

Desafios e Soluções na construção do DWH

Igor Telles, Sylvio Mello, Maria Luiza C. Wuillaume, Felipe Vilela

10/06/2024

1 Introdução

A integração correta dos modelos de dados foi um dos principais desafios enfrentados durante o desenvolvimento do nosso Data Warehouse (DWH). Diferentes abordagens foram adotadas por outros modelos, como os de Siqueira e Antonny, que divergiam em vários aspectos do modelo planejado pelo nosso grupo, Syma.

2 Modelo de Dados do DWH

Nossa primeira etapa foi construir o modelo de dados do DWH com o objetivo de preservar o máximo de colunas relevantes para análises e tomadas de decisão. Assim, incluímos no DWH algumas colunas que não estavam presentes no modelo de dados original da nossa empresa, mas que eram utilizadas em outras empresas e consideradas relevantes.

3 Processo de ETL

O processo de Extração, Transformação e Carga (ETL) foi inteiramente implementado em Python, utilizando o sistema de arquivos como área de staging. O esquema seguido foi o seguinte:

3.1 Extração

Cada empresa possui um modelo de dados distinto, armazenado em diferentes bancos de dados. Os dados dessas empresas são extraídos e convertidos para o formato CSV.

3.2 Transformação

Um script Python realiza a transformação dos dados extraídos para o formato exigido pelas tabelas de fatos e dimensões do DWH.

3.3 Carga

O mesmo script Python utilizado na transformação também realiza a carga dos dados transformados no DWH.

4 Conclusão

Integrar corretamente os modelos de dados de diferentes fontes foi um desafio significativo, mas ao adotar uma abordagem que considerasse a relevância das colunas para análises e decisões, conseguimos desenvolver um modelo de dados robusto para o nosso DWH. O uso de Python para o processo de ETL permitiu uma maior flexibilidade e controle sobre as etapas de extração, transformação e carga.