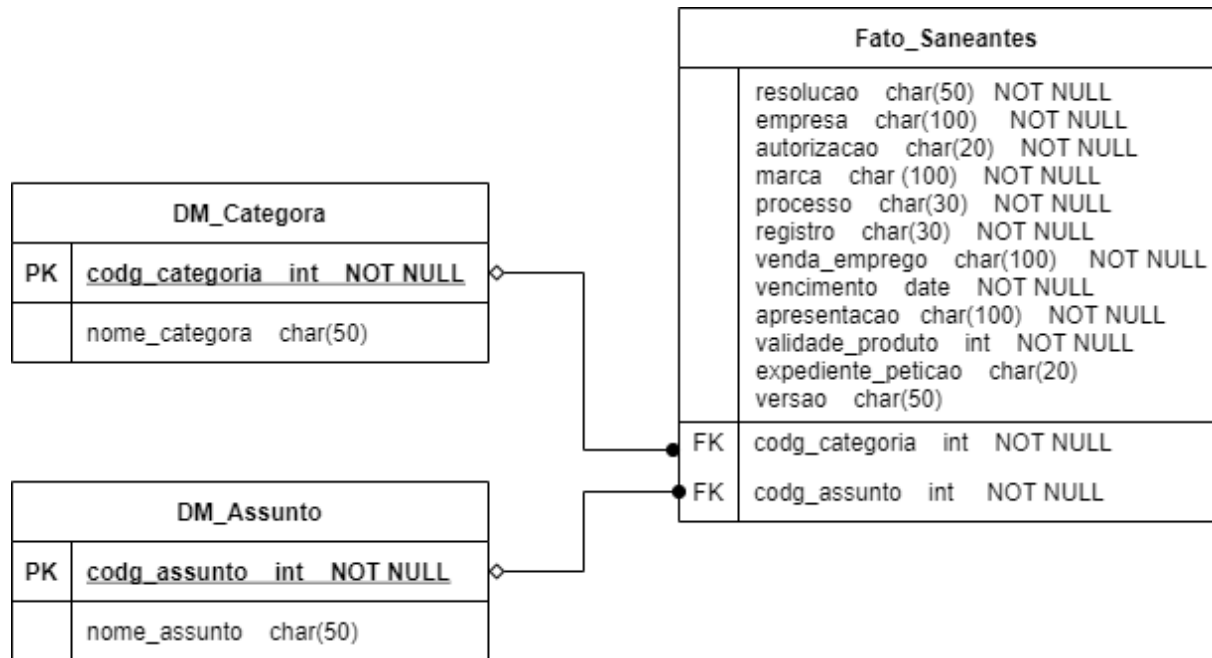
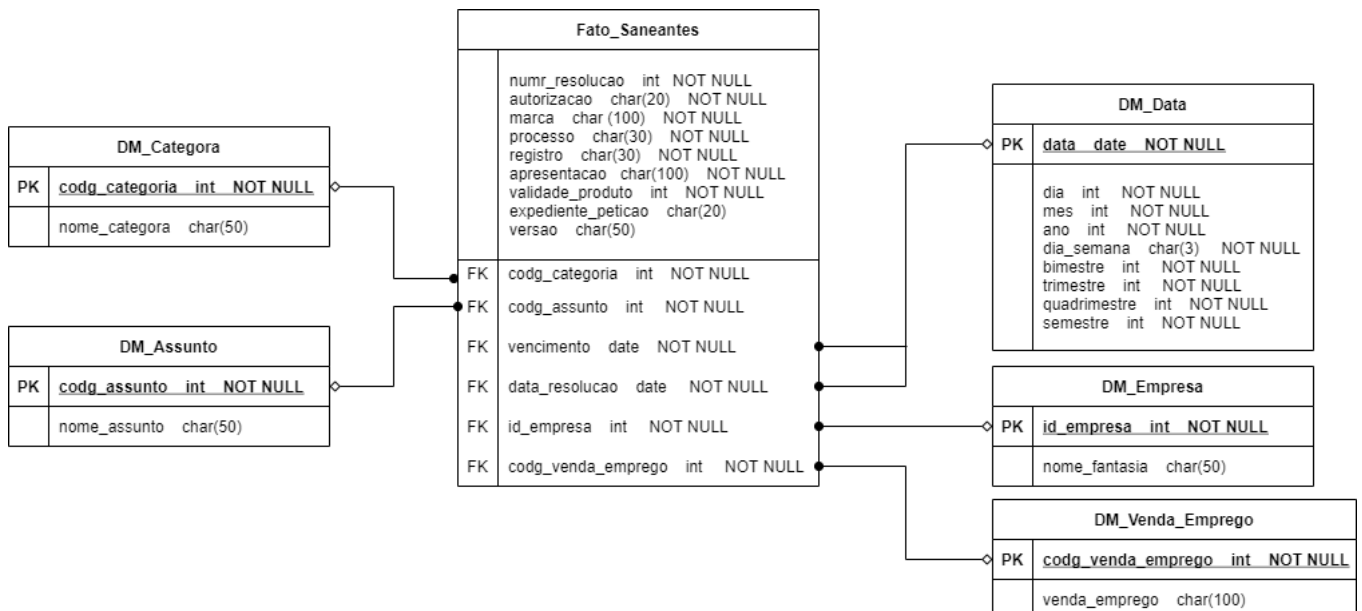


# Modelagem

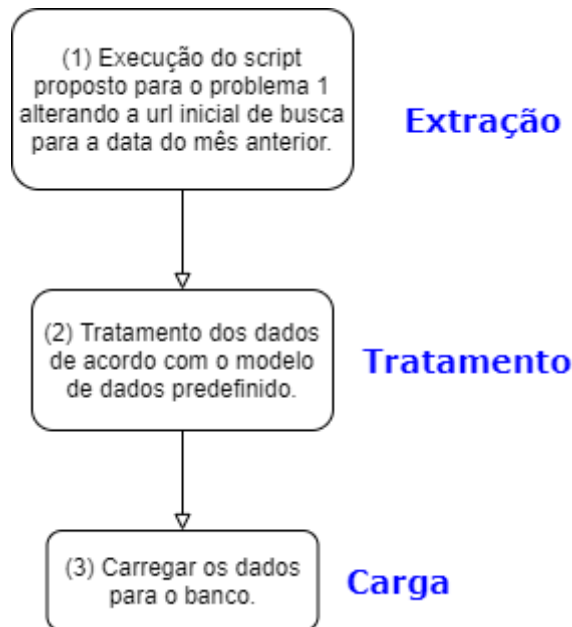
## 1.1 Modelo de dados



## 1.2 Sugestões



# Processo



Considerando o scrip de coleta dos dados proposto para o problema 1, a etapa de extração seria composta pela execução do script sendo necessário apenas alterar a url de busca das publicações para o período do mês anterior. Essa etapa pode ser realizada através de uma passagem de parâmetro do mês desejado, e o script seria responsável pela construção da url automaticamente.

Com os dados coletados, o passo (2) seria responsável por organizar os dados de acordo com o modelo de dados predefinido. Assim, os campos e seus valores seriam reformulados seguindo o modelo e os dados estariam prontos para a etapa (3), onde são carregados para o banco de dados ou outro local de registro dos mesmos, como arquivos. No caso do uso de um banco de dados, seria necessário construir o banco antes de realizar a primeira execução dos três processos indicados.

Para a realização das etapas de tratamento e carga, diversas ferramentas podem ser utilizadas. Um próprio programa em python poderia dar continuidade ao script de extração, tratando os dados e carregando-os para o local de registro. Outras ferramentas, como Pentaho e Qlik, integram essas duas etapas e poderiam ser utilizadas também. Essas ferramentas possuem mecanismos de Lookup que auxiliam a realização de cargas incrementais, implementando inclusive versionamento de registros.