

Evento: XXVIII Seminário de Iniciação Científica  
ODS: 9 - Indústria, Inovação e Infra-estrutura

## ANÁLISE E INTEGRAÇÃO DE GRANDE VOLUME DE DADOS<sup>1</sup>

### ANALYSIS AND INTEGRATION OF BIG DATA

Eduarda Feretti Duarte<sup>2</sup>, Benjamim Zucolotto<sup>3</sup>

<sup>1</sup> Projeto de pesquisa financiado pelo CNPQ

<sup>2</sup> Bolsista de Iniciação Científica PIBIC/CNPq 2019/2020 e acadêmica do curso de Engenharia Química, UNIJUÍ

<sup>3</sup> Doutor em Física, Professor Adjunto, Departamento de Ciência Exatas e Engenharias, Orientador, UNIJUÍ

## INTRODUÇÃO

Em tempos de grandes volumes de informações sendo coletadas a todo momento, o conhecimento na área de análise de dados é essencial. Tendo em vista a busca de padrões em conjuntos de dados, faz-se necessário seu tratamento e análise. Neste trabalho, estudamos o comportamento do tempo de *Makespan* e do número de *Processed Messages* com dependência das *Threads* e *Input Rates*. Utilizamos um modelo de execução implementado por um sistema de tempo de execução, ao lidar como o nível de execução de um processo de integração. Assim, procurou-se adquirir conhecimentos científicos e tecnológicos quanto ao cenário observado no contexto de *Big Data*, para identificar e analisar grande volume de dados, interpretando e modelando matematicamente o comportamento de dados experimentais, visando a descrição, previsão e otimização dos processos de integração no contexto apresentado.

**Palavras-chave:** *Big Data*; *Makespan*; mensagens.

**Keywords:** *Big Data*; *Makespan*; messages.

## METODOLOGIA

Um processo de integração é utilizado para minimizar o tempo médio que uma mensagem leva para ser processada [1,2]. O modelo de execução consiste em um determinado processo que gera o número de mensagens processadas (*Processed Messages*) a partir de um determinado número de *Threads* e contabilizado o tempo de processamento *Makespan*, ou seja, o tempo necessário para que uma mensagem seja processada.

As variáveis do processamento foram, o número de *Thread* usadas para processamento e taxa de entrada de mensagens (*Input Rate*). Destas variáveis, foram analisados o *Makespan* e o número de *Processed Messages*.

Inicialmente foram realizados os tratamentos dos dados no software *SciDavis* e posteriormente modelados matematicamente no MATLAB. Usamos o procedimento conhecido como *net search*, ou *sampling* para encontrar uma solução que representa os dados experimentais baseado na resolução de um Problema Inverso com a estratégia que consiste em resolver um Problema Direto várias vezes [3,4].

Evento: XXVIII Seminário de Iniciação Científica  
ODS: 9 - Indústria, Inovação e Infra-estrutura

## RESULTADOS E DISCUSSÃO

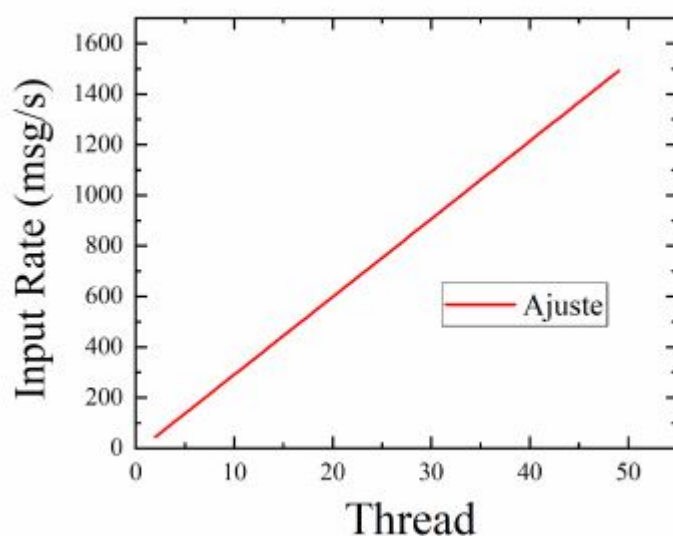
Observamos que o *Makespan* diminui com o aumento do número de *Threads* até o momento que tendem para um valor aparentemente constante para *Input Rate*. Esse comportamento de diminuição do *Makespan* possui menor inclinação quanto maior a taxa de entrada de mensagens.

Ao elevar-se o número de *Input Rate* observa-se novamente a tendência linear e posterior saturação no *Makespan* com menores inclinações para *Input Rate* mais elevadas. Entretanto, para o maior número de *Threads* o *Makespan* aumenta.

Nas mensagens processadas em função do número de *Threads* para *Input Rate* para valores mais elevados observou-se um crescimento das mensagens processadas com o aumento do número de *Threads*. Para *Input Rate* extremamente baixos observou-se o total processamento das mensagens, ou seja, todas as que entram são processadas. Ao elevar seu valor o aumento do processamento também crescente, mas mais acentuada.

O comportamento do *Makespan* em função da *Input Rate* de mensagens para diferentes números de *Threads* cresce linearmente e posteriormente possui tendência tipicamente exponencial, e o *Makespan* tende se estabilizar para menores valores de *Makespan* com o aumento do número de *Threads*.

Para taxa de entrada de mensagens máxima do crescimento que antecede o exponencial em função do número de *Threads* percebe-se um comportamento linear no regime. Logo, neste intervalo, aumentar o número de *Threads* corresponde a um aumento eficiente do processamento, como mostrado na Figura 1.



**Figura 1:** Input Rate máxima do crescimento linear que antecede o exponencial versus Thread.



Evento: XXVIII Seminário de Iniciação Científica  
ODS: 9 - Indústria, Inovação e Infra-estrutura

## CONSIDERAÇÕES FINAIS

Neste trabalho, estudamos o comportamento dos dados de um processo de integração. Percebemos que o *Makespan* possui diferentes regimes de comportamento com dependência do número de *Threads* e da *Input Rate*. Estudamos o comportamento do número de *Processed Messages* com do número de *Threads*. Avaliamos via modelagem matemática a variação do *Makespan* com número de *Threads versus Input Rate* e concluímos que sua variação considerável é limitada.

## REFERÊNCIAS BIBLIOGRÁFICAS

- [1] D. L. Freire, R. Z. Frantz, F. Roos-Frantz, and S. Sawicki. A methodology to rank enterprise application integration platforms from a performance perspective: an analytic hierarchy process-based approach. *Enterprise Information Systems*, 13(9):1292–1322, 2019.
- [2] R. Z. Frantz, R. Corchuelo, and C. Molina-Jiménez. A proposal to detect errors in enterprise application integration solutions. *Journal of Systems and Software*, 85(3):480–497, 2012.
- [3] A. J. Silva Neto e M. N. Özisik. Inverse problem of simultaneously estimating the timewise-varying strengths of two plane heat sources. *Journal of Applied Physics*, v. 73, n. 5, p. 2132-2137, 1993.
- [4] Francisco Duarte Moura Neto e Antônio José da Silva Neto. *An Introduction to Inverse Problems with Applications*. New York: Springer, 2013. 243 p.