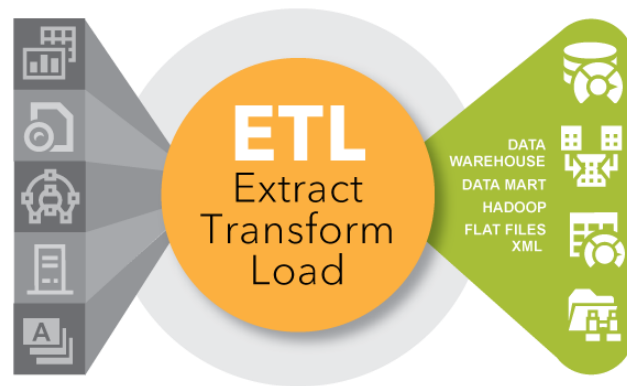


Introdução para ETL (Extract, Transform, Load)



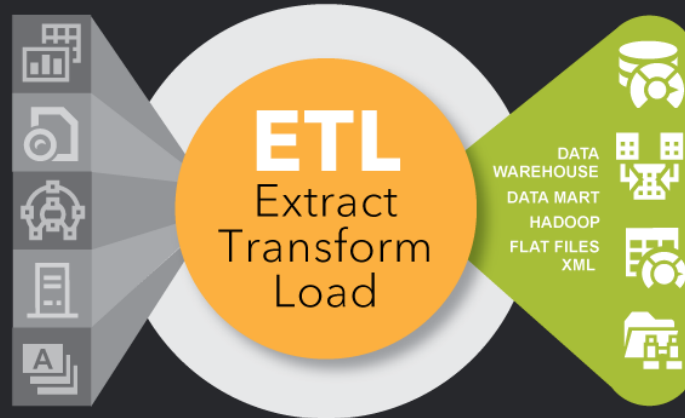
Prof. Dr. Diego Bruno

Education Tech Lead na DIO

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

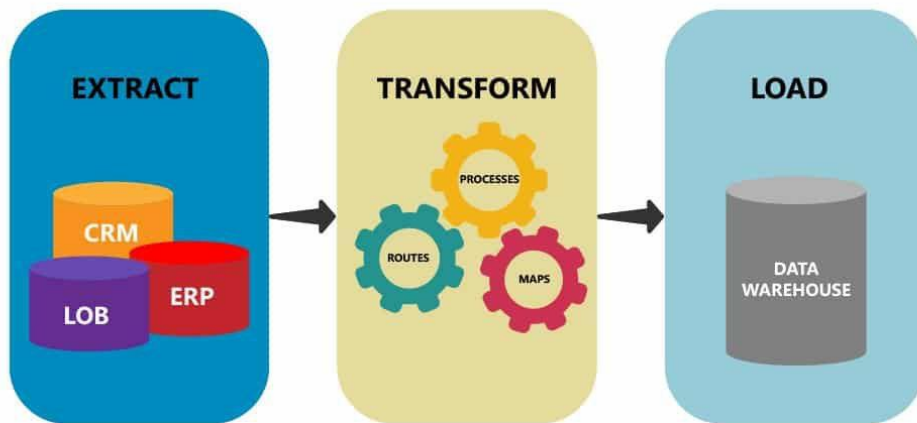
Introdução para ETLs

Prof. Dr. Diego Bruno



Mas o que é ETL?

ETL é um tipo de **data integration** em três etapas (extração, transformação, carregamento) usado para combinar dados de diversas fontes. Ele é comumente utilizado para construir um **data warehouse**.



ETL - Extract, Transform, Load

Mas o que é ETL?

ETL é um tipo de **data integration** em três etapas (extração, transformação, carregamento) usado para combinar dados de diversas fontes. Ele é comumente utilizado para construir um **data warehouse**.



EXTRACT (EXTRAIR)

- Importar Dados de diversos tipos e formatos Como:
- Pasta de Trabalho, Diversos Banco de dados, CSV, TXT, JSON, etc.



TRANSFORM (TRANSFORMAR)

- Colunas, Linhas
- Tipos de Dados
- Mesclar, Acrescentar
- Listas, Tabelas

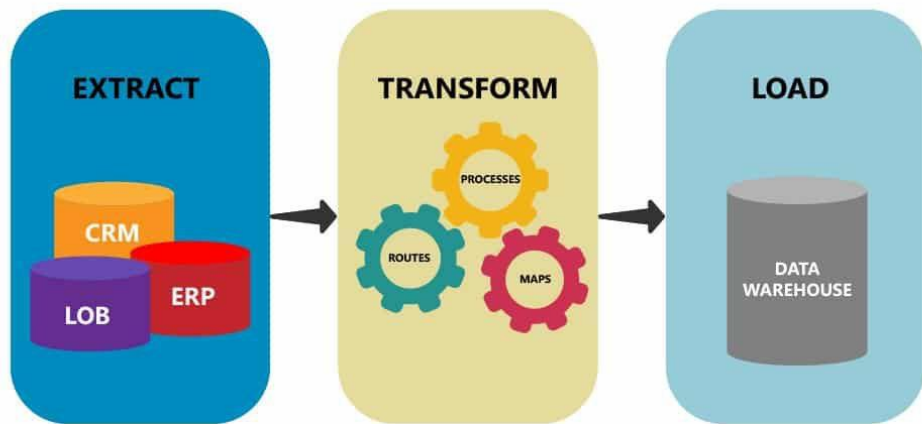


LOAD - CARREGAR

- Para o Modelo de Dados

Mas o que é ETL?

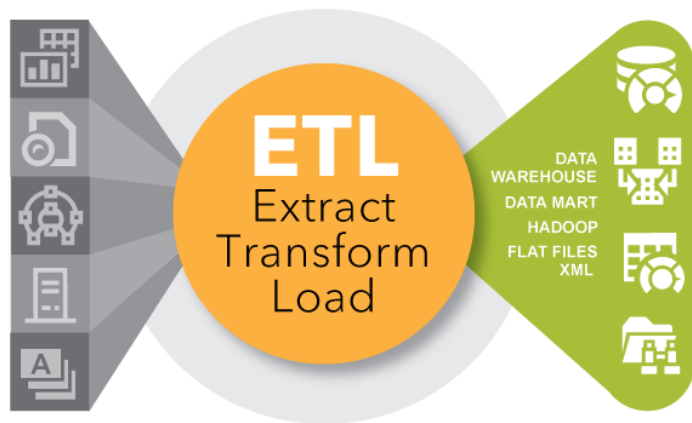
ETLs, são ferramentas de software cuja função é a extração de dados de diversos sistemas, transformação desses dados conforme regras de negócios e por fim o carregamento dos dados geralmente para um *Data Mart* e/ou *Data Warehouse*.



ETL - Extract, Transform, Load

Mas o que é ETL?

Nesse processo, os dados são retirados (extraídos) de um sistema-fonte, convertidos (transformados) em um formato que possa ser analisado, e armazenados (carregados) em nuvem ou outro sistema. Extração, carregamento, transformação (ELT) é uma abordagem alternativa, embora relacionada, projetada para jogar o processamento para o banco de dados, de modo a aprimorar a performance.



Ferramentas

Existem muitas ferramentas de ETL disponíveis no mercado como: **IBM Information Server (Data Stage)**, o **Oracle Data Integrator (ODI)**, o **Informatica Power Center**, o **Microsoft Integration Services (SSIS)**. Existe também um conjunto de Ferramentas de ETL Open Source como o **PDI – Pentaho Data Integrator** e **Talend ETL**.



Processo de ETL

O processo de extração, transformação e carregamento (ETL) abrange alguns passos importantes. Como exemplo, podemos considerar um Banco de dados de Clientes Especiais com todas as informações essenciais.



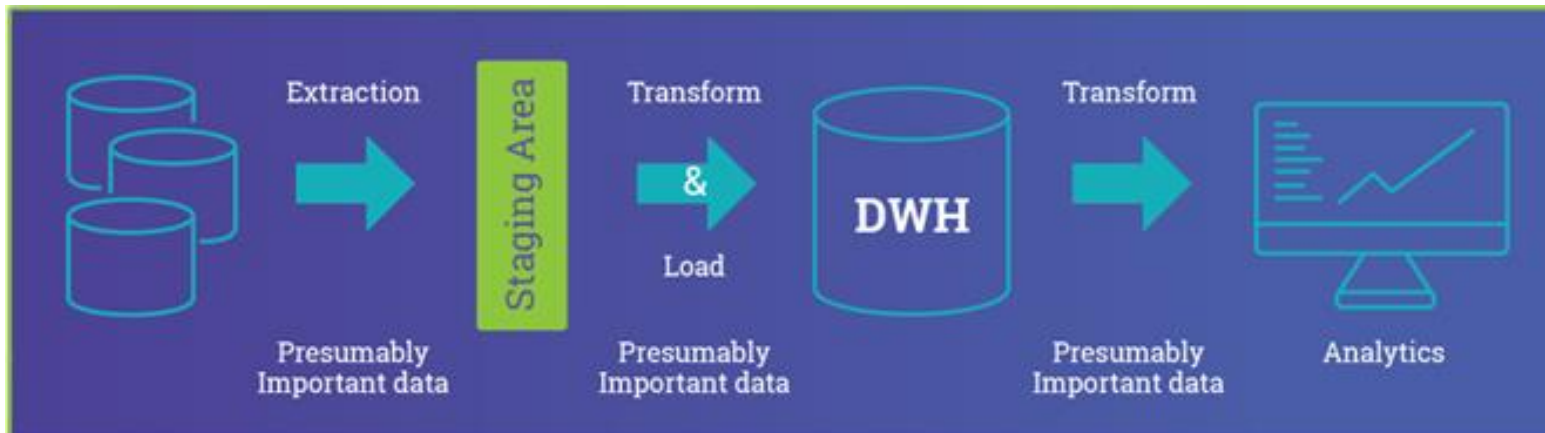
Processo de ETL

No mapeamento, a extração de origem deve conter a especificação da identidade e seus atributos detalhados, tudo armazenado numa zona temporária. Quando forem efetuadas as análises e filtrações dos dados, a nova versão poderá ser comparada com a cópia da versão prévia.



Processo de ETL

A transformação inclui limpeza, racionalização e complementação dos registros. O processo de limpeza removerá erros e padronizará as informações. O processo de complementação implicará no acréscimo de dados.



Obrigado!

Prof. Dr. Diego Bruno

