT INTO public.salarios Values(468,2022,'SE','FT','Data
Engineer',100000,'USD',100000,'US','100','US','M');
INSERT INTO public.salarios Values(469,2022,'SE','FT','Data
Scientist',140000,'USD',140000,'US','100','US','M');
INSERT INTO public.salarios Values(470,2022,'MI','FT','Data
Analyst',135000,'USD',135000,'US','100','US','M');
INSERT INTO public.salarios Values(471,2022,'MI','FT','Data
Analyst',50000,'USD',50000,'US','100','US','M');
INSERT INTO public.salarios Values(472,2022,'SE','FT','Data
Scientist',220000,'USD',220000,'US','100','US','M');
INSERT INTO public.salarios Values(473,2022,'SE','FT','Data
Scientist',140000,'USD',140000,'US','100','US','M');
INSERT INTO public.salarios Values(474,2022,'MI','FT','Data
Scientist',140000,'GBP',183228,'GB','0','GB','M');
INSERT INTO public.salarios Values(475,2022,'MI','FT','Data Scientist',70000,'GBP',91614,'GB','0','GB','M');
INSERT INTO public.salarios Values(476,2022,'SE','FT','Data
Scientist',185100,'USD',185100,'US','100','US','M');
INSERT INTO public.salarios Values(477,2022,'SE','FT','Machine Learning
Engineer',220000,'USD',220000,'US','100','US','M');
INSERT INTO public.salarios Values(478,2022,'MI','FT','Data
Scientist',200000,'USD',200000,'US','100','US','M');
INSERT INTO public.salarios Values(479,2022,'MI','FT','Data
Scientist',120000,'USD',120000,'US','100','US','M');
INSERT INTO public.salarios Values(480,2022,'SE','FT','Machine Learning
Engineer',120000,'USD',120000,'AE','100','AE','S');
INSERT INTO public.salarios Values(481,2022,'SE','FT','Machine Learning
Engineer',65000,'USD',65000,'AE','100','AE','S');

```




```

INSERT INTO public.salarios Values(482,2022,'EX','FT','Data
Engineer',324000,'USD',324000,'US','100','US','M');
INSERT INTO public.salarios Values(483,2022,'EX','FT','Data
Engineer',216000,'USD',216000,'US','100','US','M');
INSERT INTO public.salarios Values(484,2022,'SE','FT','Data
Engineer',210000,'USD',210000,'US','100','US','M');
INSERT INTO public.salarios Values(485,2022,'SE','FT','Machine Learning
Engineer',120000,'USD',120000,'US','100','US','M');
INSERT INTO public.salarios Values(486,2022,'SE','FT','Data
Scientist',230000,'USD',230000,'US','100','US','M');
INSERT INTO public.salarios Values(487,2022,'EN','PT','Data
Scientist',100000,'USD',100000,'DZ','50','DZ','M');
INSERT INTO public.salarios Values(488,2022,'MI','FL','Data
Scientist',100000,'USD',100000,'CA','100','US','M');
INSERT INTO public.salarios Values(489,2022,'EN','CT','Applied Machine Learning
Scientist',29000,'EUR',31875,'TN','100','CZ','M');
INSERT INTO public.salarios Values(490,2022,'SE','FT','Head of
Data',200000,'USD',200000,'MY','100','US','M');
INSERT INTO public.salarios Values(491,2022,'MI','FT','Principal Data
Analyst',75000,'USD',75000,'CA','100','CA','S');
INSERT INTO public.salarios Values(492,2022,'MI','FT','Data
Scientist',150000,'PLN',35590,'PL','100','PL','L');
INSERT INTO public.salarios Values(493,2022,'SE','FT','Machine Learning
Developer',100000,'CAD',78791,'CA','100','CA','M');
INSERT INTO public.salarios Values(494,2022,'SE','FT','Data
Scientist',100000,'USD',100000,'BR','100','US','M');
INSERT INTO public.salarios Values(495,2022,'MI','FT','Machine Learning
Scientist',153000,'USD',153000,'US','50','US','M');
INSERT INTO public.salarios Values(496,2022,'EN','FT','Data
Engineer',52800,'EUR',58035,'PK','100','DE','M');
INSERT INTO public.salarios Values(497,2022,'SE','FT','Data
Scientist',165000,'USD',165000,'US','100','US','M');
INSERT INTO public.salarios Values(498,2022,'SE','FT','Research
Scientist',85000,'EUR',93427,'FR','50','FR','L');
INSERT INTO public.salarios Values(499,2022,'EN','FT','Data
Scientist',66500,'CAD',52396,'CA','100','CA','L');
INSERT INTO public.salarios Values(500,2022,'SE','FT','Machine Learning
Engineer',57000,'EUR',62651,'NL','100','NL','L');

```



```

INSERT INTO public.salarios Values(501,2022,'MI','FT','Head of Data',30000,'EUR',32974,'EE','100','EE','S');
INSERT INTO public.salarios Values(502,2022,'EN','FT','Data
Scientist',40000,'USD',40000,'JP','100','MY','L');
INSERT INTO public.salarios Values(503,2022,'MI','FT','Machine Learning
Engineer',121000,'AUD',87425,'AU','100','AU','L');
INSERT INTO public.salarios Values(504,2022,'SE','FT','Data
Engineer',115000,'USD',115000,'US','100','US','M');
INSERT INTO public.salarios Values(505,2022,'EN','FT','Data
Scientist',120000,'AUD',86703,'AU','50','AU','M');
INSERT INTO public.salarios Values(506,2022,'MI','FT','Applied Machine Learning
Scientist',75000,'USD',75000,'BO','100','US','L');
INSERT INTO public.salarios Values(507,2022,'MI','FT','Research
Scientist',59000,'EUR',64849,'AT','0','AT','L');
INSERT INTO public.salarios Values(508,2022,'EN','FT','Research
Scientist',120000,'USD',120000,'US','100','US','L');
INSERT INTO public.salarios Values(509,2022,'MI','FT','Applied Data
Scientist',157000,'USD',157000,'US','100','US','L');
INSERT INTO public.salarios Values(510,2022,'EN','FT','Computer Vision Software
Engineer',150000,'USD',150000,'AU','100','AU','S');
INSERT INTO public.salarios Values(511,2022,'MI','FT','Business Data
Analyst',90000,'CAD',70912,'CA','50','CA','L');
INSERT INTO public.salarios Values(512,2022,'EN','FT','Data
Engineer',65000,'USD',65000,'US','100','US','S');
INSERT INTO public.salarios Values(513,2022,'SE','FT','Machine Learning
Engineer',65000,'EUR',71444,'IE','100','IE','S');
INSERT INTO public.salarios Values(514,2022,'EN','FT','Data Analytics
Engineer',20000,'USD',20000,'PK','0','PK','M');
INSERT INTO public.salarios Values(515,2022,'MI','FT','Data
Scientist',48000,'USD',48000,'RU','100','US','S');
INSERT INTO public.salarios Values(516,2022,'SE','FT','Data Science
Manager',152500,'USD',152500,'US','100','US','M');
INSERT INTO public.salarios Values(517,2022,'MI','FT','Data
Engineer',62000,'EUR',68147,'FR','100','FR','M');
INSERT INTO public.salarios Values(518,2022,'MI','FT','Data
Scientist',115000,'CHF',122346,'CH','0','CH','L');
INSERT INTO public.salarios Values(519,2022,'SE','FT','Applied Data
Scientist',380000,'USD',380000,'US','100','US','L');
INSERT INTO public.salarios Values(520,2022,'MI','FT','Data
Scientist',88000,'CAD',69336,'CA','100','CA','M');
INSERT INTO public.salarios Values(521,2022,'EN','FT','Computer Vision
Engineer',10000,'USD',10000,'PT','100','LU','M');

```

```

INSERT INTO public.salarios Values(522,2022,'MI','FT','Data
Analyst',20000,'USD',20000,'GR','100','GR','S');
INSERT INTO public.salarios Values(523,2022,'SE','FT','Data Analytics
Lead',405000,'USD',405000,'US','100','US','L');
INSERT INTO public.salarios Values(524,2022,'MI','FT','Data
Scientist',135000,'USD',135000,'US','100','US','L');
INSERT INTO public.salarios Values(525,2022,'SE','FT','Applied Data
Scientist',177000,'USD',177000,'US','100','US','L');
INSERT INTO public.salarios Values(526,2022,'MI','FT','Data
Scientist',78000,'USD',78000,'US','100','US','M');
INSERT INTO public.salarios Values(527,2022,'SE','FT','Data
Analyst',135000,'USD',135000,'US','100','US','M');
INSERT INTO public.salarios Values(528,2022,'SE','FT','Data
Analyst',100000,'USD',100000,'US','100','US','M');
INSERT INTO public.salarios Values(529,2022,'SE','FT','Data
Analyst',90320,'USD',90320,'US','100','US','M');
INSERT INTO public.salarios Values(530,2022,'MI','FT','Data Analyst',85000,'USD',85000,'CA','0','CA','M');
INSERT INTO public.salarios Values(531,2022,'MI','FT','Data Analyst',75000,'USD',75000,'CA','0','CA','M');
INSERT INTO public.salarios Values(532,2022,'SE','FT','Machine Learning
Engineer',214000,'USD',214000,'US','100','US','M');
INSERT INTO public.salarios Values(533,2022,'SE','FT','Machine Learning
Engineer',192600,'USD',192600,'US','100','US','M');
INSERT INTO public.salarios Values(534,2022,'SE','FT','Data
Architect',266400,'USD',266400,'US','100','US','M');
INSERT INTO public.salarios Values(535,2022,'SE','FT','Data
Architect',213120,'USD',213120,'US','100','US','M');
INSERT INTO public.salarios Values(536,2022,'SE','FT','Data
Analyst',112900,'USD',112900,'US','100','US','M');
INSERT INTO public.salarios Values(537,2022,'SE','FT','Data
Engineer',155000,'USD',155000,'US','100','US','M');
INSERT INTO public.salarios Values(538,2022,'MI','FT','Data
Scientist',141300,'USD',141300,'US','0','US','M');
INSERT INTO public.salarios Values(539,2022,'MI','FT','Data
Scientist',102100,'USD',102100,'US','0','US','M');
INSERT INTO public.salarios Values(540,2022,'SE','FT','Data
Analyst',115934,'USD',115934,'US','100','US','M');
INSERT INTO public.salarios Values(541,2022,'SE','FT','Data
Analyst',81666,'USD',81666,'US','100','US','M');

```

```

INSERT INTO public.salarios Values(542,2022,'MI','FT','Data
Engineer',206699,'USD',206699,'US','0','US','M');
INSERT INTO public.salarios Values(543,2022,'MI','FT','Data Engineer',99100,'USD',99100,'US','0','US','M');
INSERT INTO public.salarios Values(544,2022,'SE','FT','Data
Engineer',130000,'USD',130000,'US','100','US','M');
INSERT INTO public.salarios Values(545,2022,'SE','FT','Data
Engineer',115000,'USD',115000,'US','100','US','M');
INSERT INTO public.salarios Values(546,2022,'SE','FT','Data
Engineer',110500,'USD',110500,'US','100','US','M');
INSERT INTO public.salarios Values(547,2022,'SE','FT','Data
Engineer',130000,'USD',130000,'US','100','US','M');
INSERT INTO public.salarios Values(548,2022,'SE','FT','Data
Analyst',99050,'USD',99050,'US','100','US','M');
INSERT INTO public.salarios Values(549,2022,'SE','FT','Data
Engineer',160000,'USD',160000,'US','100','US','M');
INSERT INTO public.salarios Values(550,2022,'SE','FT','Data
Scientist',205300,'USD',205300,'US','0','US','L');
INSERT INTO public.salarios Values(551,2022,'SE','FT','Data
Scientist',140400,'USD',140400,'US','0','US','L');
INSERT INTO public.salarios Values(552,2022,'SE','FT','Data
Scientist',176000,'USD',176000,'US','100','US','M');
INSERT INTO public.salarios Values(553,2022,'SE','FT','Data
Scientist',144000,'USD',144000,'US','100','US','M');
INSERT INTO public.salarios Values(554,2022,'SE','FT','Data
Engineer',200100,'USD',200100,'US','100','US','M');
INSERT INTO public.salarios Values(555,2022,'SE','FT','Data
Engineer',160000,'USD',160000,'US','100','US','M');
INSERT INTO public.salarios Values(556,2022,'SE','FT','Data
Engineer',145000,'USD',145000,'US','100','US','M');
INSERT INTO public.salarios Values(557,2022,'SE','FT','Data Engineer',70500,'USD',70500,'US','0','US','M');
INSERT INTO public.salarios Values(560,2022,'SE','FT','Analytics
Engineer',205300,'USD',205300,'US','0','US','M');
INSERT INTO public.salarios Values(561,2022,'SE','FT','Analytics
Engineer',184700,'USD',184700,'US','0','US','M');
INSERT INTO public.salarios Values(562,2022,'SE','FT','Data
Engineer',175100,'USD',175100,'US','100','US','M');
INSERT INTO public.salarios Values(563,2022,'SE','FT','Data
Engineer',140250,'USD',140250,'US','100','US','M');
INSERT INTO public.salarios Values(564,2022,'SE','FT','Data
Analyst',116150,'USD',116150,'US','100','US','M');

```



```
INSERT INTO public.salarios Values(565,2022,'SE','FT','Data Engineer',54000,'USD',54000,'US','0','US','M');
INSERT INTO public.salarios Values(566,2022,'SE','FT','Data Analyst',170000,'USD',170000,'US','100','US','M');
INSERT INTO public.salarios Values(567,2022,'MI','FT','Data Analyst',50000,'GBP',65438,'GB','0','GB','M');
INSERT INTO public.salarios Values(568,2022,'SE','FT','Data Analyst',80000,'USD',80000,'US','100','US','M');
INSERT INTO public.salarios Values(569,2022,'SE','FT','Data Scientist',140000,'USD',140000,'US','100','US','M');
INSERT INTO public.salarios Values(570,2022,'SE','FT','Data Scientist',210000,'USD',210000,'US','100','US','M');
INSERT INTO public.salarios Values(571,2022,'SE','FT','Data Scientist',140000,'USD',140000,'US','100','US','M');
INSERT INTO public.salarios Values(572,2022,'SE','FT','Data Analyst',100000,'USD',100000,'US','100','US','M');
INSERT INTO public.salarios Values(573,2022,'SE','FT','Data Analyst',69000,'USD',69000,'US','100','US','M');
INSERT INTO public.salarios Values(574,2022,'SE','FT','Data Scientist',210000,'USD',210000,'US','100','US','M');
INSERT INTO public.salarios Values(575,2022,'SE','FT','Data Scientist',140000,'USD',140000,'US','100','US','M');
INSERT INTO public.salarios Values(576,2022,'SE','FT','Data Scientist',210000,'USD',210000,'US','100','US','M');
INSERT INTO public.salarios Values(577,2022,'SE','FT','Data Analyst',150075,'USD',150075,'US','100','US','M');
INSERT INTO public.salarios Values(578,2022,'SE','FT','Data Engineer',100000,'USD',100000,'US','100','US','M');
INSERT INTO public.salarios Values(579,2022,'SE','FT','Data Engineer',25000,'USD',25000,'US','100','US','M');
INSERT INTO public.salarios Values(580,2022,'SE','FT','Data Analyst',126500,'USD',126500,'US','100','US','M');
INSERT INTO public.salarios Values(581,2022,'SE','FT','Data Analyst',106260,'USD',106260,'US','100','US','M');
INSERT INTO public.salarios Values(582,2022,'SE','FT','Data Engineer',220110,'USD',220110,'US','100','US','M');
INSERT INTO public.salarios Values(583,2022,'SE','FT','Data Engineer',160080,'USD',160080,'US','100','US','M');
INSERT INTO public.salarios Values(584,2022,'SE','FT','Data Analyst',105000,'USD',105000,'US','100','US','M');
```

```

INSERT INTO public.salarios Values(585,2022,'SE','FT','Data
Analyst',110925,'USD',110925,'US','100','US','M');
INSERT INTO public.salarios Values(586,2022,'MI','FT','Data Analyst',35000,'GBP',45807,'GB','0','GB','M');
INSERT INTO public.salarios Values(587,2022,'SE','FT','Data
Scientist',140000,'USD',140000,'US','100','US','M');
INSERT INTO public.salarios Values(588,2022,'SE','FT','Data Analyst',99000,'USD',99000,'US','0','US','M');
INSERT INTO public.salarios Values(589,2022,'SE','FT','Data
Analyst',60000,'USD',60000,'US','100','US','M');
INSERT INTO public.salarios Values(590,2022,'SE','FT','Data
Architect',192564,'USD',192564,'US','100','US','M');
INSERT INTO public.salarios Values(591,2022,'SE','FT','Data
Architect',144854,'USD',144854,'US','100','US','M');
INSERT INTO public.salarios Values(592,2022,'SE','FT','Data
Scientist',230000,'USD',230000,'US','100','US','M');
INSERT INTO public.salarios Values(593,2022,'SE','FT','Data
Scientist',150000,'USD',150000,'US','100','US','M');
INSERT INTO public.salarios Values(594,2022,'SE','FT','Data Analytics
Manager',150260,'USD',150260,'US','100','US','M');
INSERT INTO public.salarios Values(595,2022,'SE','FT','Data Analytics
Manager',109280,'USD',109280,'US','100','US','M');
INSERT INTO public.salarios Values(596,2022,'SE','FT','Data
Scientist',210000,'USD',210000,'US','100','US','M');
INSERT INTO public.salarios Values(597,2022,'SE','FT','Data
Analyst',170000,'USD',170000,'US','100','US','M');
INSERT INTO public.salarios Values(598,2022,'MI','FT','Data
Scientist',160000,'USD',160000,'US','100','US','M');
INSERT INTO public.salarios Values(599,2022,'MI','FT','Data
Scientist',130000,'USD',130000,'US','100','US','M');
INSERT INTO public.salarios Values(600,2022,'EN','FT','Data Analyst',67000,'USD',67000,'CA','0','CA','M');
INSERT INTO public.salarios Values(601,2022,'EN','FT','Data Analyst',52000,'USD',52000,'CA','0','CA','M');
INSERT INTO public.salarios Values(602,2022,'SE','FT','Data
Engineer',154000,'USD',154000,'US','100','US','M');
INSERT INTO public.salarios Values(603,2022,'SE','FT','Data
Engineer',126000,'USD',126000,'US','100','US','M');
INSERT INTO public.salarios Values(604,2022,'SE','FT','Data
Analyst',129000,'USD',129000,'US','0','US','M');
INSERT INTO public.salarios Values(605,2022,'SE','FT','Data
Analyst',150000,'USD',150000,'US','100','US','M');
INSERT INTO public.salarios Values(606,2022,'MI','FT','AI
Scientist',200000,'USD',200000,'IN','100','US','L');

```



Digital College

ENSINO DE HABILIDADES DIGITAIS

digitalcollege.com.br