

Projeto: Análise da População Mundial com Big Data em Python

Introdução

A compreensão das dinâmicas populacionais é essencial para planejamento urbano, políticas públicas, saúde e logística. Grande parte dos dados relevantes (censos, mobilidade, satélite, redes sociais, dados econômicos) tem volume e variedade que demandam arquiteturas Big Data. Este projeto propõe utilizar ferramentas de Python para coletar, processar e analisar grandes volumes de dados populacionais, produzindo insights açãoáveis sobre tendências globais de crescimento, migração e distribuição etária.

Justificativa

- Os dados demográficos tradicionais podem ser limitados e esparsos no tempo; fontes alternativas permitem análises mais completas e quase em tempo real.
- Governos e organizações precisam de pipelines escaláveis para planejamento de recursos, políticas de saúde e logística urbana.
- O projeto demonstra habilidades práticas em Big Data e análise de dados em grande escala, utilizando Python, PySpark e visualizações interativas.

Objetivos

Objetivo Geral:

Construir um pipeline de Big Data em Python para coletar, processar e analisar dados populacionais, gerando relatórios e dashboards com insights globais sobre a população.

Objetivos Específicos:

- Coletar dados de diversas fontes (censos, APIs, mobilidade, satélite).
- Processar e limpar os dados utilizando PySpark.
- Realizar agravações e análises de tendências populacionais.
- Gerar visualizações interativas e relatórios resumidos.
- Documentar e validar todo o pipeline de dados.

Resultados Esperados

- Base de dados limpa, estruturada e escalável.
- Dashboards interativos com indicadores populacionais por país, região e faixa etária.
- Relatórios com projeções e tendências populacionais.
- Código modular e documentado, pronto para integração em ambientes de Big Data.

Cronograma de Atividades

Semana	Atividades
1	Definicao do escopo, levantamento de fontes de dados, setup do ambiente de desenvolvimento
2	Coleta inicial de dados (censos, mobilidade, satelite)
3	Implementacao inicial do pipeline de ingestao e limpeza de dados
4	Validacao de qualidade e transformacao de dados
5	Agregacoes e analises preliminares com PySpark
6	Criacao de notebooks para analise exploratoria de dados (EDA)
7	Desenvolvimento inicial do dashboard interativo
8	Integracao do pipeline com dashboards e testes de performance
9	Modelagem de projecao populacional simples
10	Testes finais e documentacao tecnica do projeto
11	Deploy do pipeline e dashboard em ambiente de teste
12	Entrega final: relatorio, apresentacao e demonstracao dos resultados