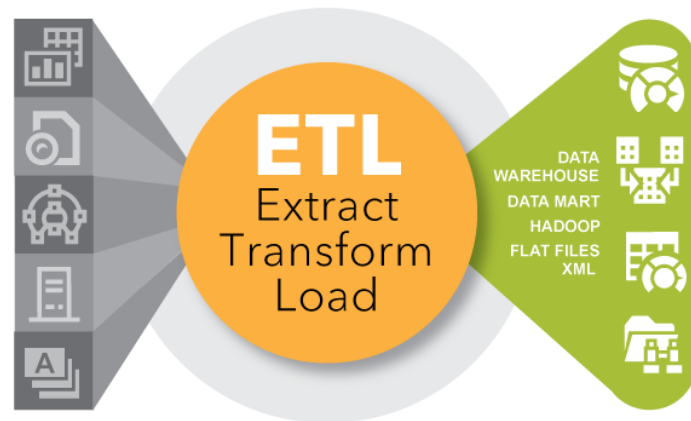


Ferramentas Aplicadas para ETL

Prof. Dr. Diego Bruno

Education Tech Lead na DIO

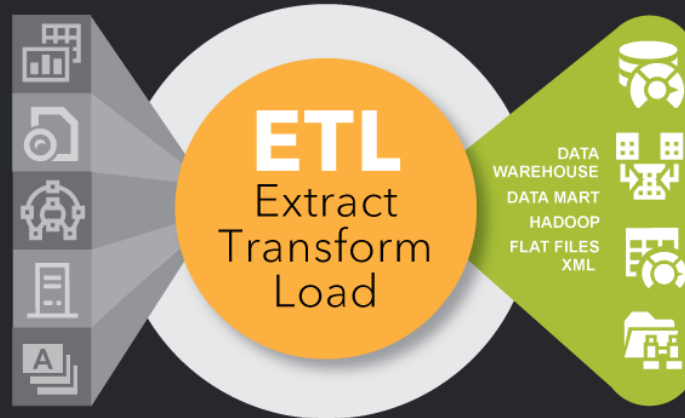
Doutor em Robótica e *Machine Learning* pelo ICMC-USP



Ferramentas de ETLs

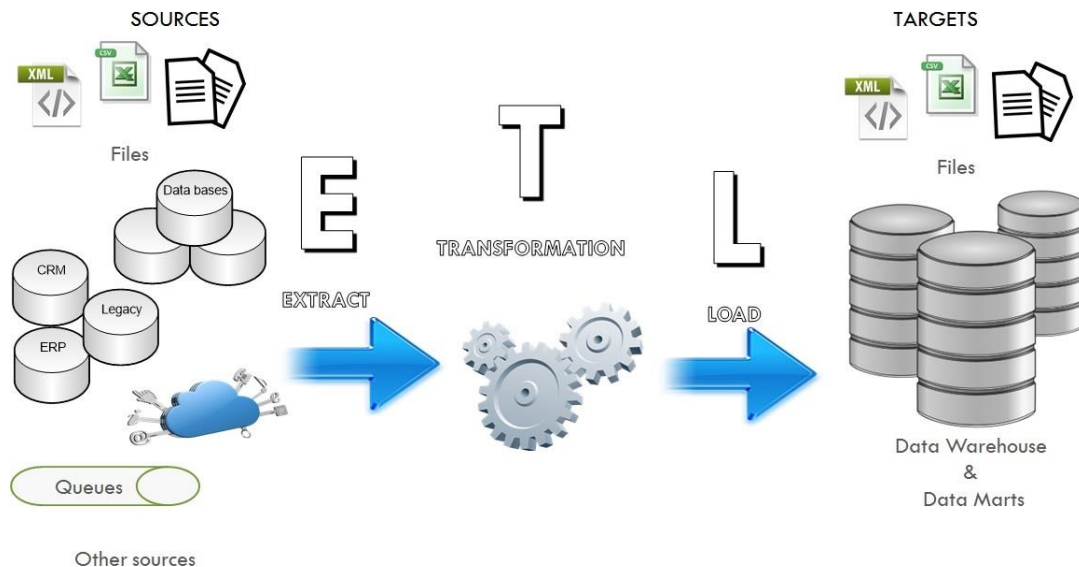
Prof. Dr. Diego Bruno

Machine Learning



Ferramentas Utilizadas

As ferramentas são *softwares* utilizados para facilitar o processo de integração de dados. Inicialmente muito usados em projetos de **Data Warehouse** e **Business Intelligence** em geral, ultimamente tem sido utilizados em processos de integração de software, bancos de dados, etc.



Ferramentas Utilizadas

Existem diversas ferramentas de ETL, como :

IBM Data Stage – <https://www.cetax.com.br/datastage-ibm-ferramenta-de-etl/> Informatica

Power Center – <https://www.cetax.com.br/power-center-informatica-ferramenta-de-etl/> SSIS

Sql Server Integration Services – <https://www.cetax.com.br/ssis-sql-server-integration->

Talend ETL – <https://www.cetax.com.br/criando-job-simples-no-talend/>



ETL para BIG Data

Hoje com o crescimento dos projetos de Big Data aumenta-se mais ainda a necessidade de fazer ETL entre plataformas heterogêneas, para isso, projetos como o *Hadoop*, possuem ferramentas próprias para carga de dados, como :

SQOOP – Ferramenta para movimentar dados dentre bancos de dados relacionais e o ambiente *Hadoop*.

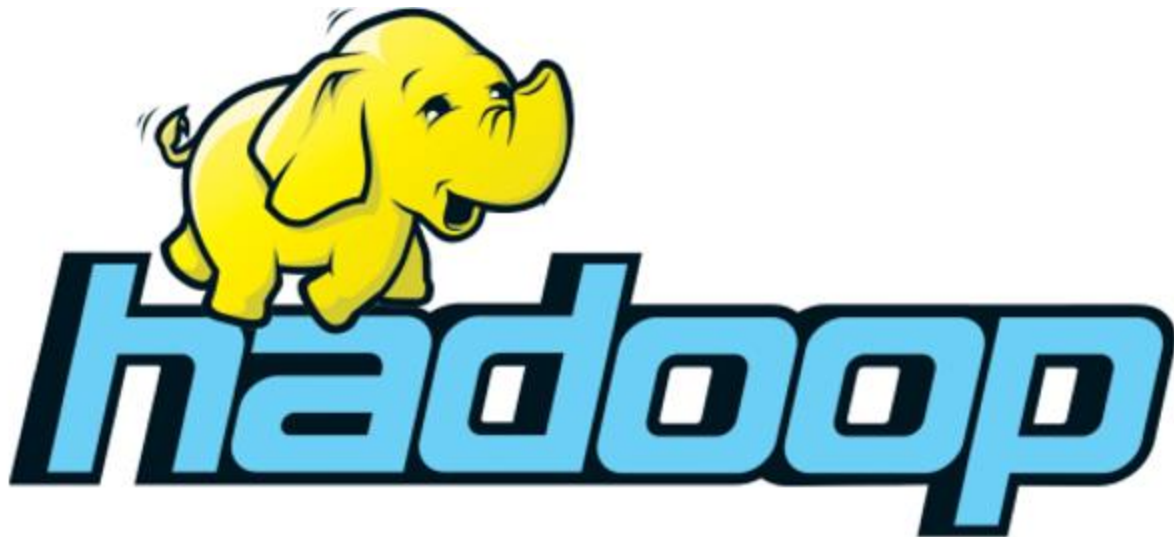
HIVE – Ambiente de SQL sobre um cluster Hadoop.

PIG – Ferramenta de Script para transformação e processamento de dados.

SPARK – Framework de processamento em memória.

Mas o que é Hadoop?

Hadoop é uma plataforma de software em Java de computação distribuída voltada para clusters e processamento de grandes volumes de dados, com atenção a tolerância a falhas.



ETL para BIG Data

Mesmo com todas as possibilidades acima, vemos as ferramentas de ETL se adaptando para **Big Data** ou gerando códigos para serem rodados nessas ferramentas do Ecosistema Hadoop.

Em Big Data, o processo de carga também é conhecido como Ingestão de Dados, que geralmente é a primeira parte da carga (Extract) a parte mais simples do processo, que consiste em extrair dados dos sistemas de origem e trazer para dentro do Data Lake ou ambiente de dados utilizado.

Obrigado!

Prof. Dr. Diego Bruno
Machine Learning

