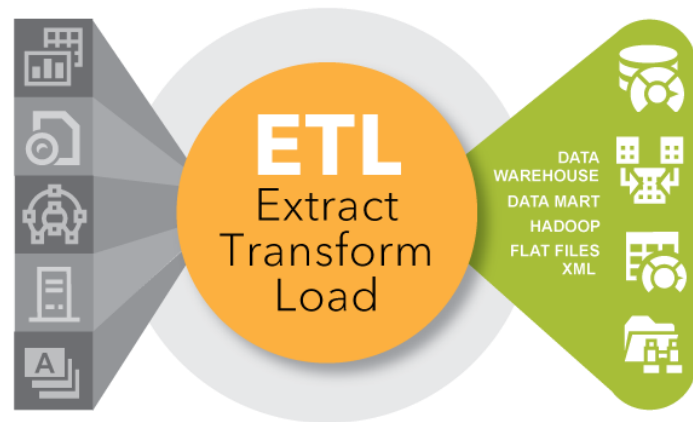


# Etapas para um Processo ETL

**Prof. Dr. Diego Bruno**

Education Tech Lead na DIO

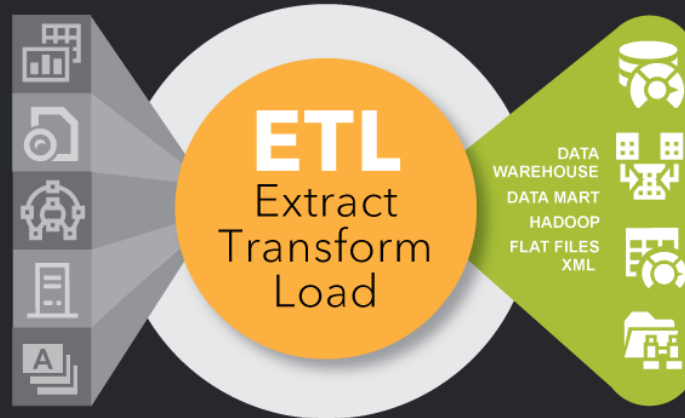
Doutor em Robótica e *Machine Learning* pelo ICMC-USP



# Etapas de ETLs

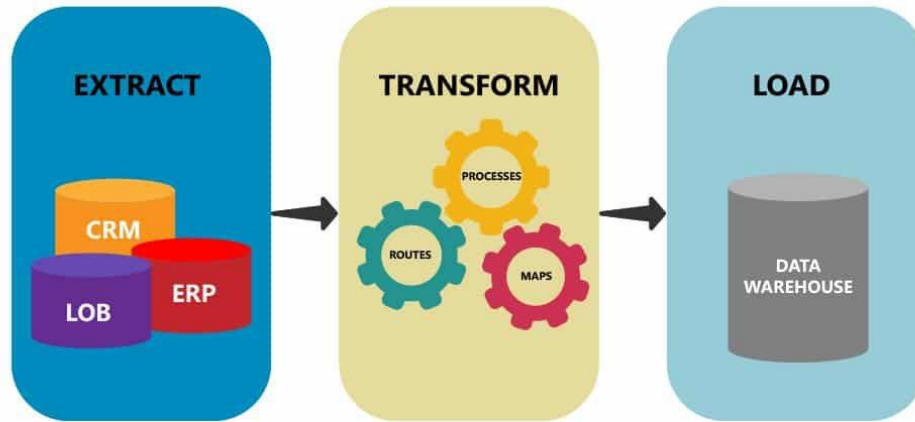
Prof. Dr. Diego Bruno

*Machine Learning*



# Temos 3 Etapas:

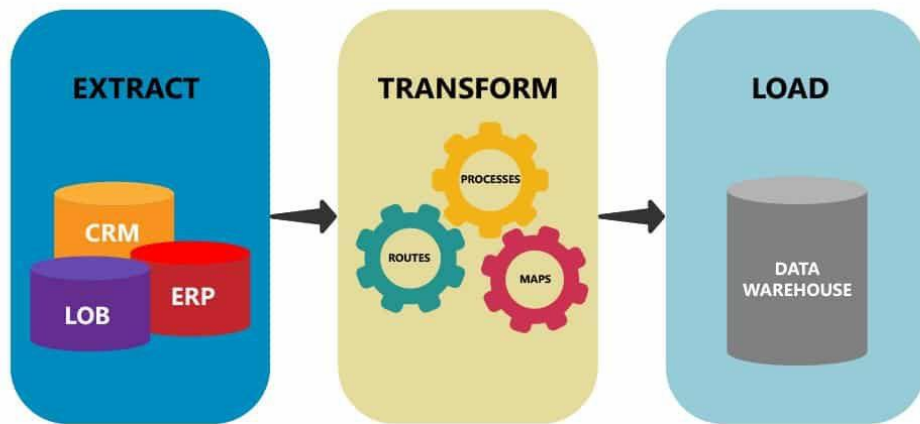
- 1) *Extract*
- 2) *Transform*
- 3) *Load*.



ETL - Extract, Transform, Load

# 1) Extract

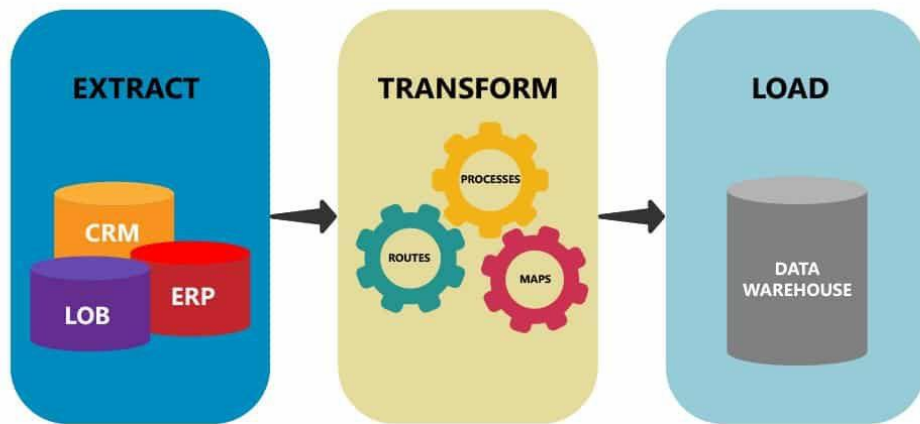
O **processo de Extração de dados** consiste em se comunicar com outros sistemas ou bancos de dados para capturar os dados que serão inseridos no destino, seja uma *Staging Area* ou outro sistema.



ETL - Extract, Transform, Load

## 2) Transform

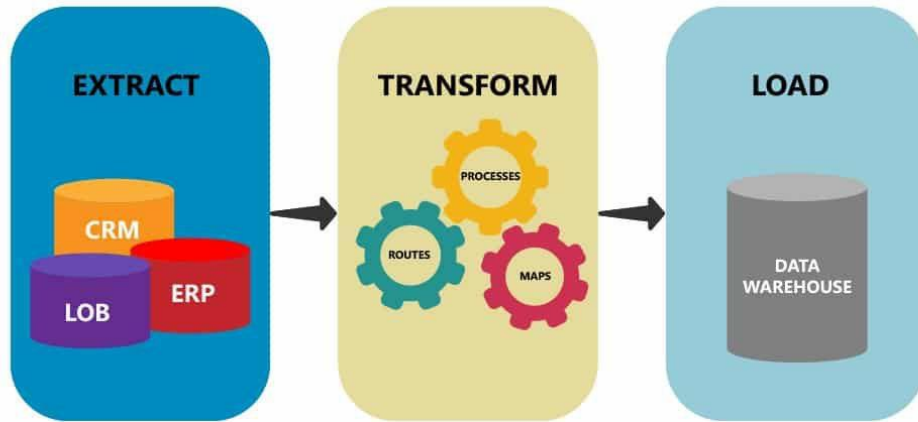
O processo de **Transformação de Dados** é composto por várias etapas : padronização, limpeza, qualidade. Dados vindos de sistemas diferentes tem padrões diferentes seja de nomenclatura ou mesmo de tipos de dados ( VARCHAR2 Oracle ou VARCHAR Sql Server.



ETL - Extract, Transform, Load

### 3) Load

O **processo de Load** é a etapa final onde os dados são lidos das áreas de **staging** e preparação de dados, carregados no **Data Warehouse ou Data Mart Final**.

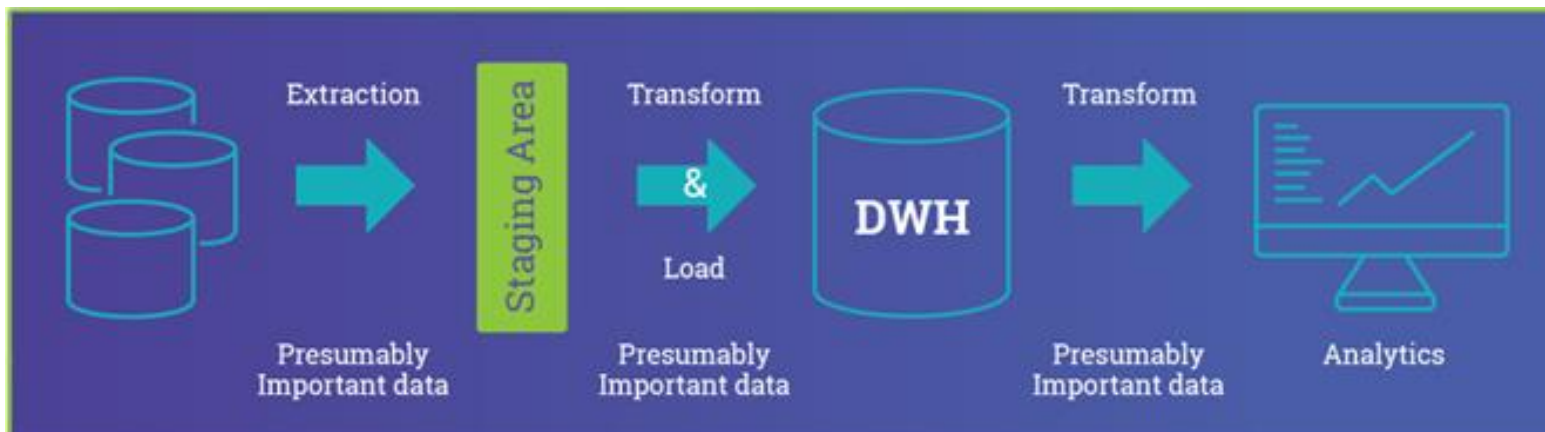


ETL - Extract, Transform, Load

# Algumas vantagens para ETL

## Garantia significativa da qualidade dos dados

A Ferramentas de ETL, através de sequências de operações e instruções tem condições de solucionar problemas de maior complexidade.



# Algumas vantagens para ETL

## Funcionalidade de execução

Uma ferramenta de ETL já possui suas funções específicas, sendo necessária apenas a atenção no fluxo de dados.

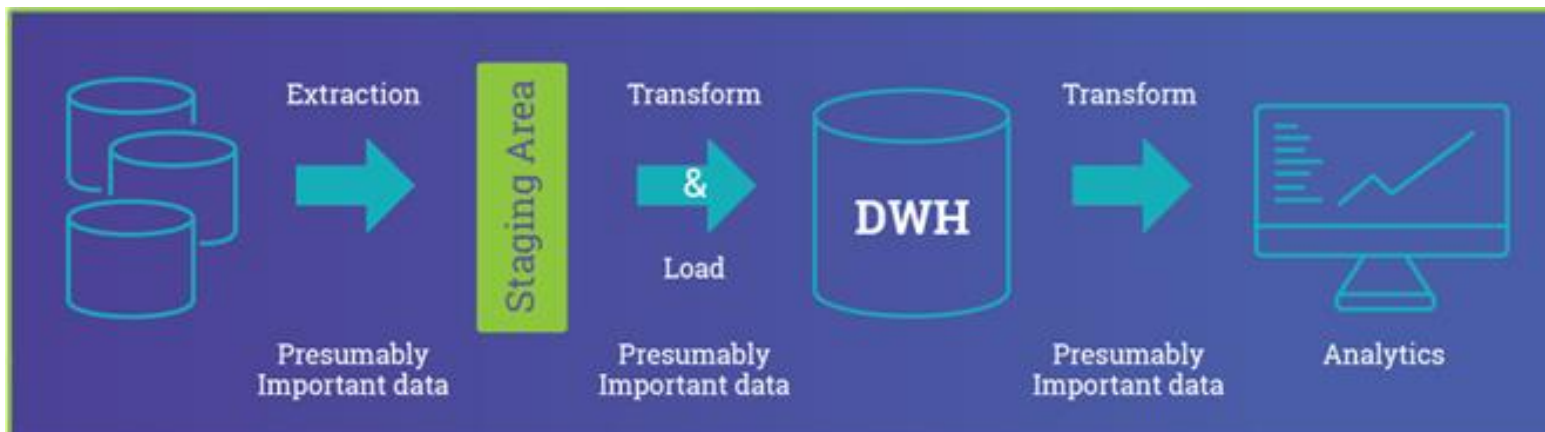




# Algumas vantagens para ETL

## Desenvolvimento das cargas

Mesmo que o usuário não seja técnico poderá desenvolver uma rotina de carga em uma ferramenta de ETL, devido a facilidade e rapidez para codificação.



# Algumas vantagens para ETL

## Manutenção das cargas

As tarefas de manutenção de uma rotina de carga são mais simples de realizar em relação à manutenção de código.



# Algumas vantagens para ETL

## Metainformação

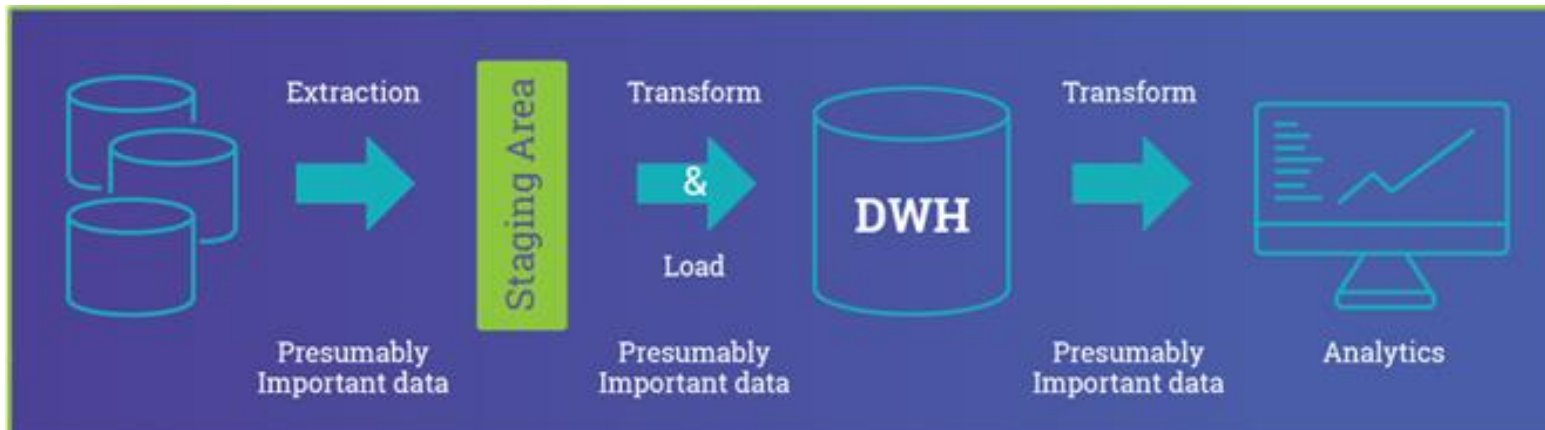
Os metadados (informações úteis para identificar, localizar, entender e gerenciar os dados) são gerados e mantidos de forma automática com a ferramenta, evitando problemas de geração de informações incorretas na finalização do processo.



# Algumas vantagens para ETL

## Performance

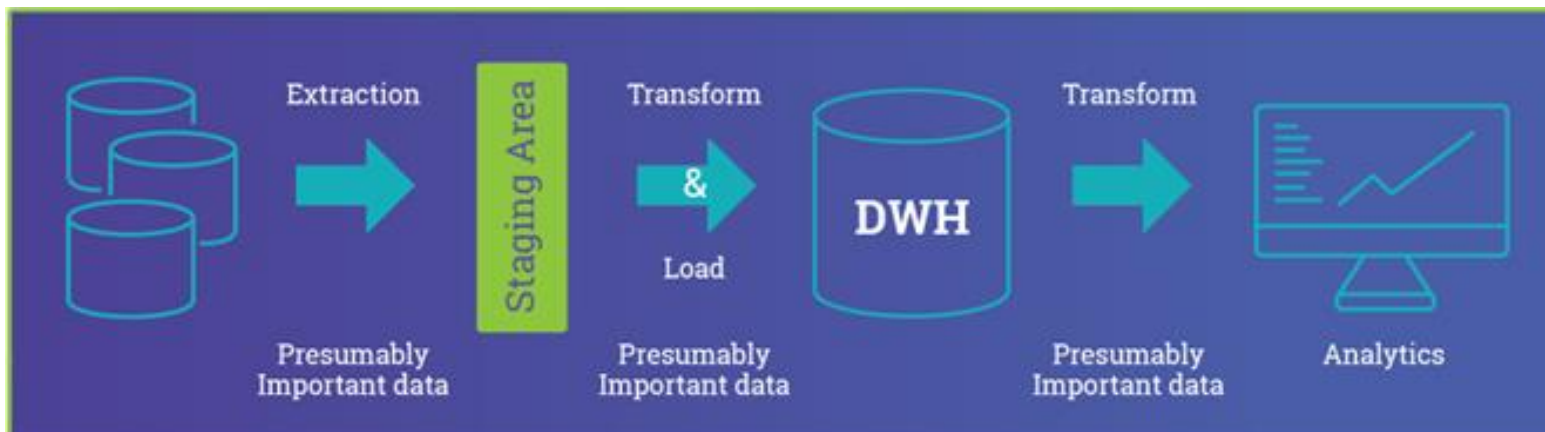
Os métodos mais usados para trabalhar com grandes volumes conseguem extrair, transformar e carregar dados com maior velocidade e menos recursos, como gravações em bloco e operações não logadas.



# Algumas vantagens para ETL

## Transferência

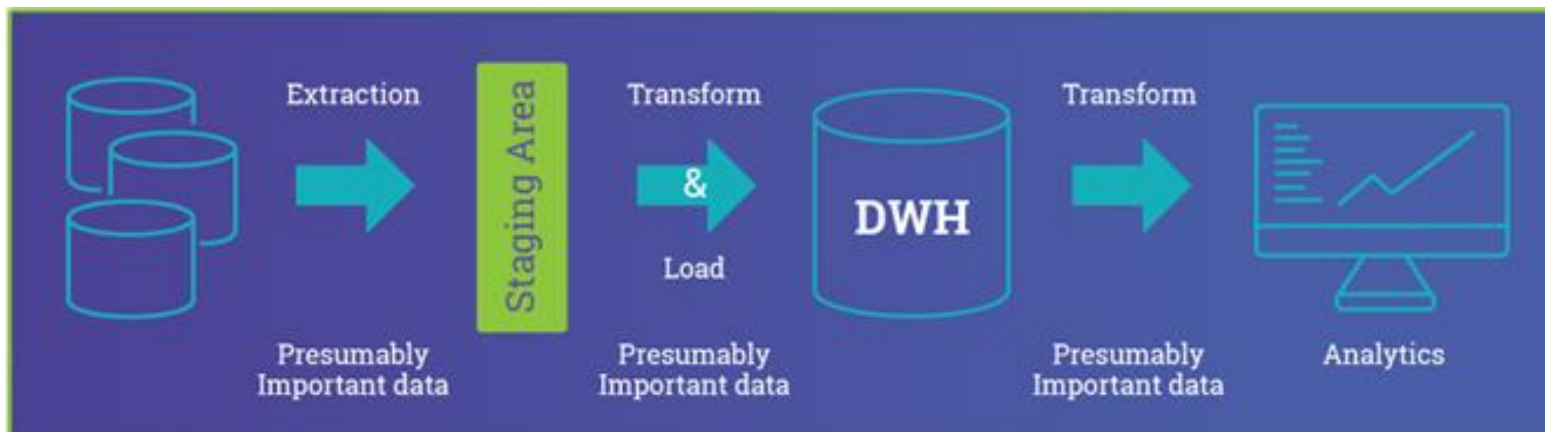
Ferramentas de ETL podem ser deslocadas de um servidor mais facilmente ou distribuídas entre vários servidores.



# Algumas vantagens para ETL

## Conectividade

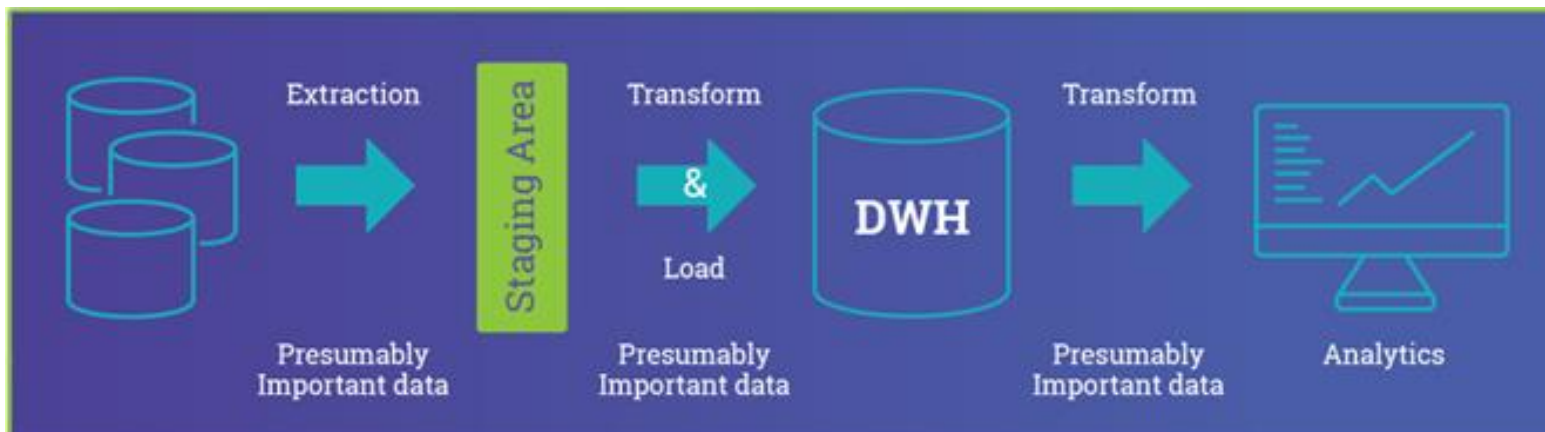
A conexão de uma ferramenta de ETL com múltiplas fontes de dados é transparente. Caso sejam precisas mais fontes como o SAP, VSAM, Mainframe ou qualquer outra, basta a aquisição do conector sem a necessidade de codificar um.



# Algumas vantagens para ETL

## Reinicialização

Ferramentas possuem a capacidade de reiniciar a carga de onde pararam sem a necessidade de codificação.



# Algumas vantagens para ETL

## Segurança e Estabilidade

É possível articular melhor a segurança tornando-a mais modular, dividindo as finalidades (criação de cargas, execução de cargas, agendamento, etc.)





# Obrigado!

Prof. Dr. Diego Bruno  
*Machine Learning*

