

# Software Module Clustering (SMC) Tools Analysis

## Grupo 5 - Metodologia Científica

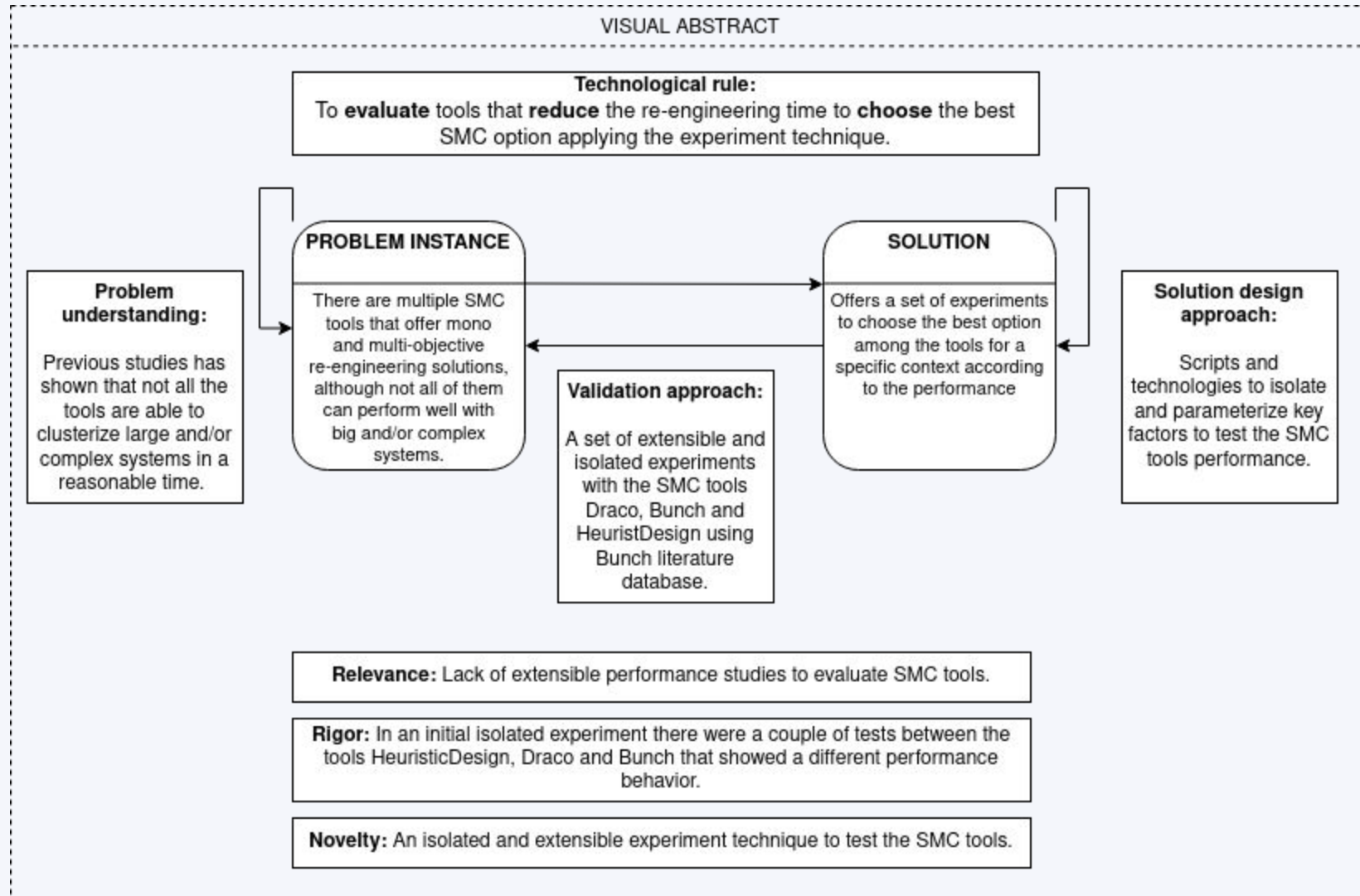
Dayanne Fernandes da Cunha - 13/0107191

João Lucas Azevedo Y. R. da Cunha - 17/0013731

Khalil Carsten do Nascimento - 15/0134495

Luis Henrique Vieira Amaral

# Visual Abstract



# Motivação

- Várias ferramentas de clusterização;
- Incerteza da melhor opção para determinados casos;
- Algumas ferramentas são incapazes de clusterizar sistemas muito grandes e/ou complexos em tempo hábil;
- Faltam formas fáceis de reproduzir a comparação de ferramentas.

## **Problem understanding:**

Previous studies has shown that not all the tools are able to clusterize large and/or complex systems in a reasonable time.

## **PROBLEM INSTANCE**

There are multiple SMC tools that offer mono and multi-objective re-engineering solutions, although not all of them can perform well with big and/or complex systems.

# Objetivo

- Oferecer um template parametrizável para executar experimentos, a fim de comparar determinadas métricas sob a óptica de contextos específicos.

## Technological rule:

To **evaluate** tools that **reduce** the re-engineering time to **choose** the best SMC option applying the experiment technique.

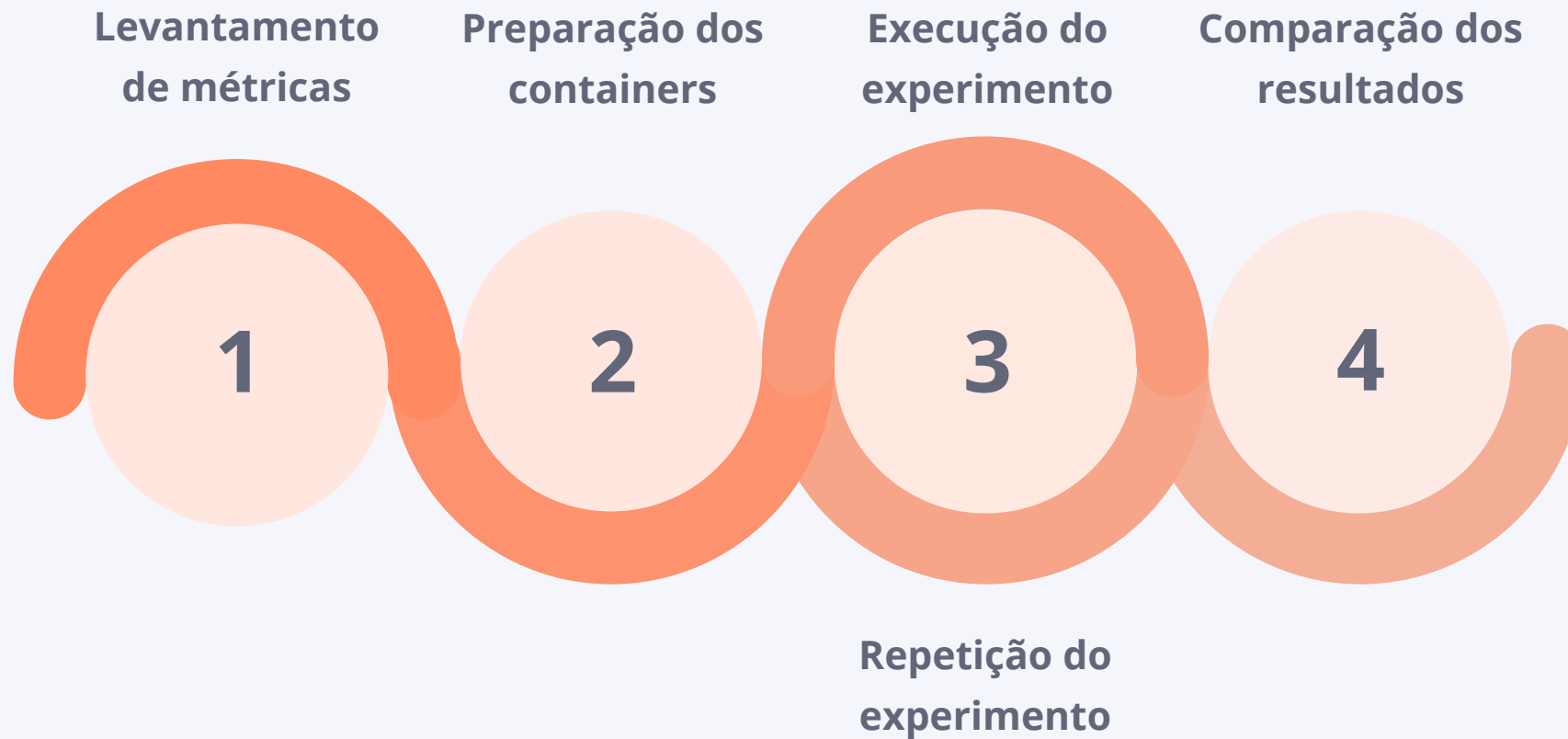
## SOLUTION

Offers a set of experiments to choose the best option among the tools for a specific context according to the performance

## Solution design approach:

Scripts and technologies to isolate and parameterize key factors to test the SMC tools performance.

# Metodologia



# Resultados

- Para cada ferramenta foi criado um container;
- Cada container executou 5 MDGs com tamanhos variados, entre 500Kb e 1Mb;
- Para cada MDG em cada ferramenta foram coletadas duas métricas utilizando a ferramenta time do UNIX. Foi possível coletar informações relacionadas ao tempo de execução (tabela do slide 7) e memória máxima utilizada por experimento (tabela do slide 8).
- Utilizamos o limite de recurso de 3 cores de CPU para ferramenta e 3 Gb de memória RAM por container;
- Apenas uma máquina foi utilizada na execução deste experimento;
- Esta máquina host possui 4 cores, 16 Gb de RAM e 4.2 GHz.

# Resultados

Arquivo	DCT				Bunch	
	Mono		Multi		Mono	
	1ª tentativa	2ª tentativa	1ª tentativa	2ª tentativa	1ª tentativa	2ª tentativa
cas	<b>0,80</b>	4,05	<b>0,95</b>	12,26	<b>7,07</b>	12,85
druid	<b>1,61</b>	7,83	<b>2,03</b>	6,71	<b>2,54</b>	11,10
hibernate-orm	<b>2,61</b>	3,32	<b>3,22</b>	3,88	28,24	<b>15,93</b>
jmonkeyengine	<b>5,06</b>	7,65	<b>5,57</b>	8,85	<b>8,27</b>	9,96
jooq	<b>2,87</b>	2,88	<b>3,88</b>	4,19	24,08	<b>17,19</b>

Comparação entre os tempos de sistema em segundos

# Resultados

Arquivo	DCT				Bunch	
	Mono		Multi		Mono	
	1ª tentativa	2ª tentativa	1ª tentativa	2ª tentativa	1ª tentativa	2ª tentativa
cas	199,78	<b>196,13</b>	<b>322,26</b>	322,46	1131,19	<b>1130,35</b>
druid	<b>276,81</b>	277,78	419,63	<b>418,94</b>	1144,30	<b>1136,36</b>
hibernate-orm	<b>342,61</b>	347,02	<b>567,01</b>	569,45	1155,14	<b>1141,66</b>
jmonkeyengine	401,19	<b>400,09</b>	<b>682,20</b>	694,81	<b>1157,98</b>	1158,35
jooq	<b>336,58</b>	338,91	<b>553,62</b>	555,75	1156,60	<b>1143,94</b>

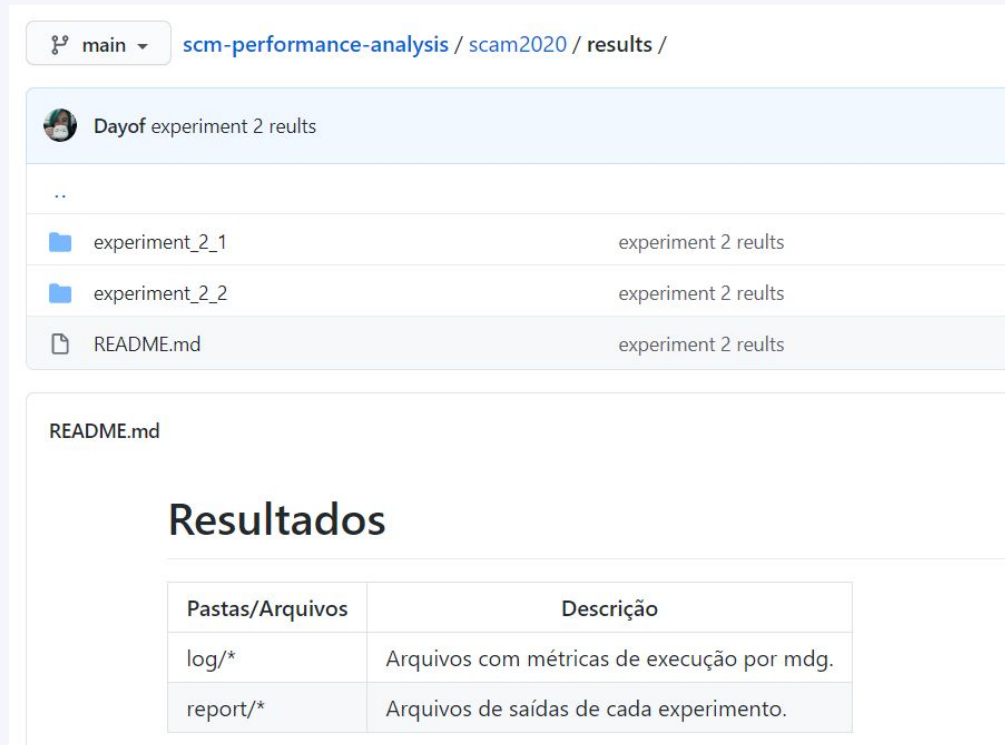
Comparação entre os picos de memória em *megabytes* (*maximum resident set size*)



# Resultados

As métricas na íntegra podem ser consultadas neste repositório:

<https://github.com/luisamaralh/scm-performance-analysis/tree/main/scam2020/results>



main scm-performance-analysis / scam2020 / results /

Dayof experiment 2 results

..

experiment\_2\_1 experiment 2 results

experiment\_2\_2 experiment 2 results

README.md experiment 2 results

README.md

## Resultados

Pastas/Arquivos	Descrição
log/*	Arquivos com métricas de execução por mdg.
report/*	Arquivos de saídas de cada experimento.

As métricas foram coletadas utilizando a ferramenta **time** do UNIX em modo verboso. (<https://man7.org/linux/man-pages/man1/time.1.html>)

# Dificuldades

Devido ao limite de tempo para a implementação do projeto, não foi possível:

- Adição do **HDG mono objetivo**;
- Testar **cgroup** que isola os recursos nos containers;
- **Automatizar** inclusão de uma nova ferramenta para execução de testes;
- Implementar um **limite máximo de tempo** de execução por MDG;
- Automatizar as **repetições** dos experimentos por ferramenta.

# Conclusão

- A reprodutibilidade dos experimentos foi facilitada, uma vez que o protótipo do template de experimentos permite adicionar um novo experimento com outras características em *containers* isolados, sem precisar modificar a máquina *host*;
- Identificamos que pelo menos dentre os experimentos realizados a ferramenta DCT apresentou tempo de execução e memória bem reduzido quando comparado ao Bunch;
- Não foi possível alcançar o nível de isolamento esperado, já que houve uma discrepância relevante entre os tempos de execução das repetições do experimento.