**BANCO DE DADOS**

**DIMENSIONAL – STAR SCHEMA**

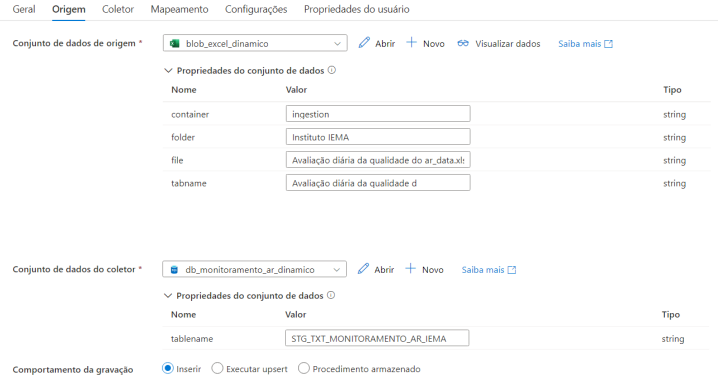
*Modelagem dimensional é a técnica de projeto lógico de banco de dados mais usada no desenvolvimento de data warehouses, embora também possa ser aplicada ao projeto de sistemas de informações operacionais. Na verdade, ela busca apresentar os dados em um formato que seja intuitivo e ao mesmo atenda a acessos com um alto desempenho (Kimball, 1998).*

*Envolve o uso de tabelas de fatos e dimensões para manter um registro de dados históricos em data warehouses. Eles são otimizados para realizar o “Selecionar” operação e são usados ​​na estrutura de design básico para construir data warehouses altamente otimizados e funcionais.*

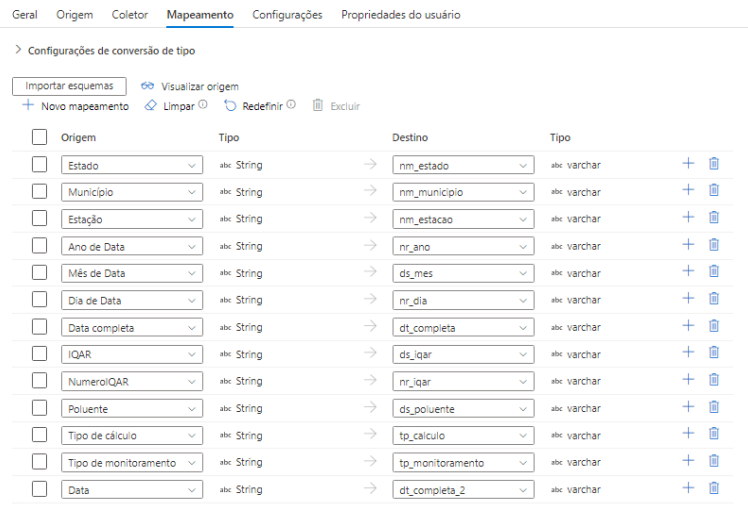
** Para a carga inicial de dados solitada, foi utilizada a fonte de dados do IEMA (Instituto de Energia e Meio Ambiente) contendo diversas bases de dados em formato .csv, para que as visualizações do Power BI fossem criadas:

<https://energiaeambiente.org.br/qualidadedoar>

Para inserção no banco de dados dessa stage, utilizamos as ferramentas Blob Storage e Azure Data Factory da Microsoft, criando um pipeline que importa do blob para o banco de dados.



Fonte: elaborado pelo grupo

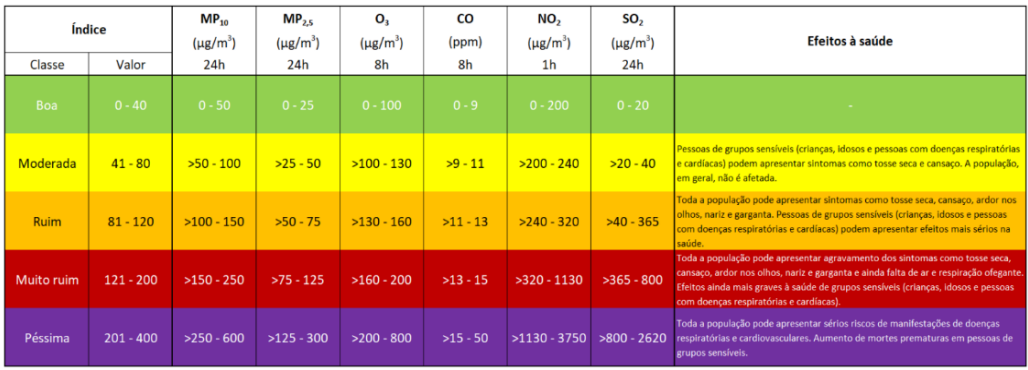


Fonte: elaborado pelo grupo

Para a inserção de dados nas dimensões DI\_POLUENTE e DI\_ESTACAO foi criado um script puxando as informações da própria stage.

Para a dimensão de DATA, foi utilizado outras 3 dimensões (DI\_MES, DI\_ANO e DI\_DIA) e depois feito um select para fazer o plano cartesiano dessas informações.

Já para a dimensão DI\_IQAR, foi criado uma tabela de acordo com o que o site IEMA descreveu sobre os índices e seus valores e depois, importado via pipeline do blob para o banco de dados.



Fonte: IEMA

Pensando-se em um modelo voltado para a agilidade e praticidade nas consultas da massiva quantidade de dados recebidos pelos dispotivos IoT ao longo do tempo, sem a necessidade de executar diversos “joins”, como normalmente ocorre nos modelos relacionais (cardinalidades e normalização dos dados), foi utilizada a modelagem dimensional Star Schema, como forma de otimizar os processos do Data Warehouse do sistema “Respire Qualidade”.

O modelo Star Schema (esquema de estrela) se dá devido à disposição em que se encontram as tabelas, sendo a tabela fato centralizada relacionando-se com diversas outras tabelas de dimensão. Nesse modelo os dados são desnormalizados para evitar joins entre tabelas, diminuindo o tempo de consultas, no entanto devido a repetição de dados, utiliza mais espaço em disco. A vantagem desse modelo é a eficiência na extração de dados, o que é um grande diferencial em se tratando de um data warehouse.

O Star Schema é o modelo mais difundido e utilizado na criação de um Data Warehouse (DW). Este foi um modelo proposto por Ralph Kimball com o objetivo de simplificar a visualização dimensional, facilitando a distinção entre as dimensões e aos fatos.

O fato de Monitoramento representa/armazena as métricas (algo que pode ser medido ou quantificado) resultantes de um evento do processo de negócio (monitoramento periódico dos índices de qualidade do ar por cada estação, identificando pressão, umidade, temperatura, nível de poluente, etc). Já as dimensões representam os contextos para análise de um fato, proporcionando diferentes perspectivas de análise para o usuário e normalmente interpretadas como os “filtros possíveis” para determinada tabela fato.

**Bibliografia:**

<https://energiaeambiente.org.br/qualidadedoar>

<https://www.astera.com/pt/type/blog/dimensional-modeling-guide/>

<https://medium.com/@aasouzaconsult/modelagem-multidimensional-e65f02bbd60>