



APOIO



- **LS Treinamentos**
- **Consulta BD**
- **Big Data Navigators**
- **Faculdade UNIT**
- **RDSE - Recife Data Science & Engineering**



**CENTRO
UNIVERSITÁRIO
TIRADENTES**



- + 12 anos na área de TI/Dados.
- Senior Azure Data Engineer
- MCT Microsoft
- 6 Certificações Microsoft
- Instrutora da LS Treinamentos
- Admin e escritora do blog Big Data Navigators
- Mãe de Ryan (prezadinho)
- Mãe de pet: Dogs (Charlotte e Lion) e calopsitas (Careca e Melado)

Quem sou eu?



Rayza Wilma



ETL com Pentaho e Databricks





OBJETIVO



- Entender o que é Engenharia de Dados
- Como funciona um processo de ETL?
- 2 ferramentas muito utilizadas no mercado
- Comparação das ferramentas
- Case prático



O QUE É ENGENHARIA DE DADOS?



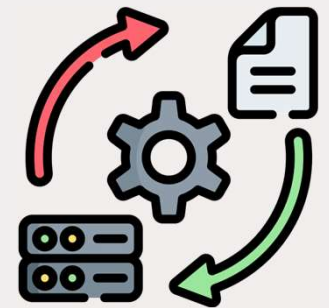
- A **Engenharia de Dados** é uma área responsável por transformar os **dados brutos** de uma empresa, em **informações** e **conhecimentos** que sejam **relevantes** para uma organização.
- Seu foco é o processo de **coleta** dos dados brutos, **tratamento**, **armazenamento** e **gerenciamento** destes dados.

O QUE É ETL/ELT?



É um termo utilizado para definir um conjunto de processos que envolvem a **extração** de dados de diversas fontes, seguida de **transformações** e **carregamento** em um destino, como um **Data Warehouse** ou **Data Lake**.

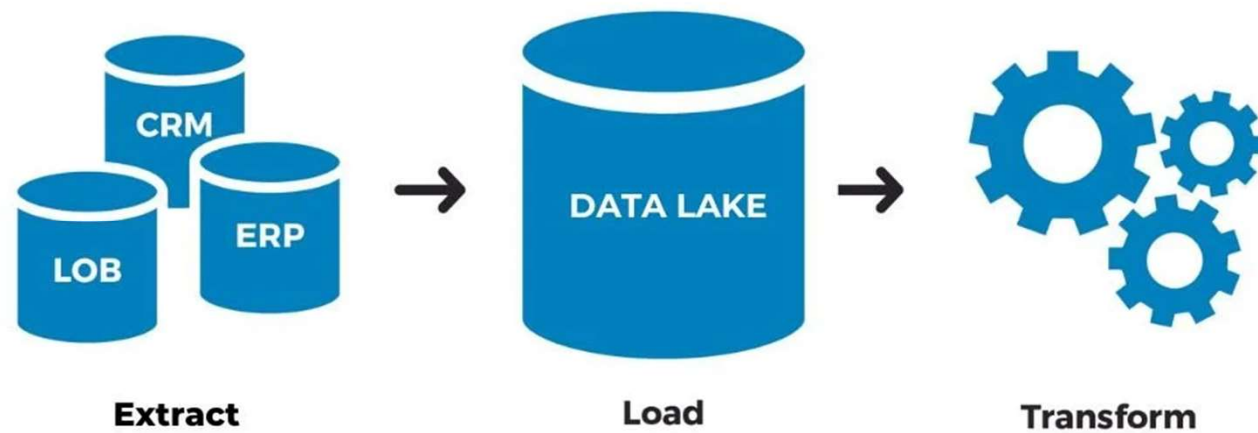
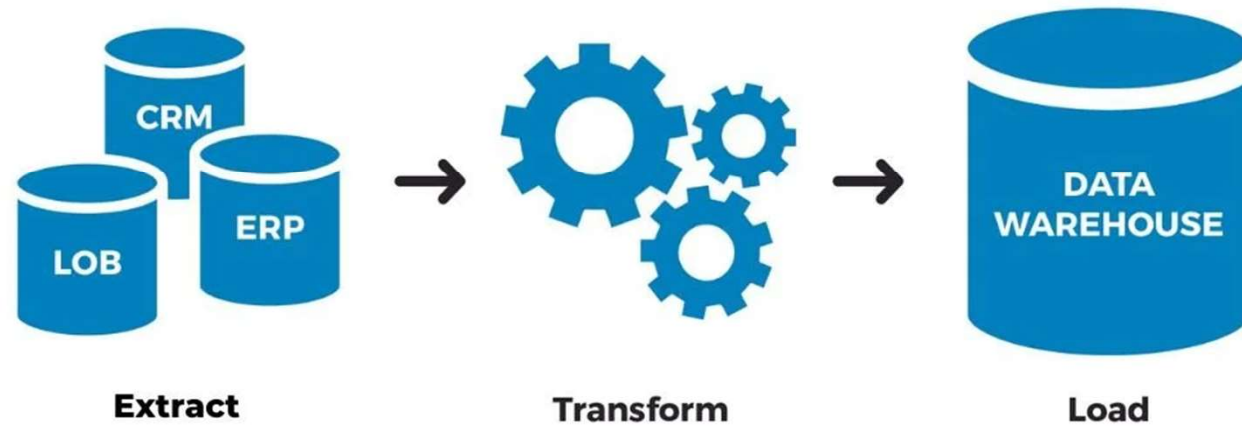
- **ETL** - Extrair, Transformar e Carregar
- **ELT** - Extrair, Carregar e Transformar



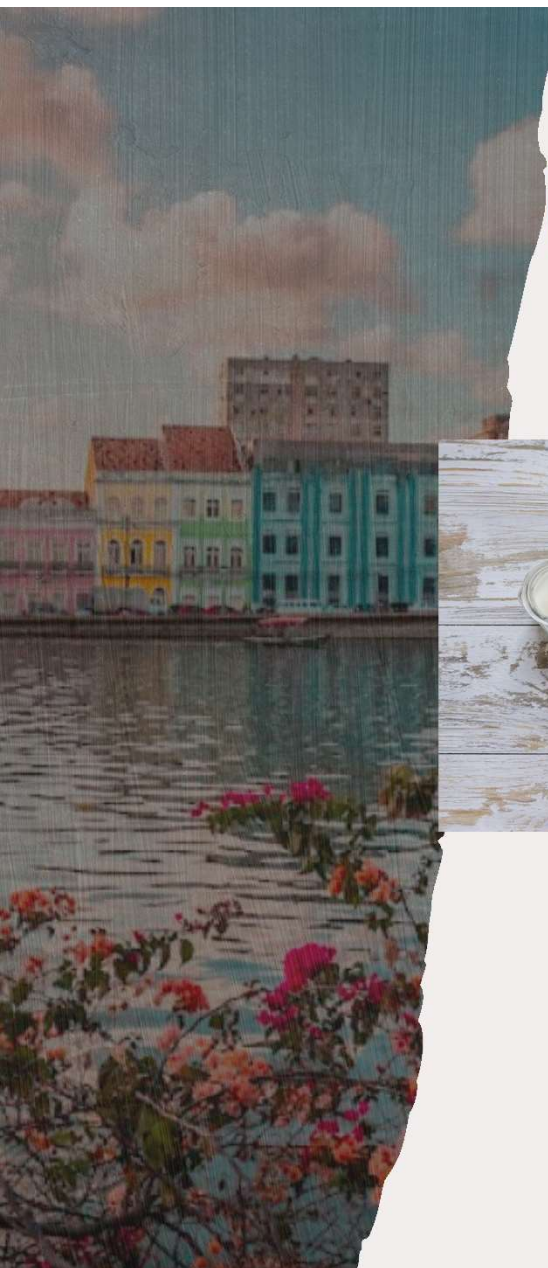
ETL

VS.

ELT



ANALOGIA





Processo de ETL



- **Fluxo de dados**
 - Transformações/Mapas/Pipeline/Notebook
- **Componentes**
 - Steps/Activity/CMD
- **Workflows/Orquestrador**
 - Trigger/ Jobs



PENTAHO



- É uma solução completa de BI.
- Trata tudo como processos e serviços integrados.
- Pentaho Data Integration, BI Server, Dashboard Designer, Report Designer, Metadata Editor, etc...





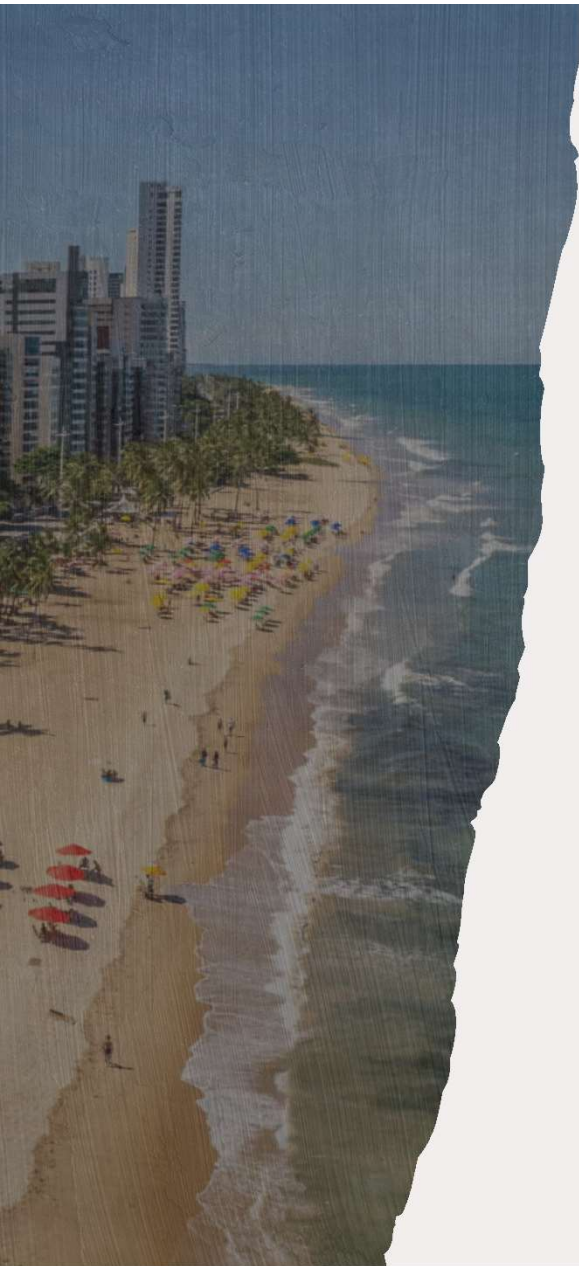
DATABRICKS



- . Plataforma de análise de dados unificada e baseada na nuvem
- . Foi projetado pelos fundadores do **Apache Spark**
- . Integração com Azure, AWS, GCP
- . Análise de dados, ingestão, preparação, visualização, construção de machine learning
- . SQL, Python, Scala, R
- . Delta Lake
- . Unity Catalog



databricks



Hands-On





PARA REFLETIR...



*“Tudo que você precisa fazer é mover as pessoas **só um pouquinho** para **mudanças** acontecerem. Não precisa ser **algo enorme**.”*

*- **Viola Davis** -*





www.bigdatanavigators.com



www.linkedin.com/in/rayzawilma/



[@bigdatanavigators](https://www.instagram.com/bigdatanavigators)



[@bigdatanavigators](https://www.youtube.com/@bigdatanavigators)