

# Projeto final do módulo Big Data!

Desenvolvimento e Avaliação de uma Arquitetura Distribuída para o Cadastro Ambiental Rural

## Objetivo

O objetivo deste trabalho é explorar as capacidades de arquiteturas de bancos de dados distribuídos para lidar com conjuntos de dados "grandes", em particular, o "Cadastro Ambiental Rural". Os alunos devem propor uma nova arquitetura, implementá-la, modelar os dados de acordo com as especificações e projetar um particionamento efetivo.

## Problema

O "Cadastro Ambiental Rural" é um conjunto de dados que contém informações sobre propriedades rurais no Brasil. O desafio é projetar uma arquitetura que possibilite o gerenciamento eficiente desses dados, permitindo consultas e análises rápidas, mesmo diante do tamanho considerável da base de dados.

## Dados

<https://dados.gov.br/dados/conjuntos-dados/cadastro-ambiental-rural1> --> Temas Ambientais

## Entregáveis

Os alunos devem seguir as seguintes etapas:

**Proposição da Nova Arquitetura:** Proponha uma arquitetura de banco de dados distribuído que seja adequada para lidar com o "Cadastro Ambiental Rural". Explique as escolhas de tecnologias, componentes e a lógica de funcionamento da arquitetura.

**Implementação da Arquitetura:** Implemente a arquitetura proposta em um ambiente de teste. Isso envolverá a configuração e integração de sistemas de gerenciamento de banco de dados distribuído, servidores, clusters e outros componentes necessários.

**Modelagem do Novo Banco de Dados:** Modele os dados do "Cadastro Ambiental Rural" de acordo com a arquitetura proposta.

**Projeto de Particionamento Adequado/efetivo:** Projete um esquema de particionamento que permita a distribuição eficiente dos dados.

## Contexto empresarial

Recupere a soma de área (em hectares) para todas as propriedades agrícolas que pertencem ao MS e MT. Ordene os resultados em ordem decrescente.

Filtre todas as propriedades que pertencem a região sudeste.

Calcule quantas propriedades foram cadastradas por ano. Apresente os resultados em ordem cronológica.

Calcule o percentual médio de área remanescente de vegetação nativa em comparação a área total da propriedade

Construa uma consulta que mostre a contagem de propriedades rurais por estado.

Faça a média de área entre todas as propriedades. Calcule quantas propriedades por estado que estão acima da média.