

**Curso: Gestão da Tecnologia da Informação**

**Professor: Vander Ribeiro**

**Nomes dos Integrantes: Maurício de Jesus | Gabriela Augusto | Gabriel de Souza |  
Guilherme Prado | Renan Macorin | Denilson Maranhão**

**Trabalho Prático: AAP 03**

---

**Baureri – São Paulo**

**2025**

## Documentação de Processos — Projeto DADOS\_RFB

Instituição: Fatec Barueri

Nome do Banco de Dados: DADOS\_RFB

### 1. Visão Geral do Projeto:

O projeto **DADOS\_RFB** tem como objetivo coletar, tratar, organizar e armazenar localmente os dados públicos de cadastro de empresas disponibilizados pela Receita Federal do Brasil (**RFB**). Esses dados permitem a realização de análises, consultas e estudos em um banco de dados relacional.

O projeto utiliza um processo de **ETL (Extract, Transform, Load)**, que compreende as etapas de extração dos dados, tratamento e carregamento no banco de dados.

O banco de dados foi nomeado como **DADOS\_RFB** e possui estrutura relacional com várias tabelas interligadas por chaves primárias e estrangeiras, sendo a principal delas o campo `cnpj_basico`.

### 2. Fluxo de Processos (ETL):

Etapa:	Descrição:
• Extração (Extract)	• Download dos arquivos públicos disponibilizados pela Receita Federal.
• Descompactação	• Extração dos arquivos compactados.

- Leitura e Parsing
  - Leitura e interpretação dos dados brutos.
- Transformação (Transform)
  - Limpeza, padronização e normalização dos dados.
- Carregamento (Load)
  - Inserção dos dados tratados no banco **DADOS\_RFB**.
- Validação
  - Verificação da integridade e consistência dos dados.
- Documentação do Modelo
  - Manutenção do diagrama **ERD** e do esquema de dados.
- Atualizações
  - Planejamento de cargas futuras de atualização de dados.

### **3. Pré-requisitos e Infraestrutura:**

- Banco de dados relacional (**MariaDB/MySQL**).
- Ambiente Python instalado.
- **Arquivo .env** configurado com as variáveis de conexão.
- Espaço de armazenamento adequado para grandes volumes de dados.

#### **4. Instruções de Uso:**

1. Criar o banco de dados com o nome **DADOS\_RFB**.
2. Configurar o **arquivo .env** com as informações de conexão.
3. Instalar as dependências do projeto.
4. Executar o script principal de **ETL**.
5. Aguardar a finalização do processo.
6. Validar os dados carregados no banco.

#### **5. Modelo de Dados:**

O projeto possui um modelo relacional estruturado com as seguintes principais tabelas:

- empresa
- estabelecimento
- socios
- simples

Além de tabelas auxiliares como **CNAE**, qualificação, natureza jurídica, municípios, países, entre outras.

A coluna `cnpj_basico` é a principal chave de relacionamento entre as tabelas.

## 6. Considerações Finais:

O projeto **DADOS\_RFB** permite o armazenamento e análise de uma grande massa de dados públicos da Receita Federal. A correta configuração do ambiente e a execução adequada do processo **ETL** garantem a integridade e confiabilidade das informações.

Recomenda-se manter registros de logs, backups do banco e atualização periódica quando houver novos dados disponíveis.