

Computação Escalável

Exercício: paralelismo através de threads

Thiago Pinheiro de Araújo

EMAp/FGV

2022.1

Escrever um programa de processamento de dados que utilize o conceito de **paralelismo através de threads** apresentado na última aula.

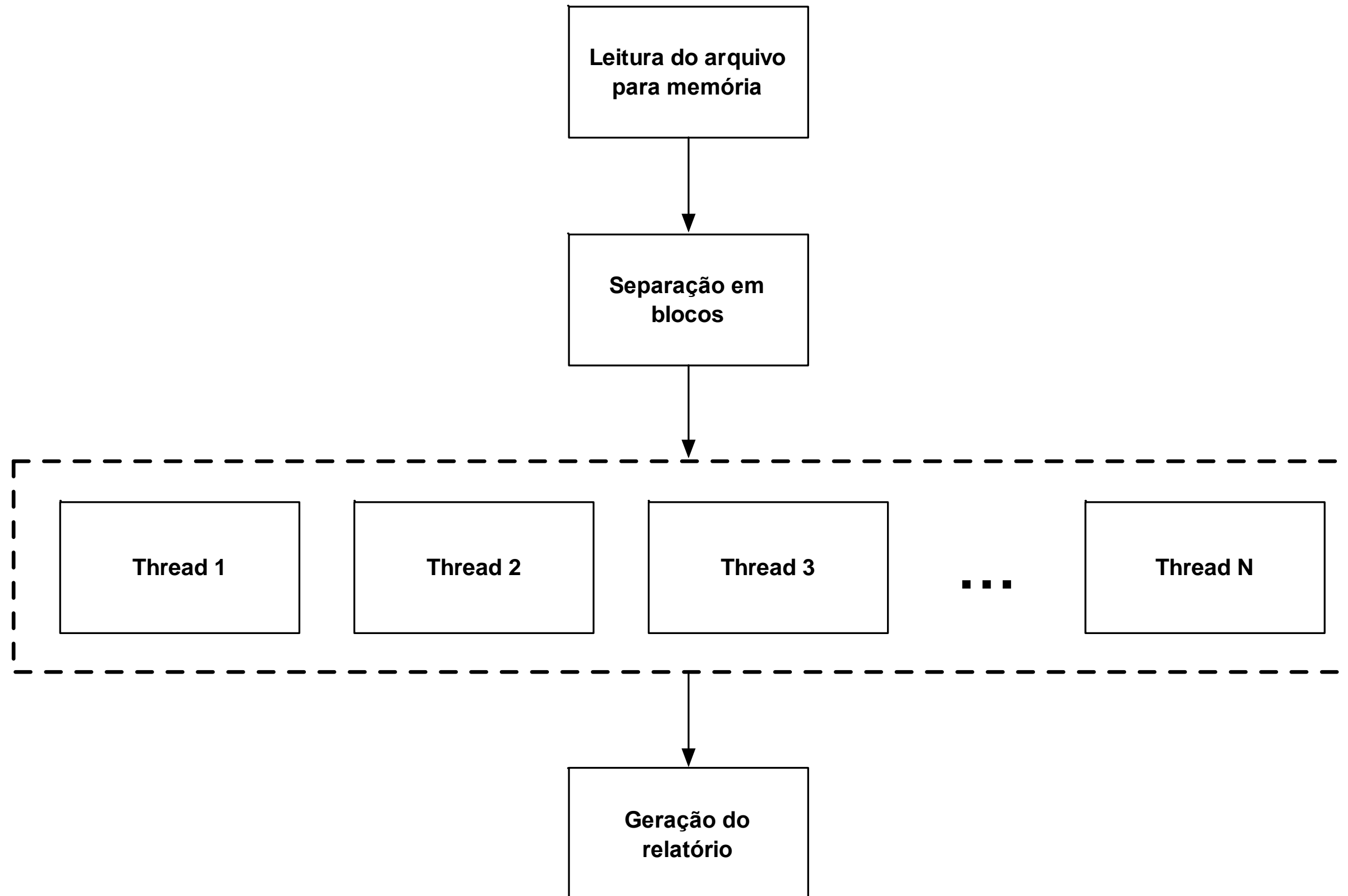
(em grupo: 4 integrantes)

Enunciado

- **Contexto:** William Shakespeare escreveu muitos livros sobre amor e ódio. Qual termo apresenta um maior número de ocorrências em suas obras?
- **Objetivo:** escrever um programa capaz de contar o número de ocorrências das palavras “*love*” e “*hate*” e responder qual foi mais utilizada
 - Os dados de entrada – livros de Shakespeare – serão carregados através de um único arquivo contendo a coletânea de livros.
 - A rotina de pesquisa pelos termos deverá ser implementada explorando paralelismo (threads).
 - O número de threads deverá configurável através de uma variável global.

Exercício

Visão geral da solução



- Ao concluir a operação o programa deverá exibir:
 - O número de threads.
 - O tamanho do bloco de cada thread.
 - O tempo de preparação (carregamento dos dados para memória e separação em blocos).
 - O tempo de execução da pesquisa.
 - O tempo total.
 - O número de ocorrências de cada palavra pesquisada.
 - Qual palavra apresenta um número maior de ocorrências.

- Inicie a execução do programa com uma thread.
- Execute repetidamente aumentando o número de threads até 100.
- Produza um gráfico tempo versus threads
 - Considerando o tempo total.
 - Considerando apenas o tempo de execução da pesquisa.
- Discuta com os demais integrantes do grupo as suas observações sobre o resultado.

Referências

- Dados
 - <https://www.gutenberg.org/ebooks/100>
 - Faça o download da versão textual.

- Leitura de arquivos (C++)
 - <https://www.cplusplus.com/doc/tutorial/files/>

- Threads (C++)
 - <https://www.cplusplus.com/reference/thread/thread/>

- Medição de tempo (C++)
 - https://en.cppreference.com/w/cpp/chrono/steady_clock/now

Exercício

Script para aumentar o volume de dados

- Utilize o script python a seguir para aumentar o volume de dados

```
with open('shakespeare.txt', 'r+b') as f:
    f.seek(0)
    content = f.read()
    for i in range(300):
        f.write(content)
```


Perguntas?

`thiago.p.araujo@fgv.br`

Atendimento: quarta-feira 8-10hs