

---

# Projeto Final de Bases de Dados Distribuídas Avançadas

---

## Mestrado em Ciência de Dados

Sistema para "Análise de Padrões de Mobilidade Urbana" com o *Stack Hadoop*

**Objetivo:** Desenvolver um sistema de análise de dados distribuído para "Análise de Padrões de Mobilidade Urbana", utilizando o stack Hadoop, de forma a aplicar os conceitos e técnicas abordados na unidade curricular.

## Parte Prática

### 1. Configuração do Ambiente:

- Utilize uma imagem Docker à sua escolha (ou desenvolva a sua) que inclua Hive e Phoenix.
- Familiarize-se com o ambiente e as ferramentas disponíveis.

### 2. Integração e Preparação de Dados:

- Selecione pelo menos dois *datasets* relacionados com mobilidade urbana (por exemplo, dados de tráfego, transportes públicos, planeamento urbano).
- Importe os *datasets* para o ambiente Hadoop.
- Realize a limpeza e garantia de qualidade dos dados.

### 3. Análise e Processamento:

- Utilize Hive e/ou Phoenix para realizar consultas e análises sobre os dados.
- Identifique padrões relevantes para a mobilidade urbana.

### 4. Visualização e Apresentação:

- Crie visualizações significativas dos resultados obtidos.
- Prepare um *dashboard* ou relatório interativo que demonstre as principais descobertas.

### 5. Produto Final:

- Desenvolva um produto (por exemplo, uma aplicação web simples ou um relatório interativo) que mostre o valor acrescentado da sua análise.
- O produto deve ser capaz de auxiliar na tomada de decisões relacionadas com mobilidade urbana.

## Parte Teórica (Relatório)

Elabore um relatório que aborde os seguintes tópicos para cada camada do stack Hadoop:

1. **HDFS (Hadoop Distributed File System)**
2. **YARN (Yet Another Resource Negotiator)**
3. **MapReduce**
4. **HBase**
5. **Phoenix**
6. **Hive**
7. **Hue**

Para cada componente, inclua:

- Uma breve descrição e a sua utilidade no contexto de bases de dados distribuídas.
- Como seria configurado num ambiente de produção.
- Relação com conceitos teóricos de bases de dados distribuídas (por exemplo, particionamento, replicação, consistência, etc.).

## Entrega Final

1. **Produto:** Demonstração prática do sistema desenvolvido, evidenciando as funcionalidades implementadas e o valor acrescentado para a análise de mobilidade urbana.
2. **Relatório:** Documento *pdf* com explicações aprofundadas tanto a parte prática (decisões tomadas, desafios encontrados, soluções implementadas) como a parte teórica (análise dos componentes do stack *Hadoop*).
3. **Apresentação:** Prepare-se para discutir o projeto, justificar escolhas técnicas e refletir sobre o impacto potencial do sistema desenvolvido. O trabalho é em grupo, mas a discussão é individual. Cada elemento deve ver como suas todas as partes desenvolvidas no projecto.

**Data de Entrega:** 06 de janeiro de 2025

**Nota:** Este projeto pretende avaliar a sua capacidade de trabalhar com ferramentas de big data num contexto prático, bem como a sua compreensão teórica dos sistemas de bases de dados distribuídas. Valorize tanto a implementação prática como a fundamentação teórica no seu trabalho.