

# Gestão de Grandes Conjuntos de Dados

## **2º Trabalho Prático**

-

Mestrado em Engenharia Informática  
Universidade do Minho

### **Grupo nº 8**

---

PG41080	João Ribeiro Imperadeiro
PG41081	José Alberto Martins Boticas
PG41091	Nelson José Dias Teixeira
PG41851	Rui Miguel da Costa Meira

13 de maio de 2020

# Conteúdo

<b>1</b>	<b>Introdução</b>	<b>3</b>
<b>2</b>	<b>Implementação</b>	<b>4</b>
2.1	A . . . . .	4
2.2	B . . . . .	4
2.2.1	B1 . . . . .	4
2.2.1.1	B11 . . . . .	4
2.2.2	B2 . . . . .	4
2.2.2.1	B21 . . . . .	4
2.2.2.2	B22 . . . . .	4
2.2.3	B3 . . . . .	4
<b>3</b>	<b>Conclusão</b>	<b>5</b>
<b>A</b>	<b>Observações</b>	<b>6</b>

# Lista de Figuras

# Capítulo 1

## Introdução

Neste trabalho prático é requerida a concretização e avaliação experimental de tarefas de armazenamento e processamento de dados através do uso da ferramenta computacional *Spark* (*batch* e *streaming*). Por forma a realizar estas tarefas, são utilizados os dados públicos do *IMDb*, que se encontram disponíveis em:

*<https://www.imdb.com/interfaces/>*

Para além destes dados, é também utilizado um gerador de *streams*, baseado nos mesmos, que simula uma sequência de votos individuais de utilizadores. Este utensílio foi desenvolvido pelo docente desta unidade curricular e encontra-se disponível na plataforma *Blackboard*.

Ao longo deste documento vão também ser expostos todos os passos tomados durante a implementação das tarefas pedidas neste projeto, incluindo as decisões tomadas pelos elementos deste grupo a nível de algoritmos e parâmetros de configuração. Para além disso são ainda apresentadas todas as instruções que permitem executar e utilizar corretamente os programas desenvolvidos. Por fim, na fase final deste manuscrito, são exibidos os objetivos atingidos após a realização das tarefas propostas.

De salientar também que durante os capítulos que se seguem são identificadas algumas alternativas para concretizar as tarefas indicadas neste trabalho prático.

## Capítulo 2

# Implementação

### 2.1 A

### 2.2 B

#### 2.2.1 B1

##### 2.2.1.1 B11

#### 2.2.2 B2

##### 2.2.2.1 B21

##### 2.2.2.2 B22

#### 2.2.3 B3

Capítulo 3

Conclusão

# Apêndice A

## Observações

- Documentação *Java* 8:  
`https://docs.oracle.com/javase/8/docs/api/`
- *Maven*:  
`https://maven.apache.org/`
- *Apache Spark*:  
`https://spark.apache.org/`
- *Docker*:  
`https://www.docker.com/`