UNIVERSIDADE FEDERAL DO RIO DE JANEIRO

ESCOLA POLITÉCNICA UFRJ

GABRIEL SANTA BARBARA RODRIGUES SOUZA - 123356007

JULIA FERNANDA TERRA SOUZA - 123460458

SAMUEL DO NASCIMENTO CAROLA - 123537899

Ordenação Externa por MergeSort

# 

Rio de Janeiro

2024

**Relatório de Implementação Ordenação Externa**

**Introdução**

Desenvolvemos um procedimento de “Ordenação Externa” para grandes “arquivos de registros”. O problema consistia em ordenar um arquivo de dados, em formato CSV, cujo tamanho supera a capacidade de memória RAM disponível para carregá-lo completamente, o que impede a sua ordenação utilizando uma função da biblioteca padrão da linguagem, que só opera sobre sequencias de dados em memória principal e não em memória secundária.

**Organização do Projeto**

Optei por desenvolver a solução em Python, organizando o código em um único arquivo bem estruturado. A classe principal `ExternalMergeSort` encapsula toda a lógica do algoritmo, enquanto funções auxiliares cuidam da geração de dados de teste e verificação dos resultados.

**Funcionamento do Algoritmo**

A implementação segue o clássico MergeSort, mas adaptado para lidar com arquivos grandes demais para a memória RAM. A primeira fase divide o arquivo em blocos menores, cada um ordenado individualmente. Na segunda fase, esses blocos são combinados gradualmente até obtermos o arquivo final ordenado.

**Desafios Encontrados**

Um dos maiores desafios foi garantir que o cabeçalho do CSV não se perdesse durante o processo. Após algumas tentativas frustradas, consegui resolver armazenando o cabeçalho separadamente e reinserindo-o em cada etapa. Outro problema foi lidar com colunas contendo tanto números quanto textos, que exigiram um tratamento especial durante as comparações.

**Aprendizados**

Este trabalho mostrou pra gente na prática como os conceitos teóricos de complexidade algorítmica se aplicam a problemas reais. Apesar das dificuldades, foi gratificante ver o algoritmo funcionando corretamente, especialmente quando testado com arquivos grandes. A experiência me deu uma nova perspectiva sobre a importância do gerenciamento de memória em aplicações que processam grandes volumes de dados.