
Projeto Final de Big Data

Licenciatura em Tecnologias Digitais IA e DSS

Sistema para "Análise de Padrões de Mobilidade Urbana" com MongoDB e PySpark

Objetivo: Desenvolver um sistema de análise de dados para "Análise de Padrões de Mobilidade Urbana", utilizando MongoDB e PySpark, aplicando os conceitos e técnicas abordados na unidade curricular de Big Data.

Parte Prática

1. Configuração do Ambiente:

- Configure um ambiente de desenvolvimento com MongoDB e PySpark.
- Familiarize-se com as ferramentas e as suas interfaces em Python.

2. Recolha e Preparação de Dados:

- Selecione pelo menos dois *datasets* relacionados com mobilidade urbana (por exemplo, dados de tráfego, transportes públicos, utilização de bicicletas partilhadas).
- Importe os *datasets* para o MongoDB.
- Utilize PySpark para realizar a limpeza e preparação inicial dos dados.

3. Análise Exploratória:

- Use PySpark para realizar uma análise exploratória dos dados.
- Identifique padrões básicos e estatísticas descritivas relevantes para a mobilidade urbana.

4. Processamento e Análise Avançada:

- Desenvolva consultas em *MongoDB* para extrair informações específicas.
- Utilize PySpark para realizar análises mais complexas, como:
 - Identificação de zonas de maior congestionamento.
 - Padrões de utilização de transportes públicos.
 - Correlações entre diferentes variáveis (por exemplo, condições meteorológicas e volume de tráfego).

5. Visualização de Dados:

- Crie visualizações significativas dos resultados obtidos utilizando bibliotecas Python como Matplotlib ou Seaborn.
- Prepare um dashboard simples que demonstre as principais descobertas.

6. Aplicação Prática:

- Desenvolva um script Python que integre MongoDB e PySpark para gerar relatórios automáticos ou previsões simples.

Relatório

Elabore um relatório que inclua:

1. Introdução:

- Contextualização do problema de mobilidade urbana.
- Objetivos do projeto.

2. Metodologia:

- Descrição dos *datasets* utilizados.
- Explicação das técnicas de preparação e análise de dados aplicadas.

3. Análise de Resultados:

- Apresentação e interpretação dos principais resultados obtidos.
- Discussão sobre os padrões de mobilidade urbana identificados.

4. Implementação Técnica:

- Visão geral da arquitetura do sistema desenvolvido.
- Desafios encontrados e soluções implementadas.

5. Conclusões e Trabalho Futuro:

- Síntese das principais descobertas.
- Sugestões para melhorias e expansões futuras do projeto.

6. Apêndice Técnico:

- Breve explicação sobre MongoDB e PySpark:
 - Vantagens e casos de uso típicos.
 - Como se complementam no processamento de big data.
- Principais comandos e funções utilizados no projeto.

Entrega Final

1. **Código:** Todos os scripts Python desenvolvidos, incluindo a integração entre MongoDB e PySpark.
2. **Relatório:** Documento detalhando todos os aspectos do projeto, conforme a estrutura acima apresentada.

3. **Apresentação:** Prepare uma apresentação de 5 minutos que demonstre o funcionamento do sistema e destaque as principais descobertas sobre padrões de mobilidade urbana, a ser realizada na discussão individual.

Data de Entrega: 6 de janeiro de 2025

Nota: Este projeto visa avaliar a sua capacidade de trabalhar com ferramentas de big data num contexto prático, aplicando os conhecimentos adquiridos na unidade curricular. Valorize tanto a implementação técnica como a análise e interpretação dos resultados obtidos.