CSC-27 – Computação Distribuída

Laboratório 2: Map Reduce

Nomes: Daniel Araujo Cavassani | Caue Marçal Guimarães | (COMP 25)

Data: 22/10/2024

PARTE 1: Trabalhando em modo sequencial

Rodando o arquivo sequencial com o comando mais básico, obtemos os resultados esperados e que poderemos usar para comparação nos próximos testes. Obs: o arquivo teste2.txt é cerca de 10x maior que o teste.txt, e foi gerado aleatoriamente.

.\wordcount -mode sequential -file files/teste.txt

```
PS C:\Users\danie\OneDrive\DanxC-Notebook-1\Documentos\Academics\ITA\8° Sem (4° COMP)\CSC-27 (Distributed Computing)\Go Labs and Activities\Lab 2
labMapReduce\wordcount> .\wordcount -mode sequential -file files/teste.txt
2024/10/22 00:17:11 Running in sequential mode.
2024/10/22 00:17:11 Running RunSequential...
2024/10/22 00:17:11 Running RunSequential...
2024/10/22 00:17:11 Running RunSequential...
2024/10/22 00:17:11 Running RunSequential...
2024/10/22 00:17:11 Fanning in file map\map-0
[Teste para ver o correto funcionamento da contagem de palavras Por exemplo a palavra teste deve ocorrer apenas tres vezes sendo que a ultima ocor
 rencia e esta testel
[{teste 1} {para 1} {ver 1} {o 1} {correto 1} {funcionamento 1} {da 1} {contagem 1} {de 1} {palavras 1} {por 1} {exemplo 1} {a 1} {palavra 1} {tes 1} {deve 1} {ocorrer 1} {apenas 1} {tres 1} {vezes 1} {sendo 1} {que 1} {a 1} {ultima 1} {correncia 1} {e 1} {esta 1} {teste 1}]
[{contagem 1} {a 2} {tres 1}]
2024/10/22 00:17:11 Fanning out file result\result-0
[{deve 1} {sendo 1}]
[{ocorrer 1} {para 1} {ver 1} {da 1} {exemplo 1}]
[{vezes 1} {ultima 1} {teste 3} {funcionamento 1} {de 1} {por 1}]
[{correto 1} {que 1} {ocorrencia 1} {e 1} {o 1} {palavras 1} {apenas 1} {esta 1}]
 2024/10/22 00:17:11 Fanning out file result\result-1
2024/10/22 00:17:11 Fanning out file result\result-2
 2024/10/22 00:17:11 Fanning out file result\result-3
2024/10/22 00:17:11 Fanning out file result\result-4
```

.\wordcount -mode sequential -file files/teste2.txt

```
PS C:\Users\danie\OneDrive\DanxC-Notebook-1\Documentos\Academics\ITA\8° Sem (4° COMP)\CSC-27 (Distributed Computing)\Go Labs and Activities\Lab 2\ labMapReduce\wordcount> .\wordcount -mode sequential -file files/teste2.txt 2024/10/22 00:43:48 Running in sequential mode. 2024/10/22 00:43:48 Running in sequential ... 2024/10/22 00:43:48 Fanning in file map\map-0 [{testar 5} {garantir 1} {após 1} {fosse 1} {a 5} {tudo 1} {tinha 1} {então 1} {havia 2} {importantes 1} {precisava 1} {fim 1} {concluído 1} {repe tições 3} {dezenas 1} {outras 1} {ecoando 1} [2024/10/22 00:43:48 Fanning out file result\result-0 [{busca 1} {claro 3} {testado 1} {em 1} {assim 1} {precisão 1} {ā 1} {muitas 1} {ser 2} {linhas 1} {essencial 1} {mais 8} {eficiência 1} {sabe 1} {exigia 1} {código 1} {vasto 1} [{história 1} {do 2} {será 2} {aparecer 1} {só 2} {se 1} {funcionava 1} {resultados 1} {final 1} {dados 1} {prepare 1} {atenção 1} {resultado 1} {para 1} {os 2} {é 1} {também 1} {sucesso 8} {talvez 1} {início 2} {importante 1} [{no 1} {começaram 1} {continuava 2} {gritavam 1} {essena 1} {acabou 1} {de 9} {paciência 1} {sempre 3} {plo 1} {um 4} {teste 22} {testes 5} {revelou 1} {vezes 1} {funcionamento 1} {frase 1} {declarou 1} {virão 1} {termina 1} {finalmente 2} {estava 1} {último 2} {quem 1} {ciclo 2} [{abós 5} {como 1} {cada 3} {que 5} {simples 1} {palavra 2} {paravam 1} {linha 1} {vez 1} {era 5} {depois 1} {palavras 3} {novo 2} {com 1} {está 1} {começando 1} {nunca 1} {chave 1} {mas 5} {começou 1} {repetida 1} {três 1} {as 1} {garantido 1} {duas 1} {universo 1} {apenas 1} {muitos 1} {sis tema 3} {próximo 1} {seria 1} {o 19} {e 7} {uma 2} {objetivo 1} {2024/10/22 00:43:48 Fanning out file result\result-1 2024/10/22 00:43:48 Fanning out file result\result-2 2024/10/22 00:43:48 Fanning out file result\result-3 2024/10/22 00:43:48 Fanning out file result\result-4 2024/10/22 00:43:48
```

Tarefa 1.3: Rode o programa com o arquivo teste.txt. Use o seguinte comando para rodar:

wordcount.exe -mode sequential -file files/teste.txt -chunksize 100 -reducejobs 2

```
PS C:\Users\danie\OneDrive\DanxC-Notebook-1\Documentos\Academics\ITA\8° Sem (4° COMP)\CSC-27 (Distributed Computing)\Go Labs and Activities\Lab 2\
labMapReduce\wordcount> .\wordcount -mode sequential -file files/teste.txt -chunksize 100 -reducejobs 2
2024/10/22 00:22:14 Running in sequential mode.
2024/10/22 00:22:14 Running in file map\map-0
2024/10/22 00:22:14 Fanning in file map\map-0
2024/10/22 00:22:14 Fanning in file map\map-1
[Teste para ver o correto funcionamento da contagem de palavras Por exemplo a palavra teste]
[{teste 1} {para 1} {ver 1} {o 1} {correto 1} {funcionamento 1} {da 1} {contagem 1} {de 1} {palavras 1} {por 1} {exemplo 1} {a 1} {palavra 1} {teste 1}
[deve ocorrer apenas tres vezes sendo que a ultima ocorrencia e esta teste]
[{deve 1} {ocorrer 1} {apenas 1} {tres 1} {vezes 1} {sendo 1} {que 1} {a 1} {ultima 1} {ocorrencia 1} {e 1} {esta 1} {teste 1}
[{ver 1} {por 1} {teste 3} {a 2} {vezes 1} {que 1} {o 1} {da 1} {palavra 1} {de 1} {sendo 1} {correncia 1} {e 1} {esta 1}
[{ver 1} {por 1} {teste 3} {a 2} {vezes 1} {que 1} {o 1} {da 1} {palavra 1} {de 1} {sendo 1} {correncia 1} {e 1} {esta 1}
[{palavras 1} {tres 1} {ultima 1} {para 1} {funcionamento 1} {contagem 1} {exemplo 1} {deve 1} {correr 1} {apenas 1} {correto 1}
[{palavras 1} {tres 1} {ultima 1} {para 1} {funcionamento 1} {contagem 1} {exemplo 1} {deve 1} {correr 1} {apenas 1} {correto 1}
[{palavras 1} {tres 1} {ultima 1} {para 1} {funcionamento 1} {contagem 1} {exemplo 1} {deve 1} {correr 1} {apenas 1} {correto 1}
[{palavras 1} {tres 1} {ultima 1} {para 1} {funcionamento 1} {contagem 1} {exemplo 1} {deve 1} {correr 1} {apenas 1} {correto 1}
[{palavras 1} {teste 3} {apenas 1} {teste 3}
[{palavras 1} {teste 3} {apenas 1} {teste 3}
[{palavras 1} {teste 3} {apenas 1} {teste 4}
[{palavras 1} {teste 3}
[{
```

Analisando o mapeamento final, podemos verificar que as palavras foram corretamente extraídas e divididas em dois chunks (map-0 e map-1), com todas convertidas para minúsculas. Além disso, a redução em si foi correta, pois a contagem de palavras também foi corretamente consolidada. Se olharmos a palavra "teste", por exemplo, ela apareceu 3 vezes (conforme o esperado), e todas as outras palavras foram contadas de forma apropriada.

Tarefa 1.4: Rode o programa com o arquivo teste.txt. Use o seguinte comando para rodar:

Mantendo reducejobs constante:

.\wordcount -mode sequential -file files/teste2.txt -chunksize 100 -reducejobs 2

```
PS C:\Users\danie\OneDrive\DanxC-Notebook-1\Documentos\Academics\ITA\8° Sem (4° COMP)\CSC-27 (Distributed Computing)\Go Labs and Activities\Lab 2\labMapReduce\wordcount>.\wordcount -mode sequential -file files/teste2.txt -chunksize 100 -reducejobs 2
 2024/10/22 00:48:38 Running in sequential mode.
 2024/10/22 00:48:38 Running RunSequential...
2024/10/22 00:48:38 Fanning in file map\map-0
 2024/10/22 00:48:38 Fanning in file map\map-1
2024/10/22 00:48:38 Fanning in file map\map-2
 2024/10/22 00:48:38 Fanning in file map\map-3
2024/10/22 00:48:38 Fanning in file map\map-4 2024/10/22 00:48:38 Fanning in file map\map-5
2024/10/22 00:48:38 Fanning in file map\map-6
2024/10/22 00:48:38 Fanning in file map\map-7
 2024/10/22 00:48:38 Fanning in file map\map-8
 2024/10/22 00:48:38 Fanning in file map\map-9
  2024/10/22 00:48:38 Fanning in file map\map-10
 2024/10/22 00:48:38 Fanning in file map\map-11
  2024/10/22 00:48:38 Fanning in file map\map-12
 2024/10/22 00:48:38 Fanning in file map\map-13
  2024/10/22 00:48:38 Fanning in file map\map-14
 2024/10/22 00:48:38 Fanning in file map\map-15
  [{declarou 1} {resultado 1 | {ecoando 1 | {a 1 | {nunca 1 | {sucesso 8 | {aparecer 1 | {no 1 | {também 1 | {chave 1 | {que 5 | {claro 3 | {com 1 | {vez 1 | {ho 1 | {também 1 | {chave 1 | {que 5 | {claro 3 | {com 1 | {vez 1 | {ho 1 | {também 1 | {chave 1 | {também 1 | {também 1 | {chave 1 | {também 1 | {também 1 | {chave 1 | {também 1 | {também 1 | {chave 1 | {também 1 | {tamb
Via 2} {testar 5} {muitos 1} {palavra 2} {duas 1} {do 2} {não 5} {e 7} {começaram 1} {linhas 1} {vasto 1} {simples 1} {exigia 1} {o 19} {mas 5} {i mportante 1} {finalmente 2} {acabou 1} {repetições 3} {garantido 1} {de 9} {gritavam 1} {assim 1} {próximo 1} {concluído 1} {teste 22} {a 5} {esse ncial 1} {universo 1} {cada 3} {precisão 1} {atenção 1} {dados 1} {frase 1} {prepare 1} {paciência 1} {muitas 1} {uma 2} {vezes 1} {só 2} {sabe 1}
 2024/10/22 00:48:39 Fanning out file result\result-0
 [{quem 1} {para 1} {resultados 1} {como 1} {testado 1} {espera 1} {é 1} {palavras 3} {objetivo 1} {seria 1} {testes 5} {será 2} {fosse 1} {virão 1} {ser 2} {começando 1} {outras 1} {continuava 2} {sempre 3} {linha 1} {último 2} {apenas 1} {revelou 1} {os 2} {garantir 1} {repetida 1} {funcion amento 1} {as 1} {depois 1} {começou 1} {após 1} {então 1} {tinha 1} {dezenas 1} {final 1} {precisava 1} {história 1} {busca 1} {eficiência 1} {código 1} {tudo 1} {um 4} {sistema 3} {funcionava 1} {importantes 1} {mais 8} {era 5} {paravam 1} {em 1} {termina 1} {está 1} {três 1} {novo 2} {fim 1} {cíclo 2} {se 1} {início 2} {pelo 1} {estava 1} {talvez 1}]
```

```
PS C:\Users\danie\OneDrive\DanxC-Notebook-1\Documentos\Academics\ITA\8° Sem (4° COMP)\CSC-27 (Distributed Computing)\Go Labs and Activities\Lab 2\
labMapReduce\wordcount> .\wordcount -mode sequential -file files/teste2.txt -chunksize 500 -reducejobs 2
2024/10/22 00:49:15 Running in sequential mode.
2024/10/22 00:49:15 Funning in file map\map-0
2024/10/22 00:49:15 Fanning in file map\map-1
2024/10/22 00:49:15 Fanning in file map\map-1
2024/10/22 00:49:15 Fanning in file map\map-2
2024/10/22 00:49:15 Fanning in file map\map-3
{acabou 1} {acabou 1} {acabou 1} {acabou 1} {repetições 3} {so 2} {finalmente 2} {duas 1} {claro 3} {dados 1} {chave 1} {exigia 1} {declarou 1} {nunca 1} {mas 5} {sucesso 8} {palavra 2} {vezes 1} {resultado 1} {precisão 1} {importante 1} {essencial 1} {também 1} {com 1} {simples 1} {paciência 1} {ciclo 2} {começando 1} {sempre 3} {precisava 1} {será 2} {talvez 1} {espera 1} {apenas 1} {continuava 2} {mais 8} {pelo 1} {eficiência 1} {teste 5} {linha 1} {virão 1} {para 1} {sistema 3} {depois 1} {com 0} {os 2} {é 1} {código 1} {revelou 1} {as 1} {história 1} {era 5} {palavras 3} {im portantes 1} {em 1} {um 4} {termina 1} {ser 2} {funcionamento 1} {busca 1} {fin 1} {funcionava 1} {dezenas 1} {está 1} {começou 1} {estava 1} {entava 1} {entava 2} {in 1} {todo 1} {se 1} {resultados 1} {três 1} {após 1} {fosse 1} {os 2} {início 2} {garantir 1} {quem 1} {testado 1} {outras 1} {objetivo 1} {final 1} {todo 1} {se 1} {resultados 1} {três 1} {após 1} {fosse 1} {code 2} {início 2} {garantir 1} {quem 1} {testado 1} {outras 1} {objetivo 1} {final 1} {todo 1} {se 1} {resultados 1} {três 1} {após 1} {fosse 1} {code 2} {início 2} {garantir 1} {quem 1} {testado 1} {code 2} {trava 1} {code 2} {tr
```

.\wordcount -mode sequential -file files/teste2.txt -chunksize 1000 -reducejobs 2

```
PS C:\Users\danie\OneDrive\DanxC-Notebook-1\Documentos\Academics\ITA\8° Sem (4° COMP)\CSC-27 (Distributed Computing)\Go Labs and Activities\Lab 2\ labMapReduce\wordcount> .\wordcount -mode sequential -file files/teste2.txt -chunksize 1000 -reducejobs 2
2024/10/22 00:49:48 Running in sequential mode.
2024/10/22 00:49:48 Running RunSequential...
2024/10/22 00:49:48 Fanning in file map\map-0
2024/10/22 00:49:48 Fanning in file map\map-0
2024/10/22 00:49:48 Fanning in file map\map-1
[{concluido 1} {mas 5} {vasto 1} {cada 3} {frase 1} {o 19} {aparecer 1} {a 5} {muitas 1} {teste 22} {testar 5} {não 5} {finalmente 2} {resultado 1} {duas 1} {nunca 1} {no 1} {de 9} {linhas 1} {declarou 1} {havia 2} {uma 2} {vez 1} {sucesso 8} {prepare 1} {vezes 1} {acabou 1} {exigia 1} {e 7} {paciência 1} {gritavam 1} {chave 1} {que 5} {palavra 2} {atenção 1} {começaram 1} {ecoando 1} {assim 1} {precisão 1} {sabe 1} {muitos 1} {claro 3} {2024/10/22 00:49:48 Fanning out file result\result-0
[{as 1} {apenas 1} {resultados 1} {funcionamento 1} {sempre 3} {fosse 1} {testes 5} {testado 1} {história 1} {para 1} {é 1} {objetivo 1} {será 2} {termina 1} {começando 1} {revelou 1} {continuava 2} {eficiência 1} {linha 1} {estava 1} {sistema 3} {após 1} {paravam 1} {viñão 1} {incio 2} {pe 10 1} {seria 1} {dezenas 1} {precisava 1} {cilo 2} {se 1} {um 4} {final 1} {código 1} {os 2} {ser 2} {quem 1} {repetida 1} {talvez 1} {então 1} {fin 1} {espera 1} {começou 1} {busca 1} {tinha 1} {três 1} {outras 1} {mais 8} {palavras 3} {tudo 1} {último 2} {com 1} {em 1} {garantir 1} {depo is 1} {importantes 1} {está 1} {era 5} {funcionava 1} {novo 2}]
```

Variando chunksize, observamos que quanto menor o valor, mais fragmentado o arquivo se torna, resultando em um maior número de operações de map. Dessa forma, aumentar o chunksize trouxe benefícios ao melhorar a eficiência, pois menos operações de map são necessárias. Além disso, independentemente do tamanho do chunksize, a contagem de palavras permaneceu precisa.

Mantendo chunksize constante:

.\wordcount -mode sequential -file files/teste2.txt -chunksize 500 -reducejobs 2

```
PS C:\Users\danie\OneDrive\DanxC-Notebook-1\Documentos\Academics\ITA\8° Sem (4° COMP)\CSC-27 (Distributed Computing)\Go Labs and Activities\Lab 2\
labMapReduce\wordcount> .\wordcount -mode sequential -file files/teste2.txt -chunksize 500 -reducejobs 2
2024/10/22 00:54:14 Running in sequential mode.
2024/10/22 00:54:14 Fanning in file map\map-0
2024/10/22 00:54:14 Fanning in file map\map-1
2024/10/22 00:54:14 Fanning in file map\map-1
2024/10/22 00:54:14 Fanning in file map\map-2
2024/10/22 00:54:14 Fanning in file map\map-3
[{concluido 1} {vez 1} {importante 1} {mas 5} {simples 1} {começaram 1} {sucesso 8} {precisão 1} {duas 1} {claro 3} {garantido 1} {testar 5} {o 19} {com 1} {a 5} {paciência 1} {palavra 2} {só 2} {gritavam 1} {que 5} {uma 2} {também 1} {nunca 1} {muitos 1} {linhas 1} {declarou 1} {vezes 1} {dados 1} {cada 3} {atenção 1} {teste 22} {prepare 1} {havia 2} {chave 1} {repetições 3} {finalmente 2} {resultado 1} {sabe 1} {muitas 1} {acabou 1} {2024/10/22 00:54:14 Fanning out file result\result-0} {testes 21} {se 1} {termina 1} {três 1} {quem 1} {os 2} {importantes 1} {linha 1} {ciclo 2} {fosse 1} {funcioname not 1} {objetivo 1} {história 1} {fim 1} {as 1} {em 1} {univa 2} {um 4} {resultados 1} {outras 1} {novo 2} {depois 1} {palavras 3} {começando 1} {começanal 1} {cocando 1} {coca
```

.\wordcount -mode sequential -file files/teste2.txt -chunksize 500 -reducejobs 5

```
PS C:\Users\danie\OneDrive\DanxC-Notebook-1\Documentos\Academics\ITA\8° Sem (4° COMP)\CSC-27 (Distributed Computing)\Go Labs and Activities\Lab 2\labMapReduce\wordcount> .\wordcount -mode sequential -file files/teste2.txt -chunksize 500 -reducejobs 5
2024/10/22 00:54:39 Running in sequential mode.
2024/10/22 00:54:40 Fanning in file map\map-0
2024/10/22 00:54:40 Fanning in file map\map-0
2024/10/22 00:54:40 Fanning in file map\map-2
2024/10/22 00:54:40 Fanning out file result\tresult-0
[{setar 5} {a 5} {importantes 1} {após 1}]
{seba 1} {muitas 1} {assim 1} {precisão 1} {exigia 1} {essencial 1} {em 1} {eficiência 1} {à 1} {código 1} {vasto 1} {testado 1} {busca 1} {sabe 1} {muitas 1} {assim 1} {cspara 1} {so 2} {o 2} {aparecer 1} {é 1} {resultado 1} {final 1} {dados 1} {funcionava 1} {também 1} {história 1} {do 2} {tal vez 1} {atenção 1} {importante 1} {para 1} {resultados 1} {sucesso 8} {início 2} {[erevelou 1} {funcionamento 1} {continuava 2} {sempre 3} {gritavam 1} {de 9} {testes 5} {finalmente 2} {pelo 1} {frase 1} {termina 1} {no 1} {um 4} {testes 2} {quem 1} {estava 1} {espera 3} {gritavam 1} {de 9} {testes 5} {muitos 1} {acabou 1} {virão 1} {2024/10/22 00:54:40 Fanning out file result\result-1} {seria 1} {com 1} {sistema 3} {simples 1} {duas 1} {as 1} {garantido 1} {apenas 1} {que 5} {e 7} {depois 1} {começou 1} {três 1} {nunca 1} {vez 1} {o 19} {uma 2} {palavras 3} {paravam 1} {universo 1} {próximo 1} {chave 1} {não 5} {novo 2} {cada 3} {li nha 1} {começon 1} {palavra 2} {como 1} {objetivo 1} {está 1} {começon 0} {color 40 Fanning out file result\result-2} {color 54:40 Fanning out file result\
```

.\wordcount -mode sequential -file files/teste2.txt -chunksize 500 -reducejobs 10

```
6 C:\Users\danie\OneDrive\DanxC-Notebook-1\Documentos\Academics\ITA\8° Sem (4° COMP)\CSC-27 (Distributed Computing)\Go Labs and Activities\Lab
 labMapReduce\wordcount> .\wordcount -mode sequential -file files/teste2.txt -chunksize 500 -reducejobs 10
2024/10/22 00:54:56 Running in sequential mode.
 2024/10/22 00:54:56 Running RunSequential...
2024/10/22 00:54:56 Fanning in file map\map-0
 2024/10/22 00:54:56 Fanning in file map\map-1
 2024/10/22 00:54:56 Fanning in file map\map-2
  2024/10/22 00:54:56 Fanning in file map\map-3
 [{havia 2} {testar 5} {a 5} {repetições 3} {ecoando 1} {concluído 1}]
   024/10/22 00:54:57 Fanning out file result\result-0
[{testado 1} {em 1} {busca 1} {eficiência 1} {ser 2} {mais 8} {código 1}]
[{aparecer 1} {só 2} {também 1} {dados 1} {prepare 1} {importante 1} {atenção 1} {resultado 1} {sucesso 8} {do 2}]
[{estava 1} {último 2} {virão 1} {sempre 3} {espera 1} {um 4} {testes 5} {continuava 2} {termina 1} {ciclo 2} {revelou 1} {funcionamento 1} {pelo
  L} {quem 1}]
[{garantido 1} {nunca 1} {mas 5} {não 5} {palavra 2} {uma 2} {cada 3} {com 1} {vez 1} {o 19} {simples 1} {e 7} {duas 1} {universo 1} {chave 1} {priverso 1} {chave 1} {c
[{gainties 1} {que 5} {muitos 1}]

2024/10/22 00:54:57 Fanning out file result\result-1

[{precisava 1} {tinha 1} {fosse 1} {então 1} {importantes 1} {após 1} {outras 1} {tudo 1} {fim 1} {garantir 1} {dezenas 1}]

[{muitas 1} {essencial 1} {claro 3} {assim 1} {exigia 1} {linhas 1} {vasto 1} {precisão 1} {à 1} {sabe 1}]

2024/10/22 00:54:57 Fanning out file result\result-2
[{para 1} {funcionava 1} {os 2} {resultados 1} {será 2} {início 2} {final 1} {história 1} {talvez 1} {se 1} {é 1}] 2024/10/22 00:54:57 Fanning out file result\result-3
 2024/10/22 00:54:57 Fanning out file result\result-4
[{no 1} {teste 22} {começaram 1} {gritavam 1} {declarou 1} {acabou 1} {de 9} {paciência 1} {finalmente 2} {vezes 1} {frase 1}] 2024/10/22 00:54:57 Fanning out file result\result-5
 2024/10/22 00:54:57 Fanning out file result\result-6
 [{sistema 3} {depois 1} {começou 1} {três 1} {começando 1} {repetida 1} {como 1} {novo 2} {está 1} {era 5} {paravam 1} {linha 1} {seria 1} {palavr
  s 3} {objetivo 1} {as 1} {apenas 1}]
 2024/10/22 00:54:57 Fanning out file result\result-7
 2024/10/22 00:54:57 Fanning out file result\result-8
2024/10/22 00:54:57 Fanning out file result\result-9
```

Variando reducejobs, observamos que um maior número de jobs distribui mais eficientemente o processamento das palavras entre os jobs de redução. Além disso, essa distribuição não afeta significativamente a contagem de palavras em si, mas sim como as tarefas são divididas entre os jobs. Dessa forma, essa variação é útil em contextos onde há muitos workers disponíveis, pois garante que o trabalho seja distribuído de maneira mais equilibrada, permitindo um cenário de maior paralelização.

PARTE 1: Trabalhando em modo distribuído

Tarefa 2.2: Alteração do código para lidar com operações que falham

• Criar um channel para guardar as operações falhas:

- Um novo canal failedOperationsChan foi adicionado na estrutura Master para armazenar operações de Map e Reduce que falharem durante sua execução.
- Esse canal foi inicializado no construtor newMaster() com um buffer de tamanho definido pela constante RETRY_OPERATION_BUFFER.

• Na função runOperation:

- Quando a operação falha (detectada por um erro de comunicação com o RemoteWorker), a operação é enviada para o canal failedOperationsChan, garantindo que será reprocessada.
- O worker falho também é enviado para o canal failedWorkerChan para ser tratado posteriormente.
- O comando wg.Done() foi colocado com defer no início da função para garantir que o contador de goroutines seja decrementado assim que a operação for concluída, seja ela bem-sucedida ou falha.
- Quando a operação é bem-sucedida, o RemoteWorker é colocado de volta no canal idleWorkerChan, permitindo que ele execute outra operação.

Na função schedule:

- Antes de despachar as operações iniciais, uma nova goroutine é iniciada para monitorar o canal failedOperationsChan. Essa goroutine é responsável por reprocessar as operações que falharem.
- As operações de Map ou Reduce são despachadas para os workers, e o sistema aguarda que todas as operações, incluindo as falhas reprocessadas, sejam concluídas utilizando o WaitGroup.
- A função aguarda que todas as operações sejam concluídas com wg.Wait() antes de passar para o próximo estágio (Reduce).

Tarefa 2.3: Um worker falho e outro normal. Emule falha numa operação de map. OBS: optei por manter o pg1342.txt por apresentar um resultado mais legal do que o teste.txt

.\wordcount -mode distributed -type master -file files/pg1342.txt -chunksize 102400 -reducejobs 5

```
√ TERMINAL

 PS C:\Users\danie\OneDrive\DanxC-Notebook-1\Documentos\Academics\ITA\8° Sem (4° COMP)\CSC-27 (Distr
ibuted Computing)\Go Labs and Activities\Lab 2\labMapReduce\wordcount> .\wordcount -mode distribute
 d -type master -file files/pg1342.txt -chunksize 102400 -reducejobs 5
 2024/10/22 03:02:17 Running in distributed mode.
 2024/10/22 03:02:17 NodeType: master
 2024/10/22 03:02:17 Reduce Jobs: 5
 2024/10/22 03:02:17 Address: localhost
 2024/10/22 03:02:17 Port: 5000
 2024/10/22 03:02:17 File: files/pg1342.txt
 2024/10/22 03:02:17 Chunk Size: 102400
 2024/10/22 03:02:17 Running Master on localhost:5000
 2024/10/22 03:02:17 Scheduling Worker.RunMap operations
 2024/10/22 03:02:17 Accepting connections on 127.0.0.1:5000
 2024/10/22 03:02:17 Registering worker '0' with hostname 'localhost:50001'
 2024/10/22 03:02:17 Running Worker.RunMap (ID: '0' File: 'map\map-0' Worker: '0')
 2024/10/22 03:02:17 Running Worker.RunMap (ID: '1' File: 'map\map-1' Worker: '0')
 2024/10/22 03:02:17 Running Worker.RunMap (ID: '2' File: 'map\map-2' Worker: '0')
 2024/10/22 03:02:17 Operation Worker.RunMap '2' Failed. Error: read tcp 127.0.0.1:64248->127.0.0.1:
 50001: wsarecv: An existing connection was forcibly closed by the remote host.
 2024/10/22 03:02:17 Removing worker 0 from master list.
 2024/10/22 03:02:18 Registering worker '1' with hostname 'localhost:50002'
 2024/10/22 03:02:18 Running Worker.RunMap (ID: '3' File: 'map\map-3' Worker: '1')
 2024/10/22 03:02:19 Running Worker.RunMap (ID: '2' File: 'map\map-2' Worker: '1')
 2024/10/22 03:02:19 Running Worker.RunMap (ID: '4' File: 'map\map-4' Worker: '1')
 2024/10/22 03:02:19 Running Worker.RunMap (ID: '5' File: 'map\map-5' Worker: '1')
 2024/10/22 03:02:19 Running Worker.RunMap (ID: '6' File: 'map\map-6' Worker: '1')
 2024/10/22 03:02:19 7x Worker.RunMap operations completed
 2024/10/22 03:02:20 Scheduling Worker.RunReduce operations
 2024/10/22 03:02:20 Running Worker.RunReduce (ID: '0' File: 'reduce\reduce-0' Worker: '1')
 2024/10/22 03:02:20 Running Worker.RunReduce (ID: '1' File: 'reduce\reduce-1' Worker: '1')
 2024/10/22 03:02:20 Running Worker.RunReduce (ID: '2' File: 'reduce\reduce-2' Worker: '1')
 2024/10/22 03:02:20 Running Worker.RunReduce (ID: '3' File: 'reduce\reduce-3' Worker: '1')
 2024/10/22 03:02:20 Running Worker.RunReduce (ID: '4' File: 'reduce\reduce-4' Worker: '1')
 2024/10/22 03:02:20 5x Worker.RunReduce operations completed
 2024/10/22 03:02:20 Closing Remote Workers.
 2024/10/22 03:02:20 Done.
```

Note que, nesse processo do master, há uma falha na operação de map de ID 2, que é seguida pela remoção do worker 0 (falho). Logo após, o worker 1 entra em ação, completando as tarefas faltantes e retrabalhando na tarefa de map de ID 2, que foi deixada sem ser completada pelo worker 0 falho.

Tarefa 2.4: Um worker falho e outro normal. Emule falha numa operação de reduce. OBS: optei por manter o pg1342.txt por apresentar um resultado mais legal do que o teste.txt

.\wordcount -mode distributed -type master -file files/pg1342.txt -chunksize 102400 -reducejobs 5

```
✓ TERMINAL

 PS C:\Users\danie\OneDrive\DanxC-Notebook-1\Documentos\Academics\ITA\8° Sem (4° COMP)\CSC-27 (Distr
 ibuted Computing)\Go Labs and Activities\Lab 2\labMapReduce\wordcount> .\wordcount -mode distribute
d -type master -file files/pg1342.txt -chunksize 102400 -reducejobs 5
 2024/10/22 03:09:46 Running in distributed mode.
 2024/10/22 03:09:46 NodeType: master
 2024/10/22 03:09:46 Reduce Jobs: 5
 2024/10/22 03:09:46 Address: localhost
 2024/10/22 03:09:46 Port: 5000
 2024/10/22 03:09:46 File: files/pg1342.txt
 2024/10/22 03:09:46 Chunk Size: 102400
 2024/10/22 03:09:46 Running Master on localhost:5000
 2024/10/22 03:09:46 Scheduling Worker.RunMap operations
 2024/10/22 03:09:46 Accepting connections on 127.0.0.1:5000
 2024/10/22 03:09:46 Registering worker '0' with hostname 'localhost:50002'
 2024/10/22 03:09:46 Running Worker.RunMap (ID: '0' File: 'map\map-0' Worker: '0') 2024/10/22 03:09:46 Running Worker.RunMap (ID: '1' File: 'map\map-1' Worker: '0')
 2024/10/22 03:09:47 Running Worker.RunMap (ID: '2' File: 'map\map-2' Worker: '0')
 2024/10/22 03:09:47 Registering worker '1' with hostname 'localhost:50001'
 2024/10/22 03:09:47 Running Worker.RunMap (ID: '3' File: 'map\map-3' Worker: '1')
 2024/10/22 03:09:47 Running Worker.RunMap (ID: '4' File: 'map\map-4' Worker: '0')
 2024/10/22 03:09:47 Running Worker.RunMap (ID: '5' File: 'map\map-5' Worker: '1')
 2024/10/22 03:09:47 Running Worker.RunMap (ID: '6' File: 'map\map-6' Worker: '0')
 2024/10/22 03:09:47 7x Worker.RunMap operations completed
 2024/10/22 03:09:47 Scheduling Worker.RunReduce operations
 2024/10/22 03:09:47 Running Worker.RunReduce (ID: '0' File: 'reduce\reduce-0' Worker: '1')
 2024/10/22 03:09:47 Running Worker.RunReduce (ID: '1' File: 'reduce\reduce-1' Worker: '0')
 2024/10/22 03:09:47 Running Worker.RunReduce (ID: '2' File: 'reduce\reduce-2' Worker: '0')
 2024/10/22 03:09:48 Running Worker.RunReduce (ID: '3' File: 'reduce\reduce-3' Worker: '0')
 2024/10/22 03:09:48 Running Worker.RunReduce (ID: '4' File: 'reduce\reduce-4' Worker: '0')
 2024/10/22 03:09:48 Operation Worker.RunReduce '0' Failed. Error: read tcp 127.0.0.1:65126->127.0.0
  .1:50001: wsarecv: An existing connection was forcibly closed by the remote host.
 2024/10/22 03:09:48 Removing worker 1 from master list.
 2024/10/22 03:09:48 5x Worker.RunReduce operations completed
 2024/10/22 03:09:48 Running Worker.RunReduce (ID: '0' File: 'reduce\reduce-0' Worker: '0')
 2024/10/22 03:09:48 Closing Remote Workers.
 2024/10/22 03:09:48 Done.
```

Note que, nesse processo do master, há uma falha na operação de reduce do ID 0, que é seguida pela remoção do worker 1 (falho). Logo após, o worker 0, que já estava funcionando, continua processando os reduces faltantes, completando a tarefa de reduce do ID 0 que foi deixada sem ser completada pelo worker 1 falho.

Tarefa 2.4: Dois workers falhos e três workers normais. O arquivo de teste continuará sendo o pg1342.txt, visto que ele é complexo o suficiente

.\wordcount -mode distributed -type master -file files/pg1342.txt -chunksize 102400 -reducejobs 5

```
| Fighth | Proceedings | Proceedings | Procedure | Pro
```

Aqui, o worker 0 falha após 3 tentativas na operação de map do ID 2, e é removido. Logo após, o worker 1 entra em ação, e refaz o id 2 e continua com outros.

```
| PERMINAL
| PS C: Ulsers India (One Orive Danx - Hotebook - 1 Documentos Academics \ ITA\8" Sem (4" COMP)\CSC-27 (Distributed Computing)\ Go Labs and Activities \ Lab 2\lab \ Appendix and \ Computing \ Computi
```

```
2004/18/22 83:16:37 Registering worker '2' with hostsame 'localhost:59082'
2004/18/22 83:16:37 Running kiorker, Runbap (II): '18' File: 'mplwap-17' Worker: '2')
2004/18/22 83:16:37 Running kiorker, Runbap (II): '18' File: 'mplwap-18' Worker: '2')
2004/18/22 83:16:37 Running kiorker, Runbap (II): '18' File: 'mplwap-19' Worker: '2')
2004/18/22 83:16:37 Running kiorker, Runbap (II): '28' File: 'mplwap-29' Worker: '2')
2004/18/22 83:16:37 Running kiorker, Runbap (II): '28' File: 'mplwap-29' Worker: '2')
2004/18/22 83:16:37 Running kiorker, Runbap (II): '28' File: 'mplwap-21' Worker: '2')
2004/18/22 83:16:37 Running kiorker, Runbap (II): '28' File: 'mplwap-22' Worker: '2')
2004/18/22 83:16:37 Running kiorker, Runbap (II): '28' File: 'mplwap-22' Worker: '2')
2004/18/22 83:16:37 Running kiorker, Runbap (II): '28' File: 'mplwap-24' Worker: '2')
2004/18/22 83:16:37 Running kiorker, Runbap (II): '28' File: 'mplwap-25' Worker: '2')
2004/18/22 83:16:37 Running kiorker, Runbap (II): '28' File: 'mplwap-26' Worker: '2')
2004/18/22 83:16:37 Running kiorker, Runbap (II): '28' File: 'mplwap-26' Worker: '2')
2004/18/22 83:16:37 Running kiorker, Runbap (II): '28' File: 'mplwap-28' Worker: '2')
2004/18/22 83:16:37 Running kiorker, Runbap (II): '28' File: 'mplwap-38' Worker: '2')
2004/18/22 83:16:37 Running kiorker, Runbap (II): '28' File: 'mplwap-38' Worker: '2')
2004/18/22 83:16:37 Running kiorker, Runbap (II): '38' File: 'mplwap-38' Worker: '2')
2004/18/22 83:16:37 Running kiorker, Runbap (II): '38' File: 'mplwap-38' Worker: '2')
2004/18/22 83:16:37 Running kiorker, Runbap (II): '38' File: 'mplwap-38' Worker: '2')
2004/18/22 83:16:37 Running kiorker, Runbap (II): '38' File: 'mplwap-38' Worker: '2')
2004/18/22 83:16:37 Running kiorker, Runbap (II): '38' File: 'mplwap-38' Worker: '2')
2004/18/22 83:16:37 Running kiorker, Runbap (II): '38' File: 'mplwap-38' Worker: '2')
2004/18/22 83:16:38 Running kiorker, Runbap (II): '38' File: 'mplwap-38' Worker: '2')
2004/18/22 83:16:38 Running kiorker, Runbap (II): '48' File: 'mplwap-48' Worker
```

Logo depois, os workers 2 e 3 são registrados. Todavia, o worker 3 falha após 5 tentativas, ainda nas operações de map, no ID 39. O worker 1, ainda ativo, consegue retrabalhar este ID 39, fazendo com que ele continue no fluxo de forma completa. Assim, os workers 1 e 2 permanecem ativos e completando as operações, sem deixar nenhum id passar antes de começar a próxima etapa de reduce.

```
2024/19/22 03:16:38 Running Worker.RunNap (ID: '53' File: 'map\map-54' Worker: '1')
2024/19/22 03:16:38 Running Worker.RunNap (ID: '55' File: 'map\map-54' Worker: '2')
2024/19/22 03:16:38 Running Worker.RunNap (ID: '55' File: 'map\map-56' Worker: '1')
2024/19/22 03:16:38 Running Worker.RunNap (ID: '55' File: 'map\map-56' Worker: '1')
2024/19/22 03:16:38 Running Worker.RunNap (ID: '55' File: 'map\map-56' Worker: '1')
2024/19/22 03:16:38 Running Worker.RunNap (ID: '55' File: 'map\map-56' Worker: '2')
2024/19/22 03:16:38 Running Worker.RunNap (ID: '58' File: 'map\map-58' Worker: '2')
2024/19/22 03:16:38 Running Worker.RunNap (ID: '56' File: 'map\map-59' Worker: '2')
2024/19/22 03:16:38 Running Worker.RunNap (ID: '66' File: 'map\map-69' Worker: '2')
2024/19/22 03:16:38 Running Worker.RunNap (ID: '61' File: 'map\map-61' Worker: '1')
2024/19/22 03:16:38 Running Worker.RunNap (ID: '61' File: 'map\map-63' Worker: '1')
2024/19/22 03:16:38 Running Worker.RunNap (ID: '63' File: 'map\map-63' Worker: '1')
2024/19/22 03:16:38 Running Worker.RunNap (ID: '64' File: 'map\map-64' Worker: '2')
2024/19/22 03:16:38 Running Worker.RunNap (ID: '66' File: 'map\map-66' Worker: '1')
2024/19/22 03:16:38 Running Worker.RunNap (ID: '66' File: 'map\map-66' Worker: '2')
2024/19/22 03:16:38 Running Worker.RunNap (ID: '66' File: 'map\map-66' Worker: '2')
2024/19/22 03:16:38 Running Worker.RunNap (ID: '66' File: 'map\map-66' Worker: '2')
2024/19/22 03:16:38 Running Worker.RunNap (ID: '66' File: 'map\map-66' Worker: '2')
2024/19/22 03:16:38 Running Worker.RunNap (ID: '66' File: 'map\map-66' Worker: '2')
2024/19/22 03:16:39 Running Worker.RunNap (ID: '66' File: 'map\map-66' Worker: '2')
2024/19/22 03:16:39 Running Worker.RunRaduce (ID: '1' File: 'reduce\reduce-0' Worker: '2')
2024/19/22 03:16:39 Running Worker.RunRaduce (ID: '2' File: 'reduce\reduce-0' Worker: '2')
2024/19/22 03:16:39 Running Worker.RunRaduce (ID: '4' File: 'reduce\reduce-0' Worker: '2')
2024/19/22 03:16:39 Running Worker.RunRaduce (ID: '4' File: 'reduce\reduce-0' Worker: '2')
20
```

Por fim, as 5 tarefas de reduce são completadas com sucesso pelos workers ativos 1 e 2.