



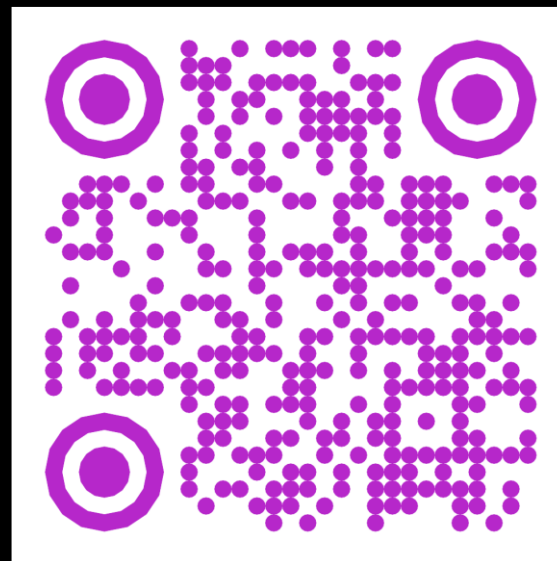
Digital College

ENSINO DE HABILIDADES DIGITAIS

Formação em Data Analytics

Unidade 2 – Módulo 1

Professor: **MSc. Alex Souza**



Sobre o curso (Unidades)

Unidade 01 – Banco de Dados

Unidade 02 – ETL

Unidade 03 – Python para Análise de Dados

Unidade 04 – Power BI Desktop



Sobre o curso (Unidades e Módulos)

Unidade 02 – Extração, Transformação e Carga - ETL	1 – Processo de Descoberta de Conhecimento
	2 – Business Intelligence
	3 – Extração, Transformação e Carga de Dados
	4 – Pipeline de Dados



Sobre o curso (Unidades e Módulos)

Unidade 03 – Python para Análise de Dados	1 – Iniciando com Python
	2 – Manipulação de Dados com Python
	3 – Pandas para Análise de Dados
	4 – Gráficos com Python

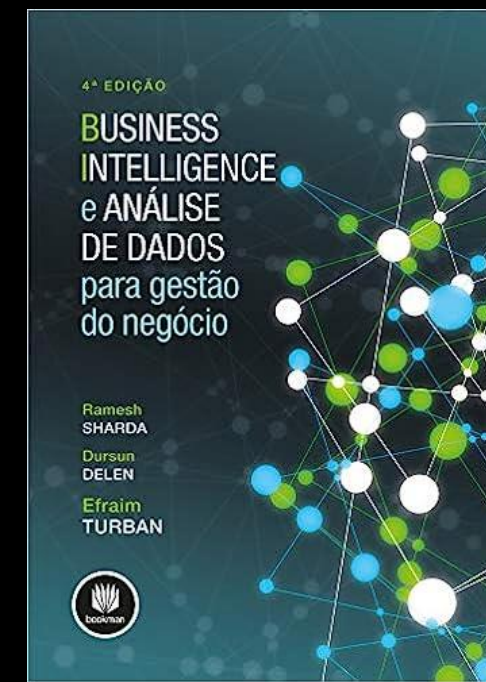
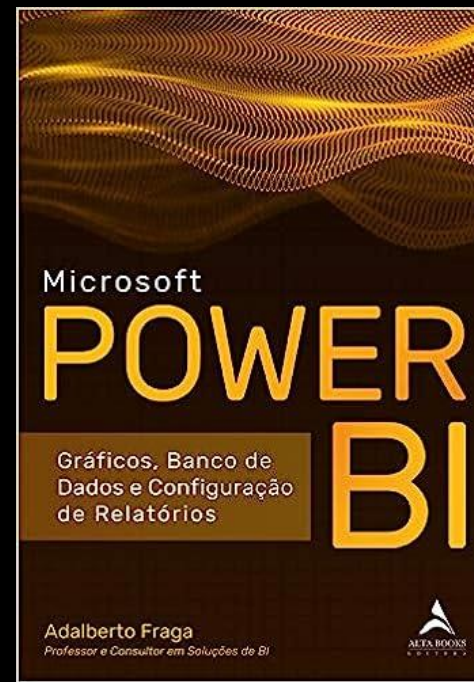
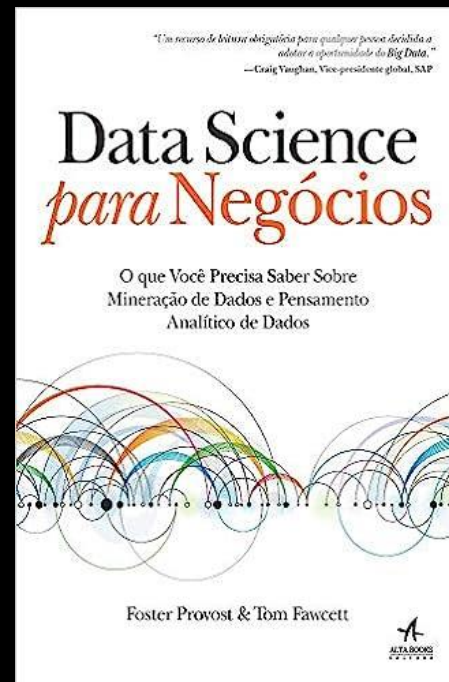
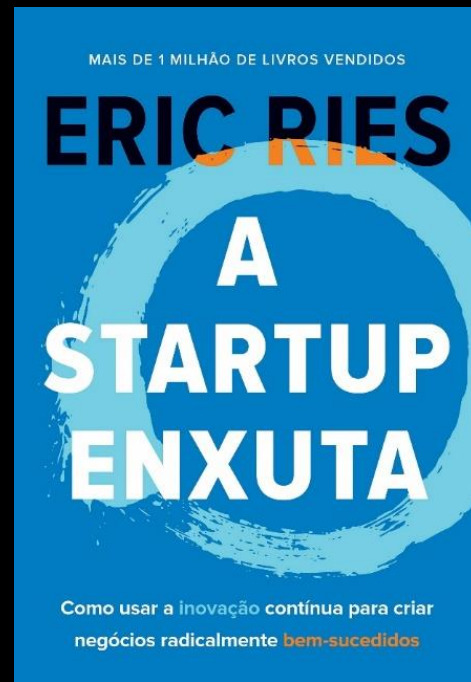
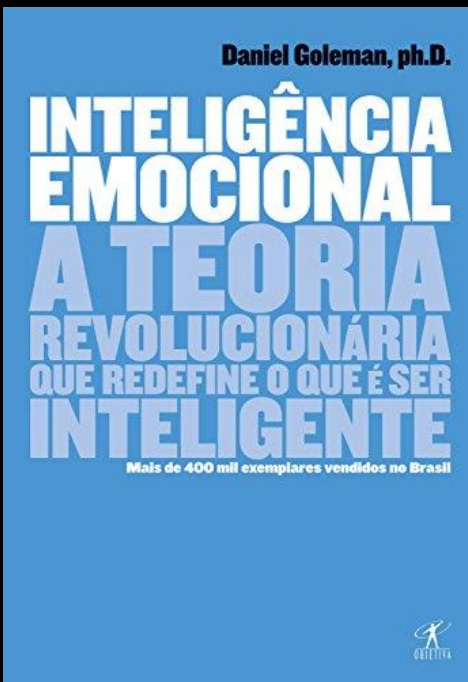
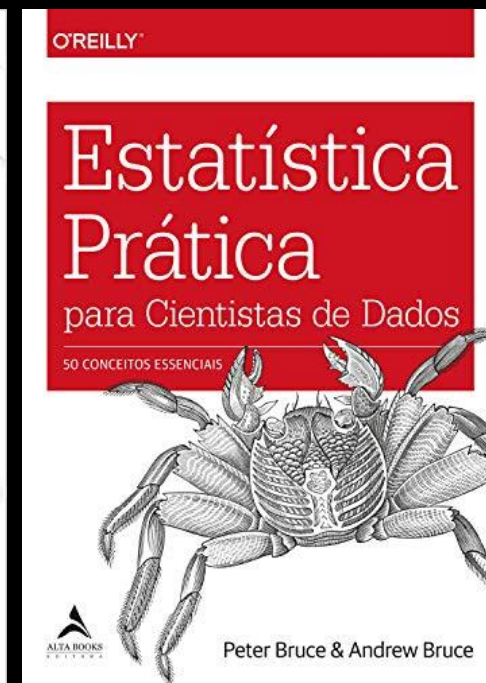
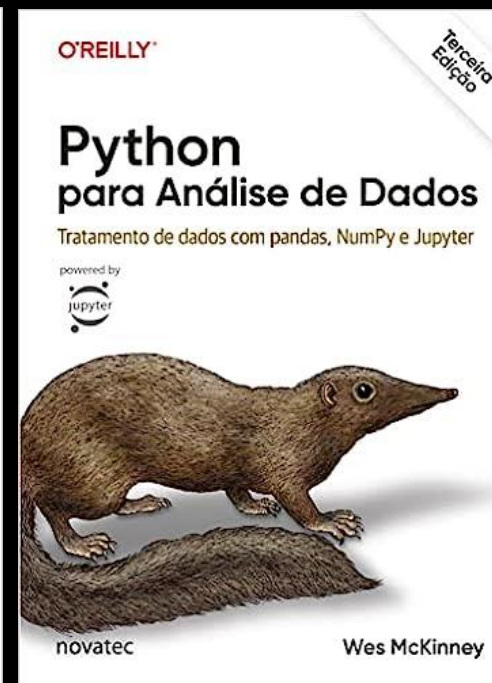
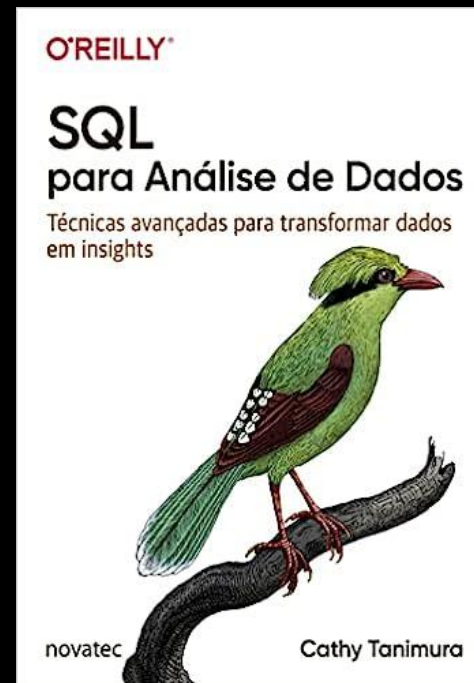
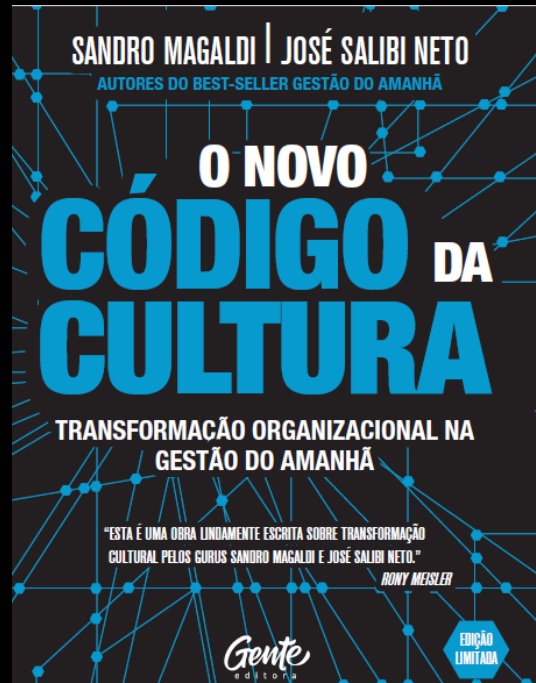


Sobre o curso (Unidades e Módulos)

Unidade 04 – Power BI Desktop	1 – Conhecendo o Power BI
	2 – Modelagem, Relacionamentos e DAX
	3 – Visualização de Dados



Livro



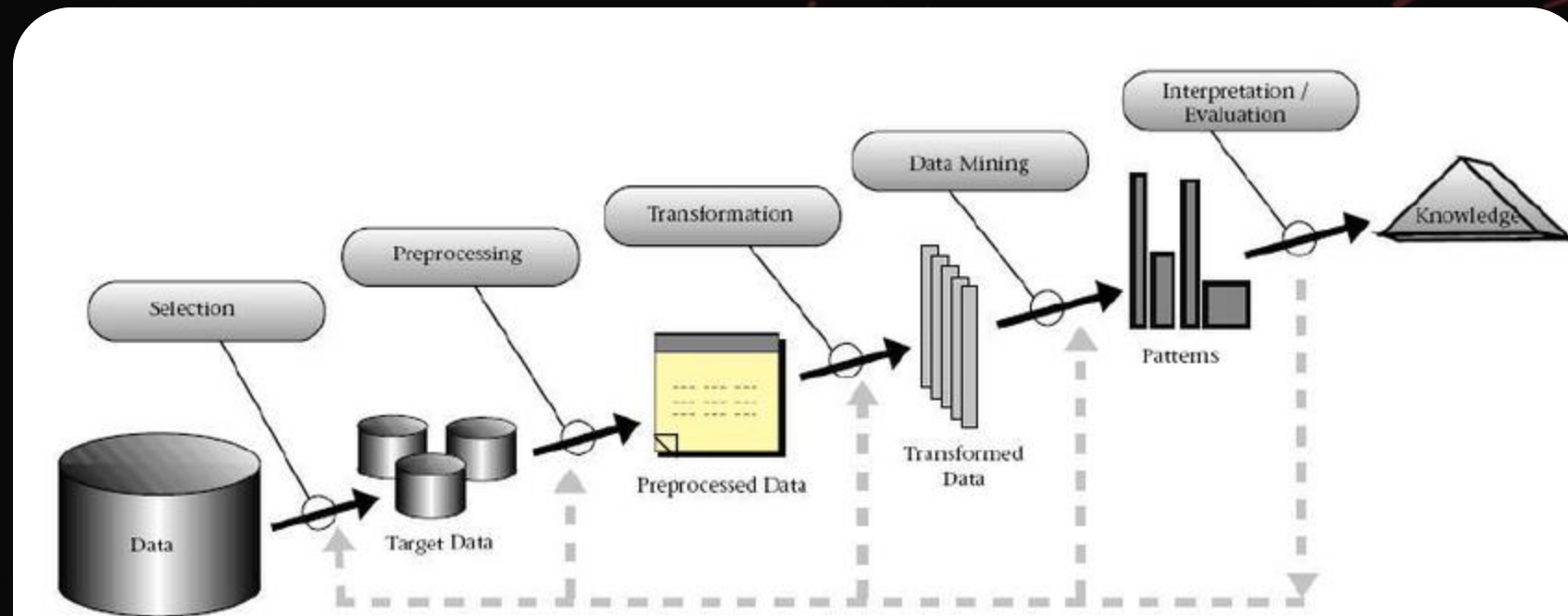


Descoberta de Conhecimento em Bancos de Dados (KDD)

KDD - *Knowledge Discovery in Databases* (Descoberta de Conhecimento em Bancos de Dados) é um processo que envolve **seleção, pré-processamento, transformação** e mineração de dados, avaliação de padrões, apresentação de resultados e utilização do conhecimento. Essa técnica ajuda a extrair informações úteis e valiosas de grandes quantidades de dados.



Etapas do processo de KDD



An overview of the steps that compose the knowledge discovery in databases (Fayyad et al. 1996)

An overview of the steps that compose the knowledge discovery in databases (Fayyad et al. 1996)

Seleção de dados

Seleção manual

Selecionar manualmente os dados que são relevantes para a análise.

Utilização de filtros

Utilizar filtros para selecionar os dados que são relevantes para análise.

Amostragem

Selecionar uma amostra representativa dos dados para análise.

Dados completos

Analisar a totalidade dos dados disponíveis.

Pré-processamento de dados

Transformação e Enriquecimento de dados

Converter dados em **formatos** adequados e **limpos**, **corrigir** os **valores** que estiverem errados. **Enriquecer** dados.

A horizontal red line with two red square markers containing the numbers 1 and 2. A vertical red line extends upwards from marker 1 to the 'Limpeza de dados' section, and another vertical red line extends upwards from marker 2 to the 'Transformação e Enriquecimento de dados' section.

1

Limpeza de dados

Remover dados **inconsistentes**, **duplicados** e **irrelevantes**.

2

Transformação de Dados

1 Normalização

Transformar os valores dos dados para uma escala comum.

2 Discretização

Converter dados contínuos em dados discretos.

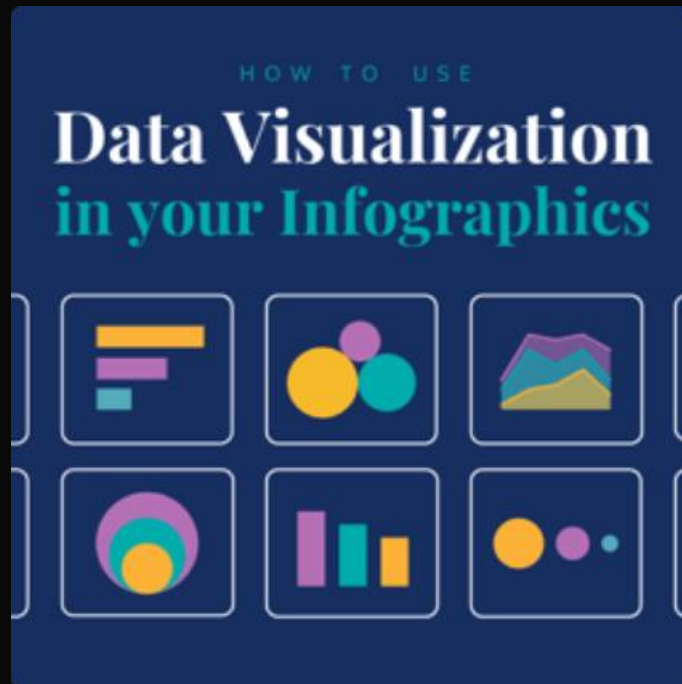
3 Agregação de dados

Agrupar dados em categorias ou grupos.

4 Redução de dados

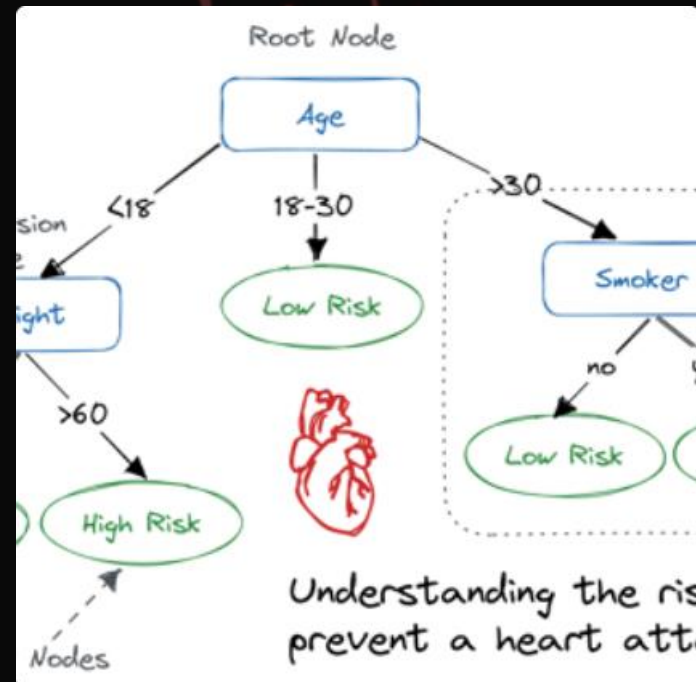
Reduzir a dimensionalidade dos dados mantendo sua importância.

Data Mining



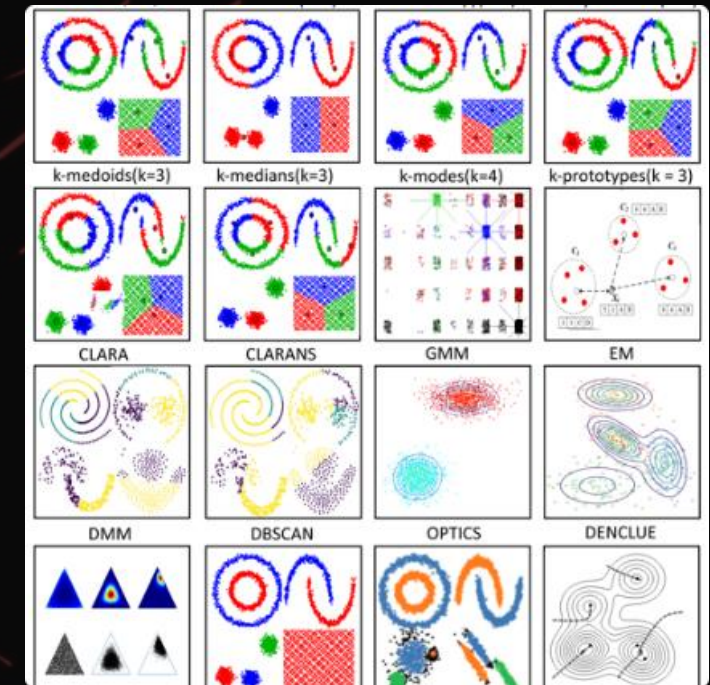
Visualização de Dados

Gráficos e mapas para ajudar a ver tendências e padrões nos dados.



Árvores de Decisão

Um modelo para ajudar na identificação de padrões e tendências



Clustering

Agrupamento de dados em conjuntos com base nas suas características comuns.

Avaliação de Padrões

Seleção de Padrões

Selecionar padrões com base no seu interesse.

Avaliação de Padrões

Verificar a validade dos padrões identificados de acordo com critérios predefinidos.

Interpretação

Compreender o significado dos padrões descobertos.

Apresentação de Resultados

Visualização de Dados

Gráficos e mapas para ajudar a ver tendências e padrões nos dados.

1

Relatórios

Incluir estatísticas, tabelas, gráficos e outros recursos visuais.

2

3

Interpretação

Apresentar o significado e as implicações dos padrões identificados.

Mais sobre KDD?



CRISP-DM:

Descobrendo Conhecimentos em Dados

CRISP-DM (*Cross Industry Standard Process for Data Mining*)



Introdução ao CRISP-DM

1 O que é o CRISP-DM?

O CRISP-DM é uma metodologia utilizada para mineração e análise de dados. Ele é frequentemente usado para ajudar as empresas a identificar insights em seus dados.

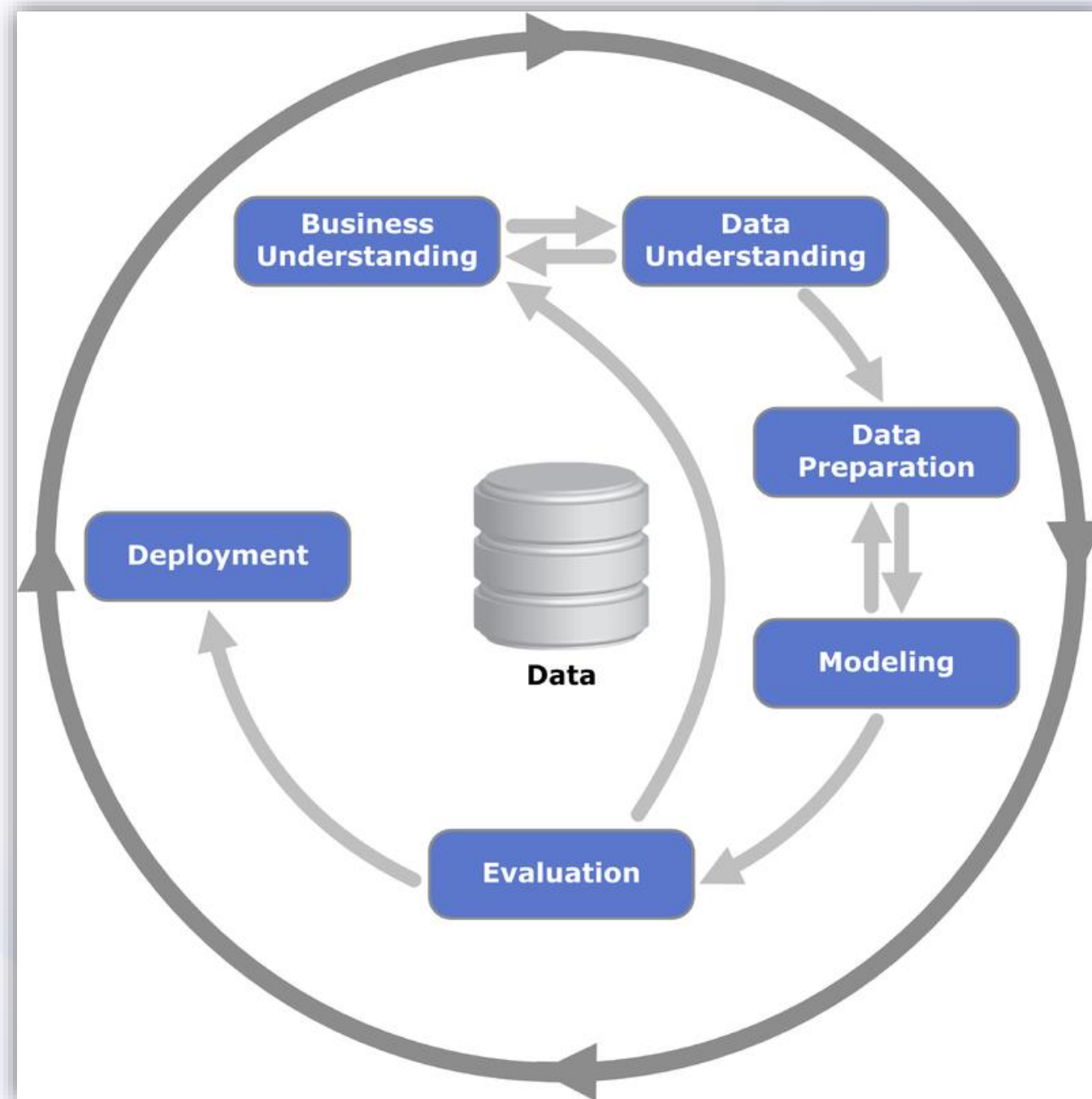
2 Qual a sua origem?

O CRISP-DM foi criado em 1996 por um consórcio composto por várias empresas líderes em mineração de dados. Tornou-se uma estrutura amplamente aceita para mineração de dados empresarial.

3 Pra que serve?

O CRISP-DM permite uma abordagem sistemática para mineração de dados, ajudando as organizações a alcançar insights de seus dados de maneira consistente e repetível.

CRISP-DM



Fases do CRISP-DM

Entendimento do Negócio

Os objetivos do projeto são definidos, analisando-se as metas da empresa, determinando as principais perguntas de negócios a serem respondidas.

Preparação dos Dados

O conjunto de dados é preparado para a mineração de dados, limpo, traduzido e transformado em um formato aceitável para a análise.

Avaliação

O modelo construído é avaliado e revisado quanto à sua precisão e eficácia.

Entendimento dos Dados

Os dados relevantes para o projeto são identificados, coletados, integrados e selecionados para posterior análise.

Modelagem

O modelo final é selecionado para auxiliar a realização do objetivo de negócios pré-determinado.

Modelagem

O modelo implantado é monitorado e mantido para garantir que ele continue a atender às necessidades de negócios em mudança



Mais sobre CRISP-DM?

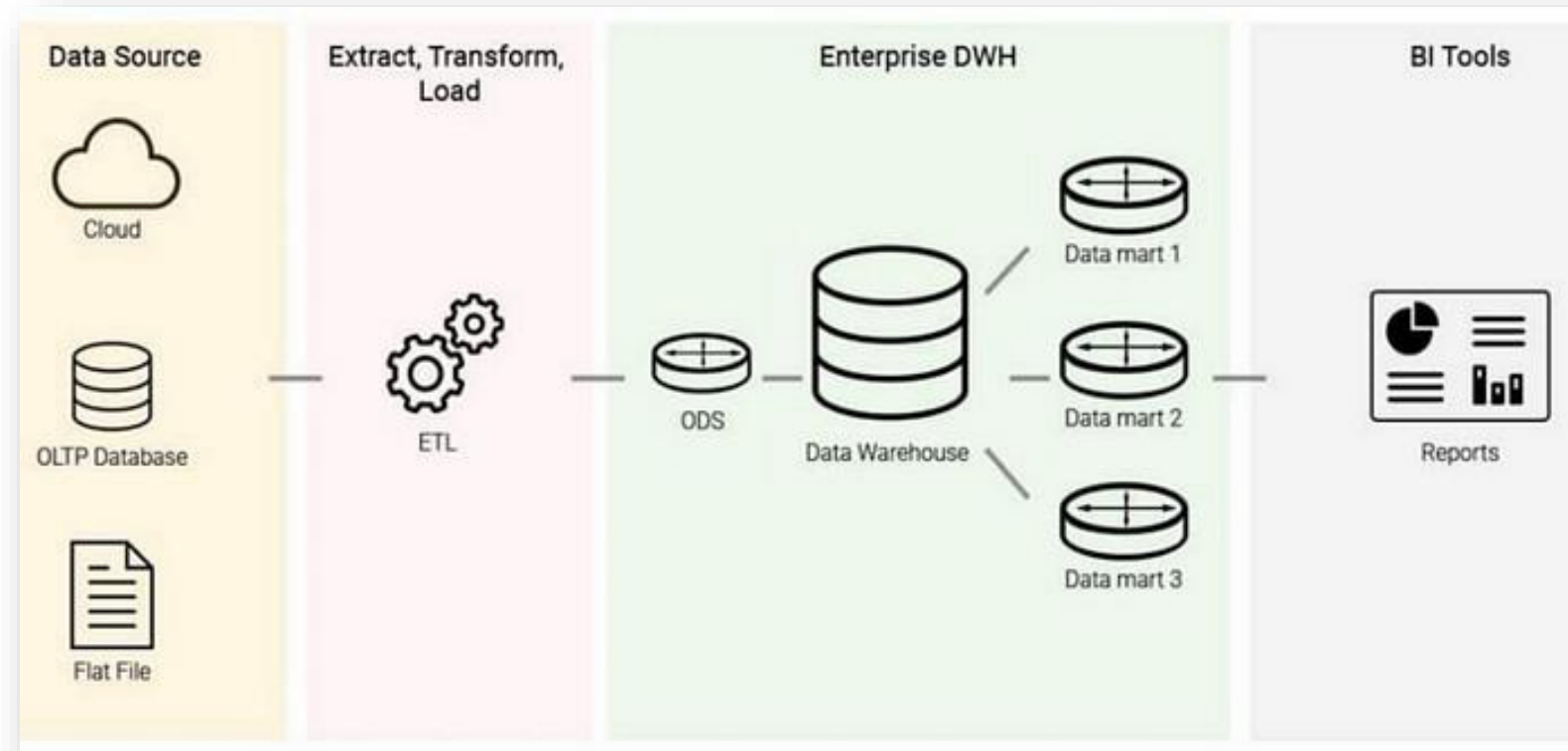


Termo usado para descrever decisões gerenciais tomadas com base em **evidências** e **fatos** (dados).

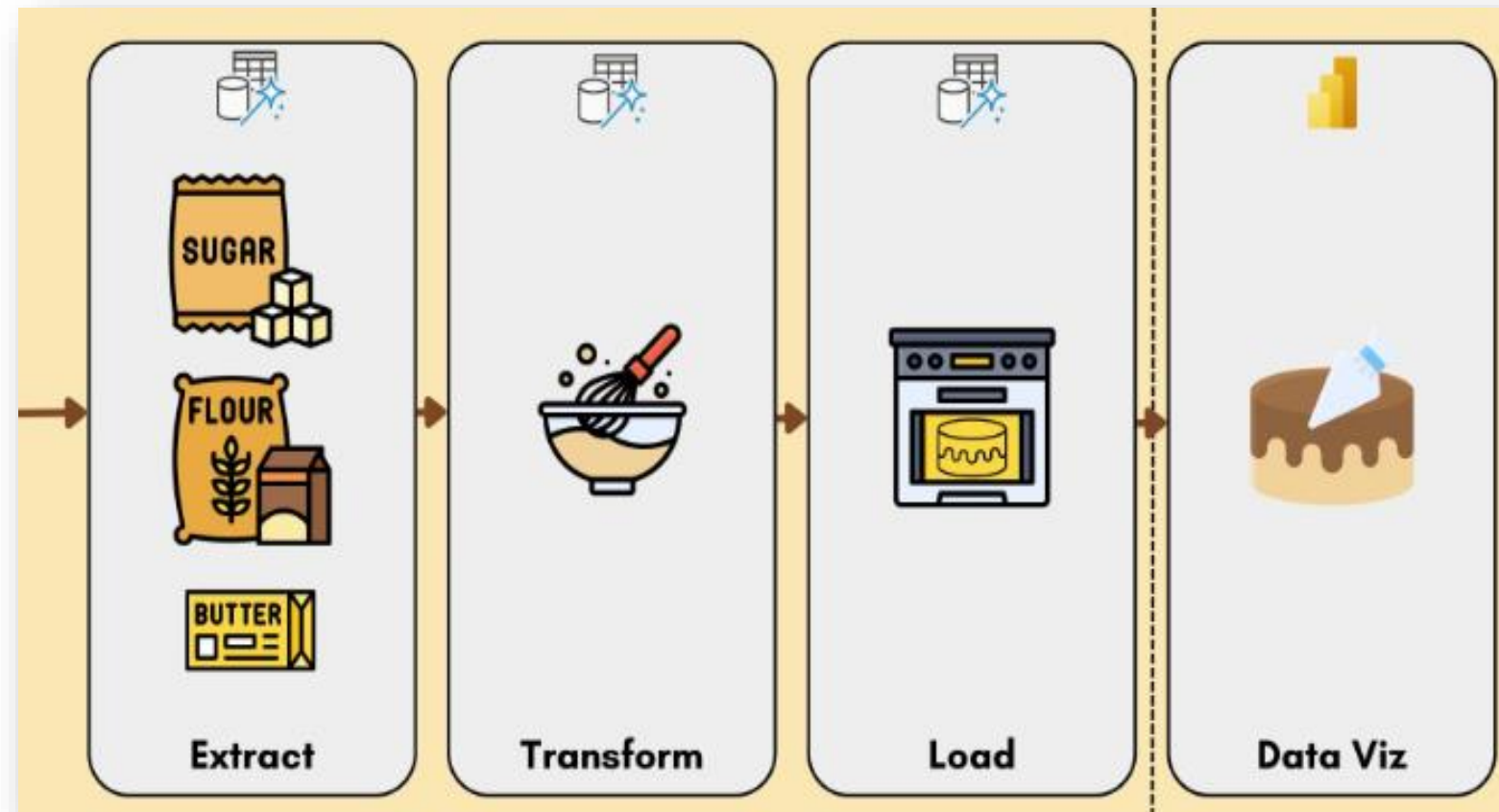
Mais sobre BI?



ETL - Extract Transform Load



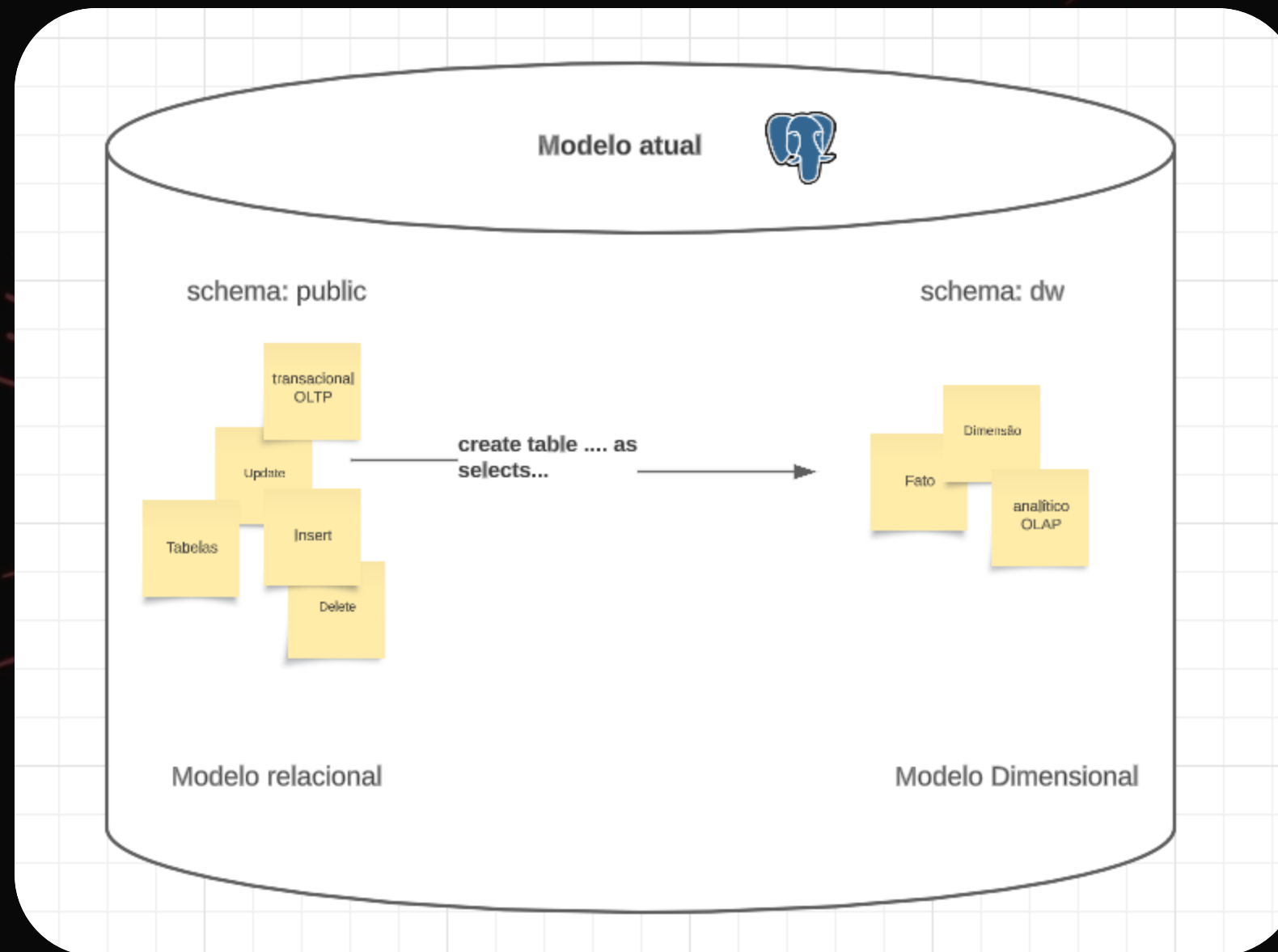
ETL - Extract Transform Load



Desenho da arquitetura atual



Lucidchart



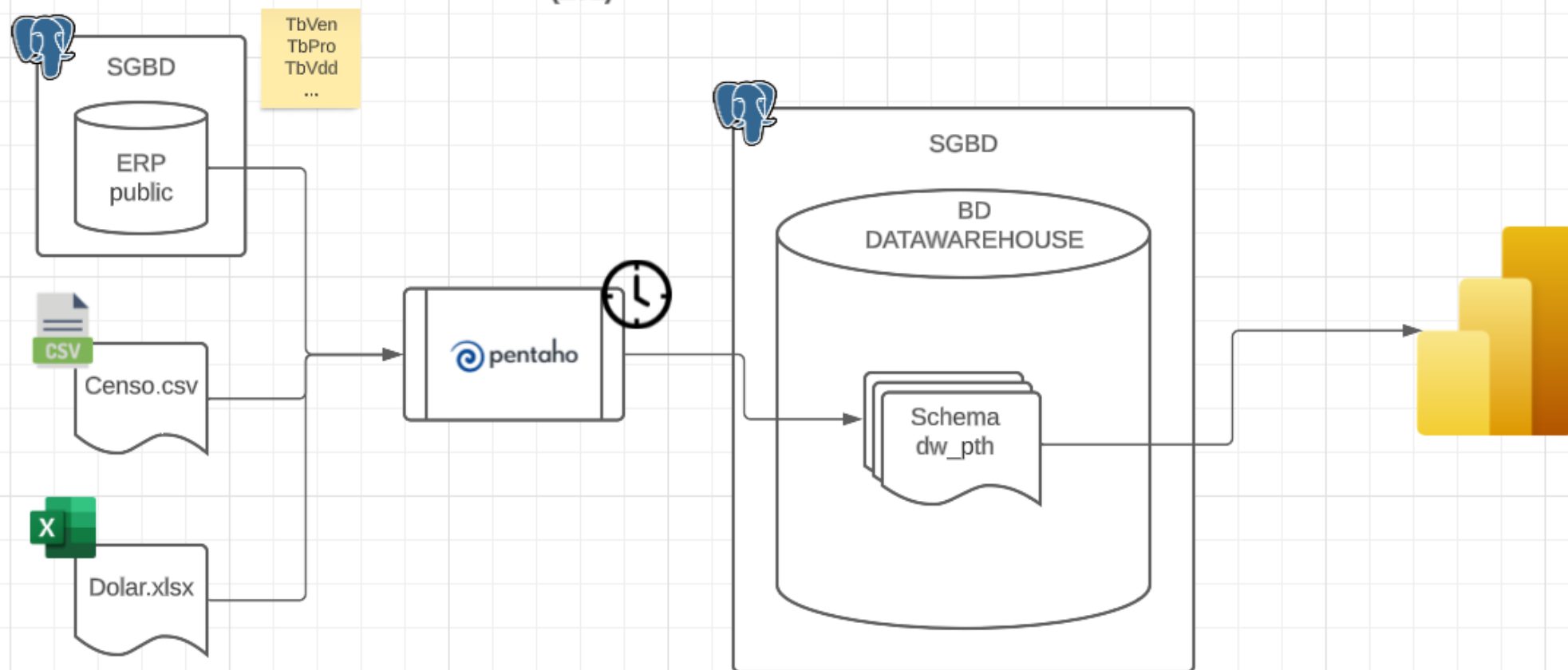
Lucidchart

Desenho da arquitetura ideal



Lucidchart

Modelo Proposto (ETL)





Pentaho ETL: Uma Explicação Completa

A extração, transformação e carregamento (ETL) de dados são cruciais para o funcionamento efetivo de empresas.

Com o Pentaho Data Integration, você pode gerenciar seus dados de maneira eficiente e confiável, economizando tempo e recursos valiosos.



Download



Funcionalidades da Ferramenta

Pentaho

Tarefas agendadas

A execução de ETL pode ser agendada para que ocorra em horários determinados, sem a intervenção do usuário.

Visualização gráfica de transformações e fluxos de dados

A ferramenta permite visualizar as alterações de dados e fluxos de trabalho, tornando mais fácil a compreensão de processos complexos.

1

Conexão fácil com várias fontes de dados

Integra-se facilmente com bancos de dados, aplicativos e serviços, além de permitir o acesso a diferentes arquivos.

2

3

Transformações completas

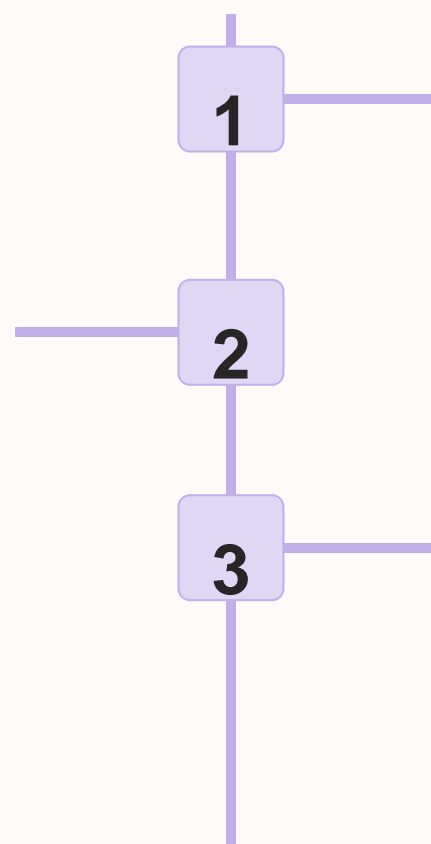
O Pentaho ETL oferece várias opções de transformação de dados, garantindo que todos os dados de fontes diferentes sejam usados de forma coesa.

4

Transformação de Dados com Pentaho Data Integration

Transformação de dados

A transformação de dados pode ser feita de várias maneiras diferentes.



Mapa de campo de entrada

O PDI mapeia campos e definições de tabelas de origem.

Alterações de campos

É possível fazer alterações de campo de acordo com o que se deseja obter e produzir.

Job

O que é Job

É uma maneira de chamar e executar transformações de dados no Pentaho ETL.

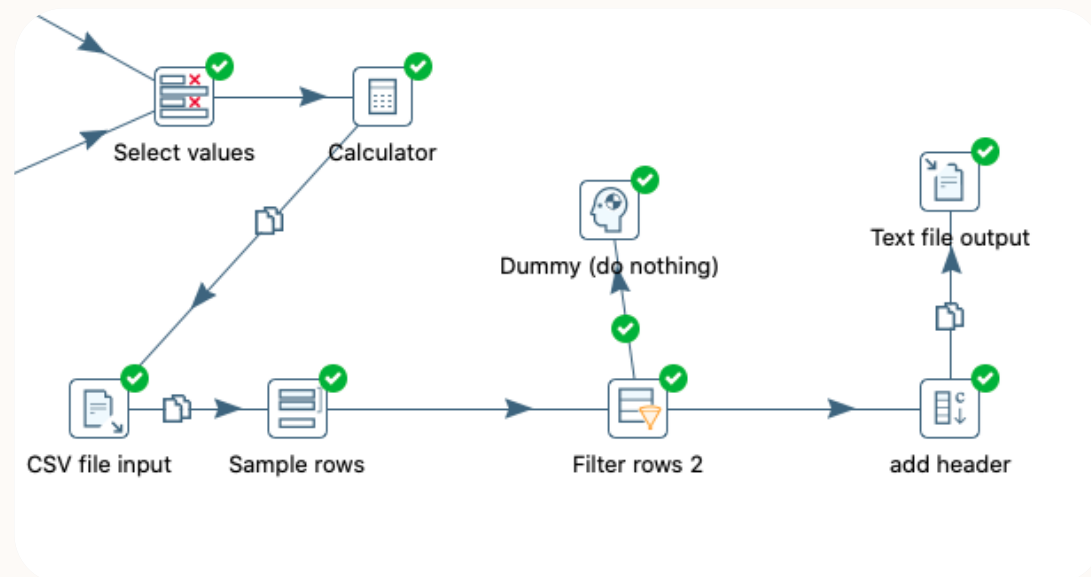
Funcionamento do Job

A execução de um Job pode ser acionada por **tempo**, **eventos** ou por outros Jobs e transformações.

Importância do Job

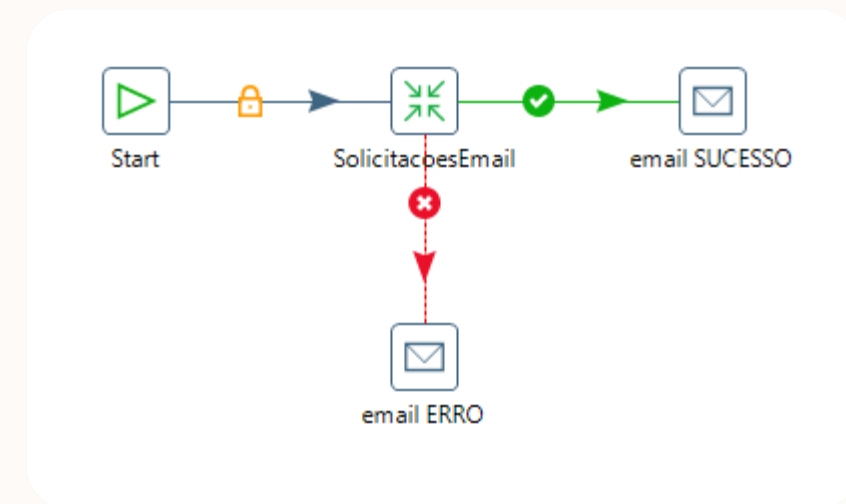
O Job é importante para garantir que as transformações de dados sejam feitas corretamente e em um tempo apropriado.

Diferença entre Transformação e Job



Transformação

Transforma um conjunto de dados por meio da aplicação de diversas regras de negócio ou lógicas.



Job

Coordena as transformações de dados e pode ser **programado para ser executado** em horários específicos.

Instalação do Pentaho



Primeiros passos no Pentaho

Transformações

Inputs



CSV file input



Table input



Microsoft Excel input

Transform / joins



Sort rows



Select values



Merge join



Value mapper



Unique rows



String operations



Add sequence

Outputs



Microsoft Excel output



Table output



Insert / update

Jobs

General



Start



Transformation



Success


Mail



Mail

Transformações

Inputs:

- **CSV file input:** Utilizado para ler arquivos CSV e transformá-los em fluxos de dados.
- **Table input:** Permite executar consultas SQL para ler dados de um banco de dados.
- **Microsoft Excel input:** ler arquivos do Excel . 

Transform / joins:

- **Sort rows:** Ordena os dados baseado em um ou mais campos.
- **Select values:** Permite selecionar, renomear e converter os tipos de dados das colunas no fluxo de dados.
- **Merge join:** Combina dois fluxos de dados baseado em uma chave de junção comum.
- **Value mapper:** Mapeia os valores de um campo para valores de saída correspondentes baseado em um dicionário predefinido.
- **Unique rows:** Remove linhas duplicadas no fluxo de dados, baseado em um ou mais campos.
- **String operations:** São utilizados para operações em campos de texto, como concatenação, substituição de caracteres, corte de strings, etc.
- **Add Sequence:** Adiciona uma sequência (um id para cada linha por exemplo)

Outputs:

- **Microsoft Excel output:** Escreve os dados processados em um arquivo do Microsoft Excel.
- **Table output:** Envia os dados processados para uma tabela em um banco de dados.
- **Insert/Update:** O step "Insert/Update" no Pentaho Data Integration é utilizado para inserir ou atualizar dados em uma tabela de banco de dados.

Jobs

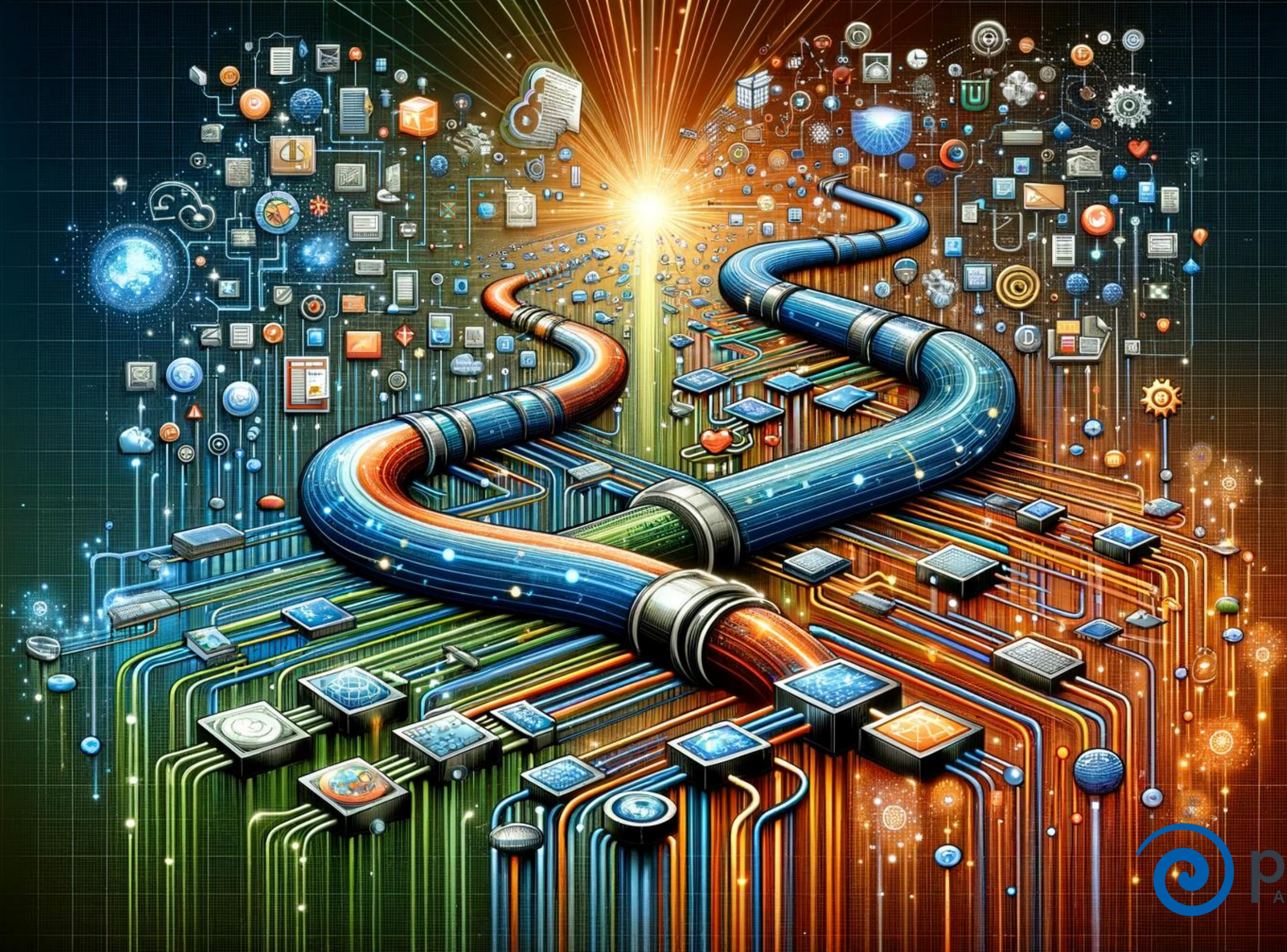
General:

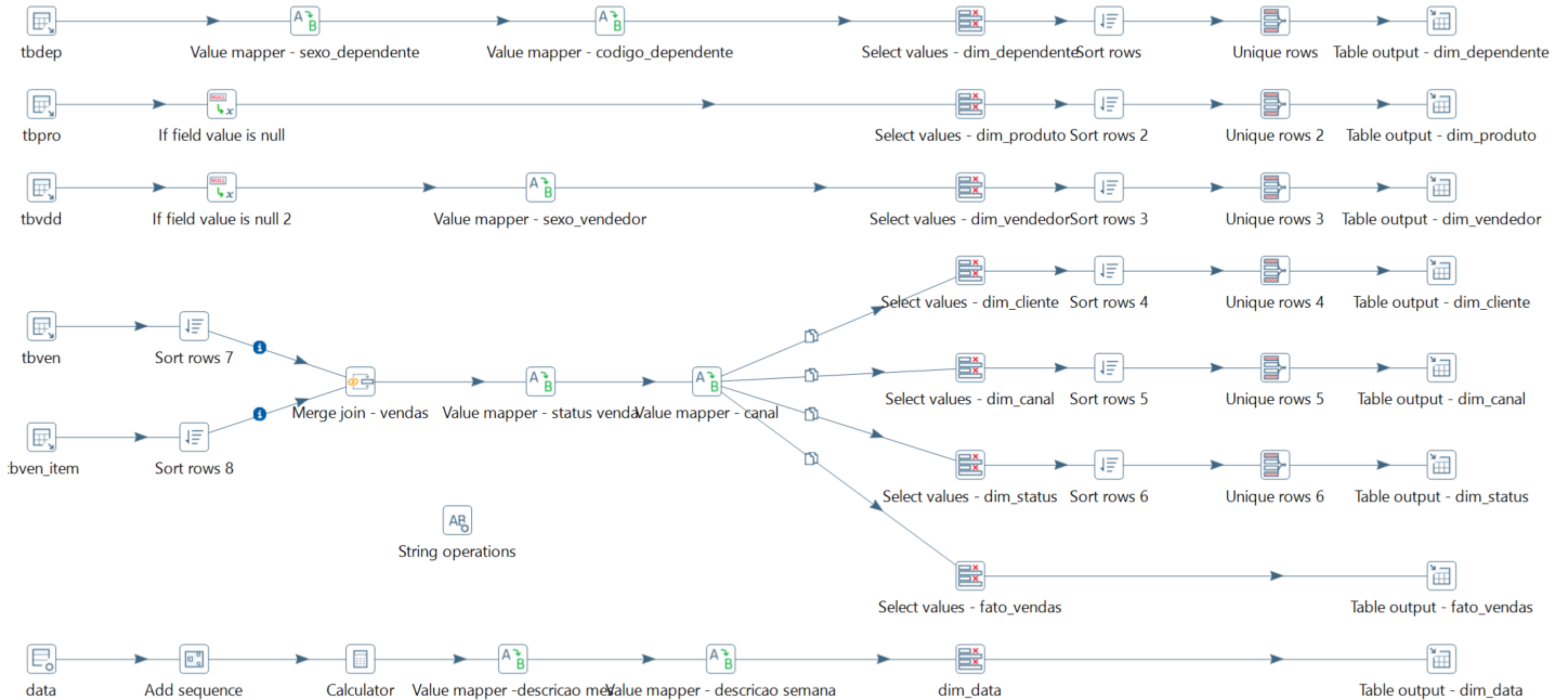
- **Start:** Indica o ponto de início de um job no PDI.
- **Transformation:** Representa a execução de uma transformação dentro do job.
- **Success:** Este ícone é normalmente usado para indicar que uma certa condição foi cumprida com sucesso ou com falha!

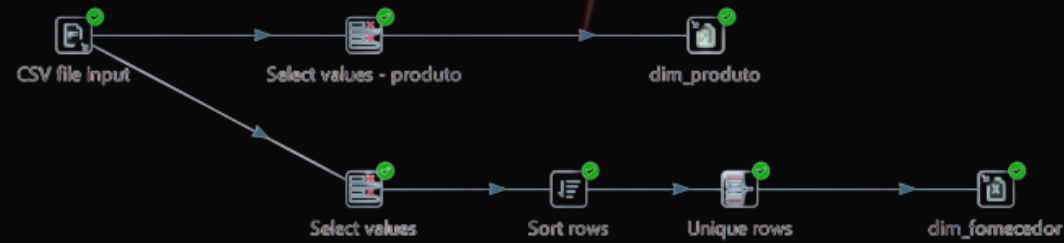
Mail:

- **Mail:** Envio de emails, que pode ser configurado para notificações ou relatórios após a conclusão de etapas do job.

Pipeline de dados







A Odisseia da Empresa Z em busca de Análises Otimizadas







Marketing



Locação Veículos



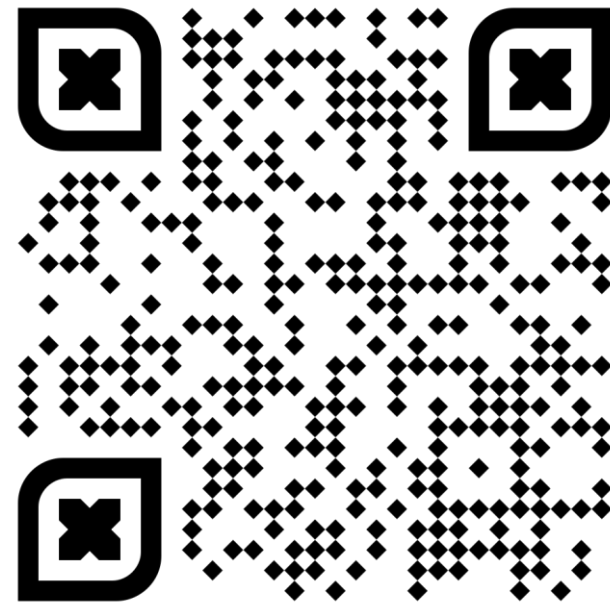
ERP

Tables (5)

>	tbdep
>	tbpro
>	tbvdd
>	tbven
>	tbven_item



Obrigado!



Obrigad@!



**Digital
College**

ENSINO DE HABILIDADES DIGITAIS

digitalcollege.com.br • @digitalcollegebr