

Instituto Superior de Engenharia de Lisboa Departamento de Engenharia de Eletrónica e Telecomunicações e de Computadores

Computação de Dados em Larga Escala – Semestre Inverno 2024/2025 Aula Prática 2

Objetivos

- Perceber os fundamentos do modelo de programação MapReduce, concretizados na infraestrutura Apache Hadoop, nomeadamente:
 - o Classes Configured, ToolRunner e interface Tool
 - o Contadores (Counters)
 - o Mappers, Reducers e Combinners
 - Cache distribuída
 - o Utilização da infraestrutura Logging
- 1. Estude os exemplos disponíveis na plataforma Moodle;
- 2. Modifique o exemplo MapReduce de contagem de palavras (Ex11-WordCount-02) de modo a perceber:
 - a) Influência dos Combinners no valor final dos contadores
 - b) Parametrização das funções map e reduce
 - c) Utilização dos mecanismos de *logging*;
- 3. Modifique o exemplo de modo que possa ser utilizado para contar n-gramas de qualquer dimensão (considere que um n-grama é uma sequência de n palavras separadas por um espaço). Para esta modificação considere que o valor de n é passado como argumento/configuração na execução da aplicação MapReduce.

Página 1 de 1 Carlos Gonçalves