

# Problema: 1

## 1.2. Você sugere algum tipo de tratamento de transformação e/ou qualidade para os dados extraídos? Quais?

### Explicação:

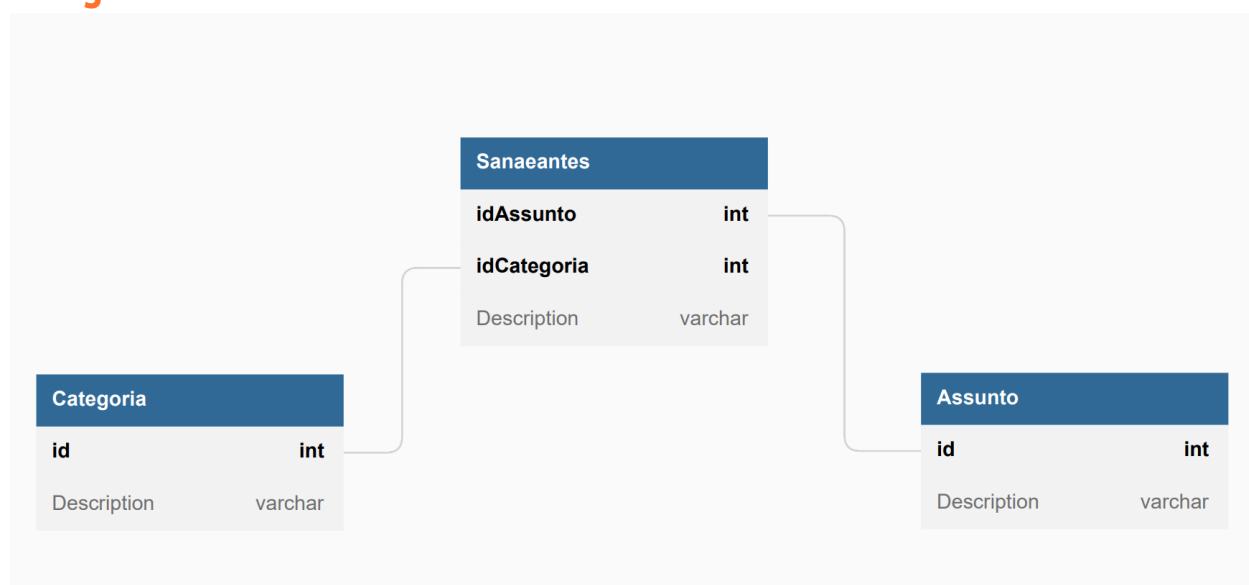
Os dados apresentados devem passar pela retirada de dados não relacionados à resolução antes de serem transferidos para uma tabela.

Como são apresentados dentro de uma requisição processada em javascript, todos os dados devem ser requisitados, atrelados a um arquivo de acesso rápido - e.g. txt por conta da quebra de linhas - e separados por uma quebra de linha, para que a leitura posterior, de cada dado, contenha um separador de formatação de arquivo

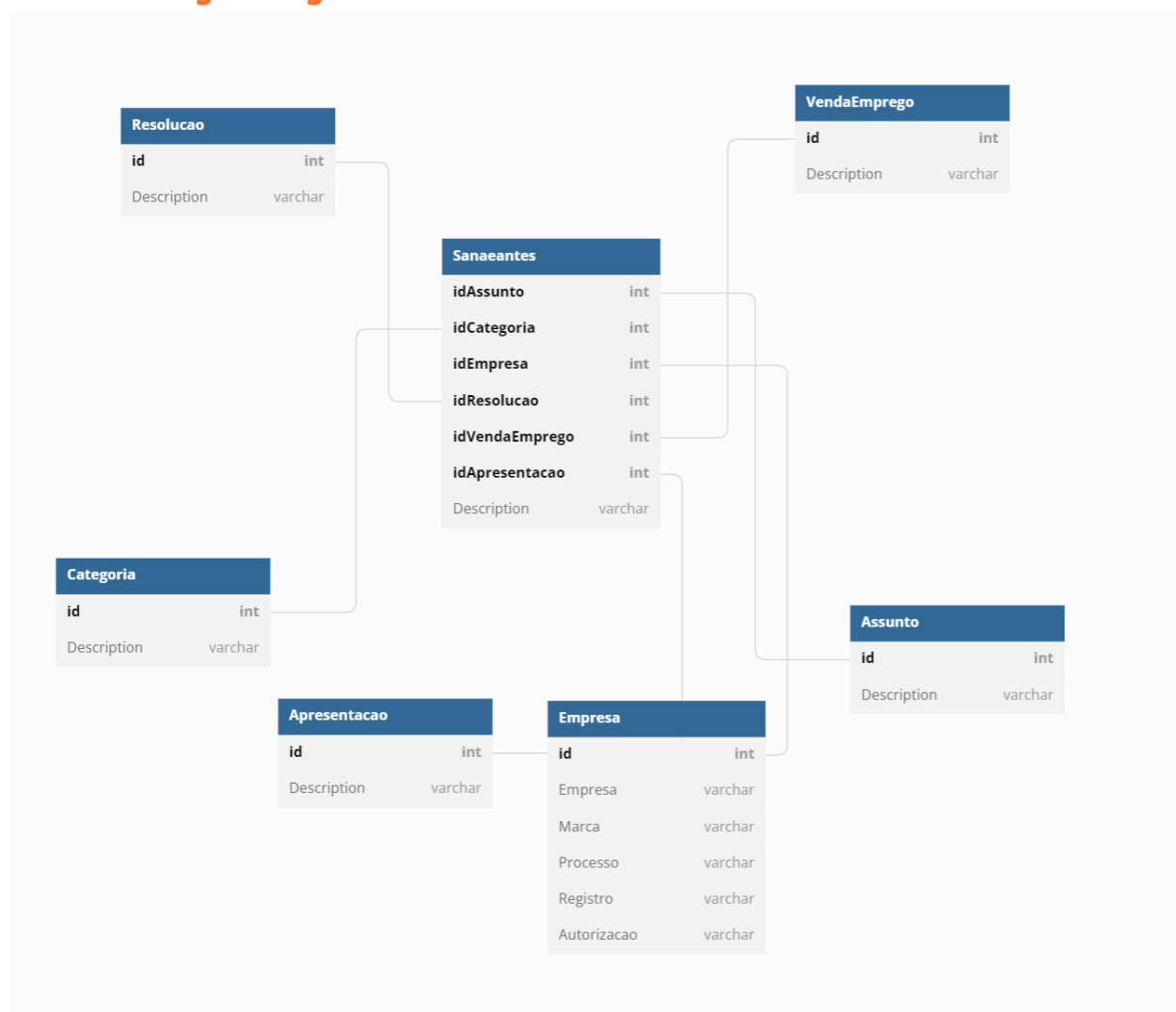
# Problema: 2

## Modelagem:

1.1. Você deve criar um modelo de dados contendo uma tabela fato para Saneantes e duas tabelas dimensão: uma para Assunto e outra para Categoria:



## 1.2. Você sugere alguma outra tabela dimensão?



Explicação:

Para toda dimensão mutável, mas dependente da empresa, os dados devem ser atrelados à tabela dimensão relacionada à empresa.

Para todas as outras categorias coletadas dos relatórios, novas dimensões devem ser adicionadas: resoluções; vendas e emprego; apresentação.



### 1.3. Considerando que este banco será em SQL que bibliotecas você utilizaria?

#### I. SQLite - Para gerenciamento de banco de dados em uma única máquina:

SQLite é um banco de dados SQL baseado em arquivo independente. Como o SQLite vem junto com o Python e pode ser usado em qualquer projeto python que será executado localmente.

#### II. MySQL - Para gerenciamento de banco de dados em servidor:

Ao contrário do SQLite, não há um módulo Python SQL padrão que você possa usar para se conectar a um banco de dados MySQL. não há um módulo SQL para python que seja utilizado, por consenso da área, para se conectar a um banco de dados MySQL. Como MySQL é um sistema de gerenciamento de banco de dados baseado em servidor, pode ter vários bancos de dados. Ao contrário do SQLite, onde criar uma conexão equivale a criar um banco de dados, um banco de dados MySQL possui a necessidade de se conectar ao servidor que contém o banco de dados e realizar o query.



**1.1. Você deve desenhar um processo de captura, transformação e carga que explique quais passos serão seguidos todos os meses para incrementar novos dados ao modelo:**

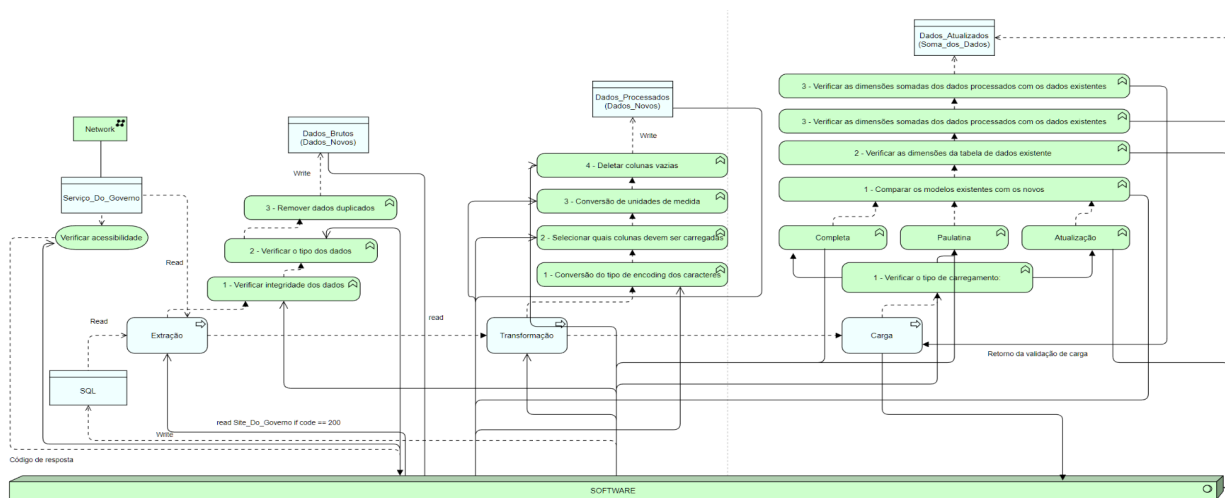


Imagem em tamanho real disponível dentro da pasta  
"Explicacoes\_PrimeiroESegundo\_Problemas\_Documentos"

### 1.2. Que ferramenta para controle deste processo você utilizaria (agendamento, monitoramento etc.)?

Ao contrário do processo de estratificação do sistema de ETL realizado na última década, sistemas de integração de consolidação de extração, transformação e carga tornaram-se epítomes de excelência na realização do processo de ETL. Utilizar um pipeline de gerenciamento de integração do controle do processo seria o ideal. Ferramentas como QuerySurge permitem integração com sistemas MySQL e PostgreSQL- utilizados em projetos Python. CData também oferece um sistema completamente customizado de integração e transformação de dados, além do gerenciamento das ferramentas de extratção e APIs de monitoramento. Integração com Redshift da amazon e Azure. Por último, mas não menos importante, temos o sistema de pipeline do Hevo.

Caso as tarefas sejam realizadas sem a utilização de pipelines, sistemas como Amazon RedShift e ferramentas proprietárias da Oracle seriam as mais indicadas.