

Análise de Dados com Base em Processamento Massivo em Paralelo

Lista de Exercícios: Modelagem Conceitual de ETL/ELT

Profa. Dra. Cristina Dutra de Aguiar

Observação:

Esta lista contém exercícios classificados como essenciais e complementares. A indicação da classificação de cada exercício é feita junto de sua definição. Recomenda-se fortemente que a lista de exercícios seja respondida antes de se consultar as respostas dos exercícios.

1. (Essencial) Descreva qual a importância de se modelar conceitualmente um *workflow* de ETL/ELT antes de implementá-lo.
2. (Essencial) Por que é interessante o uso de um modelo específico para projetar o processo de ETL/ELT (como o Modelo Intuitivo) e não o uso de um modelo de processos genérico (como o Modelo BPMN - *Business Process Model and Notation*)?
3. (Essencial) 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.

.

4. (Essencial) Considere o *workflow* de ETL modelado na Figura 1, o qual ilustra a extração de funcionários de duas bases de origem: (i) *funcionarioRelacional*, a qual representa um sistema gerenciador de banco de dados relacional; e (ii) *colaboradorJSON*, a qual representa uma coleção de documentos JSON.

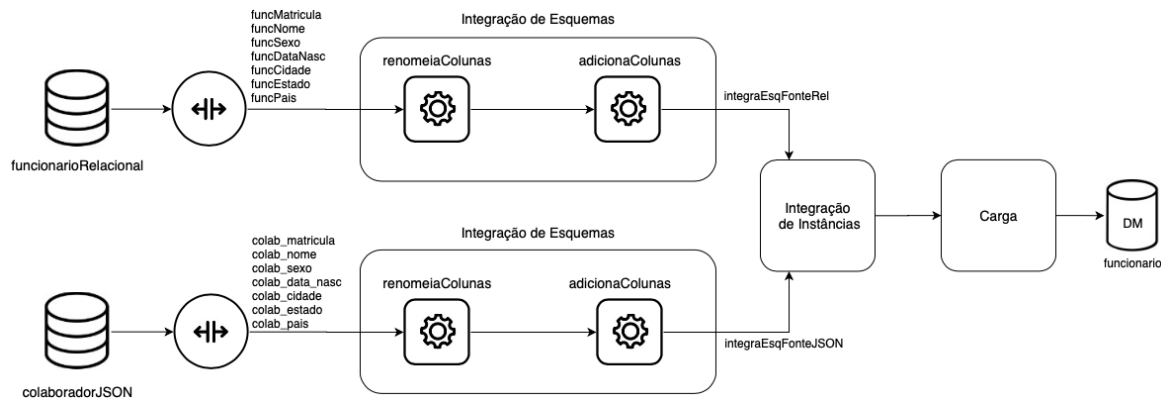


Figura 1: Visão geral do processo de ETL da **BI Solutions**.

Considere que a empresa **BI Solutions**, responsável pela manutenção do *workflow* ilustrado na Figura 1, necessita 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. Modele apenas as etapas anteriores ao subfluxo de integração de instâncias.

5. (Essencial) Considere o subfluxo relacionado à “Integração de Instâncias” da **BI Solutions**, representado tanto no diagrama conceitual da Figura 2 quanto no *workflow* da Figura 3. Esse subfluxo considera como entradas dados oriundos das fontes de dados funcionárioRelacional 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.

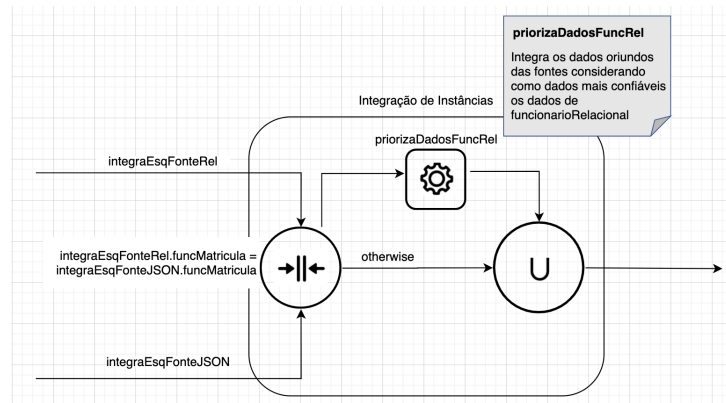


Figura 2: Diagrama conceitual para o subfluxo de “Integração de Instâncias” da **BI Solutions**.

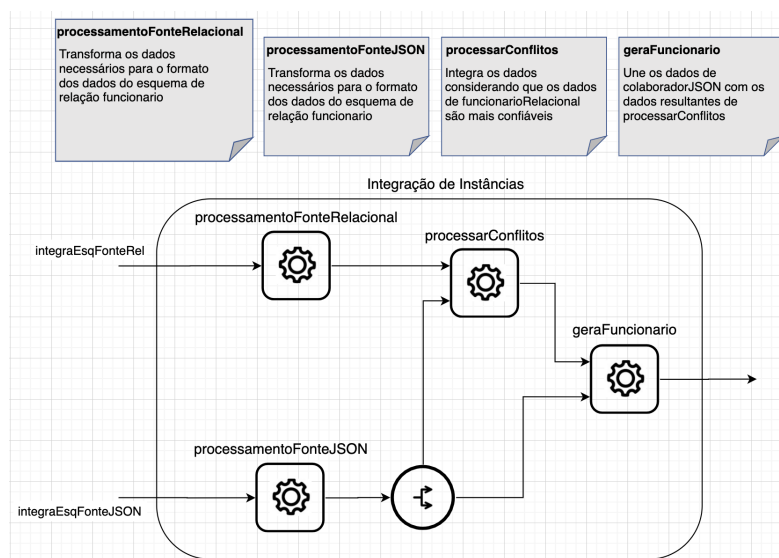


Figura 3: *Workflow* para o subfluxo de “Integração de Instâncias” da **BI Solutions**.

6. (Complementar) 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.

