



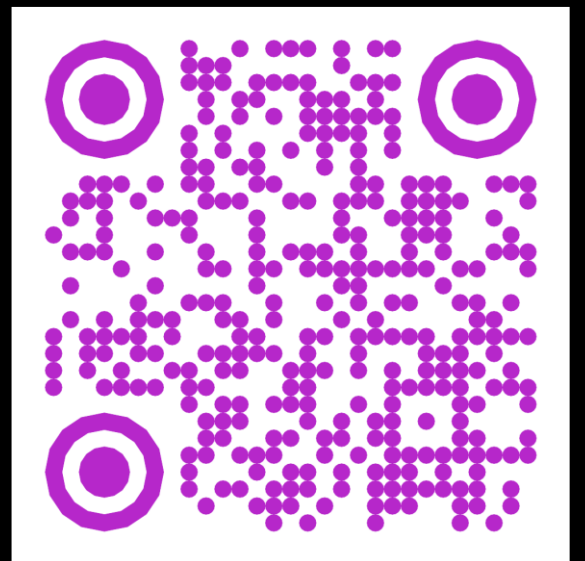
Digital College

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)

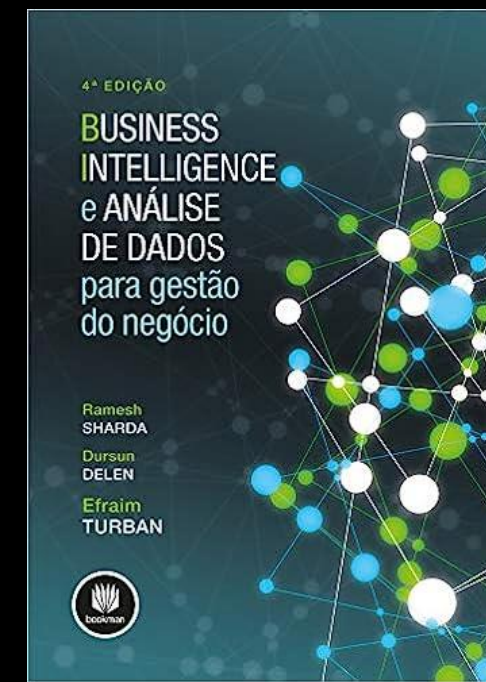
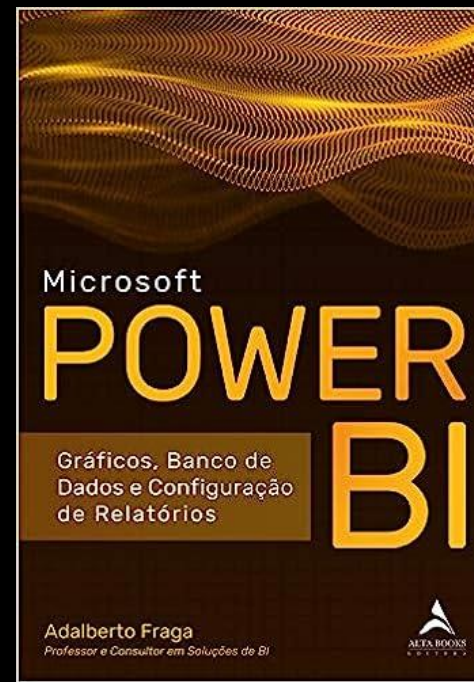
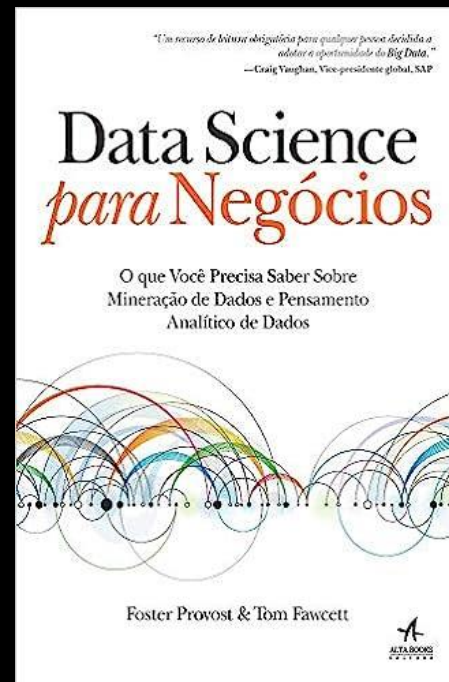
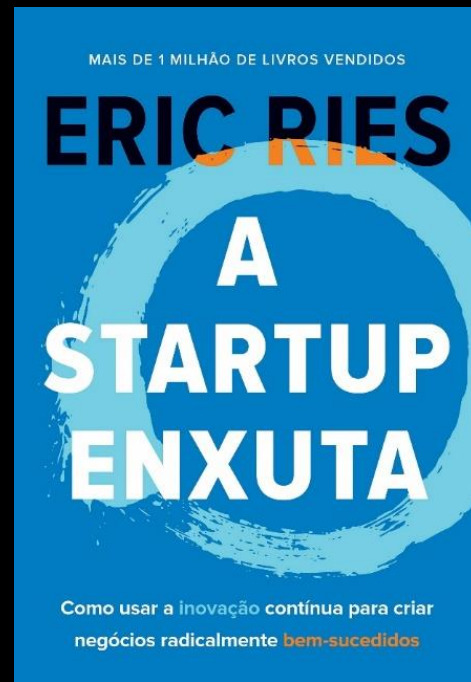
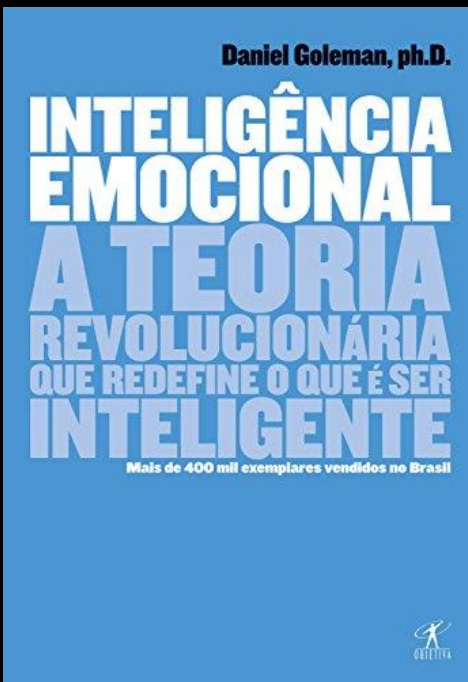
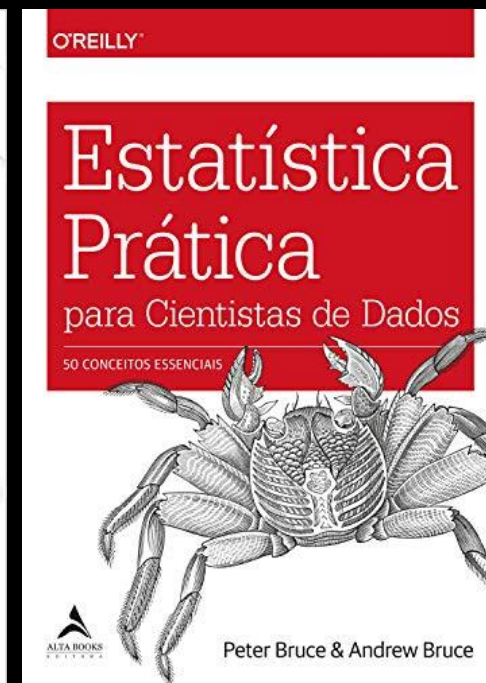
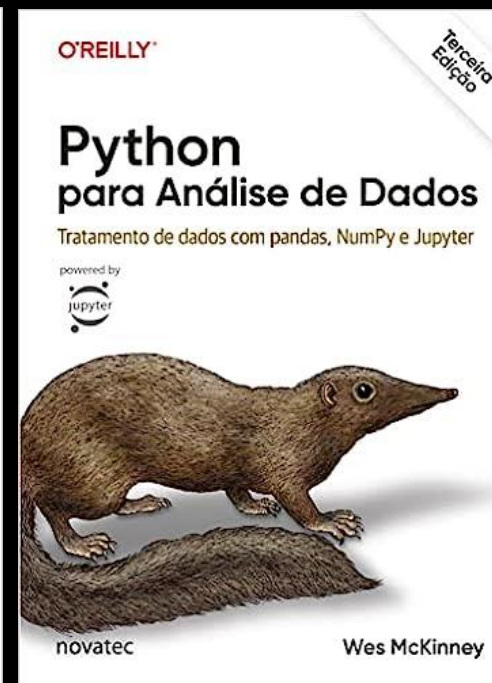
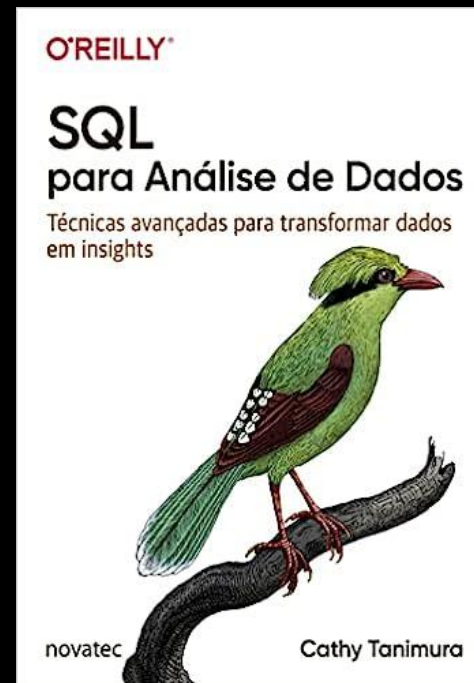
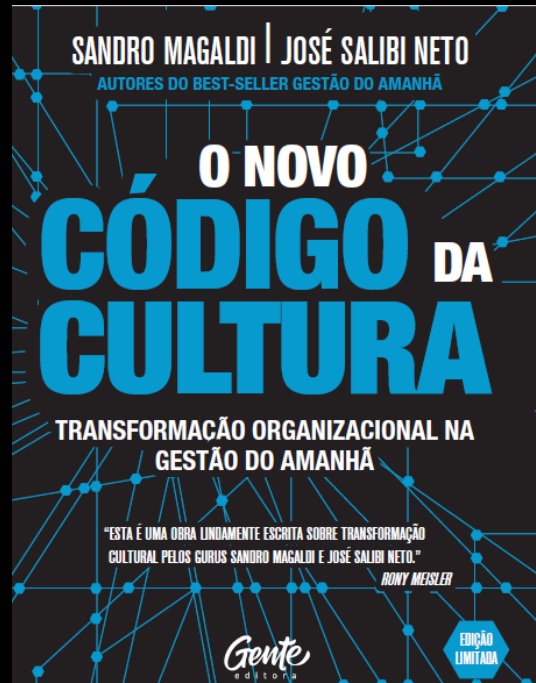
| | |
|-------------------------------|--------------------------------------|
| Unidade 04 – Power BI Desktop | 1 – Conhecendo o Power BI |
| | 2 – Modelagem, Relacionamentos e DAX |
| | 3 – Visualização de Dados |



A Odisseia da Empresa Z em busca de Análises Otimizadas



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.



Introdução ao

KI



Por que KDD?

Grandes quantidades de dados são gerados todos os dias, mas o acesso a informação valiosa é difícil devido ao volume.



Como funciona?

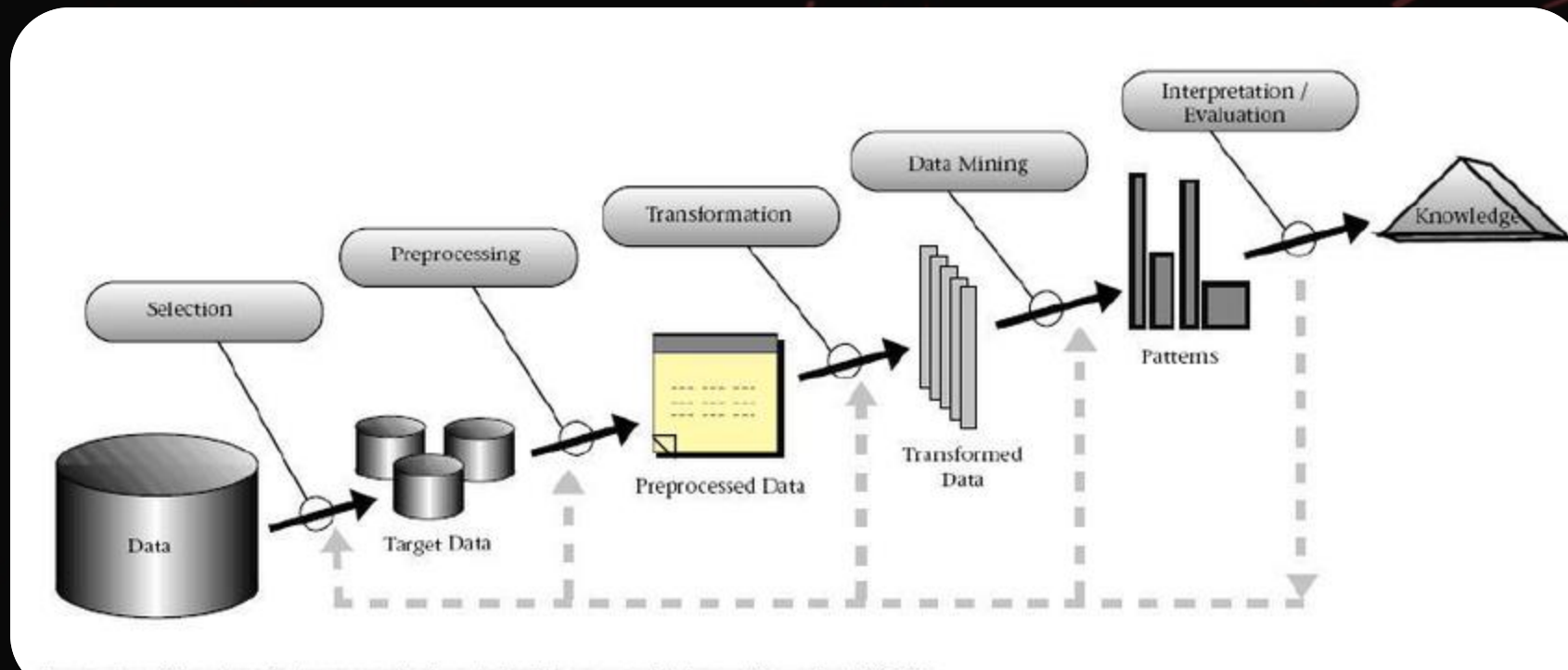
O processo de KDD envolve diversas etapas que começam com a seleção de dados e terminam com a utilização do conhecimento adquirido.



Benefícios

O KDD permite a descoberta de informações e padrões ocultos em grandes quantidades de dados, possibilitando melhores decisões e descobertas.

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

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.

1

Limpeza de dados

Remover dados inconsistentes, duplicados e irrelevantes.

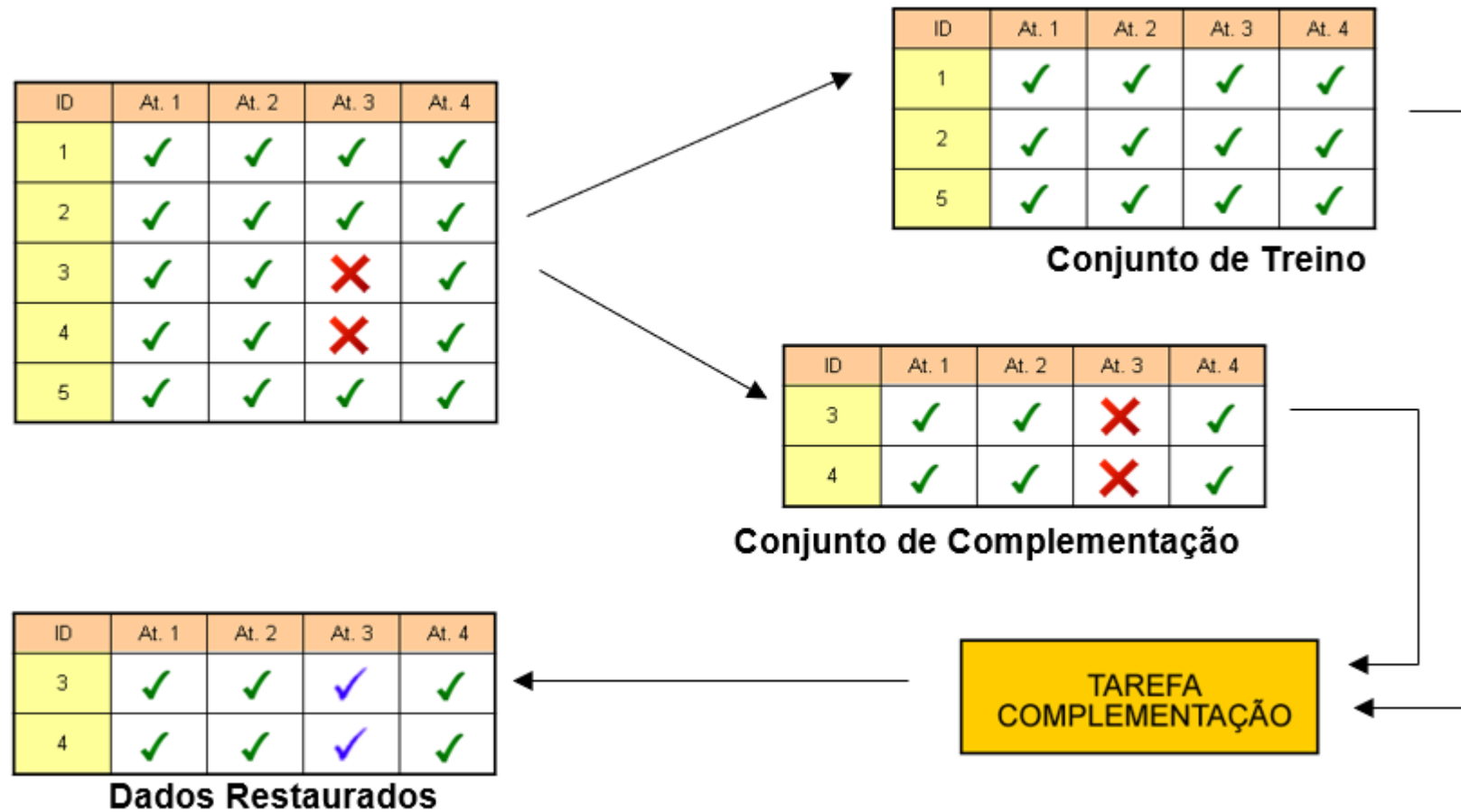
2

3

Redução de dados

Tornar os dados mais compactos sem perder informações, agrupando-os em classes ou selecionando características-chave.

Complementar valores



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

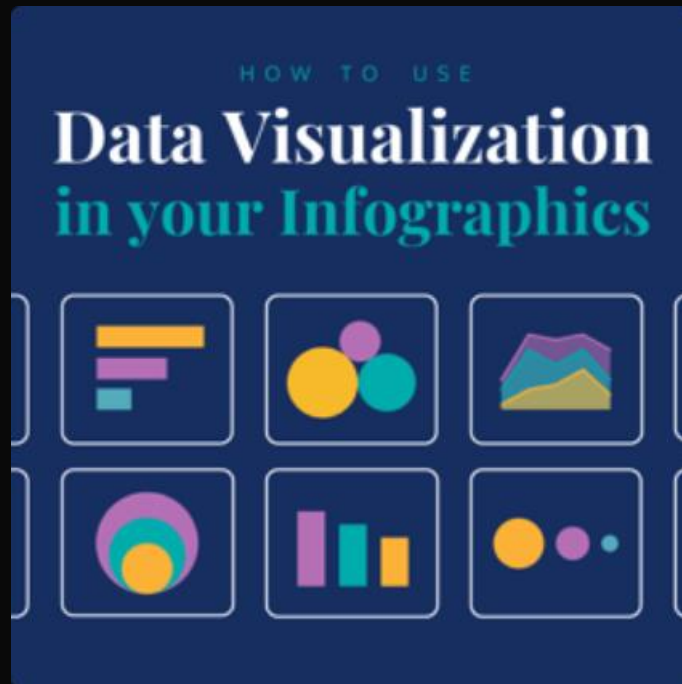
Agrupar dados em categorias ou grupos.

4

Redução de dados

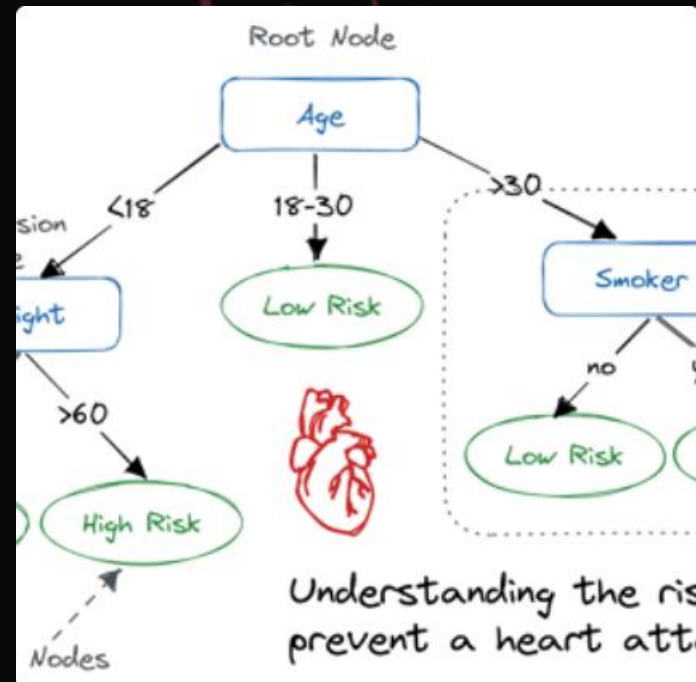
Reduzir a dimensionalidade dos dados mantendo sua importância.

Data Mining



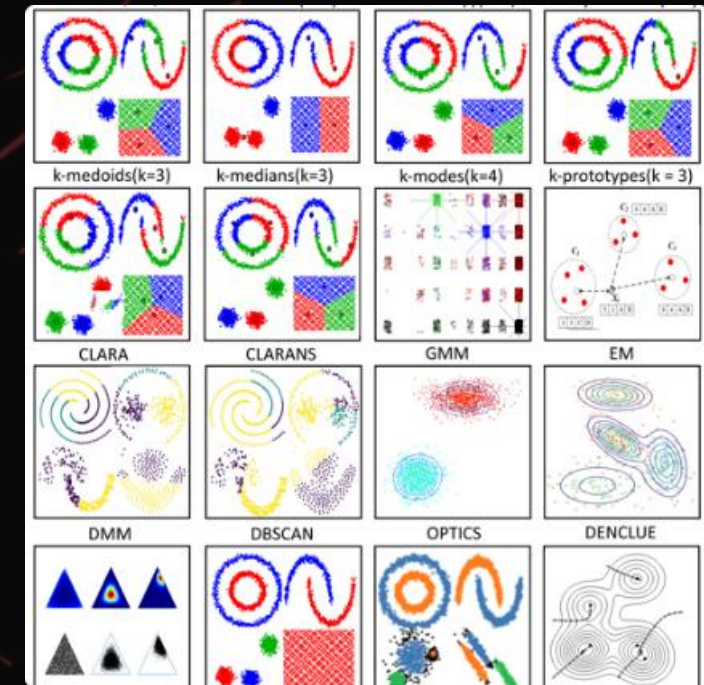
Visualização de Dados

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

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.

Utilização do

C



Máquina vs

Humano

Com a ajuda de ferramentas de KDD, máquinas podem tomar decisões mais precisas e rápidas do que humanos.



Exemplo

Analisar dados do histórico de compras dos clientes para desenvolver ofertas personalizadas e antecipar as necessidades do cliente.



Mais exemplos

Analisar dados médicos para identificar doenças em estágio inicial e desenvolver novos protocolos de tratamento

Considerações

Ferramentas

O processo de KDD ganhou impulso devido à disponibilidade de ferramentas sofisticadas, como Data Warehouses, Hadoop e SAS.

Avanços

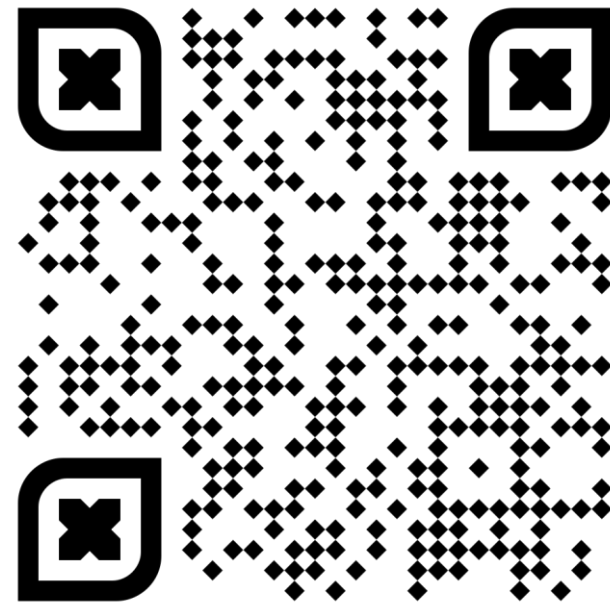
A tecnologia continua avançando, permitindo a coleta e análise de dados mais precisos em uma escala ainda maior.

Potencial

O potencial da Descoberta de Conhecimento em Bancos de Dados ainda não foi completamente explorado. Novas aplicações continuam surgindo todos os dias.

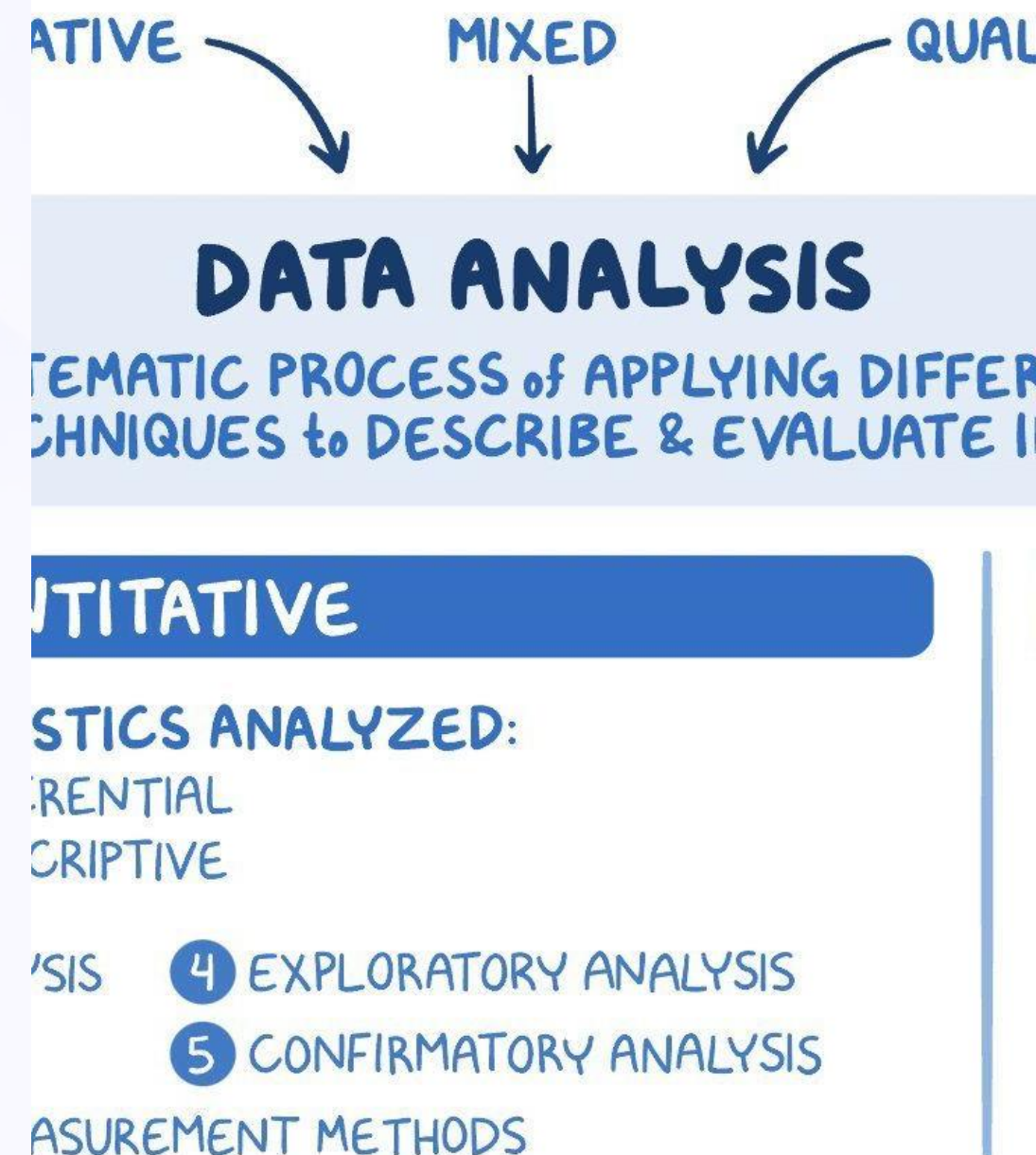
Obrigad

o!



CRISP-DM: Descobrendo Conhecimentos em Dados

O **CRISP-DM** (*Cross Industry Standard Process for Data Mining*) é uma metodologia de mineração de dados amplamente utilizada. Isso permite que as organizações obtenham insights valiosos a partir de seus dados. Aprenda como implementar esta metodologia e alcance insights importantes.



Introdução ao CRISP-

1 DM

Qual 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 valiosos 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

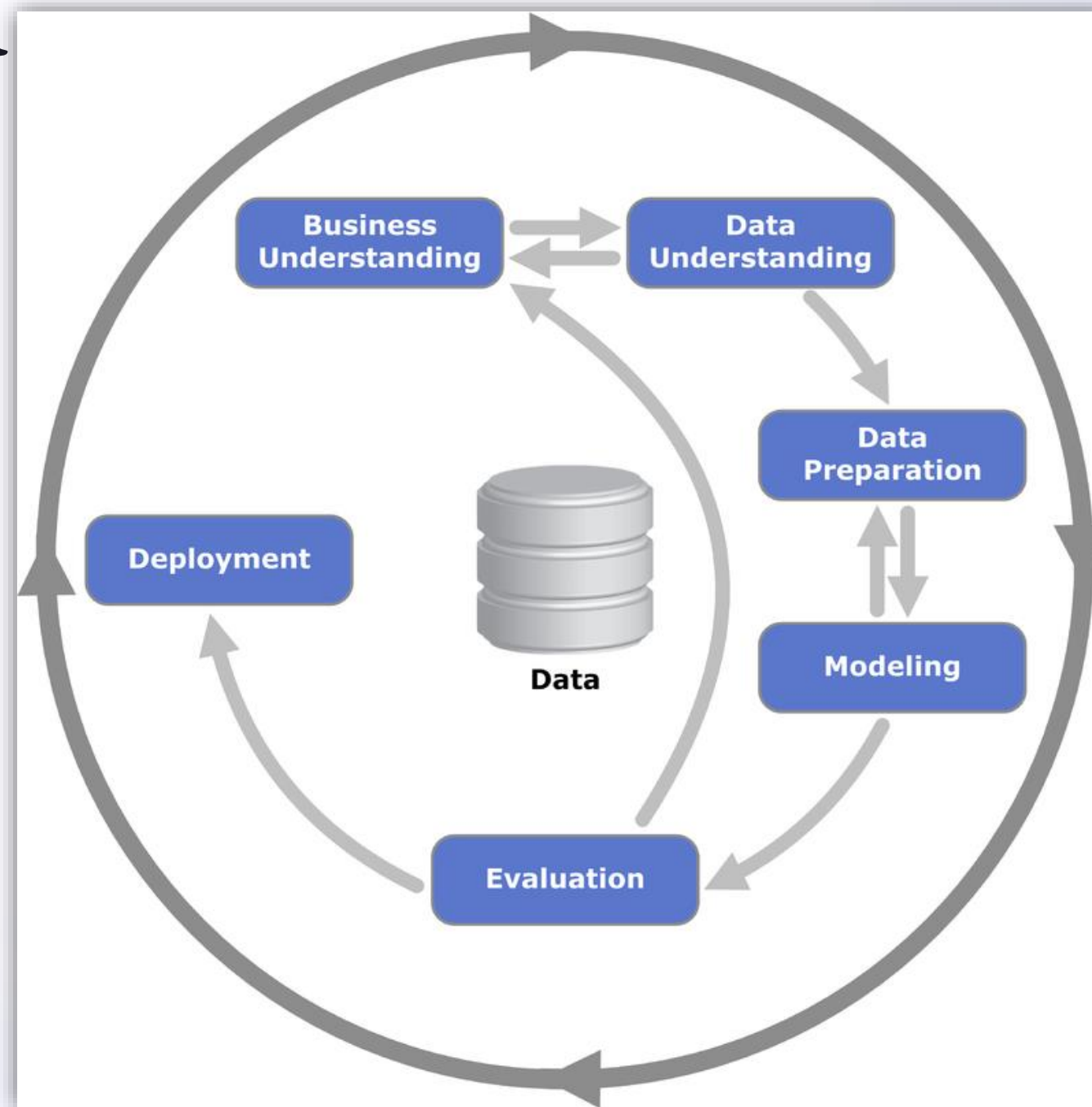
as organizações a alcançar

insights valiosos de seus

dados de maneira consistente

e repetível.

CRISP-DM



Fases do CRISP-

D



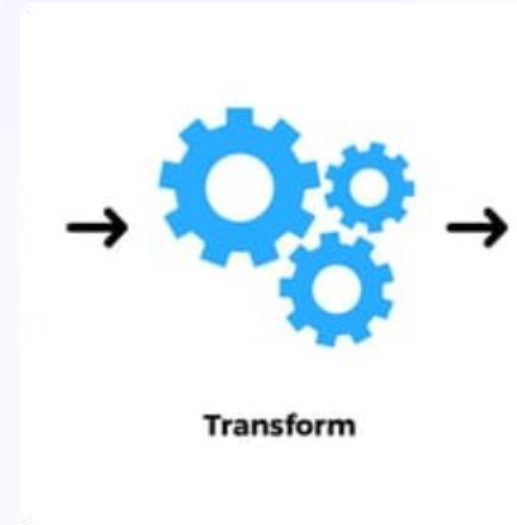
Entendimento do Negócio

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



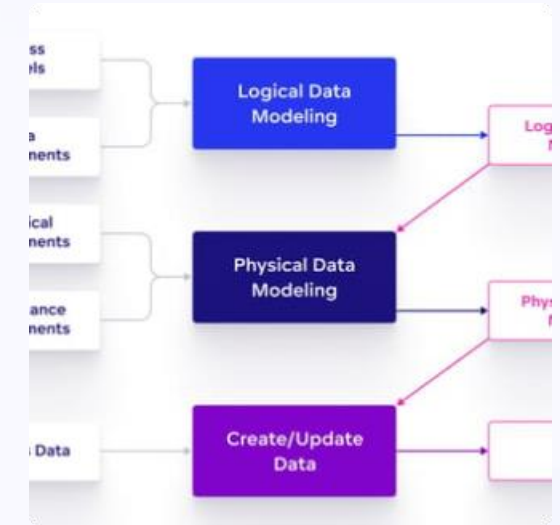
Entendimento dos Dados

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



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.

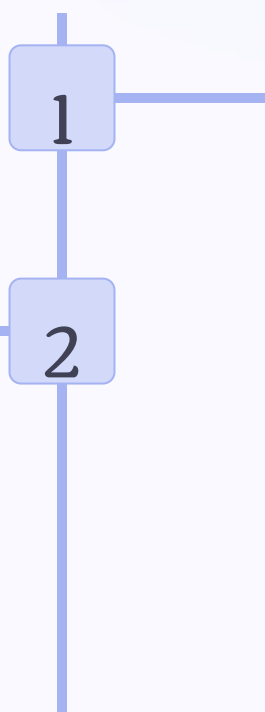


Modelagem

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

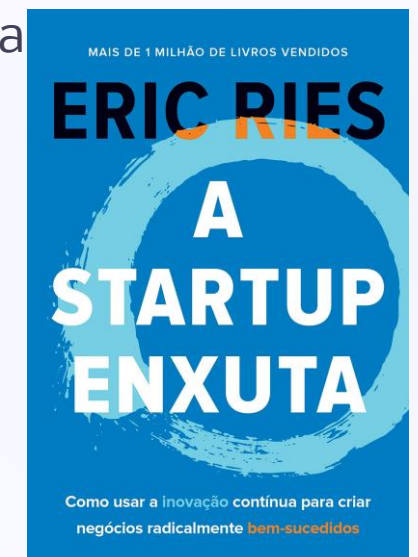
Avaliação e Implantação

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



Avaliação

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



Exemplos e casos de uso do CRISP-

Vendas e

Marketing

O CRISP-DM pode ser usado

para ajudar as empresas a

identificar clientes de alto valor e

segmentá-los em grupos com

base em sua probabilidade de

compra..

Previsão do

Tempo

O CRISP-DM pode ser aplicado

para prever o clima futuro, com

base em dados históricos e em

tempo real.

Melhorias de

Processo

O CRISP-DM pode ser usado para

identificar áreas para melhorias de

processos em empresas,

procurando por padrões e

anomalias em dados operacionais.

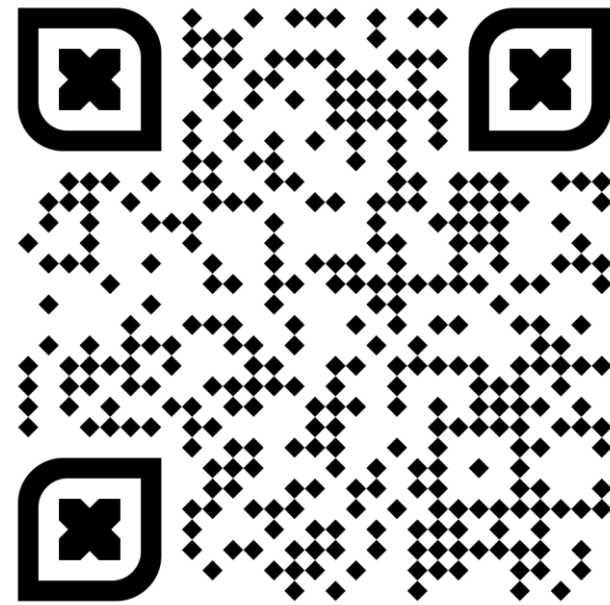
Conclusão

Ferramentas padrão, como a metodologia CRISP-DM, podem ser usadas para enfatizar a importância do gerenciamento de dados nas organizações e trazer benefícios reais e tangíveis, incluindo aumento da eficiência, compreensão do público alvo, previsão de demanda, detecção de fraudes e muito mais.



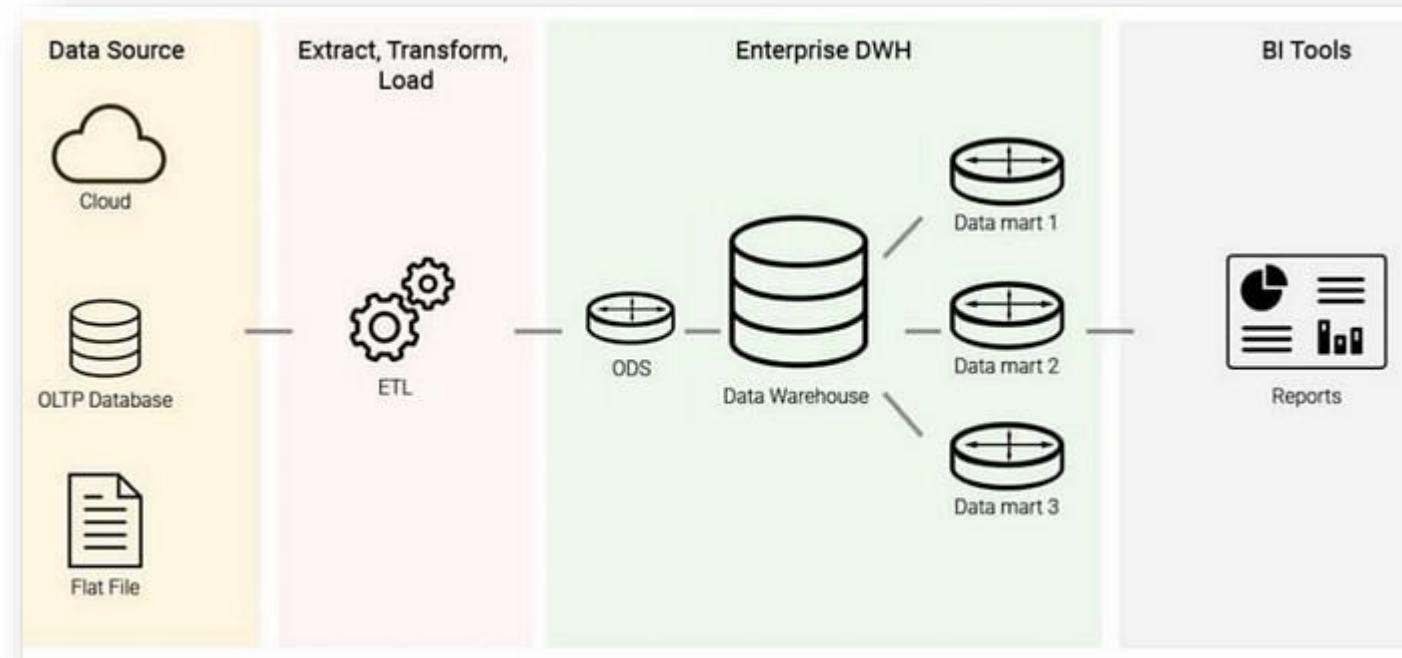
Obrigad

o!

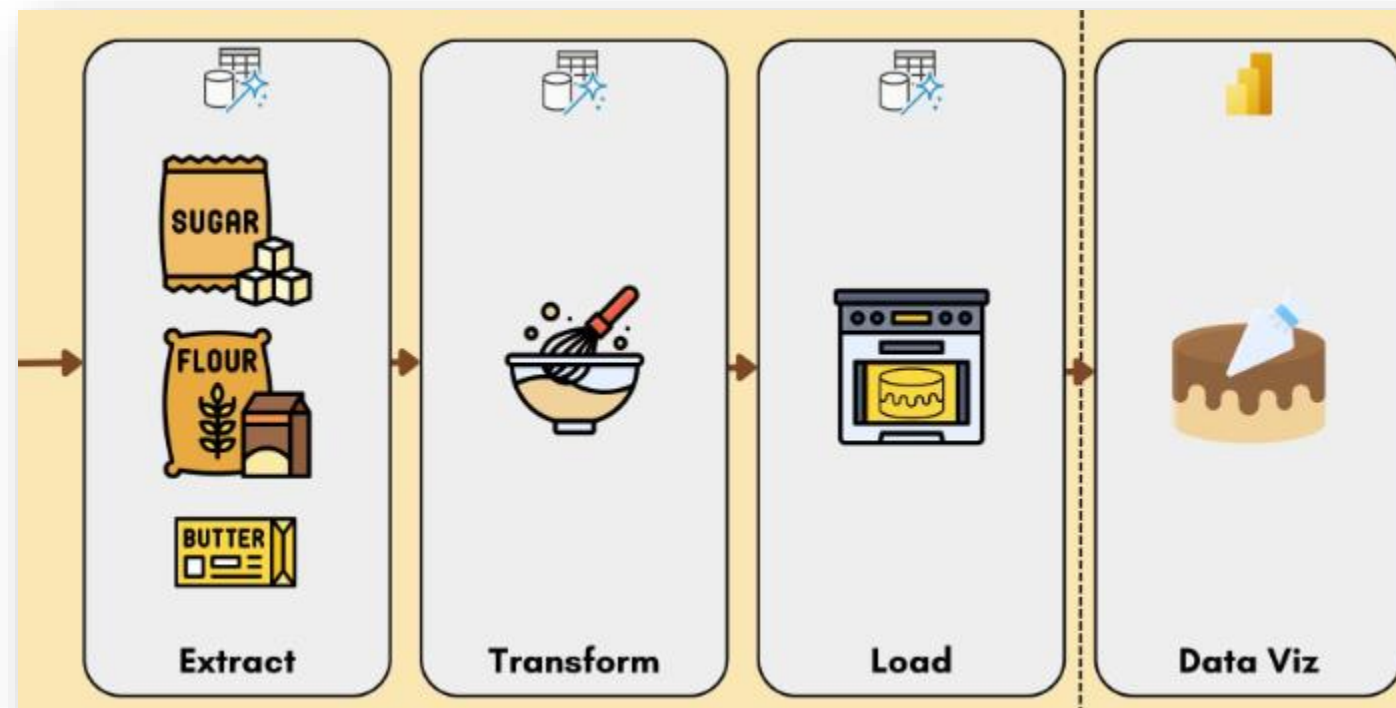




ETL - Extract Transform Load



ETL - Extract Transform Load



Desenho da arquitetura atual



Lucidchart

Desenho da arquitetura ideal



Lucidchart



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



Introdução à Ferramenta de ETL

O que é o Pentaho

ETL

Uma ferramenta de ETL de última geração usada para integrar e transformar dados de diferentes fontes em soluções empresariais.

Por que usar o Pentaho

ETL?

Garante a qualidade dos dados, aumenta a eficiência do negócio e, consequentemente, o lucro da empresa.

Vantagem

competitiva

O Pentaho ETL ajuda a manter a empresa à frente da competição, fornecendo dados precisos e em tempo real.

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

Arquitetura do Pentaho Data Integration (PDI)

Arquitetura orientada a plugins

O Pentaho ETL é baseado no conceito de plugins, tornando a arquitetura escalável e adaptável às necessidades do usuário.

Trabalho em conjunto com outras tecnologias

O PDI trabalha bem ao lado de outras tecnologias, oferecendo ainda mais possibilidades para a empresa.

Segurança

O Pentaho ETL possui várias medidas de segurança que garantem a privacidade e integridade dos dados.

Armazenamento de metadados

Armazena todos os metadados relacionados às fontes de dados em um servidor de metadados, facilitando o gerenciamento e localização de informações importantes.

Conexão com Fontes de Dados

1 Fontes de dados compatíveis

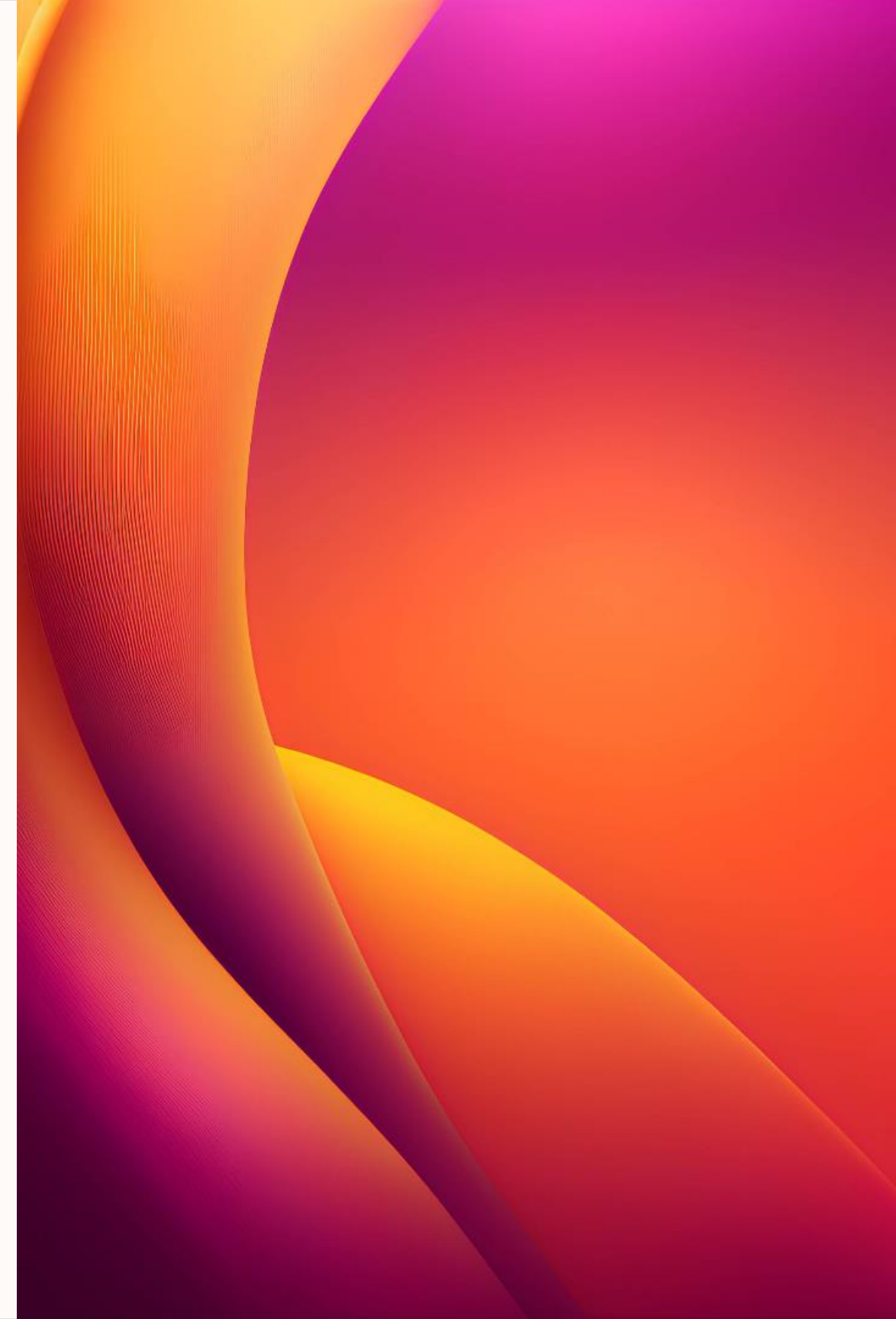
O Pentaho ETL pode se conectar a uma variedade de fontes de dados, como bancos de dados, arquivos, serviços da web e aplicativos empresariais.

3 Maior eficiência do negócio

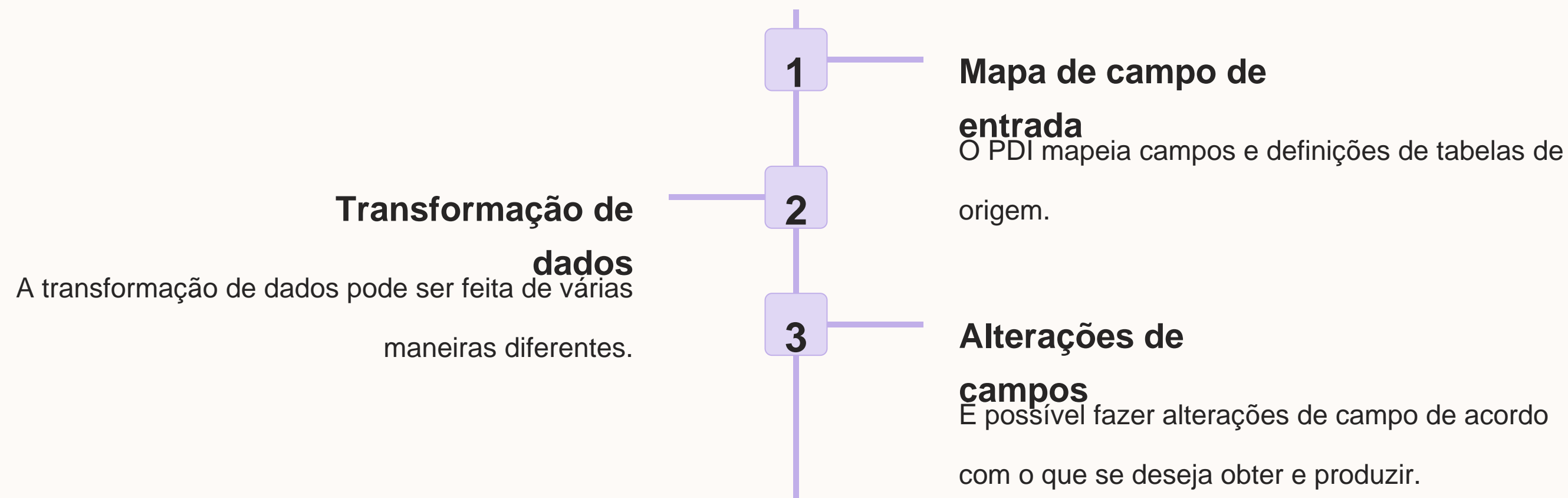
Elimina a necessidade de criar um código personalizado de conexão de dados, o que aumenta a eficiência do negócio.

2 Integração de diferentes tecnologias

Possibilita a integração de bancos de dados e tecnologias diferentes.



Transformação de Dados com Pentaho Data Integration



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.

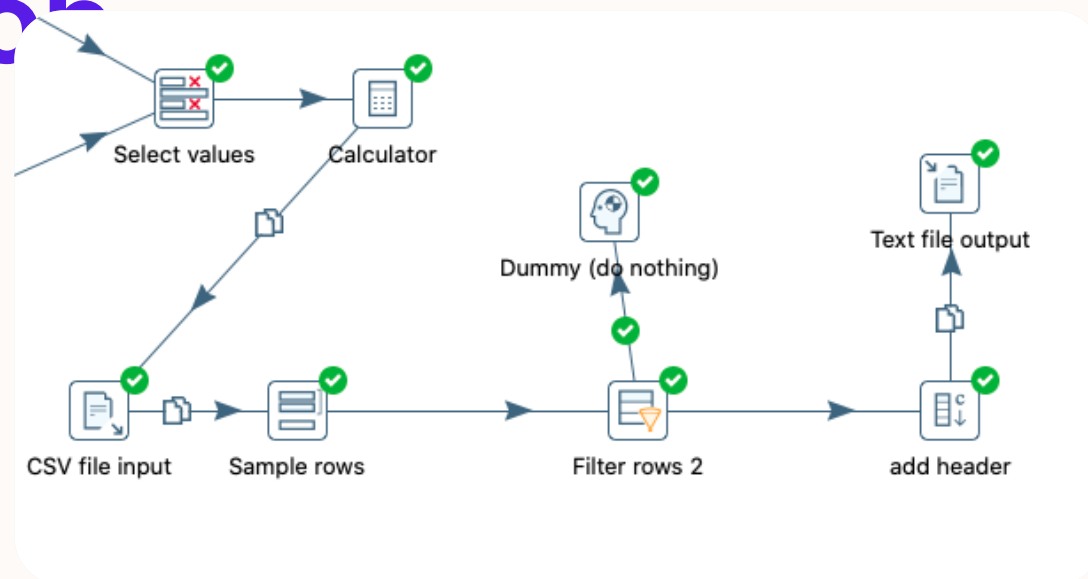
Importância do

Job

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

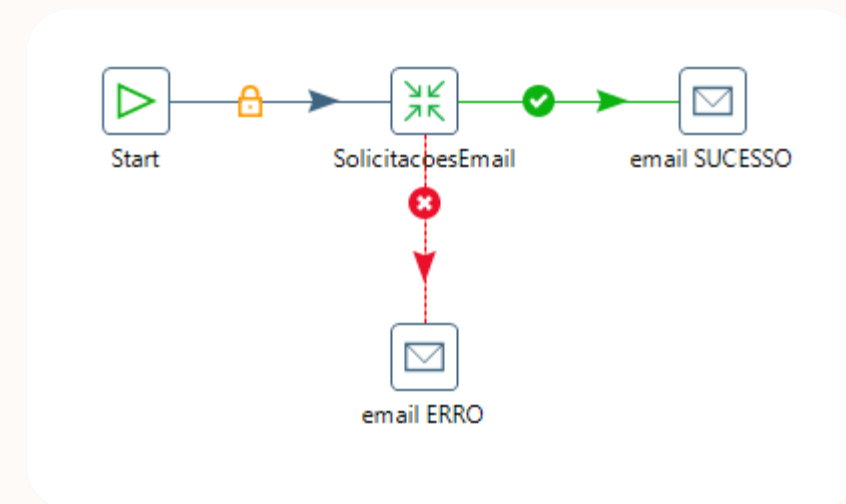
Diferença entre Transformação e

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.

Exportando dados de um arquivo Excel para um banco de dados PostgreSQL

Passos para exportação

Criar uma conexão com o banco de dados, selecionar o arquivo Excel de origem e, em seguida, mapear as colunas de origem e destino.

Dados de origem

Todos os dados de origem precisam estar em formato tabular para serem colocados em um banco de dados.

Dados de destino

Após a conexão com o banco de dados estabelecida, os dados podem ser exportados em tempo real ou programados para serem exportados periodicamente.

Principais Benefícios da Utilização do Pentaho Data Integration

Garantia de Dados Precisos

- Transformação confiável e segura dos dados
- Padronização dos dados para garantir a qualidade
- Automação de etapas do processo de ETL, evitando possíveis erros humanos.

Fluxos de Trabalho Eficientes

- Aceleração do processo de ETL
- Uso de técnicas para otimização de desempenho
- Utilização de recursos e ferramentas para maximização da produtividade

Redução dos Custos

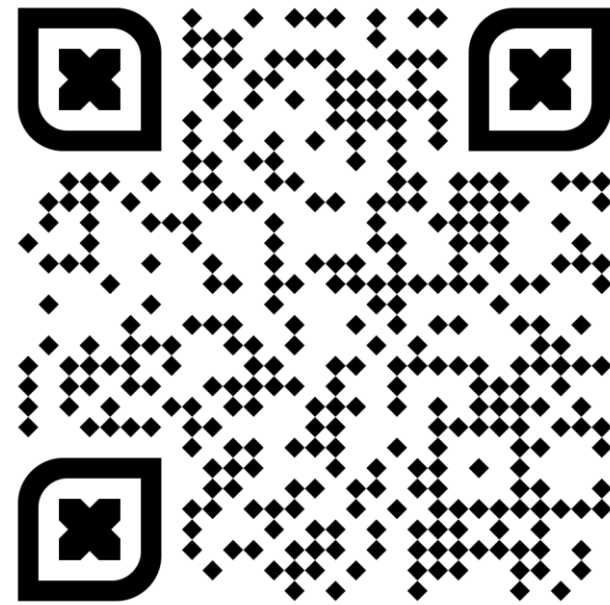
- Eliminação de customização de código carro-chefe para a conexão de dados
- Maior produtividade
- Redução significativa do tempo de desenvolvimento do ETL

Instalação do Pentaho

Primeiros passos no Pentaho

Obrigad

o!



Obrigad@!



Digital
College

ENSINO DE HABILIDADES DIGITAIS

digitalcollege.com.br • @digitalcollegebr