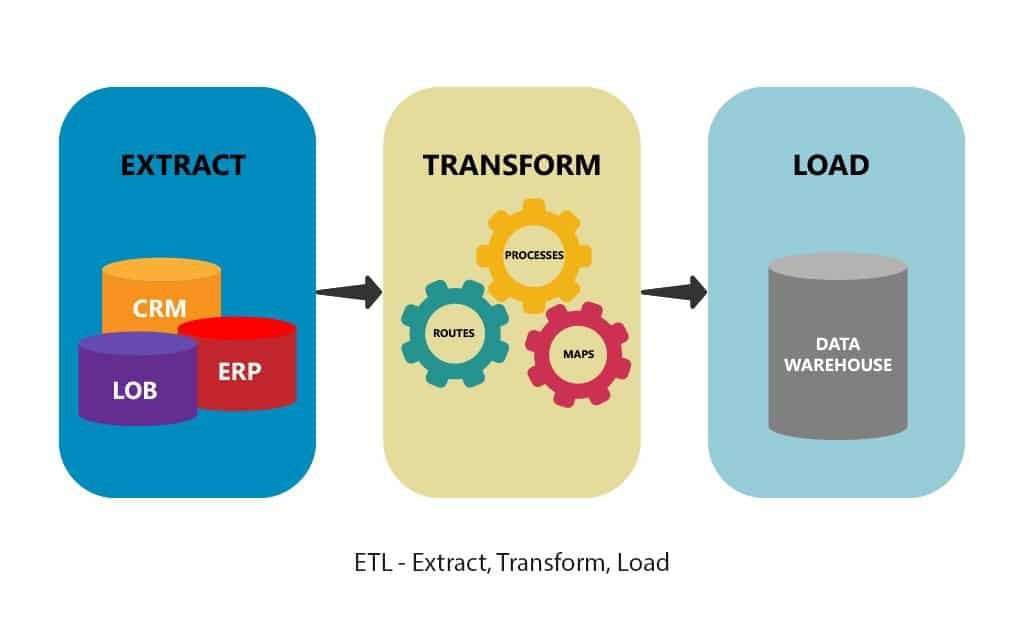
**Fundamentos de ETL**

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

****

Nesse processo, os dados são extraídos de um sistema fonte, convertidos (transformados) em um formato que possa ser analisado e armazenados (carregados) em nuvem ou outro sistema.

**Ferramentas de ETL:**

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

**Etapas de ETL:**

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.
2. Transform 🡪 O processo de transformação de dados é composto por várias etapas, sendo elas: padronização, limpeza e garantir a qualidade dos dados. Dados vindos de diferentes sistemas têm padrões diferentes (seja de nomenclatura ou mesmo de tipos de dados), portanto o processo de transformação é crucial para padronizar todos esses dados.
3. Load 🡪 O processo de carregamento é a etapa final onde os dados são lidos das áreas de staging e preparação de dados e posteriormente carregados no Data Warehouse ou *Data Mart Final*.