

Uma trajetória de pesquisa em engenharia de dados para aplicações em larga escala

Vítor Silva

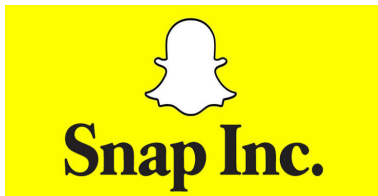
Vítor Silva



- **Engenheiro de Computação e Informação pela UFRJ**
 - ▣ Iniciação Científica com a Profa. Marta desde o primeiro período



- **Mestre e Doutor em Engenharia de Sistemas e Computação pela COPPE/UFRJ**

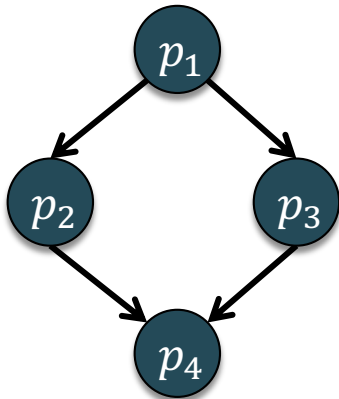


- ***Senior Engineering Technologist* na Dell EMC**
- ***Research Engineer* na Snap Inc.**

Iniciação científica

- **Iniciação Científica com a Profa. Marta Mattoso**
 - ▣ Desde o primeiro período (disciplina Intro ECI)

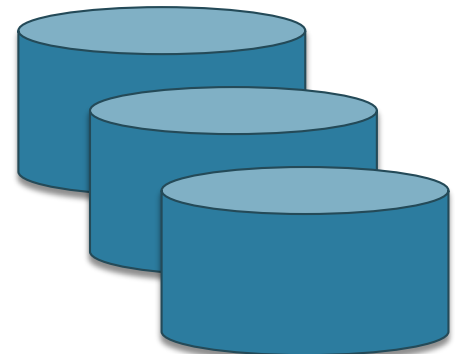
Simulações
computacionais



Ambientes de
Processamento de Alto
Desempenho (PAD)



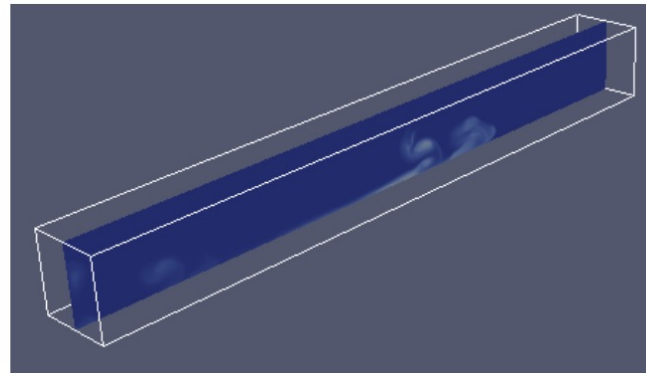
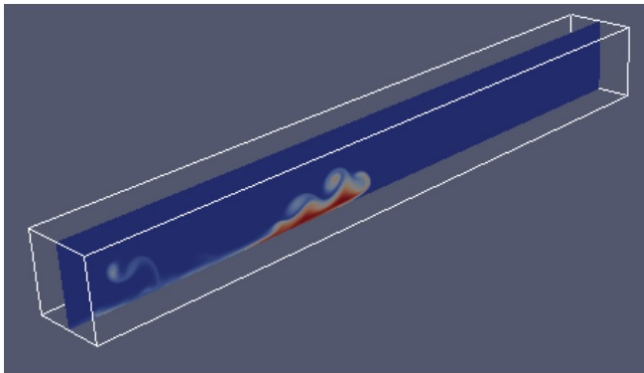
Gerência de
Banco de Dados
(BD)



Iniciação científica

- **Simulações em diversos domínios da ciência**

- Bioinformática
- Astronomia
- Dinâmica de fluidos computacionais

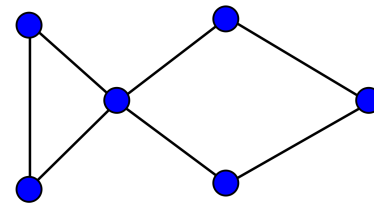


- **Terabytes de dados em centenas de milhares de recursos computacionais**

Disciplinas da graduação

□ Teoria dos grafos:

- [Prof. Daniel Figueiredo](#) e Jefferson Elbert
- Trabalho focado em construir uma biblioteca com os principais conceitos e algoritmos da disciplina
- De último a primeiro lugar! 😊



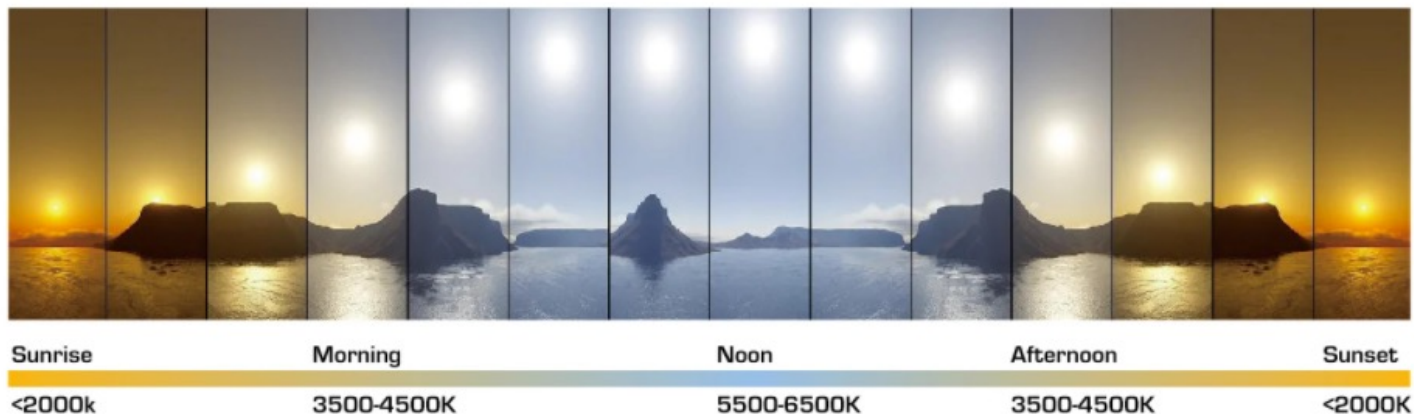
□ Computação Gráfica e Jogos e Animação:

- Prof. Claudio Esperança
- Prof. Ricardo Marroquim

Estágio supervisionado



- **Estágio em uma área nova:**
 - ▣ Computação Gráfica
 - ▣ Grupo de Pesquisa e Desenvolvimento da Rede Globo
 - ▣ Processamento de imagem e efeitos visuais em programas televisivos

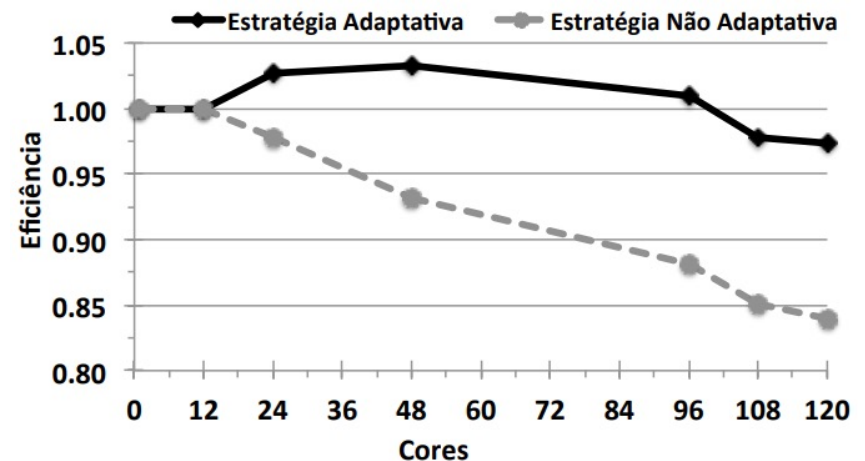
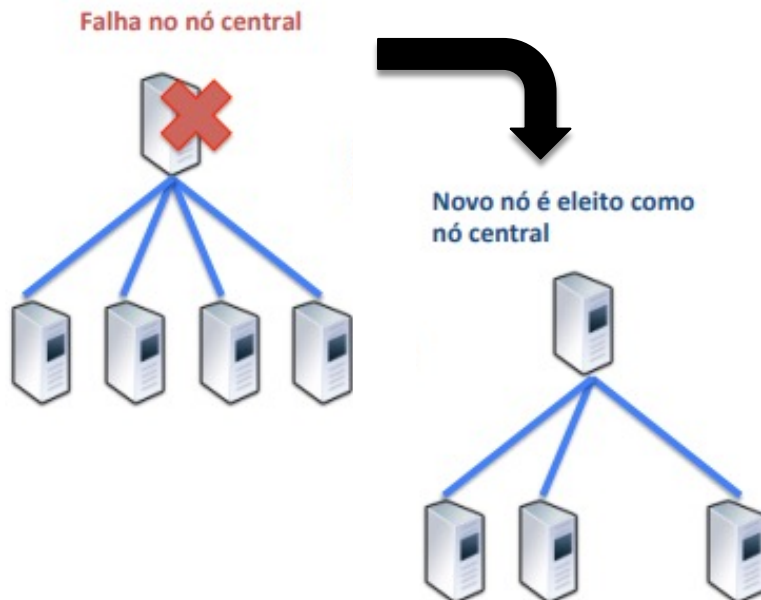


Fonte: <https://www.ledrise.eu/blog/color-temperature-explained-lr/>

Mestrado

□ Execução paralela adaptável de simulações computacionais

- ▣ Tolerância a falhas
- ▣ Arquitetura hierárquica

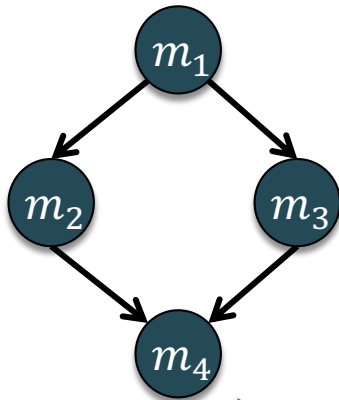


Doutorado:

Análise de dados em **larga escala**

8

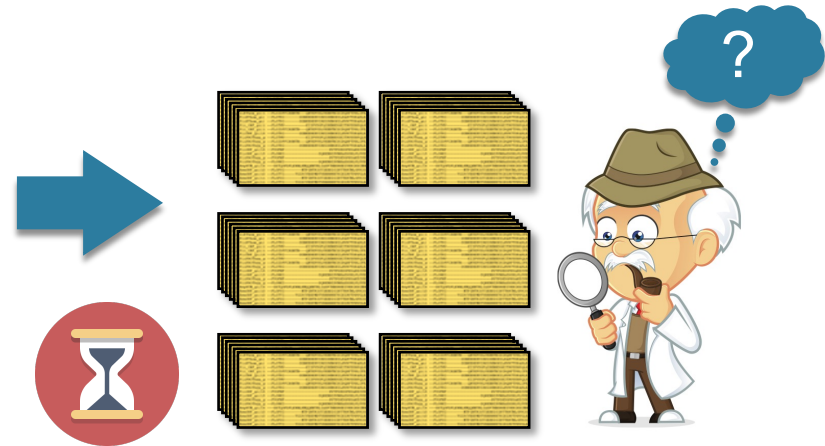
Simulações
computacionais



Ambientes de
PAD



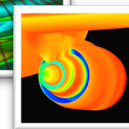
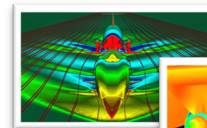
Análise de dados ad-hoc
em tempo de execução



Hipótese
científica



Parâmetros de entrada



Dados intermediários
ou resultados finais

Doutorado:

Análise de dados em **larga escala**

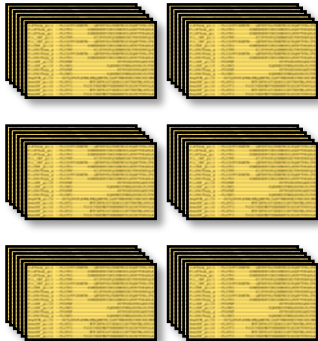
9



STAMPEDE
— SUPERCOMPUTER —

<https://www.tacc.utexas.edu/systems/stampede>

~520.000 núcleos de processamento
260TB de memória
~15PB de disco



Simulações com aproximadamente...
170.000 arquivos
5TB de dados

Experiências nessa trajetória

- **Pesquisa em Big Data no Inria**

- Montpellier, França
- [Prof. Patrick Valduriez](#)



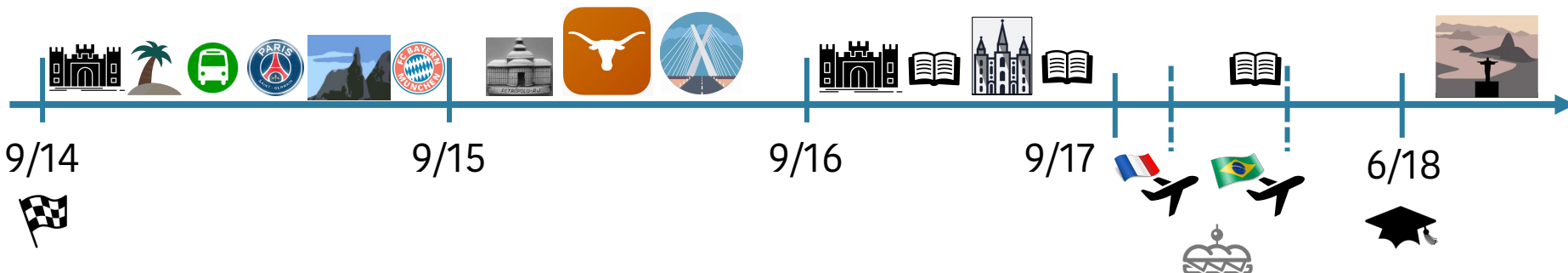
- **Projetos de pesquisa:**

- Dell, Intel, HPC4E



- **Bolsas de pesquisa:**

- CAPES, FAPERJ, CNPq, RNP



Pesquisa em engenharia de dados

□ **DELL EMC**

- ▣ Telemetria
- ▣ Aprendizado de máquina (ML)
- ▣ Gerência de dados em ambientes de PAD

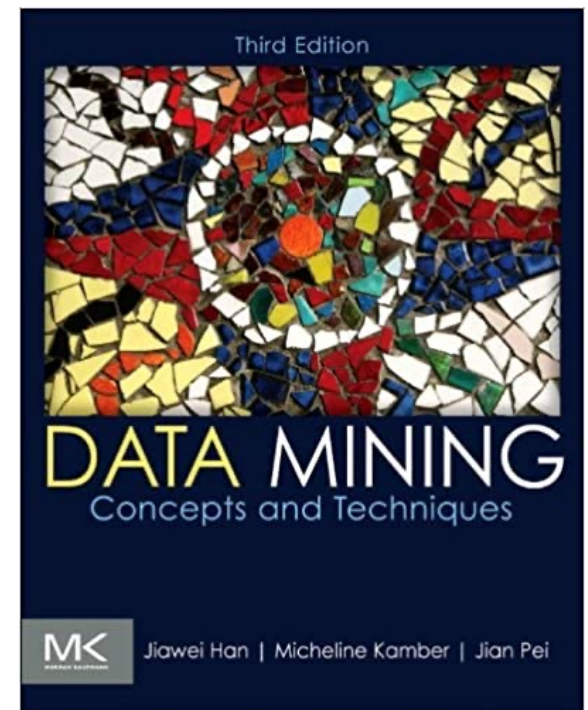
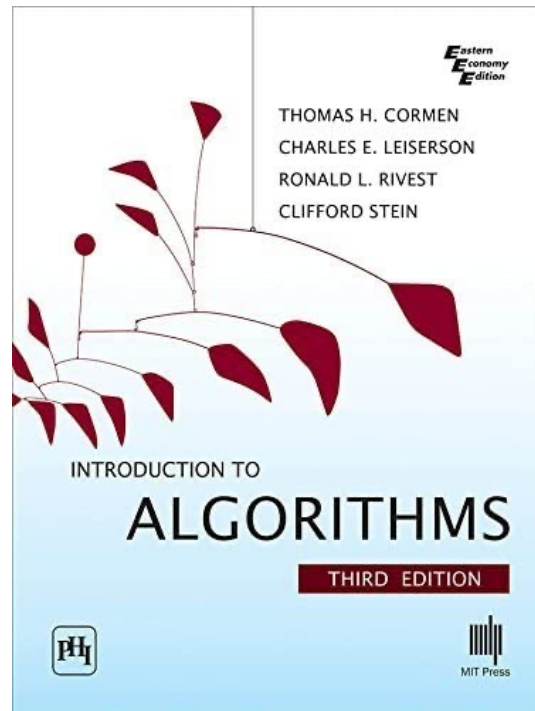
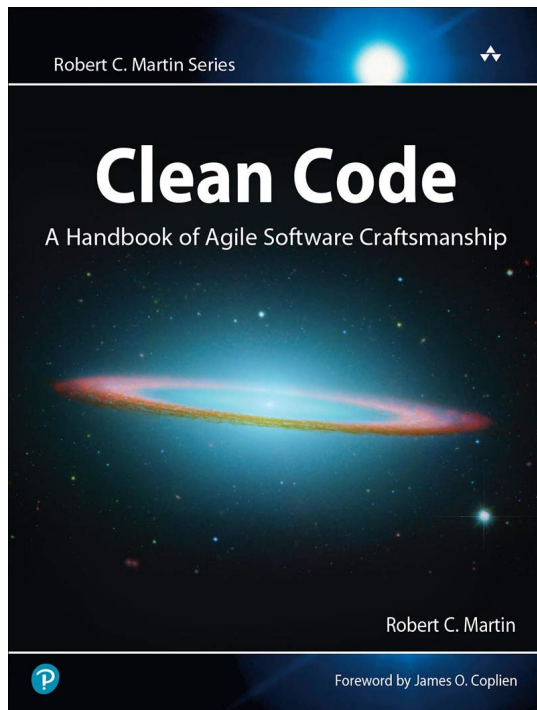


□ **Snap Inc.**

- ▣ *Computational Social Science* (Snap Research)
- ▣ Ciência de dados: ML, NLP e Séries temporais
- ▣ Infraestrutura para entregar protótipos de pesquisa em nossos produtos (ambientes de PAD)



Livros: para ler o quanto antes!



Obrigado!

Uma trajetória de pesquisa em engenharia de dados
para aplicações em larga escala

Vítor Silva