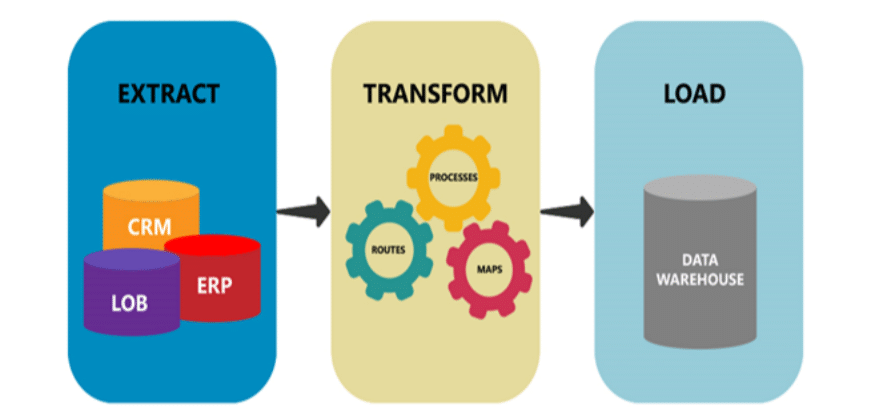
Ferramentas ETL de Código Aberto



Escolher ferramentas de engenharia de dados para compor boa estrutura para desenvolvimento é tão importante quanto a mão na massa.

Atualmente temos soluções de ETL nativa da/na/para Cloud, On-premise, de código aberto, pagas, com e sem suporte e as híbridas.

Falaremos das ferramentas ETL de código aberto mais populares, com seus recursos, e diferente de outras postagens que apresentam diversas soluções, seguiremos na lógica que uso, ‘apresentar no máximo 3 opções que goste e tenha experiência’. Aqueles que se identificam e não gostam de gastar tempo em pesquisas eternas de diversas ferramentas, para quando escolher uma, descobrir que ela já teve alterações... Boa leitura!

[**Airbyte**](https://airbyte.com/)

****

**Visão geral do Airbyte**

Inspirado no Singer, tem melhorias em diversos aspectos: A padronização da base de código para toques e destinos, propriedade centralizada do código, roteiro mais previsível, e conta com o suporte da comunidade.

**Recursos do Airbyte**

Lançado no segundo trimestre de 2020, conta com extensibilidade e flexibilidade. Uma das principais inovações é separar a etapa de transformação das etapas de extração e carregamento. Isso permite que a Airbyte se integre a outras ferramentas de transformação de dados.

A Airbyte também trouxe o conceito de Reverse ETL para a frente. O ETL reverso está se tornando cada vez mais importante para as empresas, pois ajudará a justificar os enormes custos de execução de operações de ETL para preencher data warehouses. O ETL reverso permitirá que os dados nos data warehouses sejam realimentados para os sistemas operacionais, ajudando ativamente nos insights de negócios e nas operações de negócios.

[**dbt**](https://www.getdbt.com/)

****

**Visão geral do dbt**

Iniciado como um projeto na [RJMetrics](https://blog.rjmetrics.com/" \t "_blank) em 2016 para estender os recursos de transformação do Singer pela StitchData, o dbt era de código aberto desde o início. A Fishtown Analytics pegou a base de código principal e criou seu produto. Desde então, o dbt tem sido amplamente adotado devido à sua facilidade de uso e sua capacidade de fazer transformações baseadas em SQL de forma muito eficaz, aproveitando o poder do mecanismo de modelagem Jinja2.

**Recursos de dbt**

O dbt é facilmente integrado a ferramentas de orquestração como Prefect ou Airflow e também funciona bem com qualquer ferramenta básica de extração e carregamento que queira transferir a carga de trabalho de transformação para outra ferramenta. Depende mais da linha de comando e menos da interface do usuário. É extremamente leve. Podemos executá-lo localmente instalando-o usando o Homebrew, usando pip ou executando-o em um contêiner do Docker.

Existem muitos tutoriais do YouTube e postagens de blog para aprender e conhecer o dbt, mas também há um esforço bem direcionado do dbt Labs para criar [cursos e tutoriais úteis](https://courses.getdbt.com/collections) para novos alunos. Saiba mais sobre o roteiro do dbt no [blog oficial](https://blog.getdbt.com/2022-dbt-enterprise-product-roadmap/).

[**Talend**](https://www.talend.com/products/talend-open-studio/)

****

**Visão geral do Talend Open Studio para integração de dados**

O Talend Open Studio for Data Integration é uma das ferramentas de ETL mais populares. Foi lançado pela primeira vez no ano de 2006. Depois de se tornar rapidamente uma das forças a serem reconhecidas, o TOS DI estava competindo com o Informatica PowerCenter, IBM DataStage e outros.

**Recursos do Talend Open Studio para integração de dados**

Enquanto muitas das ferramentas de ETL atuais se concentram em resolver uma etapa específica do pipeline de ETL, a Talend resolve tudo isso com recursos avançados de ETL, orquestração, privacidade de dados e segurança integrados.

A ferramenta ETL compreende dois repositórios GitHub separados — [Common Code entre Talend Products](https://github.com/Talend/tcommon-studio-se) e [TOS DI](https://github.com/pentaho/pentaho-kettle) . É uma ferramenta ETL muito sofisticada, cheia de recursos avançados de integração de dados.