

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

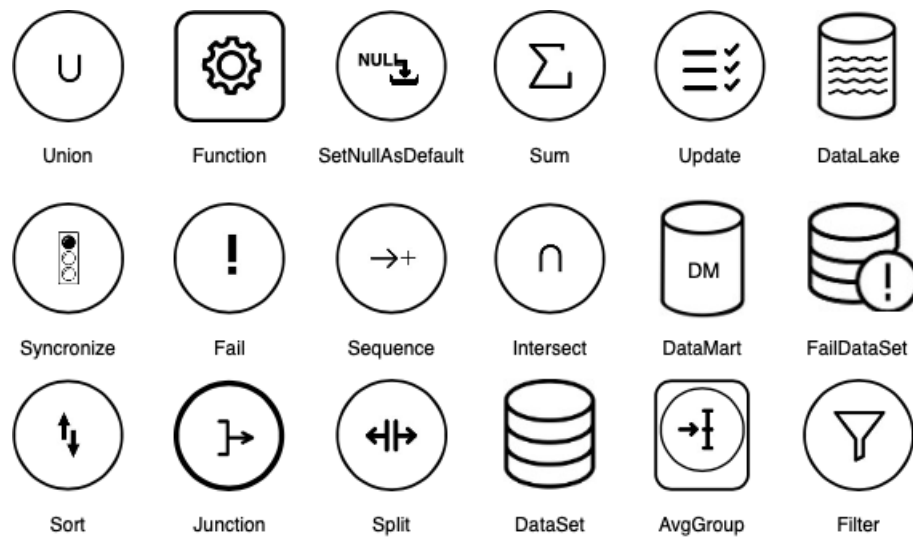
**Observação:**

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

1. Descreva qual a importância de se modelar conceitualmente um *workflow* de ETL/ELT antes de implementá-lo.
2. Considere as seguintes categorias de operadores:
  - (a) Operadores de armazenamento;
  - (b) Operadores de manipulação de dados;
  - (c) Operadores de inicialização;
  - (d) Operadores de agregação;
  - (e) Operadores de fluxo;
  - (f) Operadores especiais.

Descreva, de forma sucinta, o objetivo de cada uma das categorias supracitadas.

3. Considere os operadores ilustrados na Figura 1. Classifique-os de acordo com sua respectiva categoria de operadores.



**Figura 1:** Alguns operadores do Modelo Intuitivo.

- (a) Operadores de armazenamento:
- (b) Operadores de manipulação de dados:
- (c) Operadores de inicialização:
- (d) Operadores de agregação:
- (e) Operadores de fluxo:
- (f) Operadores especiais:

4. Considere o *workflow* de ETL modelado na Figura 2, o qual ilustra a extração de funcionários de duas bases de origem: (i) *funcRelacional*, a qual representa um sistema gerenciador de banco de dados relacional; e (ii) *colaboradorJSON*, a qual representa uma coleção de documentos JSON.

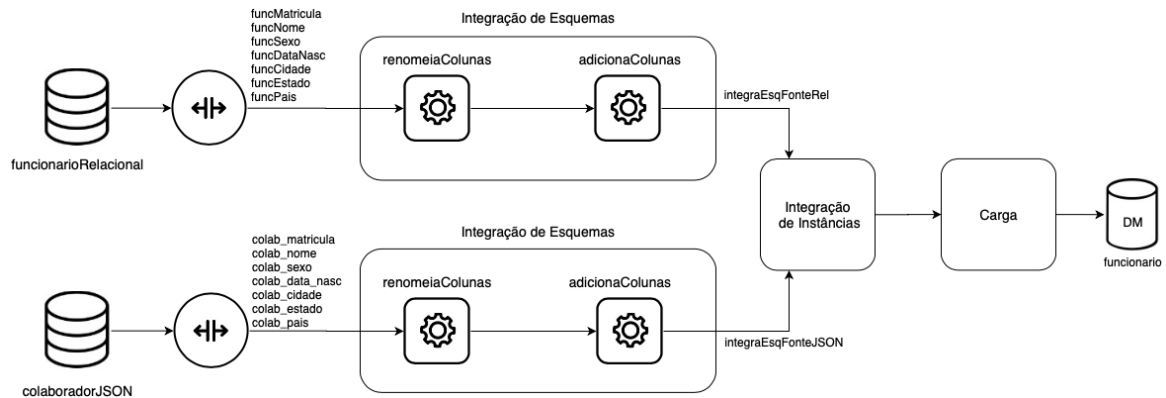


Figura 2: Visão geral do processo de ETL da BI Solutions.

Considere que a empresa **BI Solutions**, responsável pela manutenção do *workflow* ilustrado na Figura 2, necessita de incluir mais uma fonte de dados no processo de ETL. Essa nova fonte de dados, denominada *empregadoPlanilha*, representa uma planilha Excel que contém os seguintes dados de funcionários: “Matrícula do Empregado”, “Nome do Empregado”, “Sexo do Empregado”, “Data de Nascimento”, “Cidade de Residência”, “Estado de Residência”. Estenda o *workflow* de ETL para incluir essa nova fonte de dados. É necessário modelar apenas as etapas anteriores ao subfluxo de integração de instâncias.

5. Considere o subfluxo relacionado à “Integração de Instâncias” da **BI Solutions**, representado tanto no diagrama conceitual da Figura 3 quanto no *workflow* da Figura 4. Esse subfluxo considera como entradas dados oriundos das fontes de dados funcionarioRelacional e colaboradorJSON. Estenda o diagrama conceitual e o *workflow* de forma que o subfluxo relacionado à Integração de Instâncias também considere como entrada os dados oriundos da fonte de dados empregadoPlanilha.

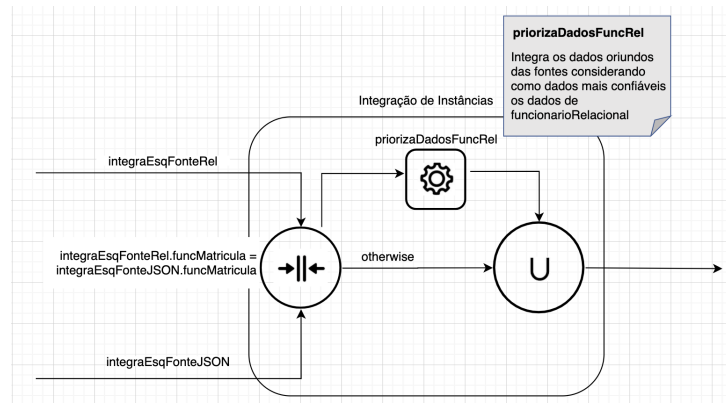


Figura 3: Diagrama conceitual para o subfluxo de “Integração de Instâncias” da BI Solutions.

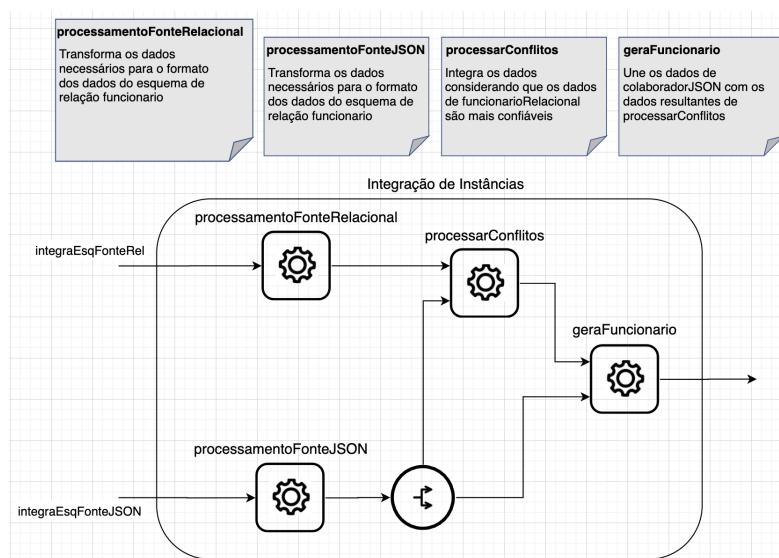


Figura 4: Workflow para o subfluxo de “Integração de Instâncias” da BI Solutions.

6. Considere um *data mart* implementado segundo o modelo relacional que armazena três conjuntos de dados, conforme descrito nas relações seguir:

(i) funcionario (funcPK, funcNome, funcEndereco, funcDataNasc)

(ii) cargo (cargoPK, nomeCargo, descricaoCargo)

(iii) funcCargo (funcPK, cargoPK, salario)

Faça um diagrama conceitual que considere desde a extração dos dados dessas relações até a geração de um relatório departamental que exibe, para cada cargo, o nome do cargo, o maior salário, o menor salário e o salário médio. Note que deve ser feito um diagrama conceitual, ou seja, não é para se fazer um *workflow* que represente a implementação.

7. Considere o exemplo do *data mart* apresentado nas aulas da disciplina, referente à folha de pagamento da empresa **BI Solutions**. Nele, são considerados dados de funcionários, datas, cargos e departamentos. Escolha uma dessas perspectivas e desenvolva um modelo completo para seu *workflow* de ETL (exceto a perspectiva de funcionários, visto que esta já foi modelada nas aulas). Considere diferentes fontes de dados e englobe as etapas de integração de esquemas, integração de instâncias e carga em seu modelo.

