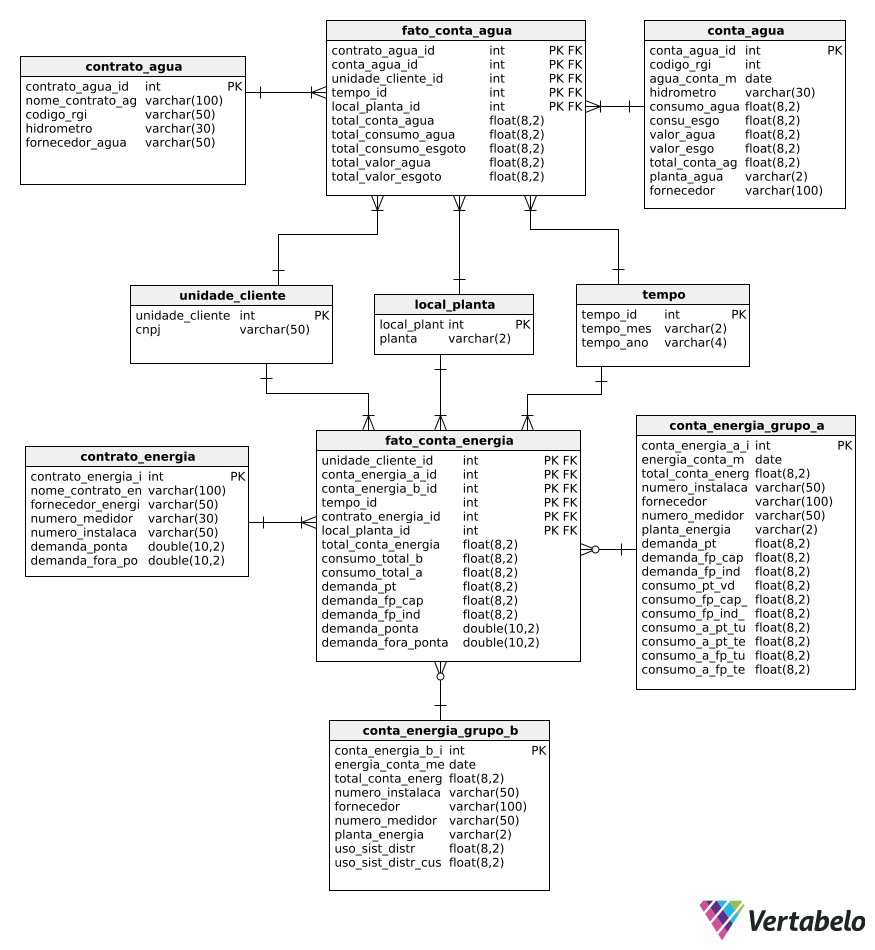
FLUXO DE DADOS – ETL

Modelagem do Banco de Dados



Extração

Nessa etapa de extração dos dados, todos os arquivos com dados brutos são oriundos de arquivos no formato .csv. Há quatro tipos de arquivos de dados que são a base para a modelagem do banco de dados.

* **Arquivo de fatura de água**: nomeado como pro\_agua.csv, contém informações acerca do consumo de água e esgoto de um cliente durante o período de um mês.
* **Arquivo de contrato de energia:** nomeado como con\_energia.csv, contém informações acerca do imóvel e serviços de energia contratados.
* **Arquivo de fatura de energia:** nomeado como pro\_energia.csv, contém informações acerca do consumo de energia de um cliente durante o período de um mês.
* **Arquivo de contrato de água:** nomeado como con\_agua.csv, contém informações acerca do imóvel e mensuração da fatura.

Transformação

Na etapa de preparação, os dados brutos de cada arquivo são tratados para garantir uma padronização de informação antes de ser carregado no banco de dados:

Faturas de água

Os dados do tipo texto que expressam o valor em R$ dos consumos das faturas, e para os dados do tipo texto que representam os valores de consumo volumétrico, são aplicadas as seguintes transformações, sequencialmente:

1. Remoção da vírgula inicial
2. Conversão do valor para tipo de números flutuantes.

Campos de aplicação:

* Valor Água R$
* Valor Esgoto R$
* Total R$
* Consumo de esgoto m3
* Consumo de água m3

Contrato de água