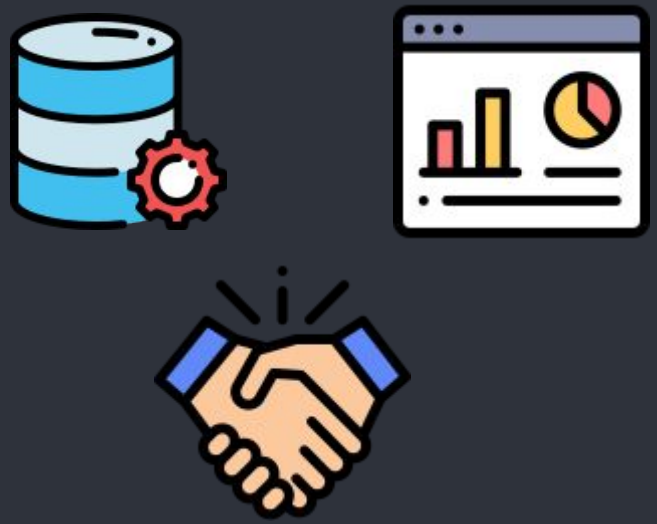
index.html

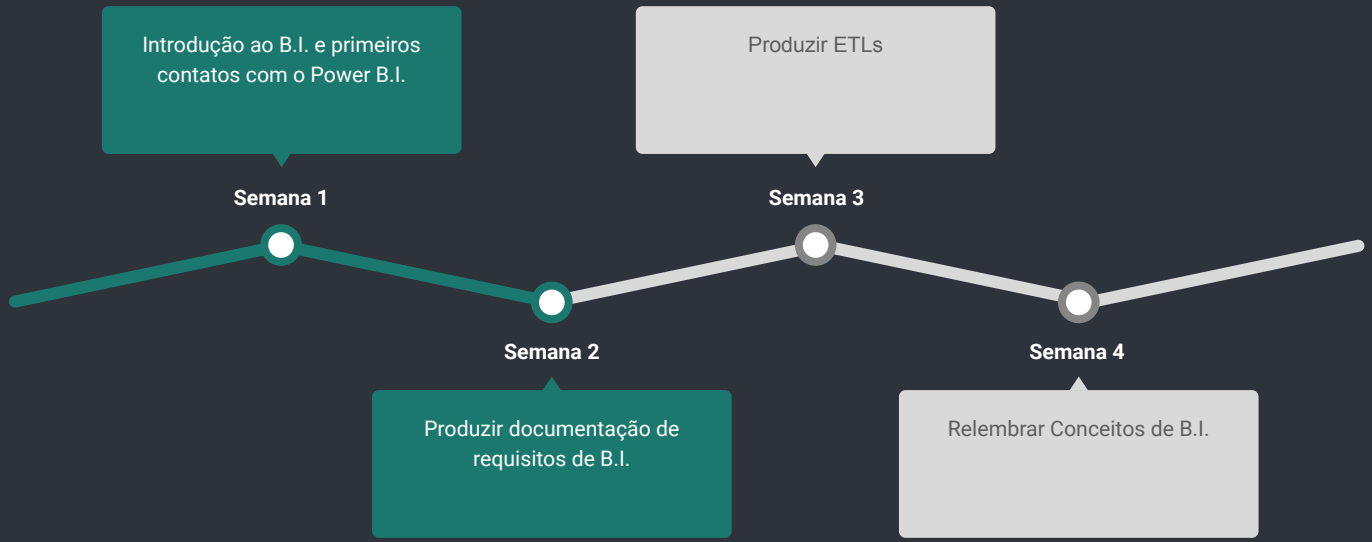
```
Business {
  [Intelligence]


  Prof. Juscelino Fernandes da Costa Junior

}
```

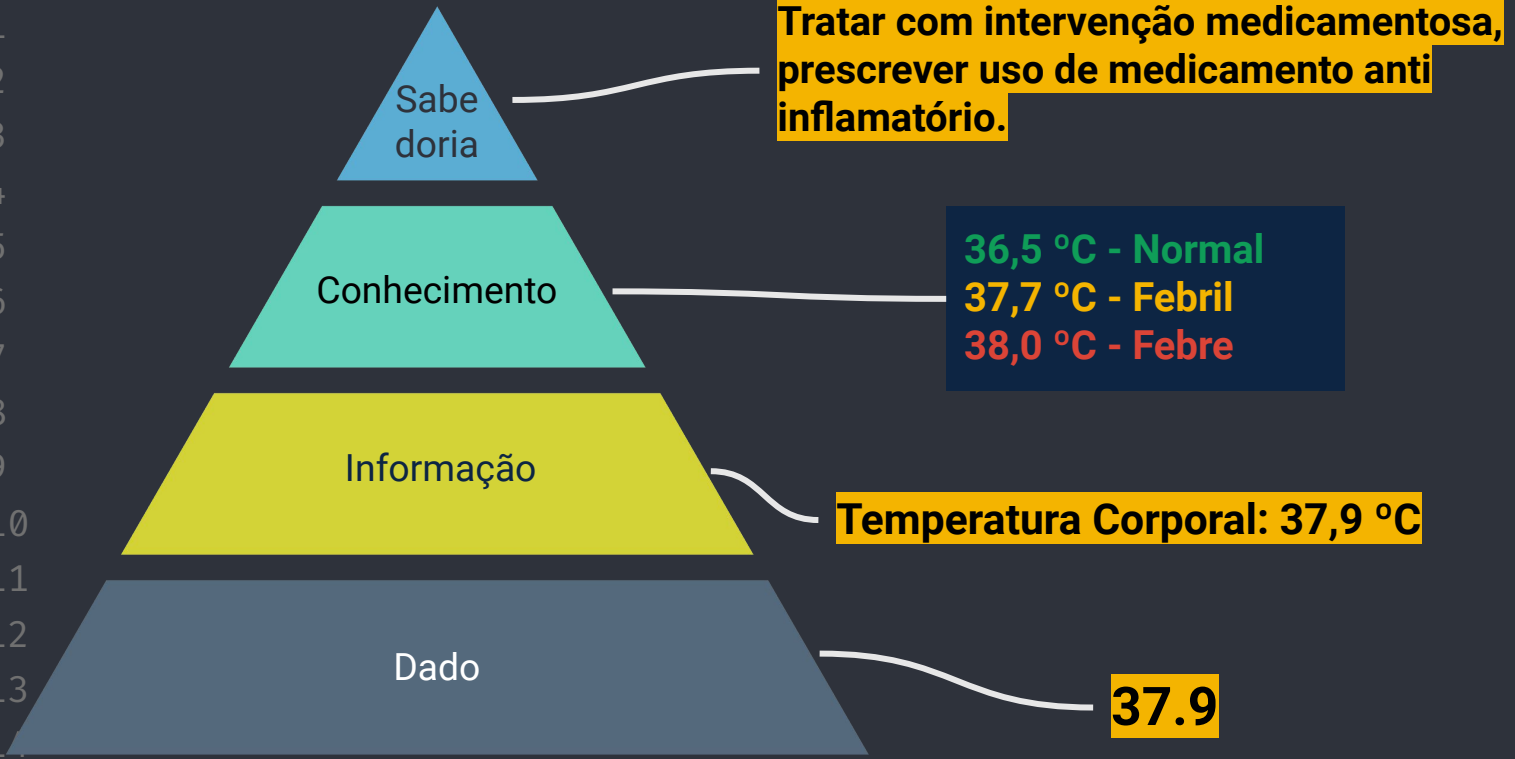
Business Intelligence;



< B.I. é o conjunto de processos e ferramentas que proporcionam o processamento e transformação de dados brutos em informações valiosas para a tomada de decisão e formação de conhecimento >

Business Intelligence;




Sabe doria
Conhecimento
Informação
Dado
Tratar com intervenção medicamentosa, prescrever uso de medicamento anti inflamatório.
36,5 ºC - Normal
37,7 ºC - Febril
38,0 ºC - Febre
Temperatura Corporal: 37,9 ºC
37.9

papeis.html

Papéis na Área de Dados {

Engenheiro de Dados

< Computação em Nuvem, Big Data, Spark, Airflow, SQL >

Report Designer

< Figma, Tableau, Superset >

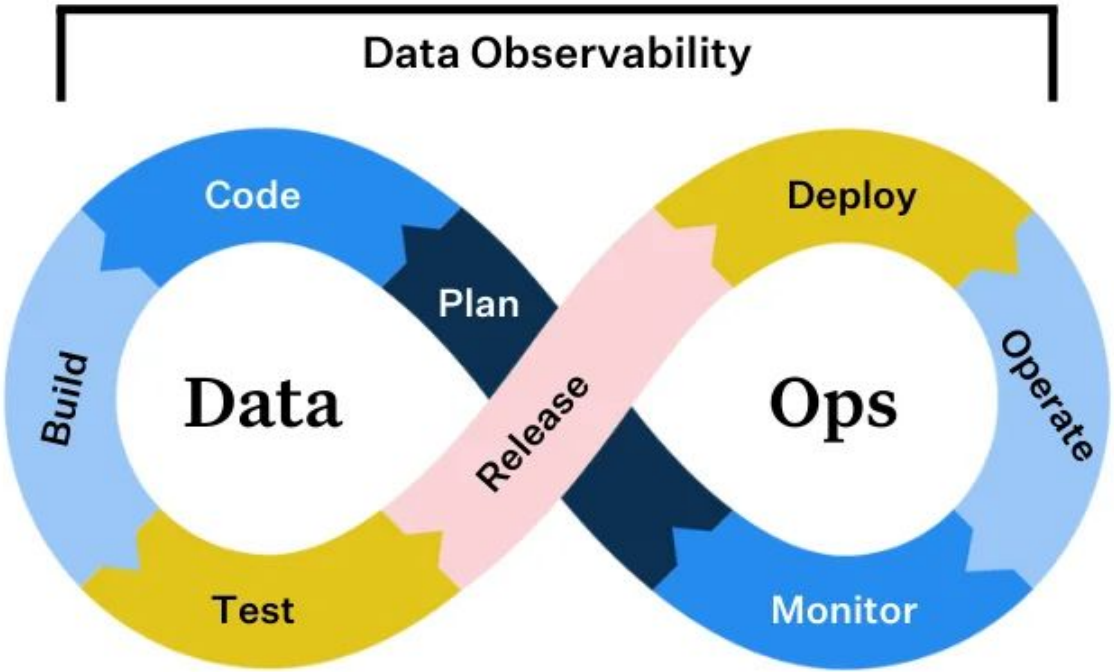
Analista de Dados

< Python, SQL, PowerBi Tableau, Qlickview, Excel>

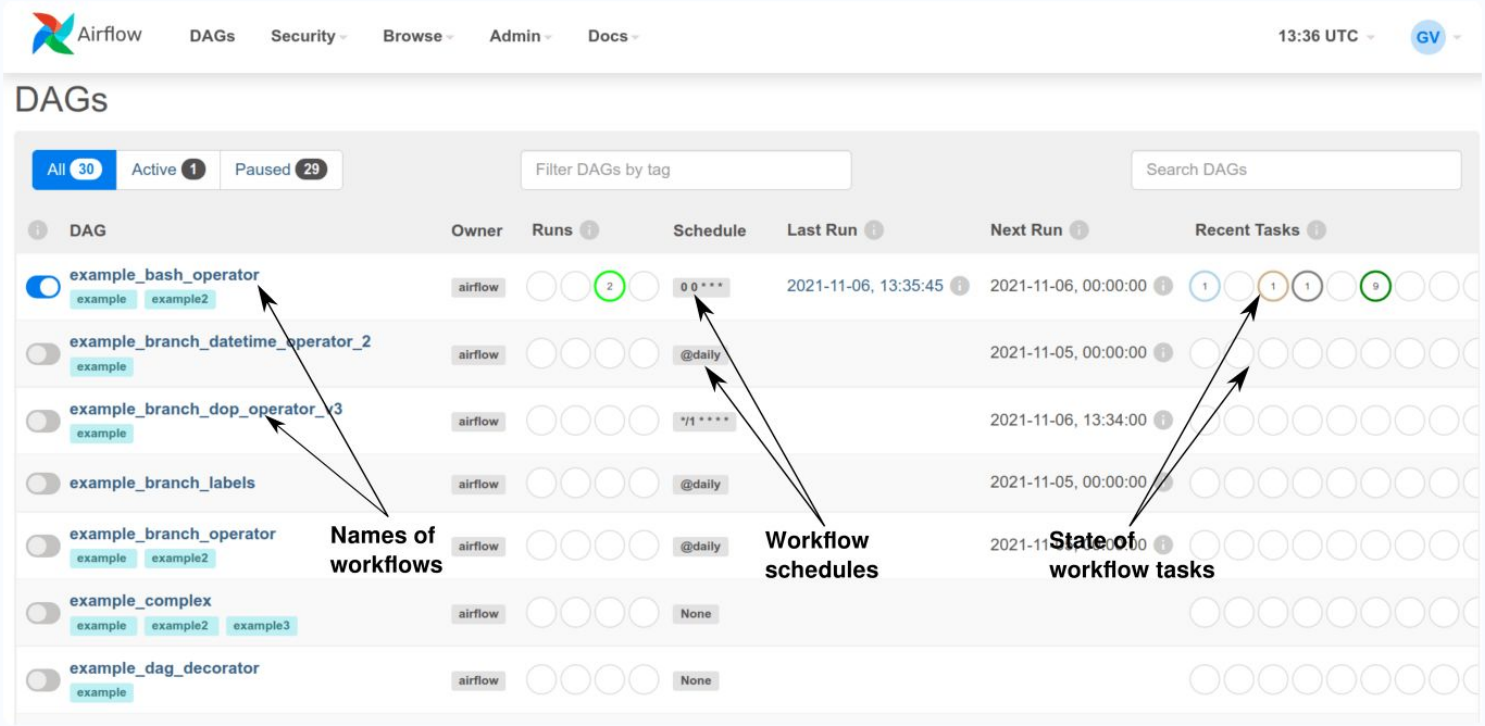
Cientista de Dados

< Python, Estatística, Machine Learning >

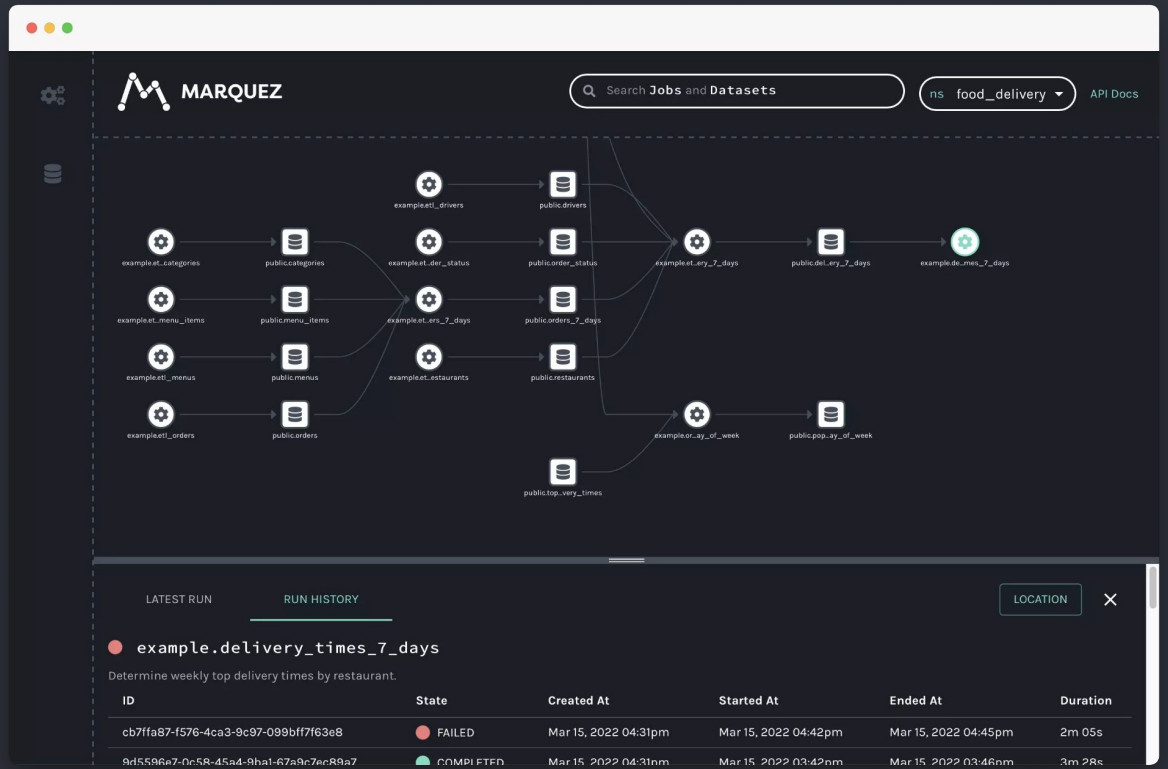
}


Data Observability
Code
Plan
Build
Data
Release
Ops
Deploy
Operate
Test
Monitor

orquestracao.json



dataLineage.json



Planejamento {

< Na etapa de **planejamento**, o objetivo é garantir que todas as partes envolvidas no ciclo de vida dos dados estejam alinhadas e que as necessidades do negócio sejam compreendidas antes de iniciar a implementação de pipelines ou sistemas. >

}

Levantar o Objetivo {

- 1. Qual o objetivo de **negócio** do cliente?
  Dica:Importante o olhar para o negócio e não para o produto final.

  Ex: Reduzir o tempo de processamento de pedidos em 30% nos próximos 12 meses, aumentando a satisfação do cliente e otimizando a eficiência operacional.

}

Levantar Indicadores {

- 2. Quais indicadores podem medir este objetivo?

  Ex:

  - Tempo Médio de Processamento de Pedidos
  - Net Promoter Score (Satisfação dos Clientes)

}

Levantar Indicadores {

- 2.1. Quais fatores influenciam para que este objetivo seja atingido?

  Ex:

  - Capacidade computacional
  - Disponibilidade de Produtos
  - Nível de Demanda
  - Disponibilidade de mão de obra
  - Automação de processos
  - Erro em transações
  - Aprovação de Pagamento
  - Disponibilidade de Sistemas
  - Conformidade de Fornecedores

}

Levantar Indicadores {

- 3. Crie métricas a partir dos indicadores propostos.

  Ex:

  - Tempo Médio de Processamento de Pedidos (Média)
  - Quantidade de Pedidos (Soma)
  - Percentual de uso dos recursos computacionais.
  - Percentual de produtos disponíveis em estoque no momento em que o pedido é realizado.
  - Percentual de fornecedores que entregam produtos dentro dos prazos e padrões acordados.

}

Estudo de Fontes de Dados {

- 4. Qual a fonte de dados disponível para medir este indicador.

  Ex:

  - Banco de Dados
  - Planilha Excel
  - Log de Sistema

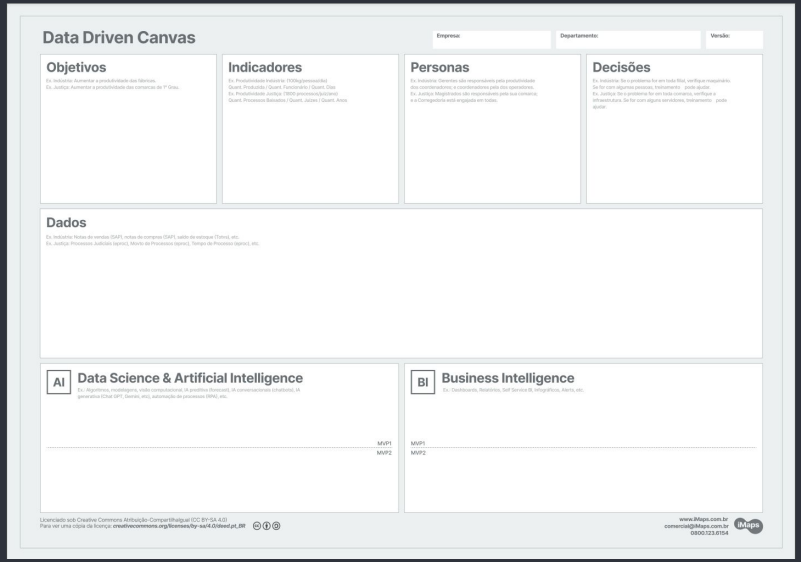
}

Documentação de Requisitos {

Data Driven Canvas

< O DDC é uma abordagem visual e bastante rápida de documentar requisitos. >

}



Documentação de Requisitos {

BUS Matrix

< O BUS Matrix ou Matriz Indicado X Dimensão é uma abordagem que confronta indicadores e dimensões em uma tabela >

}

Documentação de Requisitos {

BUS Matrix

BUSINESS PROCESSES	SHARED DIMENSIONS: Account	Customer	Data	Department	Employee	Organization	Products	Promotion	Reseller	Sales Territory
Customer Service Calls		✓	✓		✓		✓			
Customer Surveys		✓	✓				✓			
General Ledger	✓		✓	✓		✓				
Internet Sales		✓	✓				✓	✓		✓
Intventory			✓				✓			
Reseller Sales			✓		✓		✓	✓	✓	✓
Sales Plan			✓				✓			✓

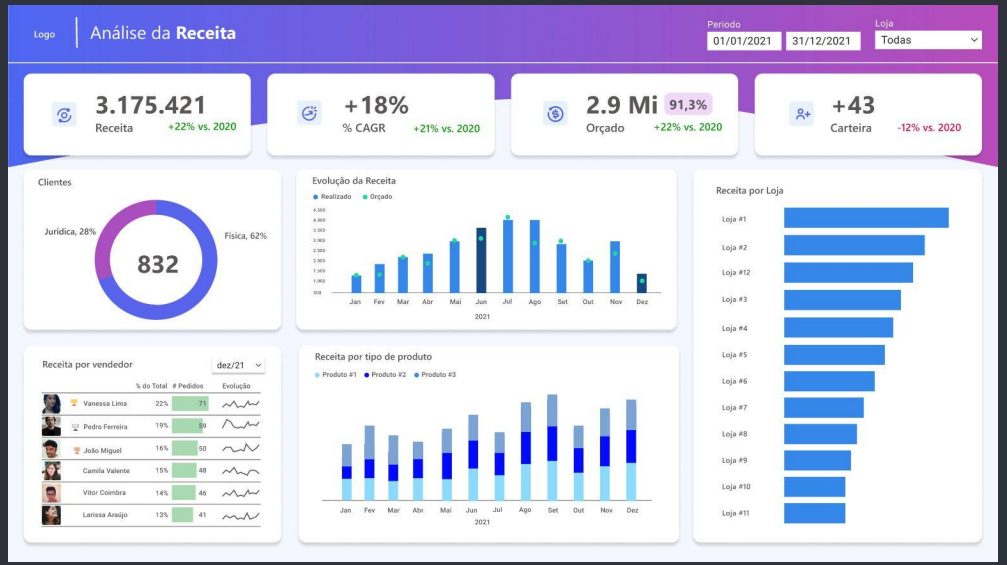
}

Documentação de Requisitos {

Prototipação

<Figma, PowerBi>

}



index.html

```
Business {
  [Intelligence]


  Prof. Juscelino Fernandes da Costa Junior

}
```

Fontes de Dados {

Estruturados

SGDB

API

Não Estruturados

Planilha

Logs

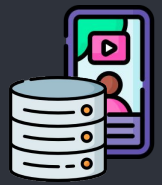
Gmail

Vídeos

Fotos

Audio

}



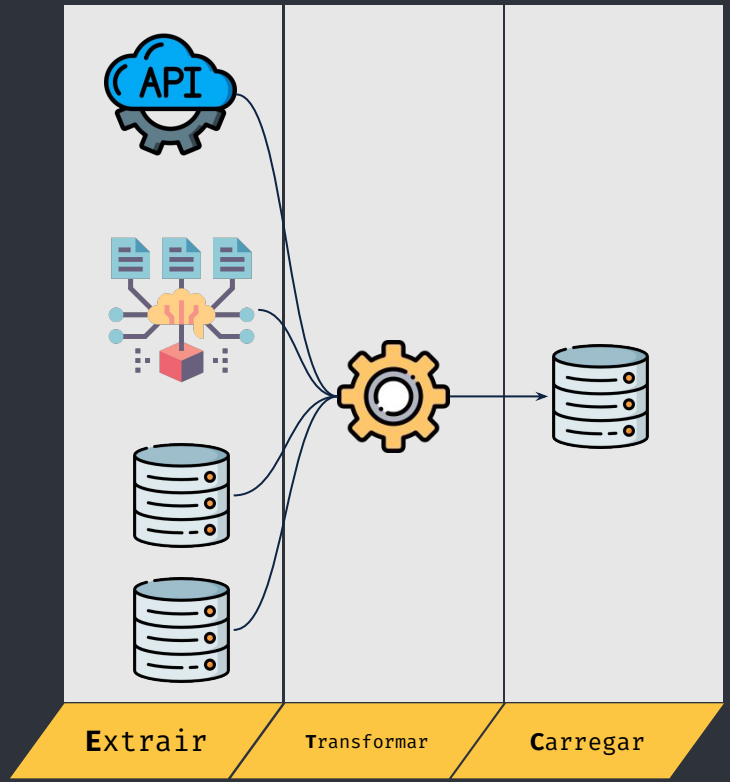
OLTP (Transacional)

OLTP é otimizado para inserção, atualização e consistência dos dados em tempo real.

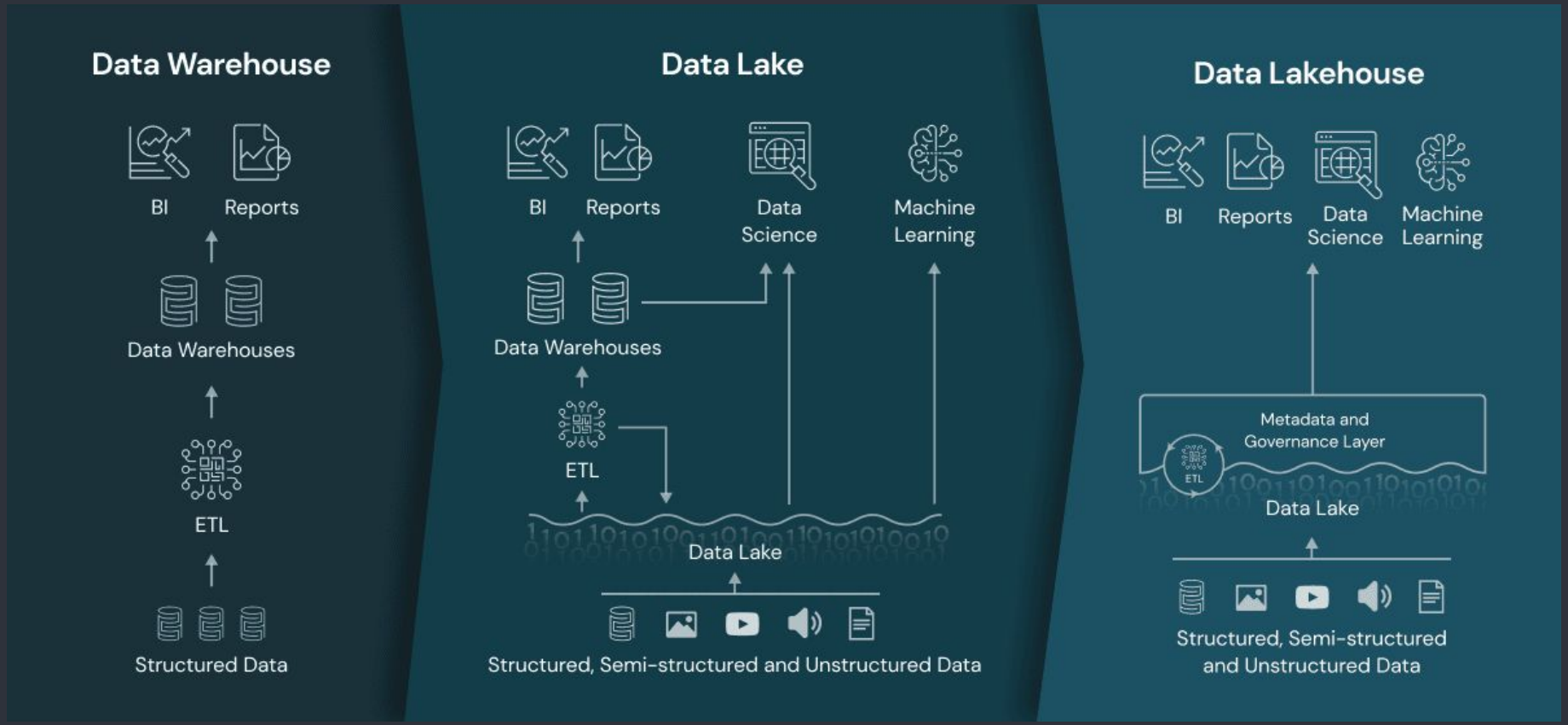
OLAP (Analítico)

O OLAP é otimizado para leitura pesada e análises de dados complexos.

< Dashboards de B.I. devem evitar consumir dados diretamente de bancos OLTP >



< **ETL** é o processo de Extrair, Transformar e Carregar dados. O intuito deste processo é extrair os dados de diversas fontes, aplicar as transformações necessárias e realizar cargas em um banco de dados OLAP, ao qual denominamos **Data Warehouse** >


Data Warehouse
BI
Reports
Data Warehouses
ETL
Structured Data
Data Lake
BI
Reports
Data Science
Machine Learning
Data Warehouses
ETL
Data Lake
Structured, Semi-structured and Unstructured Data
Data Lakehouse
BI
Reports
Data Science
Machine Learning
Metadata and Governance Layer
ETL
Data Lake
Structured, Semi-structured and Unstructured Data

Data Warehouse {

Análises Multidimensionais

O DW organiza suas tabelas em **fatos** e **dimensões**.

Fato

Uma tabela de fatos guarda **eventos numéricos** como vendas, lucros ou **quantidades**. Ela responde a "o quê, quanto e quando". É a base para análises e sempre se conecta a dimensões para dar contexto.

}

Data Warehouse {

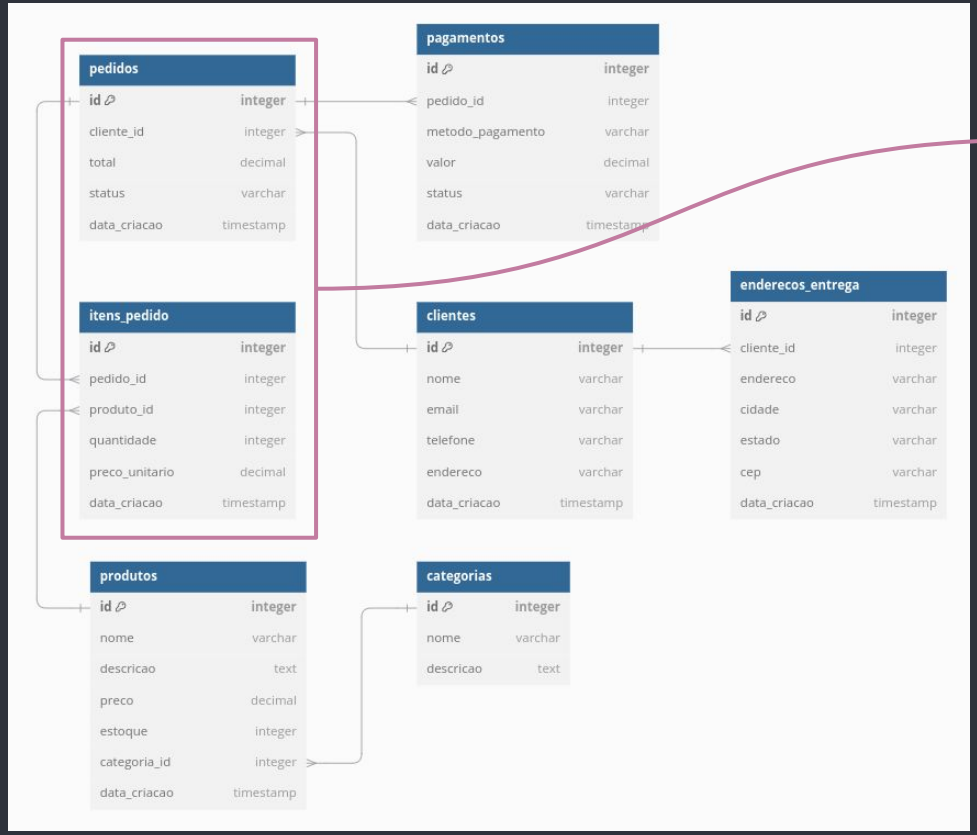
Análises Multidimensionais

O DW organiza suas tabelas em **fatos** e **dimensões**.

Dimensão

Uma tabela de dimensão organiza categorias, como tempo, produto ou cliente. Ela dá detalhes sobre o fato e permite explorar os dados por diferentes ângulos, como "quem, o quê e onde".

}



Fatos:

fVendas

fPagamentos

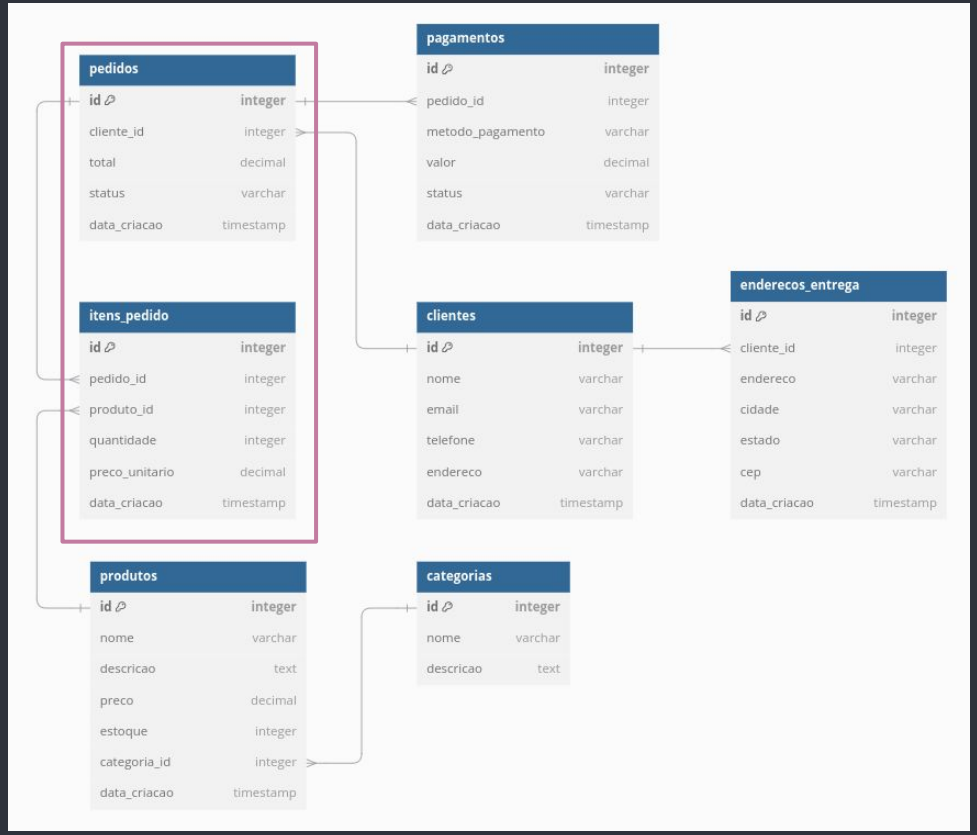
Dimenssões:

dProduto

dCategoria

dCliente

dLocalidade



Filtragens:

status = ‘VENDIDO’

Cálculos e Agregações:

Vl_venda = quantidade * preco_unitario

Transformação {

Limpeza e Tratamento

Se nulo → N/A , Remover linhas sem valores, correção ortográfica.

Agregação de Dados

Soma, Média, Mediana, Desvio Padrão, Máximo, Mínimo, Concatenação.

Classificação

Status = 1 → Pendente

Status = 2 → Concluído

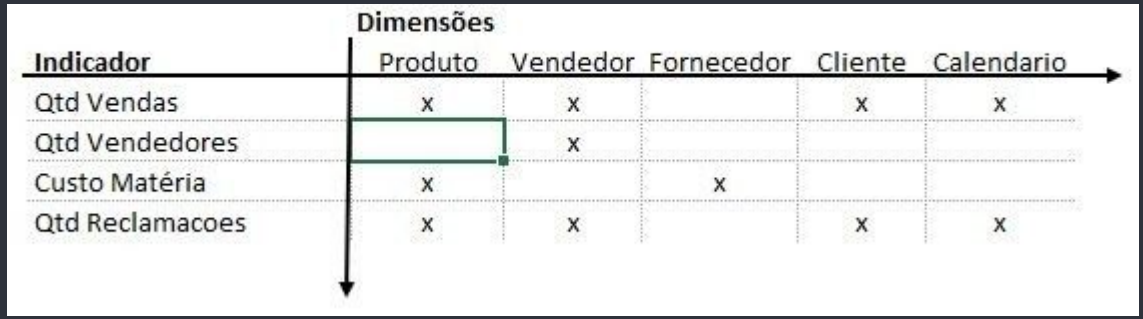
Cálculos

}

BUS Matrix {

< O BUS Matrix, é uma matriz que confronta **indicadores** e **dimensões**, facilita bastante para tratar diretamente com o cliente. >

}

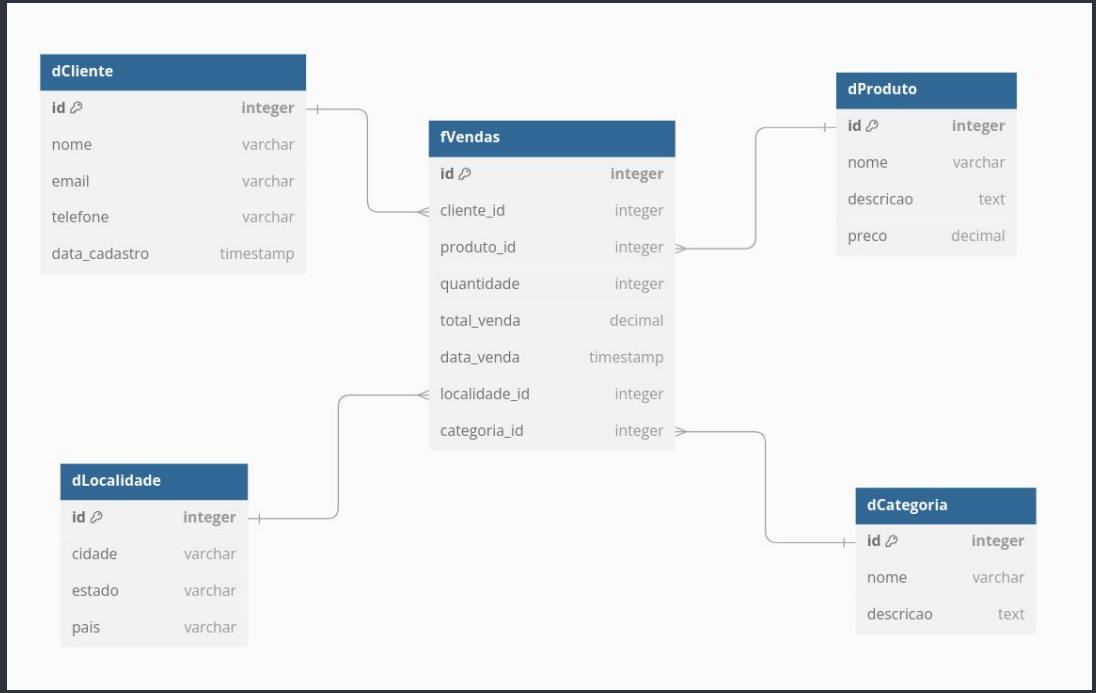


Indicador	Dimensões: Produto	Vendedor	Fornecedor	Cliente	Calendario
Qtd Vendas	x	x		x	x
Qtd Vendedores		x			
Custo Matéria	x		x		
Qtd Reclamacoes	x	x		x	x

Star Schema {

< O Star Schema ou Estrela, é um formato de modelagem de dados mais utilizado em Data Warehouses. >

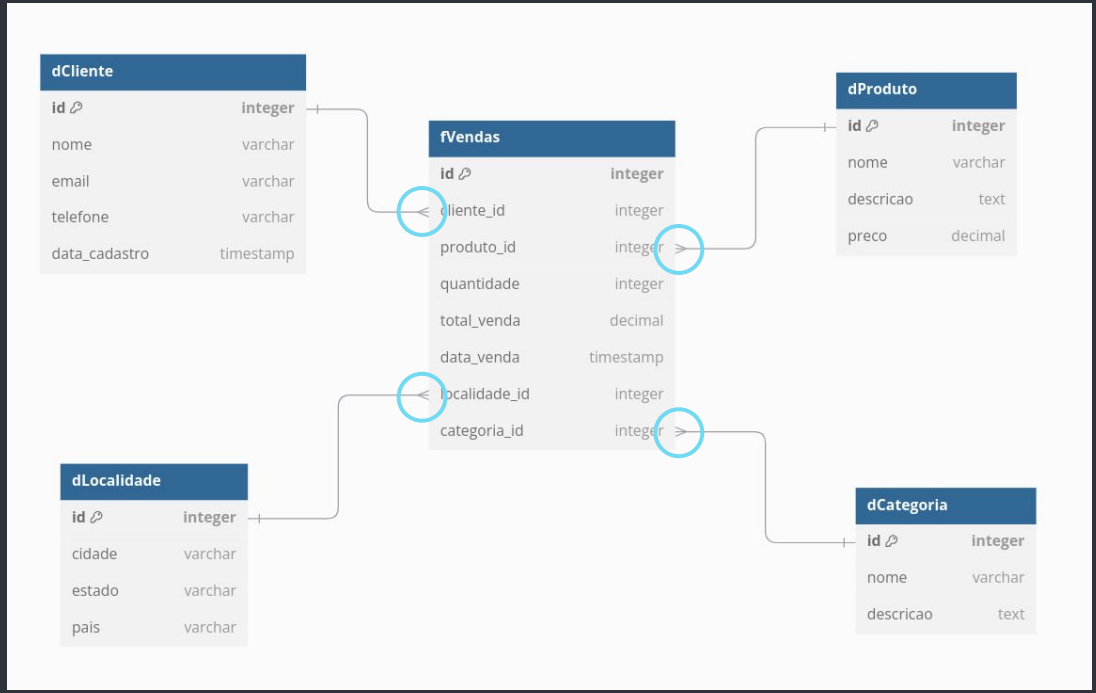
}



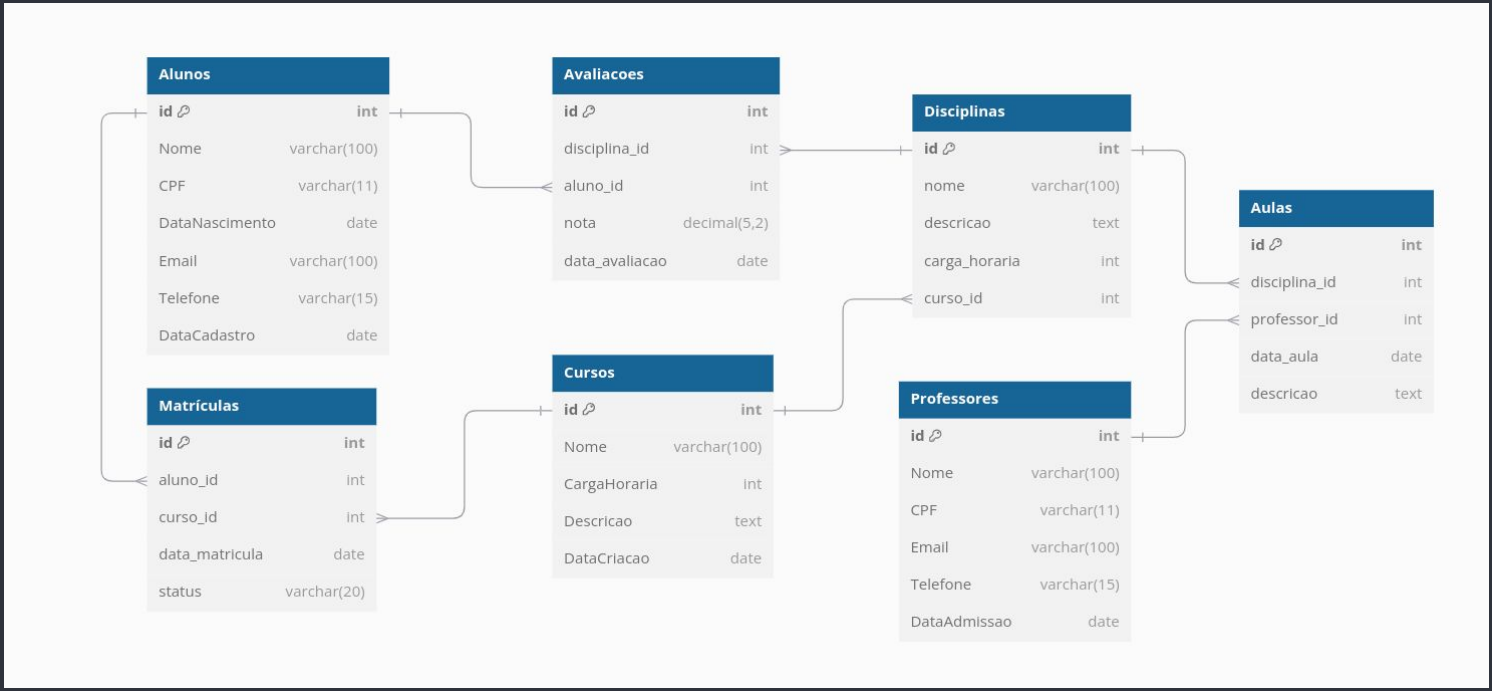
Star Schema {

< Uma característica marcante no Star Schema, é o relacionamento n:1 da fato para as dimensões. Um modelo estrela deve sempre seguir este mesmo formato. >

}



TDE {


Alunos
id int
Nome varchar(100)
CPF varchar(11)
DataNascimento date
Email varchar(100)
Telefone varchar(15)
DataCadastro date
Avaliacoes
id int
disciplina_id int
aluno_id int
nota decimal(5,2)
data_avaliacao date
Disciplinas
id int
nome varchar(100)
descricao text
carga_horaria int
curso_id int
Aulas
id int
disciplina_id int
professor_id int
data_aula date
descricao text
Matrículas
id int
aluno_id int
curso_id int
data_matricula date
status varchar(20)
Cursos
id int
Nome varchar(100)
CargaHoraria int
Descricao text
DataCriacao date
Professores
id int
Nome varchar(100)
CPF varchar(11)
Email varchar(100)
Telefone varchar(15)
DataAdmissao date

}

index.html

```
Business {
  [Intelligence]


  Prof. Juscelino Fernandes da Costa Junior

}
```

ETL < **Extração** />

/** A extração é a base de um pipeline ETL, é o processo de leitura dos dados. */

/** Vamos falar de pipeline mais adiante. */

Batch

Processamento em lote (batch) é a execução de tarefas ou cargas de trabalho de forma agrupada e programada, processando grandes volumes de dados de uma vez.

Utilização: Ideal para processamento periódico de grandes quantidades de dados, onde o tempo de resposta imediato não é essencial.

Principais Ferramentas:

- Injestores:
  - Spark
  - Airbyte
- Orquestradores:
  - Airflow
  - Jenkins

Streaming

Processamento em streaming é o tratamento de fluxos contínuos de dados em **tempo real**. Os dados são processados à medida que chegam, permitindo respostas e ações rápidas baseadas em eventos.

Utilização: Ideal para cenários onde os dados precisam ser processados continuamente.

Principais Ferramentas:

- Kafka
- Spark
- Dataflow (gcp)

Full Extraction

Extrai **todos** os dados da fonte.

Útil em situações onde os dados não mudam frequentemente, mas exige mais tempo e processamento.

Incremental

Apenas os dados que foram **alterados** ou **adicionados** desde a última extração são extraídos.

Reduz o volume de dados e otimiza o processo de ETL.

carga_full.py

```python
# Consulta para extrair todos os dados da tabela de vendas
query_full = "SELECT * FROM sales"
df_sales_full = pd.read_sql(query_full, mysql_engine)

# Carregando os dados para o Snowflake
df_sales_full.to_sql(
    'sales_snowflake',
    snowflake_engine,
    index=False,
    if_exists='replace'

)
```

Python

```python
# Definindo a data da última carga incremental
ultima_data_carga = '2024-10-10 12:00:00'

# Consulta para extrair apenas dados novos ou modificados após a última carga
query_incremental = f'SELECT * FROM sales WHERE last_modified > {ultima_data_carga}'

df_sales_incremental = pd.read_sql(query_incremental, mysql_engine)

# Carregando os dados incrementais para o Snowflake
df_sales_incremental.to_sql(
    'sales_snowflake',
    con=snowflake_engine,
    index=False,
    if_exists='append'
)
```

Python