Prova em Grupo

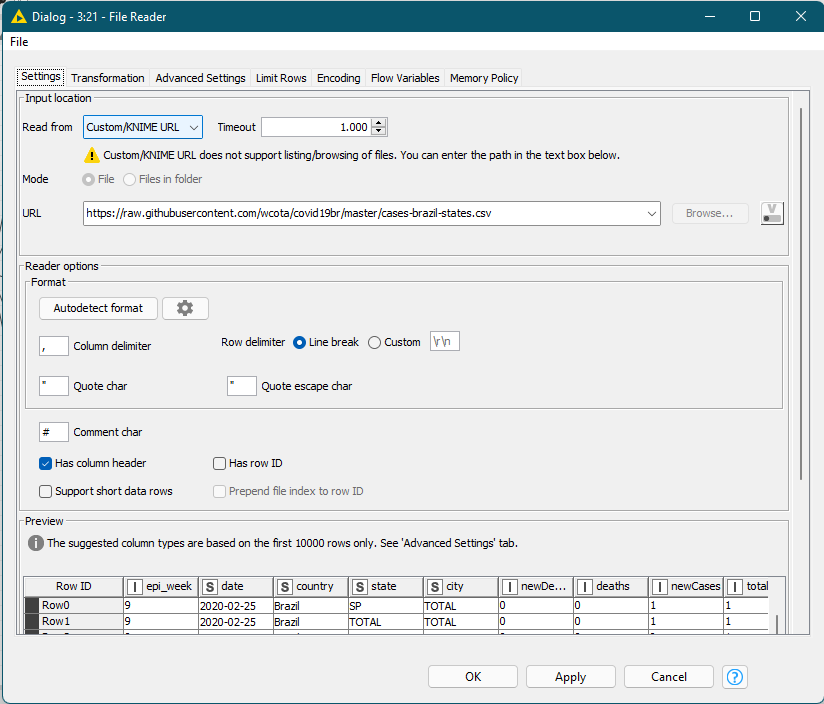
Thiago Outeiro Pereira Damasceno - DRE: 116038363  
Giovani Tricarico Barros - DRE: 118051317  
João Wendling Assayag - DRE: 118113834

# 1. Coleta de dados

Inicialmente nosso grupo faria o trabalho em grupo baseado no site “Painel COVID RJ” (<https://experience.arcgis.com/experience/38efc69787a346959c931568bd9e2cc4>) porém, os dados do site se mostraram fracos para o trabalho com isso decidimos buscar novos dados, chegando aos dados utilizados pela Johns Hopkins University, estes são os mesmos mostrados pelo google, ao realizar uma busca por “ COVID 19 Brasil”, e para tal, eles utilizam como fonte o seguinte repositorio do Github: <https://github.com/wcota/covid19br> que são dados disponibilizados pela UFV (Universidade Federal de Viçosa).

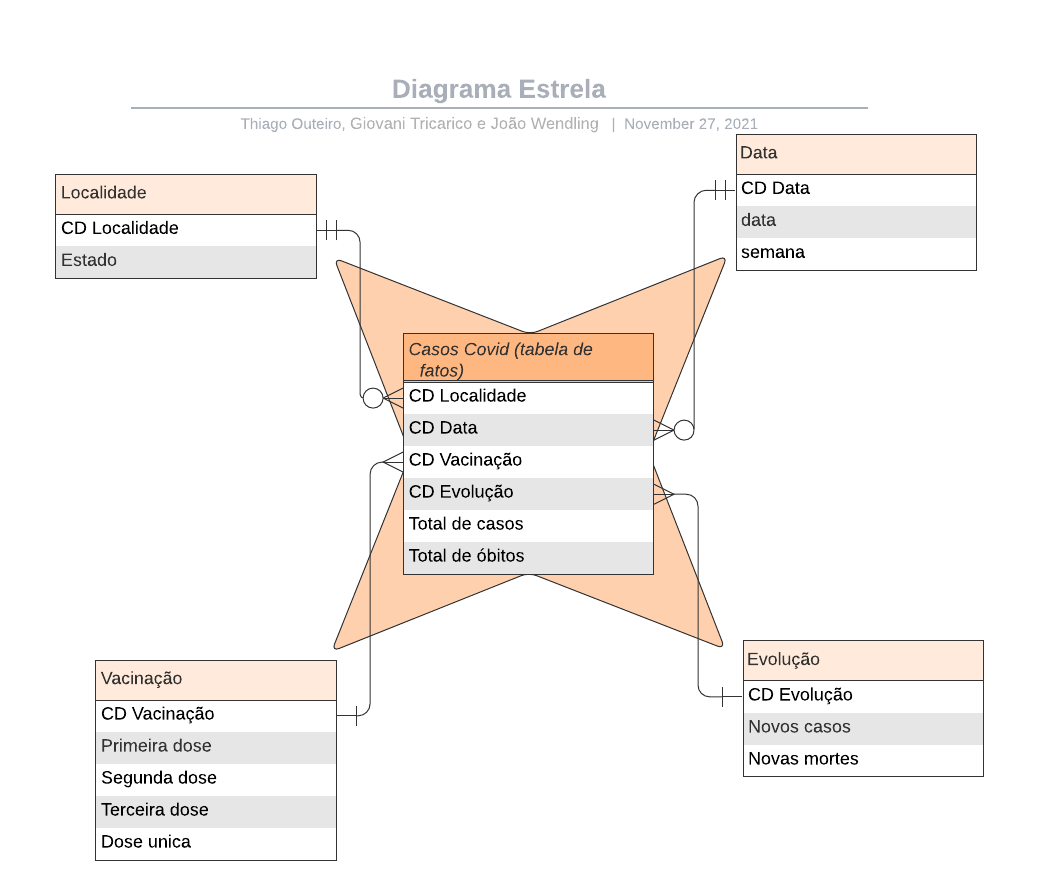
Neste repositório existem diversos tipos de dados, desde dados gerais sobre o brasil, a dados por cidades, porém para acompanhar a evloução da covid, os dados mais interessantes foram os dados por estado (<https://raw.githubusercontent.com/wcota/covid19br/master/cases-brazil-states.csv>).

Para a leitura dos dados foi utilizado o node “FileReader” com a configuração de “read from” como “ Custom/KNIME URL” e passando a URL dos dados, o resto foi deixado como auto detecção.



# 2. Diagrama Estrela

Para a criação do modelo estrela foi observado os dados presentes no arquivo, e foi montado o seguinte diagrama:



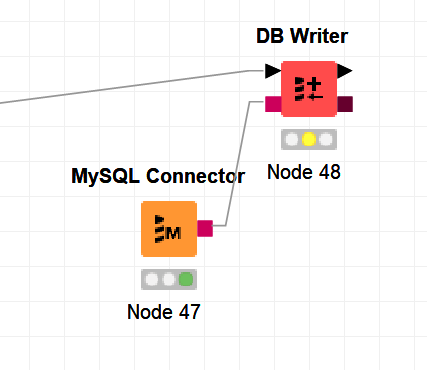
# 3. Criação do banco de dados

Para a criação do banco foi utilizado o aplicativo mySQL Workbench, e criado uma conection na própria IDE como mostrado abaixo:



Para a criação de tabelas foi utilizado o KNIME, no momento da carga dos dados, utilizando os nós MySQL Connector e DBWrite:

1. **MySQL Connector:** Utilizado para se conectar com o banco criado no MySQL WorkBench, sua saida é o banco que irá se conectar na entrada no DB Writer.
2. **DB Writer:** Utilizado para escrever no banco de dados recebido do MySQL Connector, irá criar a tabela passada nas configurações caso ela ainda não exista.

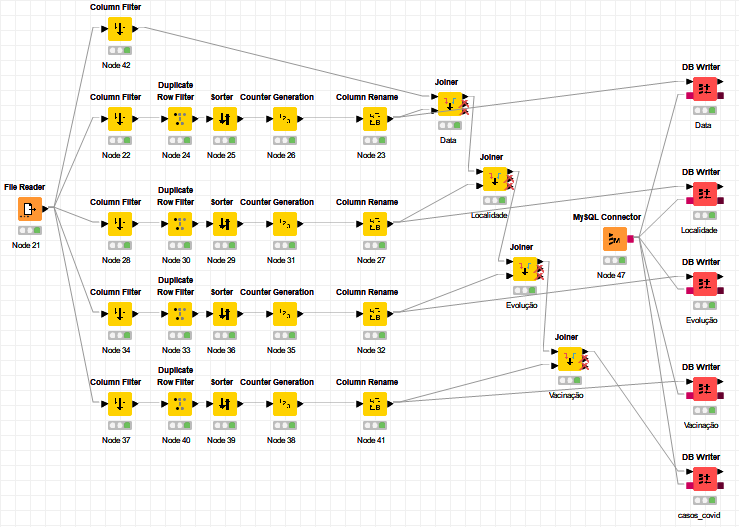


Para configurar esses nós é necessário que o banco seja criado no mySql Workbench e o schema seja criado com o seguinte script:



Após isso no node de Connector será necessário selecionar o banco ( no caso foi utilizado o próprio main) e o no writer será selecionado o schema e a tabela a ser criada.

# 4. Carregamento dos dados

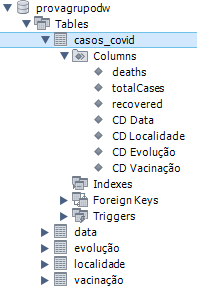
Para o carregamento dos dados foi utilizado o banco criado na questão 3, em conjunto com o KNIME obtendo o seguinte fluxo:

Para a carga de dados foram utilizados os nodes:

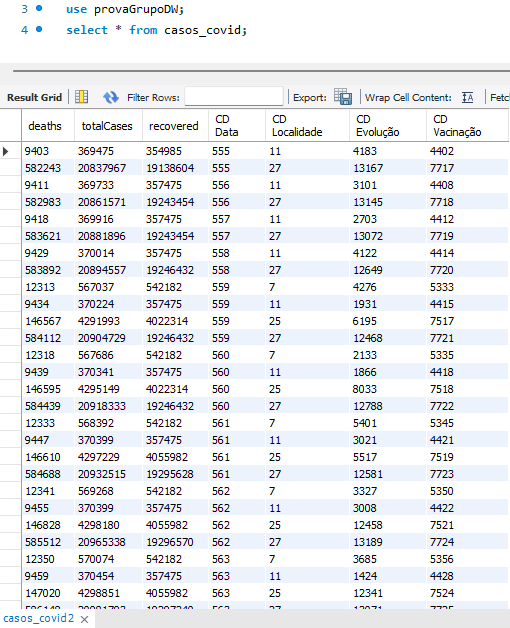
1. **Column Filter:** Utilizado para selecionar apenas as colunas necessárias para as tabelas.
2. **Duplicate Row Filter:** Utilizado nas dependencias para remover as linhas duplicadas, ficando com entradas únicas.
3. **Sorter:** Não é necessário, porém foi utilizado para a visualização ficar melhor, organizando as linhas das tabelas.
4. **Counter Generation:** Utilizado para gerar os IDs (chaves) das tabelas de dependencias
5. **Column Rename:** Utilizado para renomear as tabelas com o mesmo nome especificado no modelo.

Além destes 5 nodes, também foi utilizado o Node de Joiner, para realizar o pareamento da tabela de fatos com a chave para as tabelas de dependências. Todos os Joiners foram Inner Join, menos o último onde foi realizado o left outer join, pois como se tratam de dados de vacinas, no inicio dos casos não haviam vacinas. omo os dados eram concisos, o número de linhas entre o início e o final foi igual (17486).

Com isso na interface do banco temos:



Realizando um select na tabela de casos\_covid:



# 5. Análise de dados