



# BIG DATA

Projeto Cadastro Ambiental Rural

# ARQUITETURA

Camadas e  
tratamentos

Como  
estruturamos

Pontos de  
Dificuldade



# COMO ESTRUTURAMOS



- Separação dos processos em notebooks
- Criação de Schemas
- Arquitetura Medallion (Camadas - Bronze, Silver, Gold)
- Raw e Missing Zones
- Particionamento
- Tabelas Externas
- SQL Profiles

# CAMADAS E TRATAMENTOS

- Camadas da Arquitetura Medallion
  - Raw Zone
    - Dados importados como strings.
  - Bronze Zone
    - Tipagem e preparação inicial.
  - Silver Zone
    - Validação e qualidade dos dados.
  - Gold Zone
    - Análise e geração de insights.



# CAMADAS E TRATAMENTOS

- **Tratamentos Aplicados**
  - **Pré-Processamento de Dados**
    - Divisão da base de dados original em partes menores para facilitar o upload e o processamento.
  - **Validação e Qualidade**
    - Garantia de integridade e precisão dos dados nas diferentes camadas.
  - **Otimização de Consultas**
    - Estruturação dos dados para facilitar e acelerar as consultas e análises.





Tamanho do arquivo  
CSV

Execução parcial do  
SQL Profiler

Obstáculos no  
particionamento

**PONTOS DE DIFICULDADE**

**OBRIGADO!**