



Universidade Estadual de Campinas
Instituto de Computação



Daniel De Lucca Fonseca

The Dissertation or Thesis Title in English
or Spanish

Título da Dissertação ou Tese em Português

CAMPINAS
1500

Daniel De Lucca Fonseca

**The Dissertation or Thesis Title in English or
Spanish**

Título da Dissertação ou Tese em Português

Dissertação apresentada ao Instituto de
Computação da Universidade Estadual
de Campinas como parte dos requisitos
para a obtenção do título de Mestre em
Ciência da Computação.

Dissertation presented to the Institute
of Computing of the University of
Campinas in partial fulfillment of the
requirements for the degree of Master
in Computer Science.

Supervisor/Orientador: Prof. Dr. Edson Borin

Este exemplar corresponde à versão
da Dissertação entregue à banca
antes da defesa.

CAMPINAS
1500

Na versão final esta página será substituída pela ficha catalográfica.

De acordo com o padrão da CCPG: “Quando se tratar de Teses e Dissertações financiadas por agências de fomento, os beneficiados deverão inserir no formulário de solicitação da ficha catalográfica as informações sobre agência e número do processo pelo qual receberam o auxílio, porém essa informação não ficará visível na ficha, será utilizada para fins estatísticos. Os auxílios recebidos devem também ser citados na seção “Agradecimento” seguindo os padrões exigidos pelas agências de fomento.”

e

“caso a tese de doutorado seja feita em Cotutela, será necessário informar na ficha catalográfica o fato, a Universidade conveniente, o país e o nome do orientador.”

Na versão final, esta página será substituída por outra informando a composição da banca e que a ata de defesa está arquivada pela Unicamp.

Dedicatória

As dedicatórias devem ter apenas uma página.

[DF 1] Escrever a
dedicatória

*Vita brevis,
ars longa,
occasio praeceps,
experimentum periculosum,
iudicium difficile.*

Preciso fazer a epígrafe^{DF}

(Hippocrates)

Agradecimentos

Os agradecimentos devem ocupar uma única página.

[DF 2] Preciso escrever os agradecimentos

Resumo

O resumo deve ter no máximo 500 palavras e deve ocupar uma única página.

[DF 3] Preciso escrever o resumo

Abstract

The abstract must have at most 500 words and must fit in a single page.

[DF 4] Preciso escrever o abstract

List of Acronyms

[DF 5] Checar se os acrônimos funcionam

List of Figures

List of Tables

Contents

1	Introduction	14
2	Fundamental Concepts	16
3	Methods and Materials	17
4	Experimental Results	18
5	Conclusion	19

Chapter 1

Introduction

TODO: Começar com uma base da motivação e contexto^{DF}

TODO: Na motivação e contexto, começar pela motivação geral, explicando o problema da gestão de memória para o processamento em larga escala de dados, focando em sistemas distribuídos^{DF}

TODO: Explicar sobre o quão comum é esse tipo de aplicação usar o Dask para paralelizar cargas de trabalho^{DF}

TODO: Explicar sobre os desafios para o gerenciamento de recursos. Aprofundar a motivação para explicar que alguns casos geram um grande consumo de memória e que, nesses casos, é necessário estimar o consumo de memória para definir o tamanho do chunk^{DF}

TODO: Trazer o problem base, dos problemas que um chunk definido de maneira não adequada pode trazer: o excesso de tentativa e erro, o uso subótimo dos recursos, o overhead do escalonador^{DF} ‘

[DF 6] Vale a pena dividir em seções? Se sim, eu penso em ser: Motivação e Contexto, Background, Limitações do Auto-Chunking do Dask, Motivação para o Memory-Aware Chunking, Minha Contribuição, e Estrutura da Tese

2. Background

- Dask Overview: Provide a brief overview of Dask, its purpose, and how it facilitates parallel computing in Python.
- Emphasize Dask’s role in splitting large datasets into smaller chunks for distributed execution.
- Auto-Chunking in Dask: Introduce Dask’s auto-chunking feature.
- Explain that it was designed to break data into manageable chunks based on data size but lacks memory-awareness.
- Mention the initial use case you explored (seismic data) to demonstrate how this feature worked in specific domains.

3. Limitations of Dask’s Auto-Chunking

- Problem Identification: Explain the key limitations you identified during your work with Dask.
- Lack of memory awareness in auto-chunking: Dask only considers input data size (in MB) but does not attempt to estimate memory consumption during chunk execution.
- This approach can be particularly problematic for memory-intensive algorithms.
- Complex Algorithm Example: Highlight GST3D or other complex algorithms that exacerbate this issue.
- Explain how small chunks for such algorithms can result in unexpectedly high memory consumption, leading to task failures or inefficient execution.

4. Motivation for Memory-Aware Chunking

- **The Need for Memory Awareness:** Justify why it's essential to account for memory usage when defining chunk sizes.
- Describe the potential issues (e.g., out-of-memory errors, inefficient resource use, performance bottlenecks) that arise when memory requirements aren't considered.
- Tie this back to how efficient memory management is critical in distributed computing, especially when dealing with limited resources like worker memory.

5. Your Contribution

- **Estimation Method:** Introduce your main contribution—an approach to estimate memory usage with minimal executions.
- Briefly describe the strategy you developed, which allows for predicting memory requirements by running a few sample executions.
- **Optimal Chunk Sizing:** Explain how you leveraged this estimation to define the best chunk size based on available worker memory.
- Mention how this approach is flexible and can work across different algorithms and data processing scenarios (not just seismic).

6. Thesis Structure Overview

- Provide an overview of the structure of the thesis and how the rest of the chapters are organized, giving the reader a roadmap for the work.
- Example: “The remainder of this thesis is organized as follows. Chapter 2 reviews the related work in memory profiling and distributed computing. Chapter 3 presents the methodology used for memory estimation and optimal chunk sizing...”

[DF 7] Preciso organizar a estrutura da introdução

Chapter 2

Fundamental Concepts

[DF 8] Preciso organizar a estrutura dos conceitos fundamentais

Chapter 3

Methods and Materials

[DF 9] Preciso organizar a estrutura dos métodos e materiais

Chapter 4

Experimental Results

[DF 10] Preciso organizar a estrutura dos resultados experimentais

Chapter 5

Conclusion

[DF 11] Preciso organizar a estrutura da conclusão