

Universidade Federal de São Carlos - Sorocaba
Bacharelado em Ciência da Computação

CouchDB

Processamento Massivo de Dados - Fase Final

Bruno Morii Borges
Isabella Soares de Lima

Prof^a Dr^a Sahudy Montenegro González

14 de Junho de 2019



Descrição do Ambiente de Teste

Projeto do Teste

Implementação do Teste

Critérios de Avaliação

Análise dos Resultados

Considerações



▶ Hardware

- ▶ Intel Core i5-7200U de 4 núcleos a 2.50GHz
- ▶ 8GB de RAM
- ▶ HDD Toshiba MQ01ABD100 - 1 TB

▶ Software

- ▶ CouchDB 2.3.1 (Standalone)
- ▶ Apache JMeter
- ▶ SO Windows 10 Home

▶ Rede

- ▶ 5.50 Mb de download
- ▶ 1.50 Mb de upload



- ▶ Ferramenta: *Apache JMeter*
- ▶ Número de usuários (acessos) simultâneos: 100, 500 e 1000
- ▶ Os grupos de usuários são executados de maneira sequencial
- ▶ O BD foi deletado e populado em cada etapa do teste
- ▶ Os *scripts* de teste foram executados cada um 1 vez na etapa de Execução e 3 vezes na etapa de Validação
- ▶ Tamanho inicial do BD: 28.061 documentos
- ▶ Distribuição das operações
 - ▶ Requisições dos usuários: 75% de leitura e 25% de escrita
 - ▶ *Script* completo: 37,5% de leitura e 62,5% de escrita
- ▶ Link para o projeto: <https://github.com/IsabellaSoares/PMD-CouchDB>



- International migration: April 2019 – citizenship by visa by country of last permanent residence¹

```
{  
  "year_month",  
  "month_of_release",  
  "passenger_type",  
  "direction",  
  "citizenship",  
  "visa",  
  "country_of_residence",  
  "estimate",  
  "standard_error",  
  "status"  
}
```



Gerenciador de Cabeçalhos HTTP

Nome: Gerenciador de Cabeçalhos HTTP

Comentários:

Cabeçalhos Armazenados no Gerenciador de Cabeçalhos

Nome:	Valor
Content-Type	application/json
Accept	application/json

Gerenciador de Autorização HTTP

Nome: Gerenciador de Autorização HTTP

Comentários:

Options

☐ Clear auth on each iteration?

Autorizações Armazenadas no Gerenciador de Autorização

URL Base	Nome do Usuário	Senha	Domínio	Reino (realm)	Mechanism
	pmd	●●●			BASIC



Grupo de Usuários

Nome: Grupo de usuários

Comentários:

Ação a ser tomada depois de erro do testador

☒ Continuar ☐ Start Next Thread Loop ☐ Interromper Usuário Virtual ☐ Interromper Teste ☐ Interrompe Teste Agora

Propriedades do Usuário Virtual

Número de Usuários Virtuais (threads):

Tempo de inicialização (em segundos):

Contador de Iteração ☒ Infinito

☐ Delay Thread creation until needed

☐ Agendador

Configuração do Agendador

⚠ If Loop Count is not -1 or Forever, duration will be min(Duration, Loop Count * iteration duration)

Duração (segundos):

Atraso para início (segundos):

► Grupo de usuários

- Número de *threads*: 100, 500 ou 1000
- Tempo de inicialização (em segundos): metade do número de *threads*, totalizando sempre em 2 usuários por segundo
- Contador de iteração: 1 (Execução) ou 3 (Validação)



Configuração dos dados CSV

Nome: Configuração dos dados CSV

Comentários:

Configurar fonte de dados CSV

Nome do arquivo:	.././CSV Inserções/Insercao1 - 100.csv	Procurar...
Codificação do arquivo (encoding):	▼	
Nomes das variáveis (separados por vírgula):	year_month, month_of_release, passenger_type, direction, citizenship, visa, country_of_residence, estimate, standard_error, status, _id	
Ignore a primeira linha (usada apenas se os Nomes de variáveis não estiverem vazios):	True ▼	
Separador (usar '\t' para tabulações):	, ▼	
Permitir dados com citações?:	False ▼	
Recidir no final do arquivo (EOF)?:	True ▼	
Finalizar usuário virtual no final do arquivo?:	False ▼	
Modo de compartilhamento:	Todos os usuários virtuais ▼	

- Arquivo consumido durante as operações de escrita



- ▶ População
 - ▶ Delete do banco de dados
 - ▶ Criação
 - ▶ População
 - ▶ Indexação do campo "year_month"
- ▶ Requisições dos grupos de usuários
 - ▶ Leitura com 2 parâmetros (\$or)
 - ▶ Leitura com 2 parâmetros e filtro de campos
 - ▶ Leitura com 1 parâmetro, filtro de campos e ordenação
 - ▶ Escrita de 1 documento por cada usuário

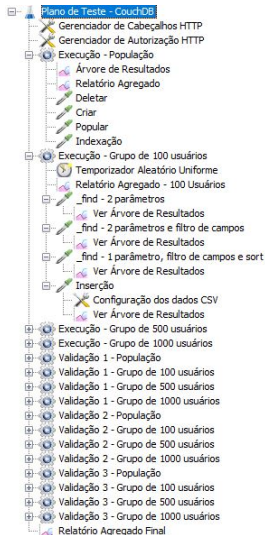
Implementação do Teste

Estrutura



Implementação do Teste

Estrutura





- ▶ Temporizador Aleatório Uniforme
 - ▶ Os *scripts* de requisições possuem temporizadores (*Thinking time*) para simular a ação real dos usuários na aplicação, ou seja, o tempo que o usuário pensa entre as ações que ele realiza no sistema.
- ▶ Ouvintes
 - ▶ Árvore de resultados: captura informações sobre cada requisição, por exemplo: quais os parâmetros enviados na requisição, tempo de resposta do servidor, conteúdo da resposta, entre outras;
 - ▶ Relatório agregado: captura informações da execução do teste, como: quantidade de usuários (amostras) que executaram cada requisição (rótulo), tempo médio em milissegundos (média), tempo mínimo em milissegundos (mín) e tempo máximo em milissegundos (máx), entre outras. Esses tempos são calculados considerando todas as execuções simultâneas.



- ▶ **Tempo médio** gasto por todos os usuários para executar uma requisição;
- ▶ **Tempo mínimo e máximo** para uma requisição;
- ▶ **Taxa de transferência (*Throughput*)**, representa o número de solicitações processadas por unidade de tempo pelo servidor. Este tempo é calculado a partir do início da primeira amostra até o final da última amostra;
- ▶ **Mediana do tempo** de resposta, indica que 50% dos usuários não levaram mais do que esse tempo para completar a requisição;
- ▶ **Percentis 90%, 95% e 99%**, indica que dada porcentagem dos usuários não demoraram mais do que esse tempo para completar a requisição

Análise dos Resultados

População - Deletar



13

Deletar		
	Execução	Validação
Média	118 ms	84.67 ms
Mediana	118 ms	84.67 ms
Percentil 90%	118 ms	84.67 ms
Percentil 95%	118 ms	84.67 ms
Percentil 99%	118 ms	84.67 ms
Tempo mínimo	118 ms	84.67 ms
Tempo máximo	118 ms	84.67 ms
Vazão	8.47 solicitações / s	11.93 solicitações / s

Análise dos Resultados

População - Criar



Criar		
	Execução	Validação
Média	1741 ms	1792.67 ms
Mediana	1741 ms	1792.67 ms
Percentil 90%	1741 ms	1792.67 ms
Percentil 95%	1741 ms	1792.67 ms
Percentil 99%	1741 ms	1792.67 ms
Tempo mínimo	1741 ms	1792.67 ms
Tempo máximo	1741 ms	1792.67 ms
Vazão	0.57 solicitações / s	0.69 solicitações / s

Análise dos Resultados

População - Popular



Popular		
	Execução	Validação
Média	8195 ms	8445.67 ms
Mediana	8195 ms	8445.67 ms
Percentil 90%	8195 ms	8445.67 ms
Percentil 95%	8195 ms	8445.67 ms
Percentil 99%	8195 ms	8445.67 ms
Tempo mínimo	8195 ms	8445.67 ms
Tempo máximo	8195 ms	8445.67 ms
Vazão	0.12 solicitações / s	0.12 solicitações / s

Análise dos Resultados

População - Indexação



Indexação		
	Execução	Validação
Média	219 ms	107.33 ms
Mediana	219 ms	107.33 ms
Percentil 90%	219 ms	107.33 ms
Percentil 95%	219 ms	107.33 ms
Percentil 99%	219 ms	107.33 ms
Tempo mínimo	219 ms	107.33 ms
Tempo máximo	219 ms	107.33 ms
Vazão	4.57 solicitações / s	9.68 solicitações / s

Análise dos Resultados

Leitura com 2 parâmetros



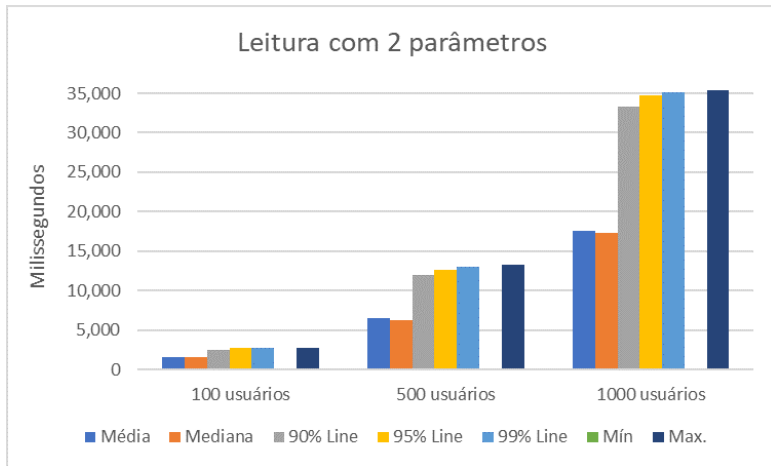
Leitura com 2 parâmetros			
	100 usuários	500 usuários	1000 usuários
Média	1,530.67 ms	6,470.67 ms	17,574.33 ms
Mediana	1,592.00 ms	6,299.00 ms	17,317.67 ms
Percentil 90%	2,520.33 ms	11,987.67 ms	33,289.00 ms
Percentil 95%	2,673.33 ms	12,653.00 ms	34,761.67 ms
Percentil 99%	2,743.67 ms	12,969.67 ms	35,114.33 ms
Tempo mínimo	65.67 ms	81.67 ms	87.00 ms
Tempo máximo	2,771.33 ms	13,239.67 ms	35,335.00 ms
Vazão	1.92 solicitações / s	1.91 solicitações / s	1.87 solicitações / s

Análise dos Resultados

Leitura com 2 parâmetros



18



Análise dos Resultados

Leitura com 2 parâmetros



19

	100 usuários	500 usuários	1000 usuários
Média	1,530.00 ms	6,470.67 ms	17,574.33 ms

+ 322.74%

+ 1,048.15%

	100 usuários	500 usuários	1000 usuários
Vazão	1.92 solicitações / s	1.91 solicitações / s	1.87 solicitações / s

- 0.66%

- 2.35%

Análise dos Resultados

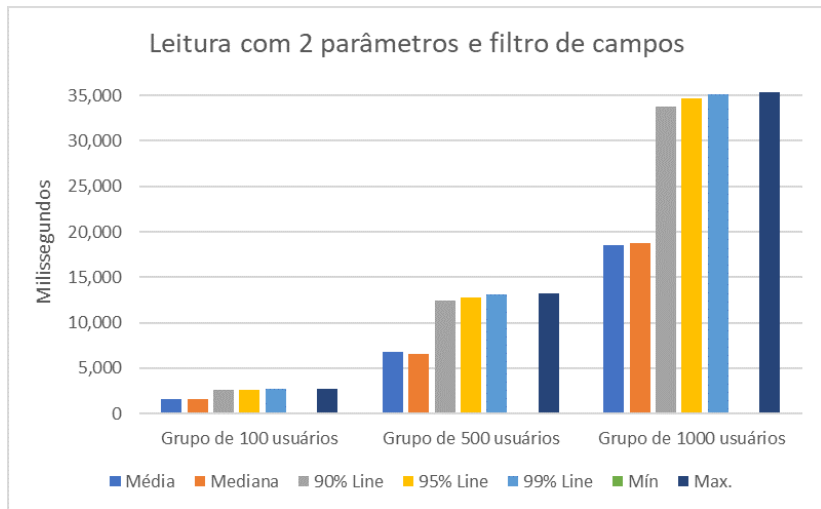
Leitura com 2 parâmetros e filtro de campos



Leitura com 2 parâmetros e filtro de campos			
	100 usuários	500 usuários	1000 usuários
Média	1,578.00 ms	6,762.67 ms	18,586.33 ms
Mediana	1,649.00 ms	6,621.33 ms	18,749.00 ms
Percentil 90%	2,567.00 ms	12,392.33 ms	33,812.67 ms
Percentil 95%	2,670.67 ms	12,777.33 ms	34,723.33 ms
Percentil 99%	2,741.67 ms	13,061.67 ms	35,083.00 ms
Tempo mínimo	61.67 ms	60.67 ms	69.00 ms
Tempo máximo	2,781.00 ms	13,245.67 ms	35,307.00 ms
Vazão	1.84 solicitações / s	1.82 solicitações / s	1.77 solicitações / s

Análise dos Resultados

Leitura com 2 parâmetros e filtro de campos



Análise dos Resultados

Leitura com 2 parâmetros e filtro de campos



	100 usuários	500 usuários	1000 usuários
Média	1,578 ms	6,7620.67 ms	18,586.33 ms

+ 328.56%

+ 1,077.84%

	100 usuários	500 usuários	1000 usuários
Vazão	1.84 solicitações / s	1.82 solicitações / s	1.77 solicitações / s

- 0.78%

- 3.86%

Análise dos Resultados

Leitura com 1 parâmetro, filtro de campos e ordenação



Leitura com 1 parâmetro, filtro de campos e ordenação

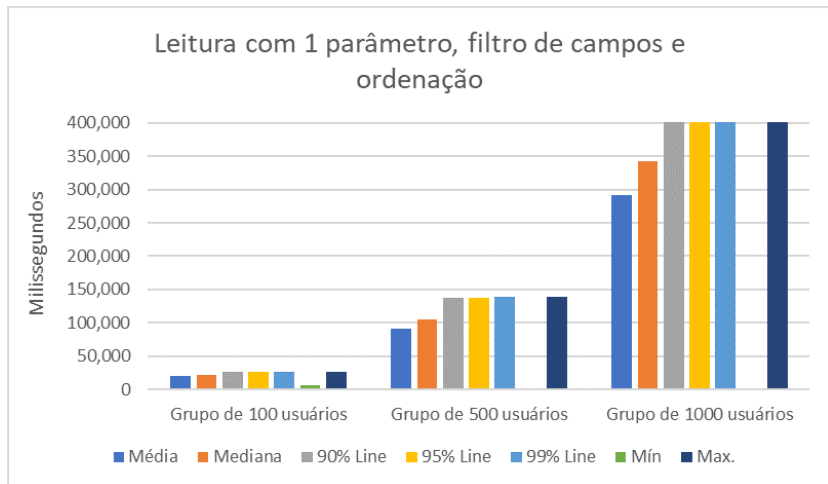
	100 usuários	500 usuários	1000 usuários
Média	19,753.67 ms	91,175.00 ms	291,379.33 ms
Mediana	21,311.00 ms	105,084.67 ms	342,336.00 ms
Percentil 90%	26,490.67 ms	136,793.00 ms	407,935.00 ms
Percentil 95%	26,695.67 ms	137,960.33 ms	410,334.67 ms
Percentil 99%	26,801.33 ms	138,791.00 ms	412,180.67 ms
Tempo mínimo	5,705.33 ms	794.00 ms	1,221.67 ms
Tempo máximo	26,877.67 ms	139,072.33 ms	412,959.00 ms
Vazão	1.40 solicitações / s	1.36 solicitações / s	1.16 solicitações / s

Análise dos Resultados

Leitura com 1 parâmetro, filtro de campos e ordenação



24



Análise dos Resultados

Leitura com 1 parâmetro, filtro de campos e ordenação



	100 usuários	500 usuários	1000 usuários
Média	19,753.67 ms	91,175.00 ms	291,379.33 ms

+ 361.56%

+ 1,375.06%

	100 usuários	500 usuários	1000 usuários
Vazão	1.40 solicitações / s	1.36 solicitações / s	1.16 solicitações / s

- 2.75%

- 17.10%

Análise dos Resultados

Inserção de 1 documento por usuário



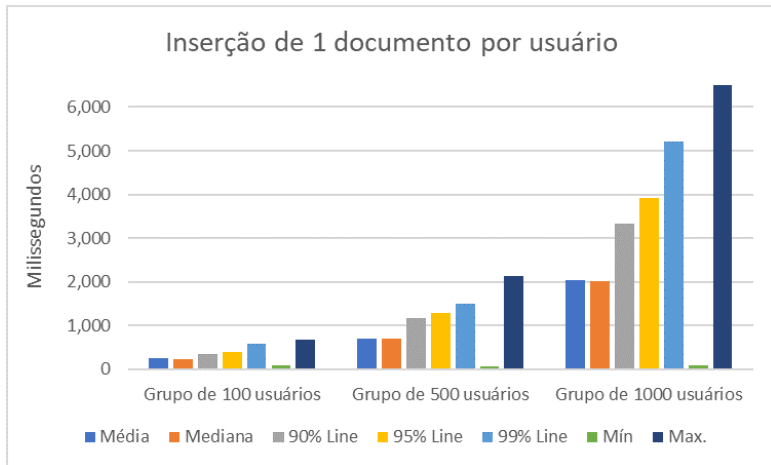
Inserção de 1 documento por usuário			
	100 usuários	500 usuários	1000 usuários
Média	245.67 ms	713.67 ms	2,043.67 ms
Mediana	234.33 ms	702.67 ms	2,020.00 ms
Percentil 90%	352.00 ms	1,166.33 ms	3,345.00 ms
Percentil 95%	401.00 ms	1,285.00 ms	3,914.00 ms
Percentil 99%	587.00 ms	1,499.67 ms	5,201.33 ms
Tempo mínimo	83.67 ms	73.67 ms	86.00 ms
Tempo máximo	677.67 ms	2,131.33 ms	6,513.00 ms
Vazão	1.58 solicitações / s	1.36 solicitações / s	1.16 solicitações / s

Análise dos Resultados

Inserção de 1 documento por usuário



27



Análise dos Resultados

Inserção de 1 documento por usuário



	100 usuários	500 usuários	1000 usuários
Média	245.67 ms	713.67 ms	2,043.67 ms

+ 190.50%

+ 731.89%

	100 usuários	500 usuários	1000 usuários
Vazão	1.58 solicitações / s	1.36 solicitações / s	1.16 solicitações / s

- 13.83%

- 26.54%



- ▶ **Cerca de 1:30 hora** de tempo de execução
- ▶ O tempo total para a **população inicial** do BD foi de **8,445.67 ms**, sendo aproximadamente **0.3 ms por documento**
- ▶ Nas operações de **leitura**, duplicando o número de usuários de 500 para 1000, o desempenho do sistema **cai pelo menos 3 vezes**
- ▶ O **tempo de indexação** para um campo é de **200ms** aproximadamente, mesmo com 28.061 documentos
- ▶ Para as operações de **leitura com 2 parâmetros** o tempo foi semelhante, **aumentando** cerca de **2 segundos** ao acrescentar filtro
- ▶ O impacto da ordenação é **extremo**
 - ▶ Leitura com 2 parâmetros e filtro de campos para 100 usuários: 1,578.00 ms
 - ▶ Leitura com 1 parâmetro, filtro de campos e ordenação para 100 usuários: 19,753.67 ms



► Leitura mais complexa

	100 usuários	500 usuários	1000 usuários
Média	19,753.67 ms	91,175.00 ms	291,379.33 ms
		+ 361.56%	+ 1,375.06%
	100 usuários	500 usuários	1000 usuários
Vazão	1.40 solicitações / s	1.36 solicitações / s	1.16 solicitações / s
		- 2.75%	- 17.10%

► Inserção de 1 documento

	100 usuários	500 usuários	1000 usuários
Média	245.67 ms	713.67 ms	2,043.67 ms
		+ 190.50%	+ 731.89%
	100 usuários	500 usuários	1000 usuários
Vazão	1.58 solicitações / s	1.36 solicitações / s	1.16 solicitações / s
		- 13.83%	- 26.54%



- ▶ Tempo médio de resposta para leitura ao duplicar o número de usuários de 500 para 1000 **triplica** ao alterar o número de usuários
- ▶ Taxa de vazão para leitura **diminui 6 vezes** ao duplicar o número de usuários
- ▶ Tempo médio de resposta ao duplicar o número de usuários de 500 para 1000 ao inserir 1 documento **aumenta 6 vezes**
- ▶ Taxa de vazão para escrita **diminui 2 vezes** ao duplicar o número de usuários
- ▶ **No entanto, a vazão para a leitura mais complexa é igual a vazão de escrita de 1 documento**
- ▶ CouchDB prioriza disponibilidade dos dados



- ▶ Temporizador afeta tempo das requisições
- ▶ Execução no modo **Standalone**
- ▶ JMeter e leitura de JSON **gigante** dos dados

A decorative graphic consisting of several overlapping, curved lines in shades of orange, peach, and light blue. The lines sweep from the left side of the frame towards the right, creating a sense of motion and elegance. Small white dots are scattered along the curves, adding a sparkling effect.

Obrigada!