



# *Big-data:* Processamento de Dados Eólicos G7

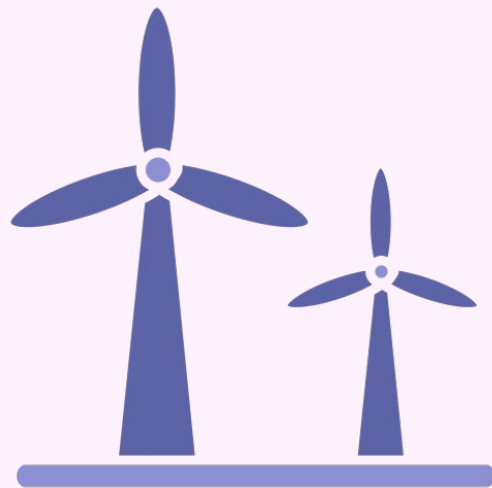
Bruno Crespo Ferreira  
Fellipe Silvério Brandão  
Gustavo Biaso Dias Pinto  
Lara Ramos Linhares  
Marco Antônio Magalhães

# Descrição do Projeto

Este projeto implementa uma solução completa de big-data para análise de constrained-off (restrições de geração) em usinas eólicas brasileiras, utilizando dados consolidados do ONS (Operador Nacional do Sistema Elétrico).

## Objetivos:

- **Processamento** dos dados consolidados via lotes.
- **Agregações temporais e espaciais** com conversão para formato Parquet.
- **Deteção de padrões e anomalias** nos eventos de constrained-off.
- **Visualização** interativa das análises.
- **Avaliação** da arquitetura proposta.



# Estrutura dos Dados

## 1. Arquivo Principal:

### **RESTRICAO\_COFF\_EOLICA\_2021\_10**

- Geração, disponibilidade, restrições por usina
- Dados consolidados de todos os meses
- Granularidade: 30 minutos

## 2. Arquivo de Detalhamento:

### **RESTRICAO\_COFF\_EOLICA\_DETAIL\_2021\_10**

- Dados de vento, geração estimada vs verificada
- Informações por unidade geradora

## 3. Período de Análise:

- Início: Outubro/2021
- Fim: Abril/2025
- Total: 42 meses de dados
- Volume: Dados consolidados de constrained-off eólico

# Apresentação dos Dados

Column Name	Description
id_subsistema	Subsystem id (N, NE, S)
nom_subsistema	System name (NORTE, NORDESTE, SUL)
id_estado	State acronym
nom_estado	State name
nom_usina	Name of the Plant or Set of Plants
id_ons	Identifier of the Plant or Group of Plants in the ONS
ceg	Unique Generation Project Code (CEG), established by ANEEL
din_instante	Datetime (every 30min)
val_geracao	Generation Value, in MWmed
val_geracaolimitada	Limited Generation Value by any restriction, in MWmed
val_disponibilidade	Verified Disponibility Value in real time, in MWmed
val_geracaoreferencia	Referenced Generation Value (or Estimated), in MWmed
val_geracaoreferenciafinal	Final Referenced Generation Value, in MWmed
cod_razaorestricao	Restriction Reason Code (REL, CNF, ENE, PAR)
cod_origemrestricao	Restriction Origin Code (LOC, SIS)

# Apresentação dos Dados – Detalhado

Column Name	Description
id_subsistema	Subsystem id (N, S, L, O)
id_estado	State acronym
nom_modalidadeoperacao	Plant Operation Mode
nom_conjuntousina	Name of the Plant Set (only for plants with Type II-C operating mode)
nom_usina	Plant's name
id_ons	Identifier of the Plant or Group of Plants in the ONS
ceg	Unique Generation Project Code (CEG), established by ANEEL
din_instante	Datetime (every 30min)
val_ventovalidado	Verified wind, in m3/s
flg_dadoventoinvalido	Indication of invalid wind data. The data is considered invalid (flag = 1) when there is a failure in the measurement received by the ONS Supervision and Control System for more than 6 minutes within the corresponding half-hour.
val_geracaoestimada	Value of the estimated generation of the plant, in MWmed, obtained from the verified wind data and the wind x power curve or, if the curve is unavailable, from the verified generation history.
val_geracaovalidada	Verified Generation Value, in MWmed

# Arquitetura da Solução

**CSV Consolidados → Processamento Parquet → Agregações → Detecção de Anomalias → Visualização**

## Pré Processamento

- Carregamento de dados consolidados
- Limpeza e normalização
- Conversão para Parquet

## Detecção de Anomalias

- Análise de constrained-off extremo:
- Variações bruscas na geração
- Padrões temporais anômalos
- Clusters espaciais

## Dashboard

- Visualizações em tempo real
- Filtros dinâmicos
- Relatórios automáticos



# Tipos de Anomalias Detectadas

## Constrained-Off Extremo

- Critério: >70% de restrição média
- Detecção: Usinas com perda significativa de geração
- Impacto: Perda de energia renovável

## Variações Bruscas na Geração

- Critério: Z-score >3 na geração
- Detecção: Instabilidade na produção
- Impacto: Instabilidade na rede

## Padrões Temporais Anômalos

- Critério: Horários com >1.5x a média de constrained-off
- Detecção: Padrões recorrentes de restrição
- Impacto: Ineficiência operacional

## Clusters Espaciais

- Critério: Estados com >3 usinas e >30% constrained-off
- Detecção: Problemas regionais de transmissão
- Impacto: Limitações de infraestrutura

# Queries SQL

## Constrained-Off Extremo

```
SELECT
    nom_usina, nom_estado, ano, mes,
    AVG(percentual_constrained) as percentual_medio,
    MAX(percentual_constrained) as percentual_max,
    COUNT(*) as registros
FROM wind_data
WHERE percentual_constrained > 50
GROUP BY nom_usina, nom_estado, ano, mes
HAVING AVG(percentual_constrained) > 70
ORDER BY percentual_medio DESC
```

## Variações Bruscas na Geração

```
WITH geracao_stats AS (
    SELECT
        nom_usina, nom_estado, ano, mes,
        AVG(val_geracao) as geracao_media,
        STDDEV(val_geracao) as geracao_std
    FROM wind_data
    GROUP BY nom_usina, nom_estado, ano, mes
)
SELECT
    w.nom_usina, w.nom_estado, w.din_instante,
    w.val_geracao, gs.geracao_media, gs.geracao_std,
    ABS(w.val_geracao - gs.geracao_media) / gs.geracao_std as z_score
FROM wind_data w
JOIN geracao_stats gs ON w.nom_usina = gs.nom_usina
    AND w.ano = gs.ano AND w.mes = gs.mes
WHERE ABS(w.val_geracao - gs.geracao_media) / gs.geracao_std > 3
ORDER BY z_score DESC
```



# Queries SQL

## Padrões Temporais

```
SELECT
    nom_usina, nom_estado, hora,
    AVG(percentual_constrained) as percentual_medio_hora,
    COUNT(*) as registros_hora
FROM wind_data
GROUP BY nom_usina, nom_estado, hora
HAVING AVG(percentual_constrained) > (
    SELECT AVG(percentual_constrained) * 1.5
    FROM wind_data
    WHERE nom_usina = wind_data.nom_usina
)
ORDER BY percentual_medio_hora DESC
```

## Clusters Espaciais

```
WITH state_anomalies AS (
    SELECT
        nom_estado, ano, mes,
        AVG(percentual_constrained) as percentual_estado,
        COUNT(DISTINCT nom_usina) as num_usinas_afetadas
    FROM wind_data
    WHERE percentual_constrained > 30
    GROUP BY nom_estado, ano, mes
)
SELECT
    nom_estado, ano, mes, percentual_estado, num_usinas_afetadas,
    CASE
        WHEN percentual_estado > 50 AND num_usinas_afetadas > 5 THEN 'CLUSTER_CRITICO'
        WHEN percentual_estado > 30 AND num_usinas_afetadas > 3 THEN 'CLUSTER_MODERADO'
        ELSE 'ISOLADO'
    END as tipo_cluster
FROM state_anomalies
WHERE percentual_estado > 30
ORDER BY percentual_estado DESC, num_usinas_afetadas DESC
```

# Obrigado!

Alguma dúvida?

**CREDITS:** This presentation template was created by [Slidesgo](#), and includes icons by [Flaticon](#), and infographics & images by [Freepik](#)

