

# **Análise de Dados com Base em Processamento Massivo em Paralelo Modelagem Conceitual de ETL/ELT: Respostas dos Exercícios**

**Profa. Dra. Cristina Dutra de Aguiar Ciferri**

**André Perez**

**Guilherme Muzzi da Rocha**

**Jadson José Monteiro Oliveira**

**João Pedro de Carvalho Castro**

**Leonardo Mauro Pereira Moraes**

**Piero Lima Capelo**

## **Observação:**

Recomenda-se fortemente que a lista de exercícios seja respondida antes de se consultar as respostas dos exercícios.

1. A anuência de que o processo de ETL/ELT é a etapa mais custosa de todo o projeto de *data warehousing* já é algo consolidado tanto na literatura quanto no mercado. Sendo assim, é importante modelá-lo conceitualmente a fim de contribuir para diminuir o esforço dos projetistas e desenvolvedores durante a implementação do *workflow*. Além disso, o esquema conceitual é um recurso precioso para a documentação das decisões tomadas na construção do processo de ETL/ELT, para a análise de impacto das alterações necessárias para atendimento de demandas que ocorrem no ciclo de vida do *data warehouse* (tais como alterações nas fontes de dados, evolução dos requisitos ou das regras de negócio, necessidade de melhoria no desempenho das consultas, correção de erros cometidos durante a fase de projeto, entre outros) e para facilitar a exploração de cenários alternativos para a solução desejada.

## 2. Descrição sucinta das categorias.

- (a) Operadores de armazenamento: Os operadores de armazenamento podem ser usados para representar áreas de armazenamento de dados, tais como repositórios, arquivos ou bases de dados.
- (b) Operadores de manipulação de dados: Os operadores de manipulação de dados são usados para representar as tarefas de transformação e de limpeza que são aplicadas aos dados extraídos das diversas fontes para torná-los compatíveis com a estrutura proposta para o *data warehouse*.
- (c) Operadores de inicialização: Os operadores de inicialização de dados servem para representar a atribuição de valores específicos para atributos de um conjunto de dados.
- (d) Operadores de agregação: Os operadores de agregação podem ser usados para representar funções que, quando aplicadas a um conjunto de dados, processam os valores de um atributo específico e retornam um único valor como resultado. Se forem definidos atributos de agrupamento, as funções processam os valores de um atributo e retornam um único valor para cada atributo de agrupamento.
- (e) Operadores de fluxo: Os operadores de fluxo de dados permitem representar uma alteração no fluxo dos dados no *workflow* de ETL, sem impactar esses dados.
- (f) Operadores especiais: Os operadores especiais representam operações que envolvem especificidades, complementando as funcionalidades dos demais operadores

## 3. Classificação dos operadores.

- (a) Operadores de armazenamento: DataLake, DataMart, FailDataSet, DataSet.
- (b) Operadores de manipulação de dados: Union, Update, Intersect, Sort, Split, Filter.
- (c) Operadores de inicialização: SetNullAsDefault, Sequence.
- (d) Operadores de agregação: Sum, AvgGroup.
- (e) Operadores de fluxo: Sincronize, Fail, Junction.
- (f) Operadores especiais: Function.



4. O diagrama de resposta está ilustrado na Figura 1.

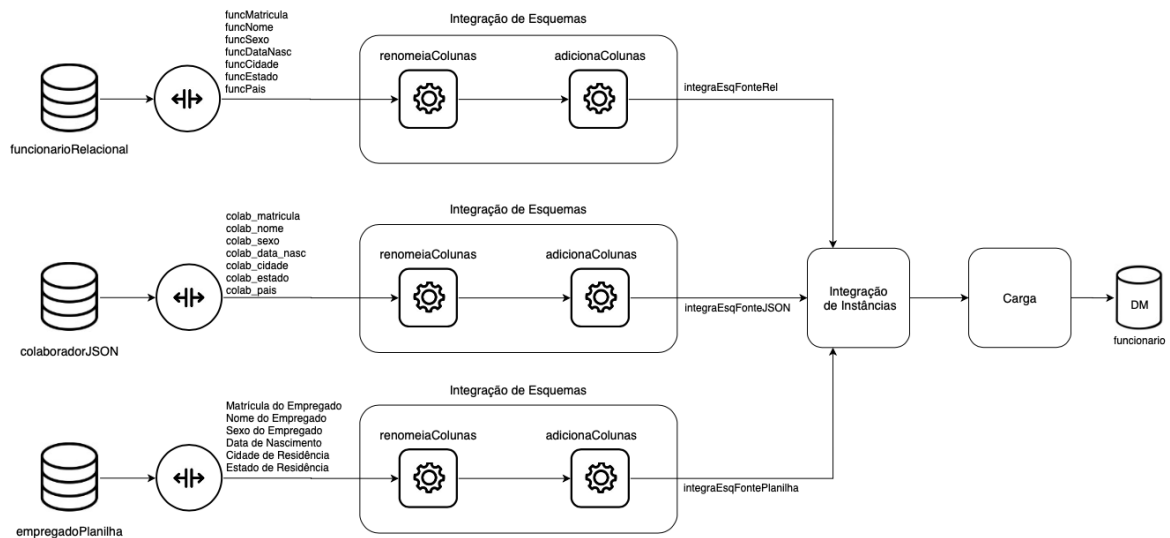


Figura 1: Resposta da questão 4.

5. Os diagramas de resposta estão ilustrados nas Figuras 2 e 3.

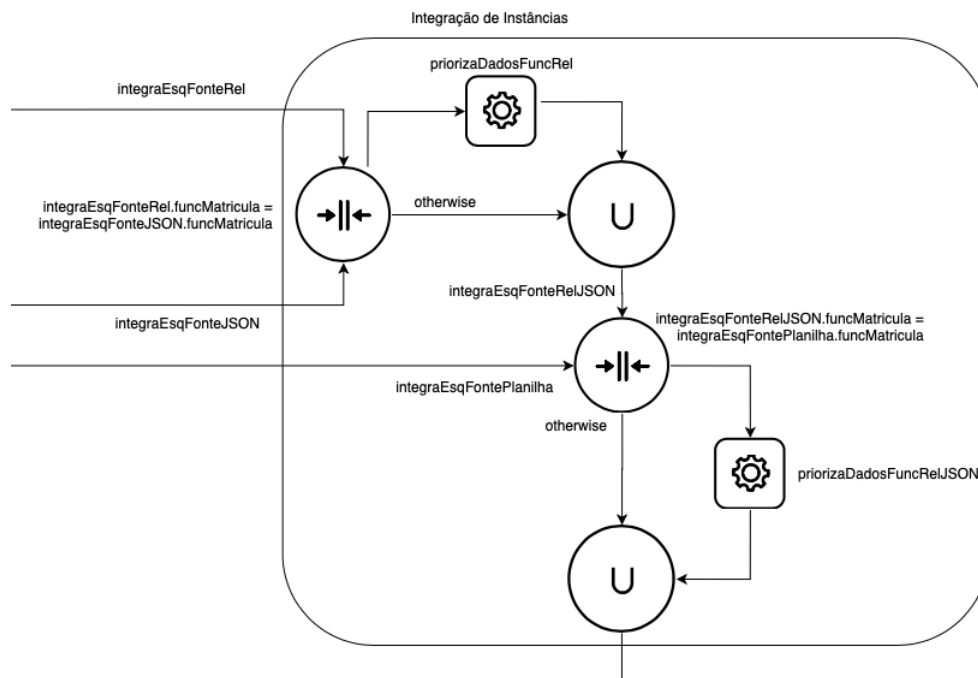


Figura 2: Resposta da primeira parte da questão 5.

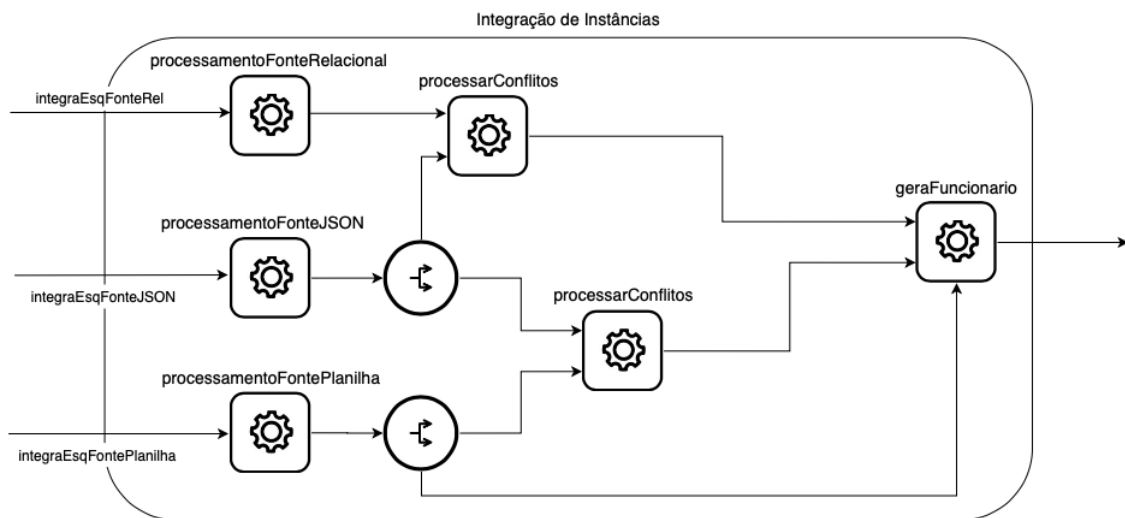


Figura 3: Resposta da segunda parte da questão 5.

6. O diagrama de resposta está ilustrado na Figura 4.

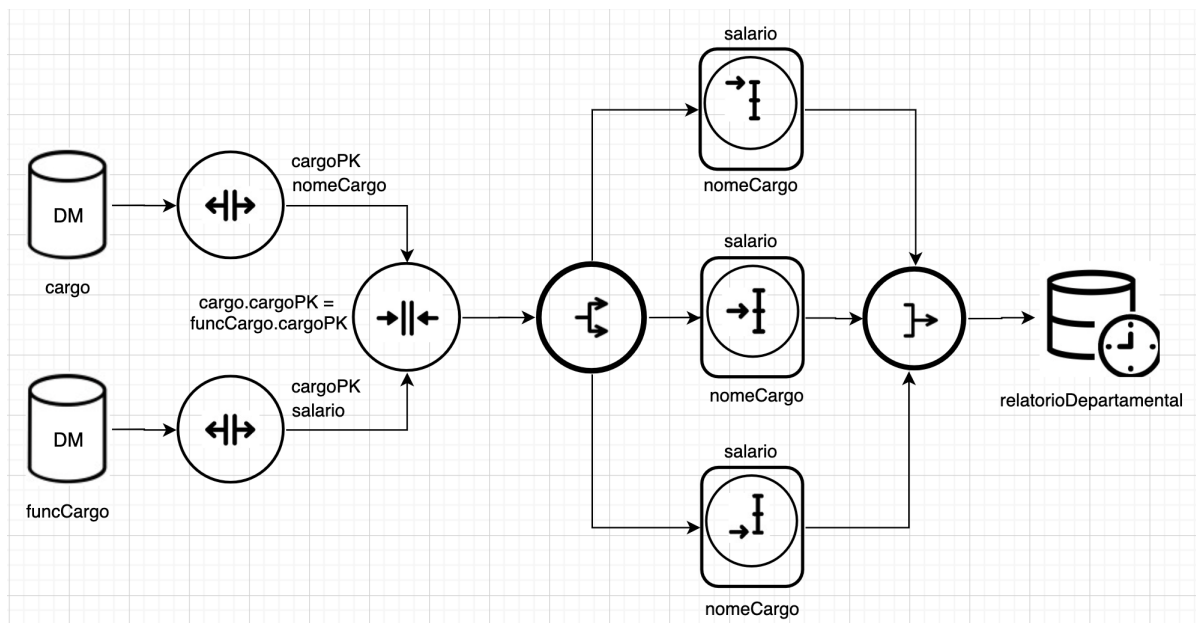


Figura 4: Resposta da questão 6.

7. Questão elaborada para gerar discussões nas tutorias. Não há uma única resposta correta.