



Centro Universitário de Brasília (CEUB)

Faculdade de Tecnologia e Ciências Sociais Aplicadas (FATECS)

Thales Rassi Porto de Matos - 22400186

Gabriel Marques da Rocha - 22451254

Gabrielle Gutierrez - 22350026

Pedro Klein - 22105154

Matheus de Moraes - 22352763

Henrique Lessa - 22402204

Documentação ETL – UF

Brasília

2025

Thales Rassi Porto de Matos

Gabriel Marques da Rocha

Gabrielle Gutierres

Pedro Klein

Matheus de Moraes

Henrique Lessa

Documentação ETL – UF

Atividade final apresentada à Faculdade de
Tecnologia e Ciências Sociais Aplicadas
(FATECS)

, do Centro Universitário de Brasília
(CEUB) como parte integrante do currículo
da disciplina Interação Humano
Computador, da graduação em Ciência da
computação

Professora responsável: Kadidja Valeria
Reginaldo de Oliveira

Brasília

2025

SUMÁRIO

IDENTIFICAÇÃO	04
OBJETIVO DO ETL	04
FONTES DE DADOS DE ENTRADA	04
TABELAS DE DESTINO IMPACTADAS	05
FLUXO DO ETL	05
MAPEAMENTO DE CAMPOS (ORIGEM → DESTINO)	07
DEPENDÊNCIAS E PRÉ-REQUISITOS	07
COMO EXECUTAR	07

IDENTIFICAÇÃO

Script: ETL_UF.py

Responsável(is): Equipe Radiologia DF

OBJETIVO DO ETL

Inserir automaticamente na tabela `unidade_da_federacao` todas as Unidades da Federação do Brasil, garantindo que a base de dados contenha sigla e nome de cada estado e do Distrito Federal.

Este ETL evita que inserções manuais sejam necessárias e garante que todos os demais ETLs possam referenciar corretamente o `id_uf` por meio de chave estrangeira.

FONTES DE DADOS DE ENTRADA

Arquivo principal:

1. Este ETL não utiliza arquivos externos. As UFs são definidas diretamente no script, na lista constante:

```
UFS = [  
    ("UF 1", "Sigla 1"),  
    ...  
    ("UF 27", "Sigla 27"),  
]
```

TABELAS DE DESTINO IMPACTADAS

1. `unidade_da_federacao`
 - Tabela contendo os nomes das unidades da federação

FLUXO RESUMIDO DO ETL

Passo 1 – Coleta das UFs existentes no banco

- Abre conexão com o banco.
- Executa: `SELECT sigla FROM unidade_da_federacao;`
- Cria um conjunto com todas as siglas já cadastradas.

Passo 2 – Identificação das UFs faltantes

- Para cada item da lista UFS:
 - Se a sigla não estiver na tabela:
 - Adiciona uma tupla (nome, sigla) na lista de inserções.

Passo 3 – Inserção das UFs

- Caso existam registros faltantes:
 - Insere em lote com `execute_values`:
 - `INSERT INTO unidade_da_federacao (nome, sigla) VALUES %s`
 - Realiza commit
 - Imprime no console quantas UFs foram inseridas
- Caso nenhuma seja nova:
 - Imprime “Nenhuma UF nova para inserir.”

Passo 4 – Encerramento

- Fecha o cursor
- Fecha a conexão com o banco
- Imprime mensagem final confirmando que o ETL foi concluído

MAPEAMENTO DE CAMPOS (ORIGEM → DESTINO):

Origem no CSV	Destino
nome (ex.: "Acre")	nome
sigla (ex.: "AC")	sigla

DEPENDÊNCIAS E PRÉ-REQUISITOS

Bibliotecas Python:

- pandas
- psycopg2
- psycopg2.extras
- config_db (função get_conn para conexão com o banco)

Tabelas:

- unidade_da_federacao (já criada no banco)

COMO EXECUTAR

- Garantir que o ambiente virtual (se houver) esteja ativado.
- Garantir que o arquivo .env esteja configurado com os parâmetros de conexão ao banco.
- Executar: python3 ETL_UF.py

Pré-condições:

- Banco PostgreSQL acessível
- Tabela unidade_da_federacao criada