

Integração e Fluxo de Dados (ETL)



Autoria do Desafio Profissional: Thiago Salhab Alves

Leitora Crítica: Cecilia Sosa Arias Peixoto

► Proposta de resolução

Para a resolução do Desafio Profissional, os seguintes passos devem ser atendidos:

Solução:

- Realize o processo de extração de dados por meio do mapeamento de dados lógicos com os dados fonte e os dados de destino, colocando o nome da tabela, nome da coluna, tipo de dados e nome do banco;
- Realize o processo de limpeza de dados utilizando algumas técnicas de preenchimento de dados, tais como descarte de toda a tupla, preenchimento manual do valor faltante, uso de uma constante global para preencher os valores faltantes, média do atributo com valor a ser substituído. No caso da remoção de ruídos, pode-se aplicar a compartimentalização (*binning*), utilizando a suavização pela mediana ou pelos valores de fronteira. É possível também utilizar regressão linear e não linear para a remoção de ruídos. Para se eliminarem *outliers*, podem-se utilizar as técnicas de agrupamento (*clustering*);
- Realize o processo de conformação de dados, combinando os dados de diferentes fontes em uma base de dados única e coerente. Para isso, podem-se utilizar técnicas de redução de dados, tais como seleção de atributos relevantes, redução de dimensão, redução de número e discretização de atributos. Deve-se criar uma tabela de metadados que suportem dados em conformidade;
- Realize o processo de carregamento de dados preparando uma estrutura de tabela dimensional, contendo o mínimo de componentes.

The background features a complex geometric pattern. It includes large, overlapping triangles in shades of blue and grey, creating a low-poly effect. A prominent yellow circle is located in the lower-left quadrant, partially overlapping a smaller blue circle. Diagonal bands of yellow and blue run across the top and bottom right corners. The overall aesthetic is modern and minimalist.

Bons estudos!