

Documentação do Projeto: Pipeline Financeiro dos Dados das Cooperativas de Crédito

1. Objetivo

O objetivo deste projeto é desenvolver uma solução completa que, a partir de um arquivo CSV, modela um banco de dados normalizado, cria um processo de ETL para automatizar a importação do CSV, conecta o Power BI a esse banco de dados e cria um painel com indicadores para visualização dos dados.

2. Detalhamento

2.1 Obtenção de Dados

Fonte: Banco Central do Brasil

Descrição: Utilizar o arquivo Balancetes das Cooperativas de crédito.

Período Base: Arquivo baixado em 09/2024 (dados de 07/2023 a 06/2024)

2.2 Modelagem do Banco de Dados

Requisitos:

- Banco de Dados normalizado contendo mais de uma tabela.
- Utilizar os últimos 12 arquivos disponibilizados pelo Banco Central.
- Utilizar SQL Server como SGBD.

2.3 Criação da ETL

Objetivo: Automatizar a carga dos dados do CSV para o banco de dados.

Tecnologia: Python

Etapas:

- Extração: Ler dados do arquivo CSV.
- Transformação: Limpeza e formatação dos dados para adequação ao banco de dados.
- Carga: Inserir os dados transformados no banco de dados SQL Server.

2.4 Conexão com Power BI e Criação do Painel

Objetivo: Visualizar os dados através de um painel no Power BI.

Indicadores:

- Saldo total dos ativos e passivos das cooperativas no período.
- Principais fontes de receita das cooperativas no trimestre/ano.
- Principais despesas enfrentados pelas cooperativas durante o período.
- Situação atual do capital social das cooperativas.

Componentes do Painel:

Filtros

Indicadores

Card Visuais

2.5 Boas Práticas

- Verificação de Fonte de Dados: Dados obtidos do Banco Central estão completos e sem erros.
- Normalização: Técnicas de normalização para evitar redundâncias e melhorar a integridade dos dados.
- Manejo de Erros: Mecanismos para capturar e tratar erros durante o processo de ETL.
- Documentação do Esquema: Esquema do banco de dados, incluindo tabelas, colunas e relacionamentos.
- Facilidade de Uso: Garanta que o painel seja fácil de usar e interpretar pelos usuários finais.

3. Tecnologias Utilizadas

- Python: Para a criação do processo de ETL.
- SQL Server: Para armazenamento dos dados.
- Oracle VM Virtual-Box: Máquina virtual armazenando a infraestrutura do SQL Server.
- Power BI: Para visualização e criação de painéis interativos.

4. Estrutura do Banco de Dados

4.1 Tabelas

tbl_Balanco: Contém dados dos balancetes e balanços patrimoniais.

tbl_Contas: Detalhamento dos Contas dos Balancetes das cooperativas.

tbl_Cooperativas: Detalhamento das cooperativas.

5. Fluxo do Processo de ETL

Extração: Carregar dados do CSV.

Transformação: Processar e limpar dados, remover duplicatas, tratar valores ausentes.

Carga: Inserir dados no banco de dados SQL Server.

6. Painel no Power BI

6.1 Visualizações

Indicadores

Card Visuais: Exibir valores totais de indicadores importantes como capital social.

6.2 Filtros

Filtro por Data: Permite seleção de períodos específicos.

7. Resultados Esperados

Painel Interativo: Para análise detalhada das finanças das cooperativas de crédito.

Automação do Processo: Redução do tempo de importação e processamento dos dados.

Visualizações Claras: Facilitar a tomada de decisão com base nos indicadores fornecidos.