CENTRO PAULA SOUZA

FATEC MAUÁ

Tecnólogo em Desenvolvimento de Software Multiplataforma

JOÃO PEDRO SASSI GRANADO

DATAWAREHOUSE

Mauá

2023

JOÃO PEDRO SASSI GRANADO

DATAWAREHOUSE

Trabalho sobre os conceitos de Datawarehouse apresentado no Tecnólogo de Desenvolvimento de Software Multiplataforma da Fatec Mauá, orientado pelo professor Prof. Anderson Vanin da disciplina Banco de Dados – Não Relacional.

Mauá

2023

ÍNDICE DE ILUSTRAÇÕES

[Figura 1: Exemplo DatawareHouse 4](#_Toc144395932)

[Figura 2: DatawareHouse 5](#_Toc144395933)

[Figura 3: Exemplos Data Mart 6](#_Toc144395934)

[Figura 4: Ilustração DatawareHouse 9](#_Toc144395935)

SUMÁRIO

[1. INTRODUÇÃO 4](#_Toc144395904)

[2. Tipos de datawarehouse 5](#_Toc144395905)

[3. ETL 5](#_Toc144395906)

[4. DATA MART 5](#_Toc144395907)

[5. ARQUITETURAS DE UM DATAWAREHOUSE 6](#_Toc144395908)

[5.1. Arquitetura global 6](#_Toc144395909)

[5.2. Arquitetura de data marts independentes 6](#_Toc144395910)

[5.3. Arquitetura de data marts integrados 6](#_Toc144395911)

[6. IMPLEMENTAÇÕES DE DATAWAREHOUSE 6](#_Toc144395912)

[6.1. Abordagem Top-Down 6](#_Toc144395913)

[6.2. Abordagem Bottom-Up 7](#_Toc144395914)

[6.3. Abordagem combinada 7](#_Toc144395915)

[7. MODELAGEM MULTIDIMENSIONAL 7](#_Toc144395916)

[7.1. Elementos de um DW 7](#_Toc144395917)

[7.2. Star Schema | Esquema Estrela 8](#_Toc144395918)

[7.3. Membros 8](#_Toc144395919)

[7.4. Dimensões 8](#_Toc144395920)

[7.5. Granularidade 8](#_Toc144395921)

[7.6. Snowflake 8](#_Toc144395922)

[7.7. Agregação 8](#_Toc144395923)

[8. CONCLUSÃO 9](#_Toc144395924)

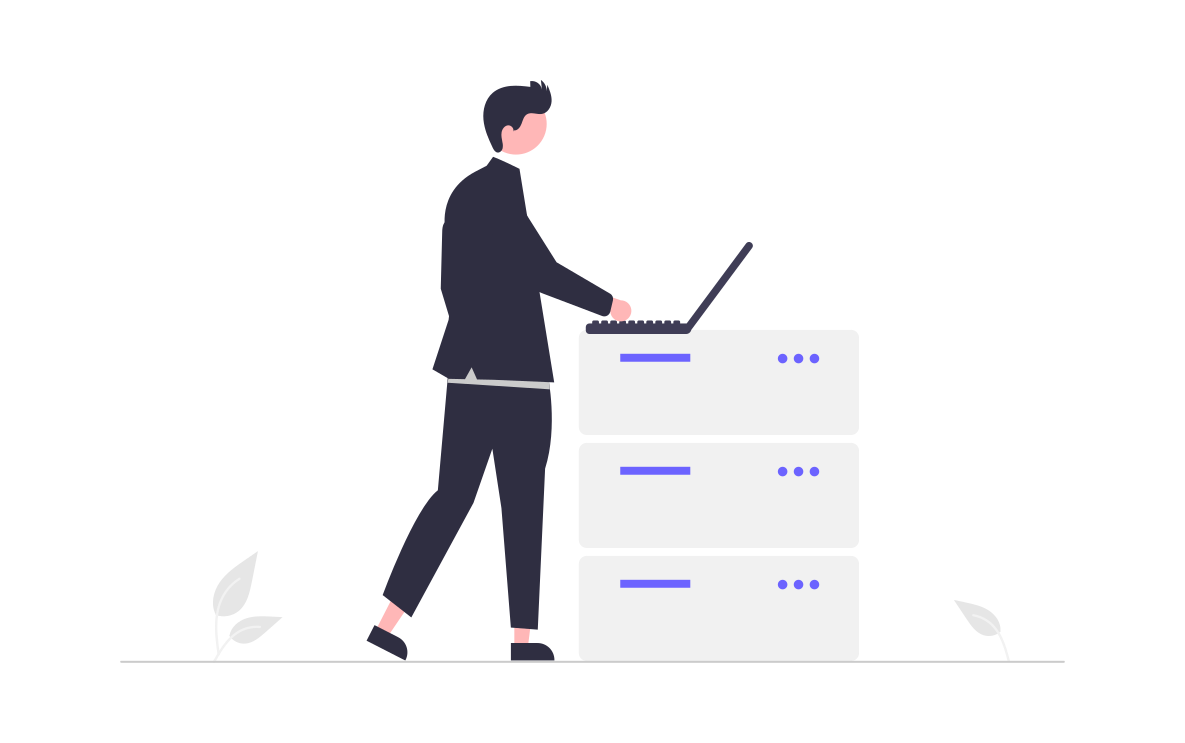
[BIBLIOGRAFIA 10](#_Toc144395925)

# INTRODUÇÃO

Se trata de um conjunto de dados, agrupado por assunto, integrados, variável em relação ao tempo e não volátil (Os dados nunca se perdem). Serve de auxílio para tomada de decisão.

Além disso, o Data Warehouse permite a redundância de dados. Também permite buscas complexas.

Figura 1: Exemplo DatawareHouse



Fonte: Portal Undraw, 2023

# Tipos de datawarehouse

Orientado ao assunto → Está sempre orientado ao redor do principal assunto da organização. Orientado por processos/funções

Integrado → Os dados são integrados. Tem uma consistência em nomes, códigos etc.

Figura 2: DatawareHouse



#### Fonte: Portal Medium, 2023

# etl

O processo ETL é fundamental no Data Warehouse e envolve as seguintes etapas:

* Extração: Dados são coletados de várias fontes, como bancos de dados, planilhas, arquivos CSV, APIs etc.
* Transformação: Os dados extraídos são limpos, formatados e transformados em um formato adequado para análise. Isso pode incluir conversão de unidades, correção de erros e enriquecimento de dados.
* Carga: Os dados transformados são carregados no Data Warehouse, onde são armazenados de maneira estruturada para permitir consultas eficientes

Os processos ETL consomem 70% do tempo de desenvolvimento em um projeto de DW.

Eles são específicos para cada organização.

# data mart

Um Data Mart é um subconjunto especializado de um Data Warehouse, projetado para atender às necessidades analíticas de um grupo específico de usuários ou de uma área funcional dentro de uma organização. Ele contém um conjunto de dados focado em um tópico ou processo de negócios específico e é otimizado para consultas e análises relacionadas a essa área.

Figura 3: Exemplos Data Mart

Diagrama

Descrição gerada automaticamente

#### Fonte: Slide usado na aula 17//08, 2023

# arquiteturas de um datawarehouse

## Arquitetura global

Utiliza um repositório comum de dados, integrado, utilizado pela organização inteira.

## Arquitetura de data marts independentes

Possui um data mart para atender a cada departamento em específico. Não se tem acesso aos data marts de outros departamentos.

## Arquitetura de data marts integrados

Possui um data mart para atender a cada departamento em específico, e os dados são compartilhados entre os data marts diferentes

# implementações de datawarehouse

## Abordagem Top-Down

Se começa do começo. Como os dados estão chegando, como vai tratar, como vai armazenar, os Data Marts. Essa abordagem quer atender às necessidades da organização como um todo, e se trata do modelo mais comum de implementação.

## Abordagem Bottom-Up

Começa de como quero ver, depois em como vou implementar. Se parte dos Data Marts até compor os DW por completo.

Por esses pontos, tem uma maior dificuldade na padronização dos dados. Apesar disso, tem-se uma implementação mais rápida manutenção mais fácil.

## Abordagem combinada

Combina características das abordagens anteriores. O planejamento geral da estruturação do DW para toda a organização. São equipes que vão trabalhando até se juntar no meio.

# modelagem multidimensional

Em nível gerencial, o foco do DW é em aplicações gerenciais. Análise histórica das infos. Aqui tem-se uma visão ampla das informações, podendo visualizar os dados em diferentes perspectivas.

Para implementar um DW é preciso um novo modelo, diferente do ER.

Na modelagem multidimensional, iremos representar os tipos de dados por uma estrutura camada **cubo de dados**

* Células contêm valores
* Lados definem as dimensões de análises

Hipercubo é quando se usa mais de 3 dimensões.

## Elementos de um DW

Fatos → Tudo aquilo que pode ser representando por valores numéricos | EX: Vendas

Dimensões → Determinam o contexto no qual os fatos são analisados | EX: Local, ano e Produto

Variáveis → Atributos numéricos que representam os fatos | EX: Valor das vendas, Unidades vendidas

## Star Schema | Esquema Estrela

É uma forma de dispor as tabelas do banco para simular um banco de dados multidimensional.

É composto por uma tabela dominante (Tabela de Fatos) rodeada de tabelas auxiliares (Tabelas de Dimensão).

## Membros

São os elementos das dimensões. Existe uma hierarquia de dimensão. EX: Cidades, Estados e Regiões formam o “Local”

## Dimensões

As dimensões representam entidades que evoluem com o tempo.

1. O histórico não é relevante. As alterações podem ser feitas diretamente no registro, sem salvar o valor anterior

2. O histórico é relevante. Existe a inserção de um novo registro na mesma entidade dimensional refletindo a mudança

3. O histórico é relevante e deseja-se analisar dados usando os status original e atual. Serão necessários os campos:

1. Status original do atributo

2. Status atual do atributo

3. Data efetiva da última alteração

## Granularidade

É o nível de detalhes das tabelas. Quanto menor o grão, maior a granularidade (Detalhamento):

## Snowflake

Esquema snowflake é uma ampliação da granularidade

## Agregação

Através da agregação cria-se entidades contendo dados sumarizados

# CONCLUSÃO

Os Data Warehouses desempenham um papel crucial na transformação de dados em insights significativos para as organizações. Eles facilitam a análise de dados históricos, a tomada de decisões informadas e a obtenção de uma compreensão profunda dos padrões de negócios. Com ferramentas de ETL, modelagem multidimensional e arquiteturas flexíveis, os Data Warehouses continuam a ser uma parte fundamental da estratégia de dados de muitas empresas para impulsionar o sucesso nos negócios. Além disso, os Data Marts oferecem uma abordagem específica e otimizada para análises de áreas funcionais ou grupos de usuários dentro da organização, contribuindo para insights mais focados e melhorando a eficácia das decisões.

Figura 4: Ilustração DatawareHouse



#### Fonte: Portal Freepik, 2023

# BIBLIOGRAFIA

VANIN, Anderson. **Slide usado em Aula sobre DatawareHouse** https://teams.microsoft.com/\_#/pdf/viewer/teams/https:~2F~2Ffatecspgov.sharepoint.com~2Fsites~2FSection\_IBD016.A914.N.139.113.20232~2FShared%20Documents~2FGeneral~2FAula%2002%20-%20IBD-016%20%E2%80%93%20BANCO%20DE%20DADOS%20-%20N%C3%83O%20RELACIONAL.pdf?threadId=19:3IUfj4PrBwUZ50rVNAVBJDw0DPwMSNVx31vZhvhT9MA1@thread.tacv2&messageId=1692299363118&baseUrl=https:~2F~2Ffatecspgov.sharepoint.com~2Fsites~2FSection\_IBD016.A914.N.139.113.20232&fileId=b38e0789-d6fe-430b-a355-98192688be5f&ctx=chiclet&viewerAction=view. Acesso: 31/08/2023

Portal Amazon AWS. **Conceitos de Data Warehouse** O que é um Data Warehouse? | Conceitos principais | Amazon Web Services Acesso em: 31/08/2023