Algoritmos e Estruturas de Dados III

Aula 3.2 – Ordenação Externa

Prof. Hayala Curto 2022



Roteiro do Conteúdo





3.1 Arquivos

- Arquivos Sequenciais
- Chaves de Ordenação
- Chave Primária
- Operações em Arquivos CRUD

3.2 Ordenação Externa

- Ordenação Externa
- Intercalação Balanceada
- Análise de Complexidade
- Otimização do Algoritmo

Ordenação Externa

- Processo de ordenação de dados em arquivos
- Adotado quando os dados a serem ordenados são maiores que a capacidade de ordenação em memória principal
- Prioriza o acesso sequencial aos arquivos (algoritmo criado na época das fitas)

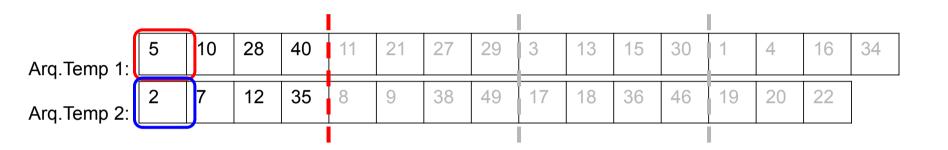
Ordenação Externa - Intercalação Balanceada

Intercalação balanceada

 Algoritmo que ordena os registros por meio da intercalação de registros de várias fontes balanceadas (arquivos temporários com tamanho aproximado semelhante)

Etapas:

- 1. Distribuição de blocos de B registros ordenados por M caminhos balanceados (arquivos temporários)
- 2. Intercalações sucessivas dos segmentos ordenados dos M caminhos



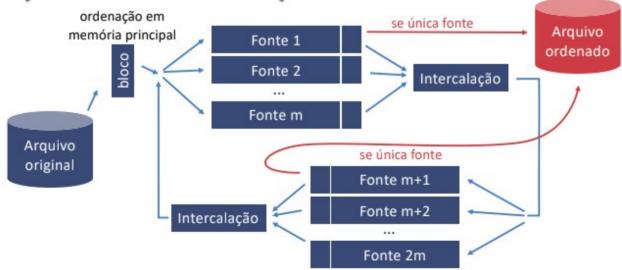
Distribuição

Arq.Temp 3:

Arq.Temp 5:

Ordenação Externa - Intercalação Balanceada

Intercalação balanceada (de m caminhos)



Execução do algoritmo



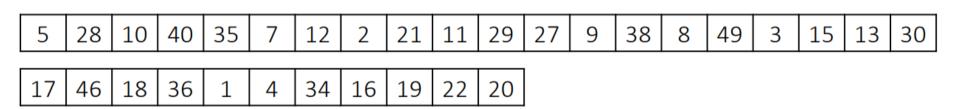
Arquivo original

5	28	10	40	35	7	12	2	21	11	29	27	9	38	8	49	3	15	13	30
17	46	18	36	1	4	34	16	19	22	20									

Observação:

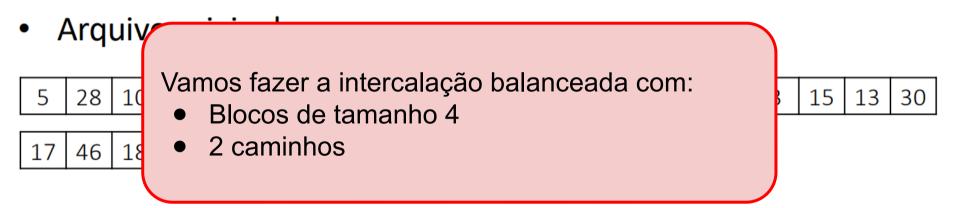
Nesse exemplo, apenas os IDs dos registros estão sendo apresentados, mas os registros são completos, isto é, contém diversos outros atributos.

Arquivo original



Etapa 1: Distribuição

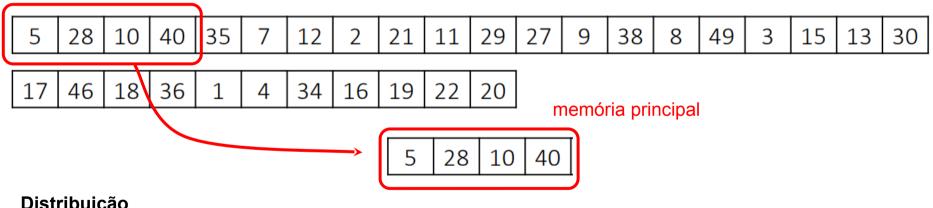
Os registros serão lidos em blocos de grande tamanho (**4 neste exemplo**) para a memória principal, ordenados em memória principal (Quicksort, Mergesort, ...) e escritos nos primeiros M caminhos (**2 neste exemplo**)



Etapa 1: Distribuição

Os registros serão lidos em blocos de grande tamanho (**4 neste exemplo**) para a memória principal, ordenados em memória principal (Quicksort, Mergesort, ...) e escritos nos primeiros M caminhos (**2 neste exemplo**)

Arquivo original

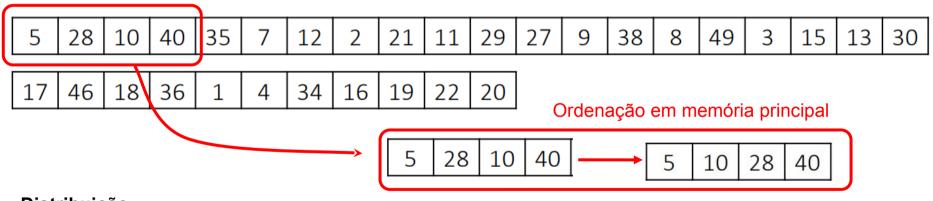


Distribuição

Arq.Temp 1:

Arq.Temp 2:

Arquivo original

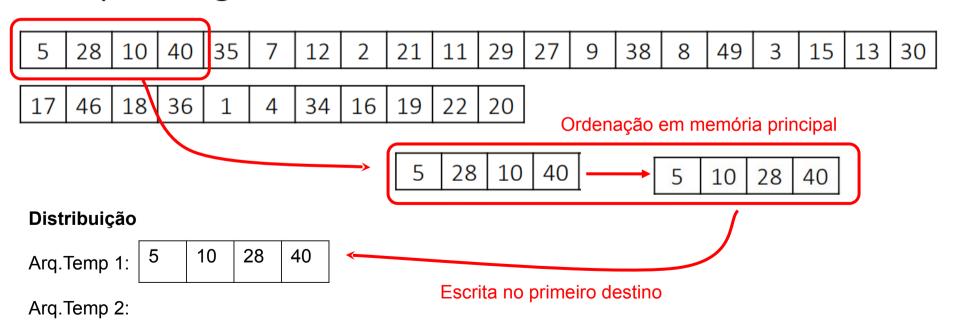


Distribuição

Arq.Temp 1:

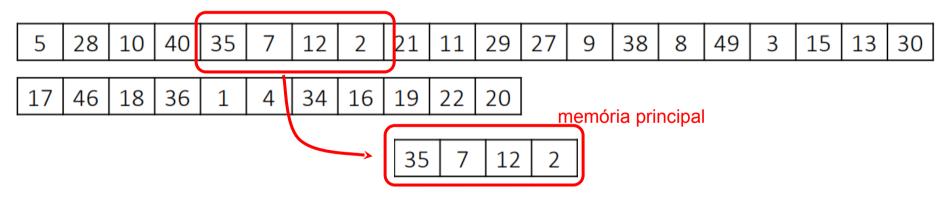
Arq.Temp 2:

Arquivo original



Hayala Curto

Arquivo original

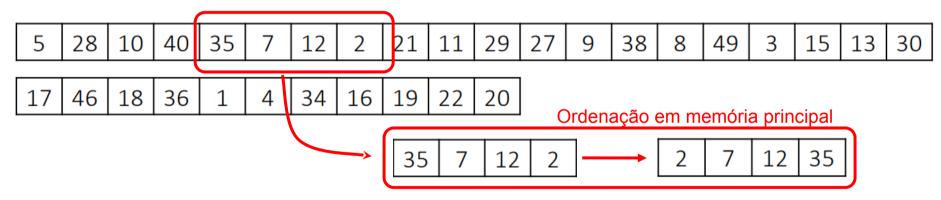


Distribuição

Arq.Temp 1: 5 10 28 40

Arq.Temp 2:

Arquivo original

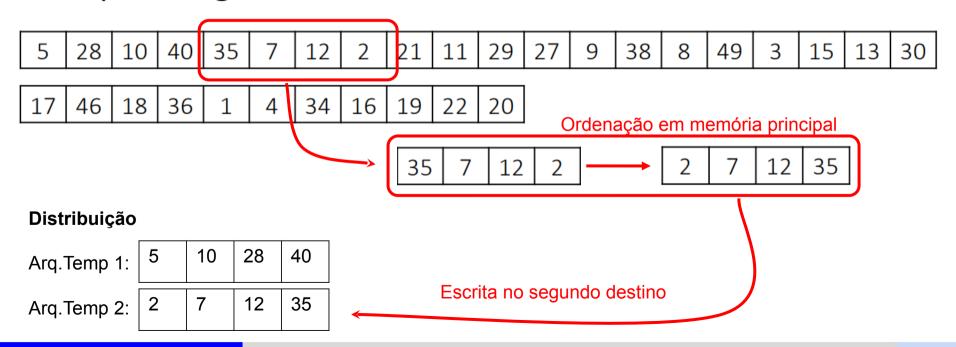


Distribuição

Arq.Temp 1: 5 10 28 40

Arq.Temp 2:

Arquivo original



Hayala Curto

10

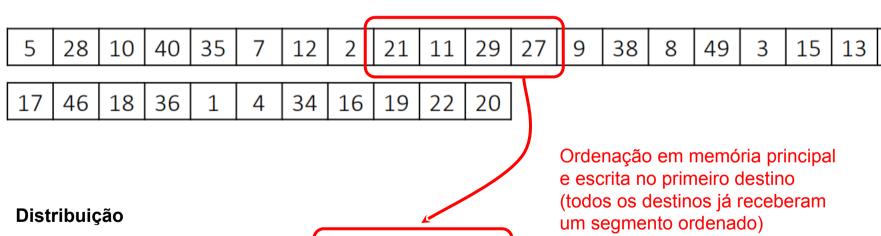
28

12

40

35

Arquivo original



27

29

21

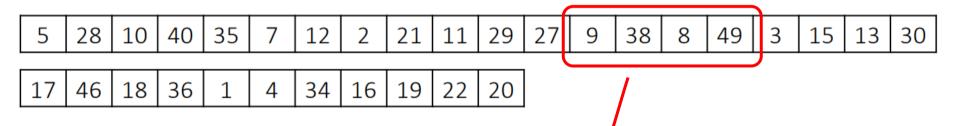
30

Arq.Temp 2:

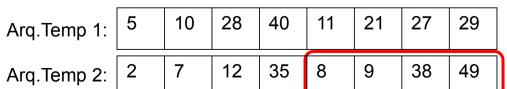
Hayala Curto

Arq.Temp 1:

Arquivo original



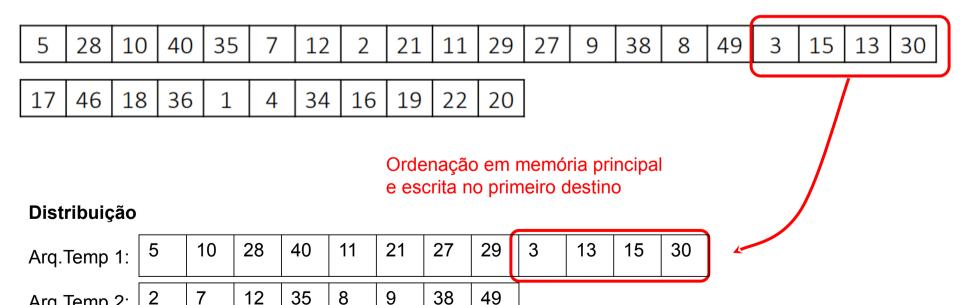
Distribuição



Ordenação em memória principal e escrita no segundo destino

Hayala Curto

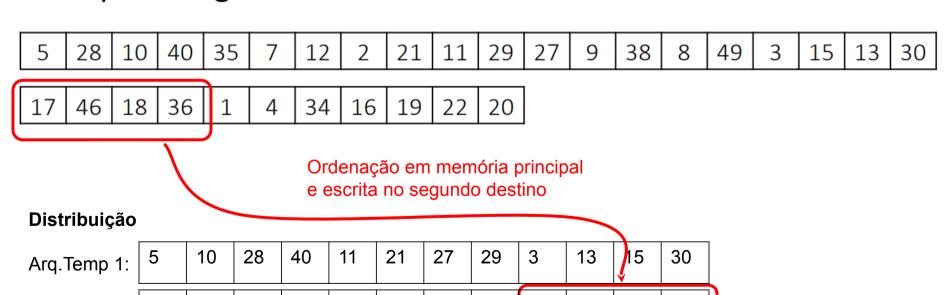
Arquivo original



Hayala Curto

Arq.Temp 2:

Arquivo original

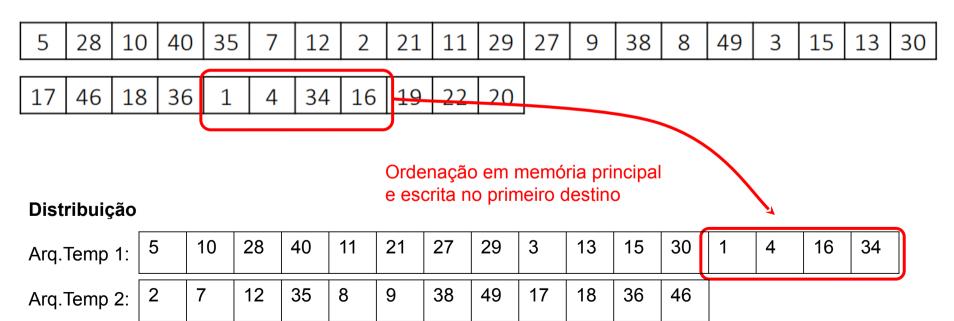


Hayala Curto

Arq.Temp 2:

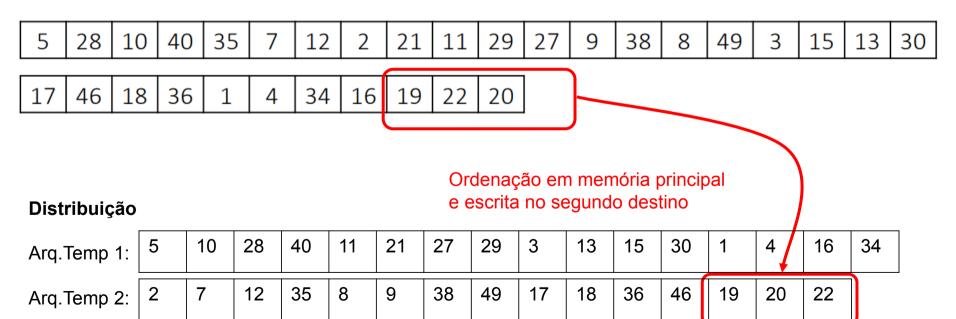
Algoritmos e Estruturas de Dados III

Arquivo original



Hayala Curto

Arquivo original



Hayala Curto

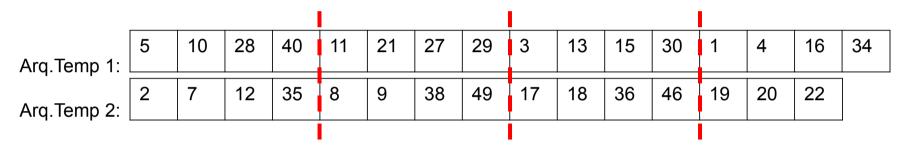
Arquivo original

5	28	10	40	35	7	12	2	21	11	29	27	9	38	8	49	3	15	13	30
17	46	18	36	1	4	34	16	19	22	20									

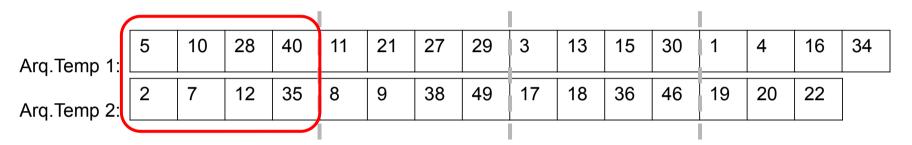
Fim da Etapa 1: Distribuição

Distribuição

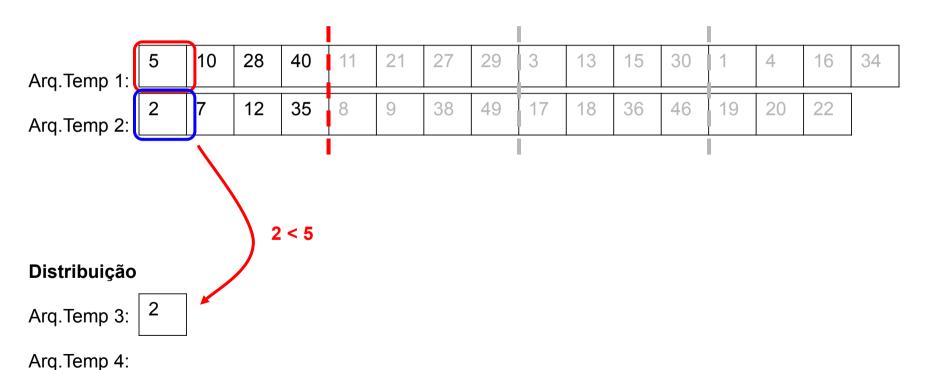
Arq.Temp 1:	5	10	28	40	11	21	27	29	3	13	15	30	1	4	16	34	
Arq.Temp 2:	2	7	12	35	8	9	38	49	17	18	36	46	19	20	22		

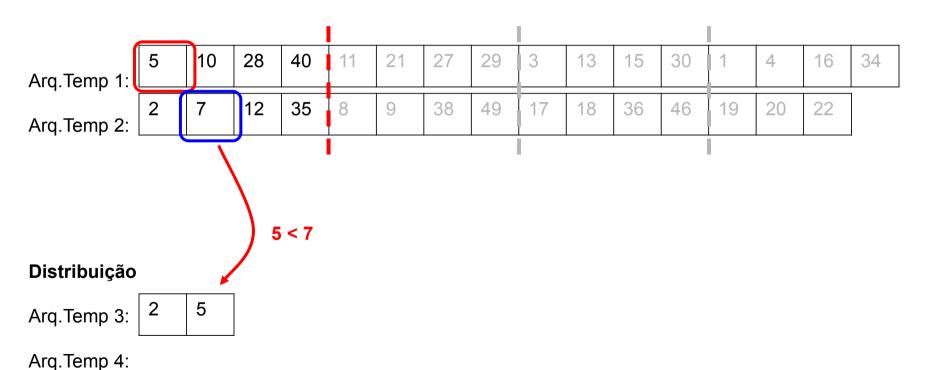


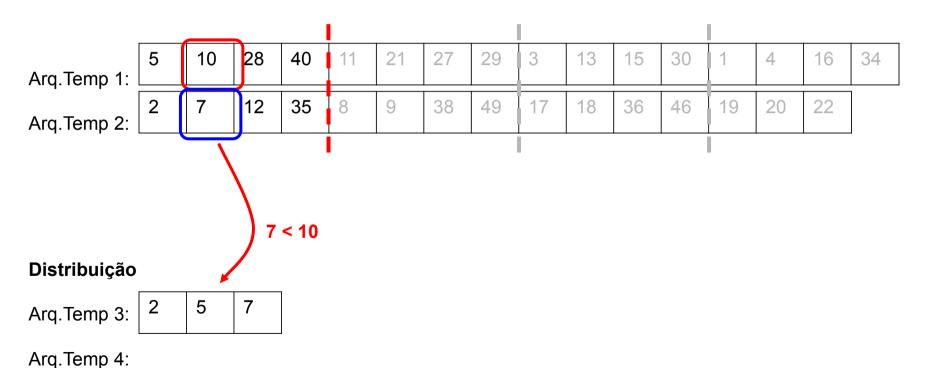
Etapa 2: Primeira intercalação

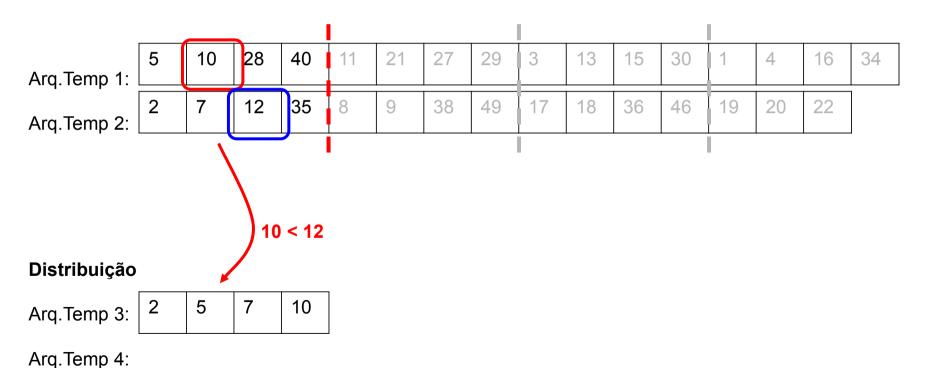


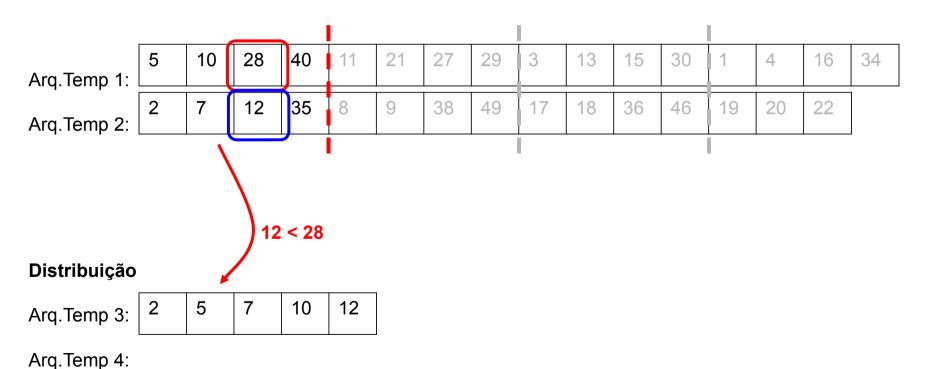
Etapa 2: Primeira intercalação

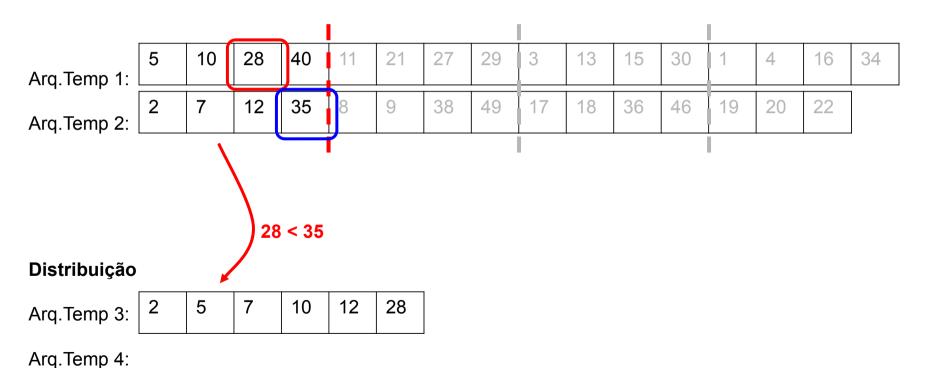


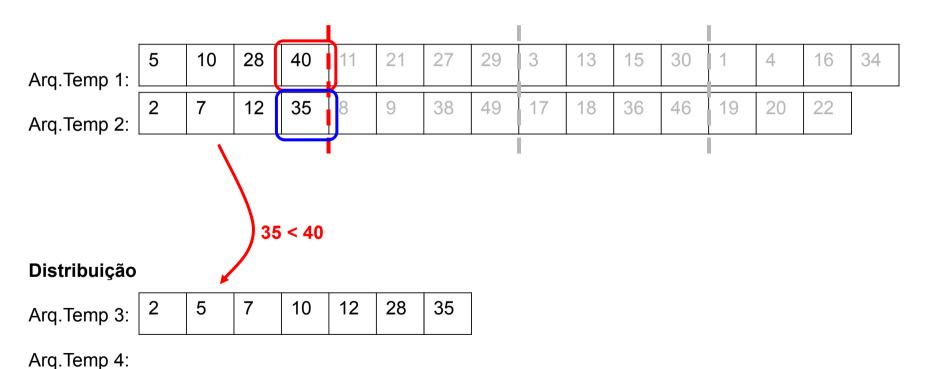


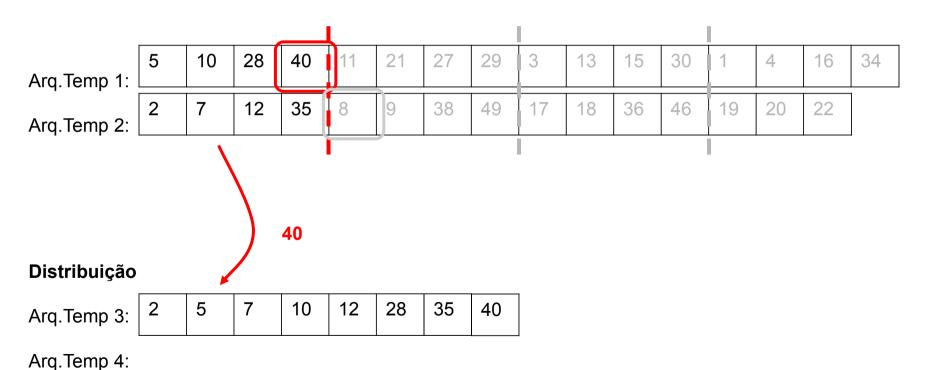


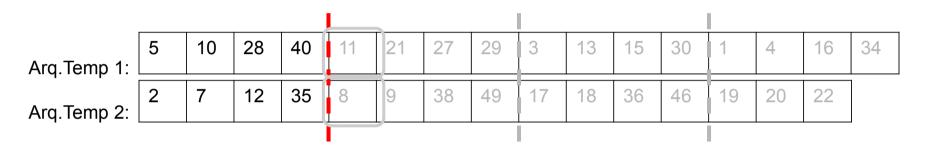










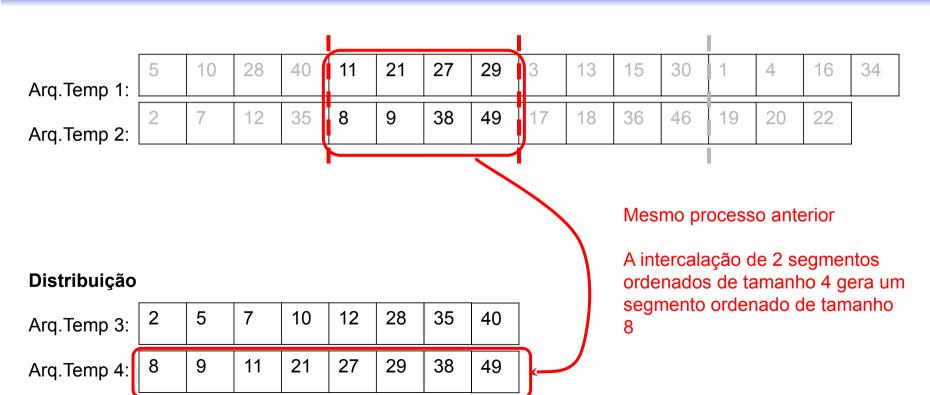


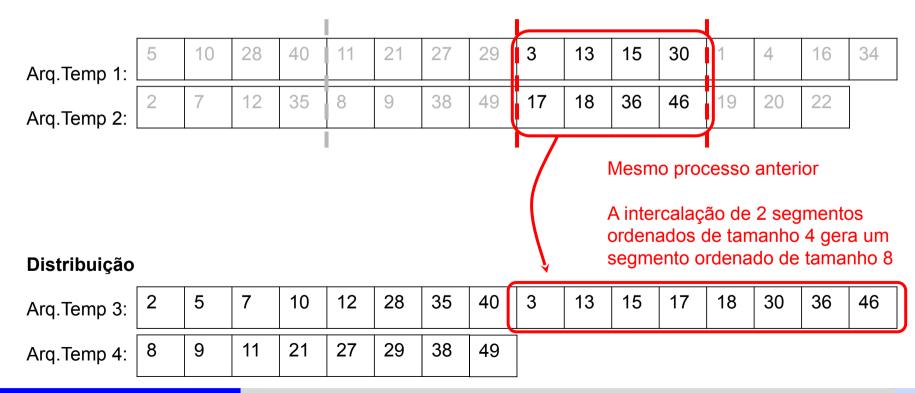
A intercalação de 2 segmentos ordenados de tamanho 4 gera um segmento ordenado de tamanho 8

Distribuição

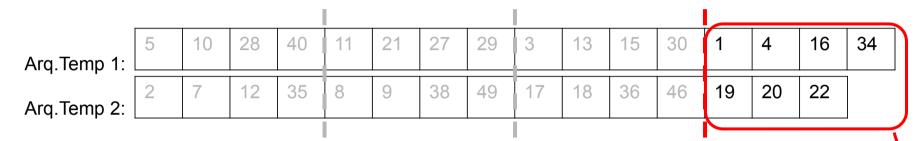
Arq.Temp 3: 2 5 7 10 12 28 35 40

Arq.Temp 4:





Hayala Curto

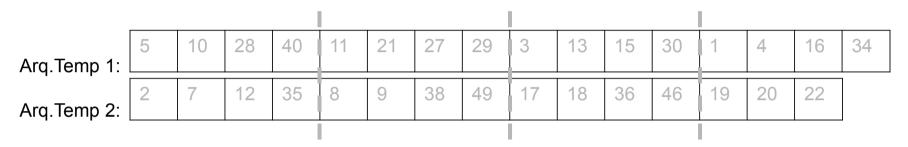


Mesmo processo anterior

A intercalação de 2 segmentos ordenados de tamanho 4 gera um segmento ordenado de tamanho 8

Distribuição

Arq.Temp 3:	2	5	7	10	12	28	35	40	3	13	15	17	18	30	36	46
Arq.Temp 4:	8	9	11	21	27	29	38	49	1	4	16	19	20	22	34	



Etapa 2: Fim da primeira intercalação

Distribuição

Arq.Temp 3:	2	5	7	10	12	28	35	40	3	13	15	17	18	30	36	46
Arq.Temp 4:	8	9	11	21	27	29	38	49	1	4	16	19	20	22	34	

Algoritmos e Estruturas de Dados III

Distribuição

Arq.Temp 3:	2	5	7	10	12	28	35	40	3	13	15	17	18	30	36	46
Arq.Temp 4:	8	9	11	21	27	29	38	49	1	4	16	19	20	22	34	

Etapa 3: Segunda intercalação

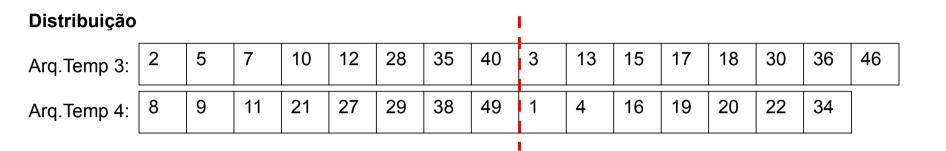
Distribuição

Arq.Temp 3:	2	5	7	10	12	28	35	40	3	13	15	17	18	30	36	46
Arq.Temp 4:	8	9	11	21	27	29	38	49	1	4	16	19	20	22	34	

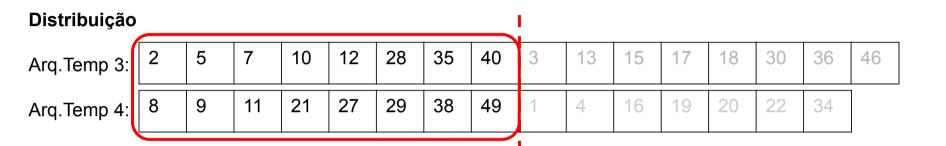
Etapa 3: Segunda intercalação

Os arquivos temporários 1 e 2 podem ser reaproveitados

Arq.Temp 1:



Arq.Temp 1:



Arq.Temp 1:



Arq.Temp 2:

Algoritmos e Estruturas de Dados III



Hayala Curto

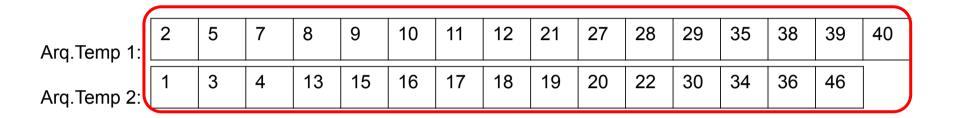
Algoritmos e Estruturas de Dados III

Arq.Temp 1:	2	5	7	8	9	10	11	12	21	27	28	29	35	38	39	40
Arq.Temp 2:	1	3	4	13	15	16	17	18	19	20	22	30	34	36	46	

Etapa 3: Fim da segunda intercalação

Arq.Temp 1:	2	5	7	8	9	10	11	12	21	27	28	29	35	38	39	40
Arq.Temp 2:	1	3	4	13	15	16	17	18	19	20	22	30	34	36	46	

Etapa 4: Terceira intercalação



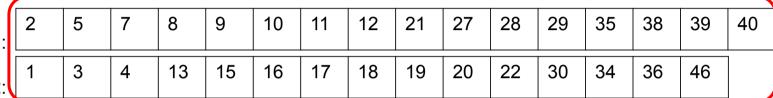
Etapa 4: Terceira intercalação

Os arquivos temporários 3 e 4 podem ser reaproveitados

Arq.Temp 3:



Arq.Temp 2:



A intercalação de 2 segmentos ordenados de tamanho 16 gera um segmento ordenado de tamanho 32

:	1	2	3	4	5	7	8	9	10	11	12	13	15	16	17	18
	19	20	21	22	27	28	29	30	34	35	36	38	40	46	49	

Arq.Temp 1: Arq.Temp 2:

:	1	2	3	4	5	7	8	9	10	11	12	13	15	16	17	18
	19	20	21	22	27	28	29	30	34	35	36	38	40	46	49	

Arq.Temp 3:

1	2	3	4	5	7	8	9	10	11	12	13	15	16	17	18
19	20	21	22	27	28	29	30	34	35	36	38	40	46	49	

Quando todos os registros ficarem em um único destino, o arquivo estará ordenado.



Estratégia 2 Intercalação com seleção por substituição



- A base da análise é o número de "passadas" (leituras) por todos os registros
 - O tempo da ordenação em memória principal é pouco significativo
- Cálculo:
 - $passadas = 1 + \left[log_m \left(\frac{N}{b} \right) \right]$
 - · Em que:
 - N = número total de registros
 - b = tamanho do bloco ordenado em memória principal
 - m = quantidade de destinos usados na intercalação
 - A primeira passada corresponde à etapa de distribuição

$$passadas = 1 + \left\lceil log_m \left(\frac{N}{b} \right) \right\rceil$$

Exemplo – Ordenação de 50.000 registros usando 3 caminhos, com capacidade de ordenação em memória principal de 150 registros.

- N = 50.000
- b = 150
- m = 3 (mas serão usados 6 arquivos temporários)

$$passadas = 1 + \left[log_m \left(\frac{N}{b}\right)\right]$$

$$passadas = 1 + \left[log_3 \left(\frac{50000}{150}\right)\right]$$

$$passadas = 1 + \left[log_3 (333,33)\right]$$

$$passadas = 1 + \left[5,29\right]$$

$$passadas = 1 + 6$$

$$passadas = 7$$

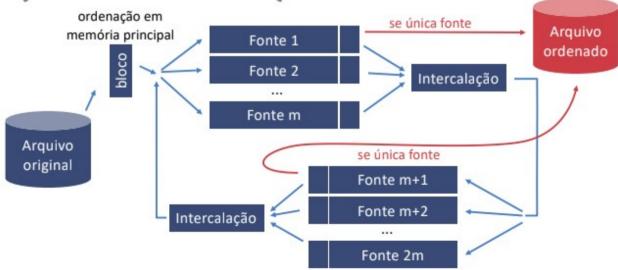
- N = 50.000
- b = 150
- m = 3 (mas serão usados6 arquivos temporários)

Otimização do Algoritmo



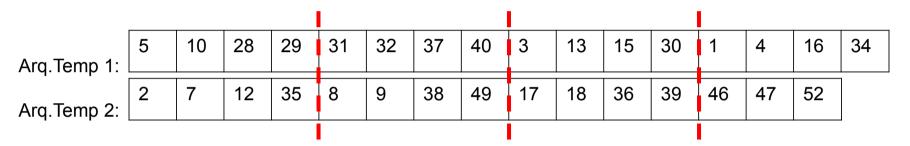
Ordenação Externa - Intercalação Balanceada

Intercalação balanceada (de m caminhos)



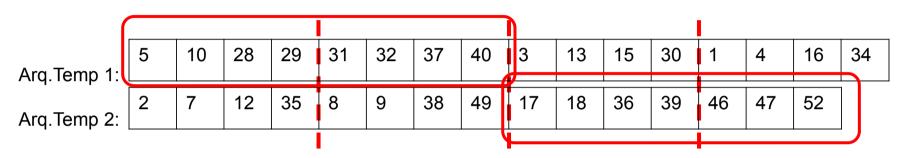
Estratégia 1 Segmento de Tamanho Variável





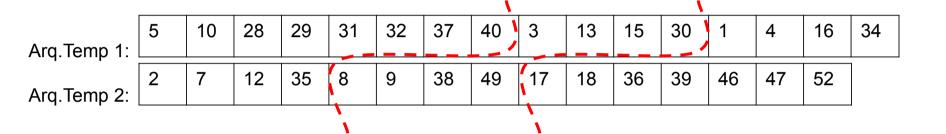
A Etapa 1 (ordenação em memória principal) já foi feita

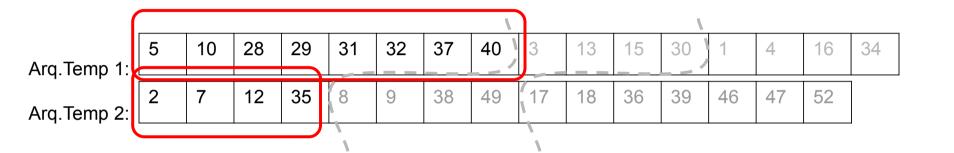
Estamos na Etapa 2: Primeira intercalação

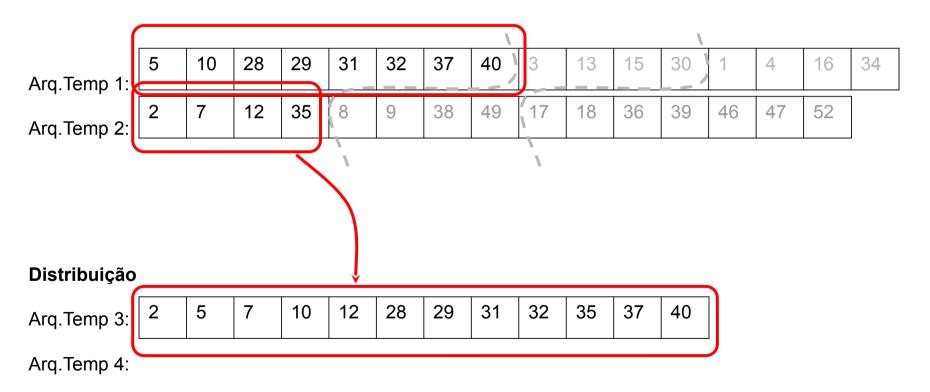


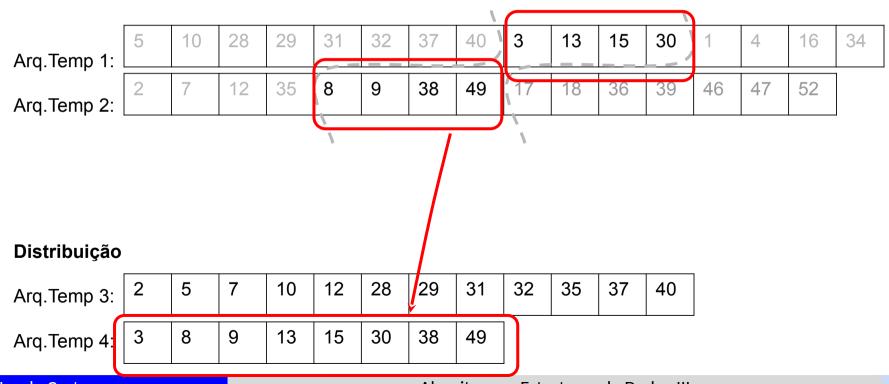
Uma primeira estratégia de otimização é tentar aproveitar a eventual ordenação entre os blocos.

O segmento só para de "crescer" quando encontramos um número menor que o anterior. Lembrando que os números dentro de cada segmento já estão ordenados



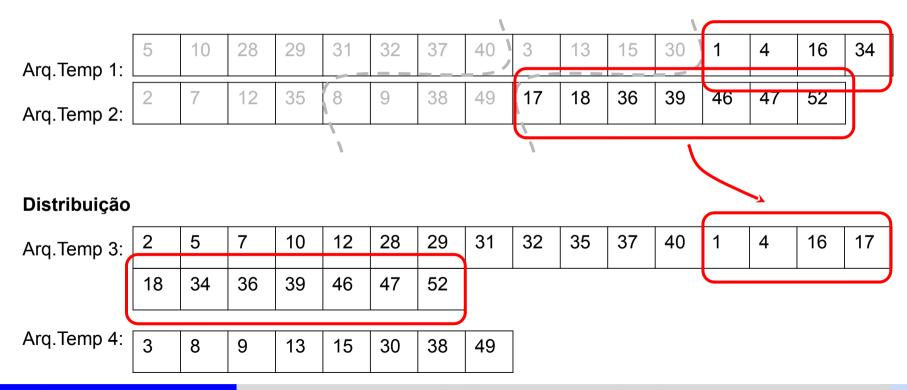




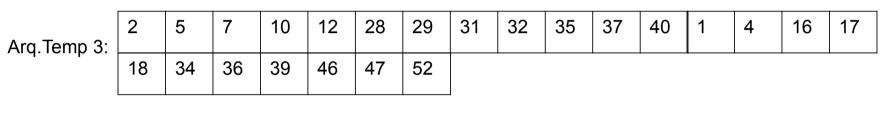


Hayala Curto

Algoritmos e Estruturas de Dados III



Hayala Curto

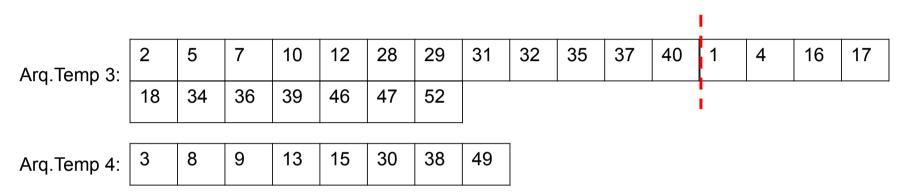


Arq.Temp 4: 3 8 9 13 15 30 38 49

Fim da primeira intercalação

Início da segunda intercalação

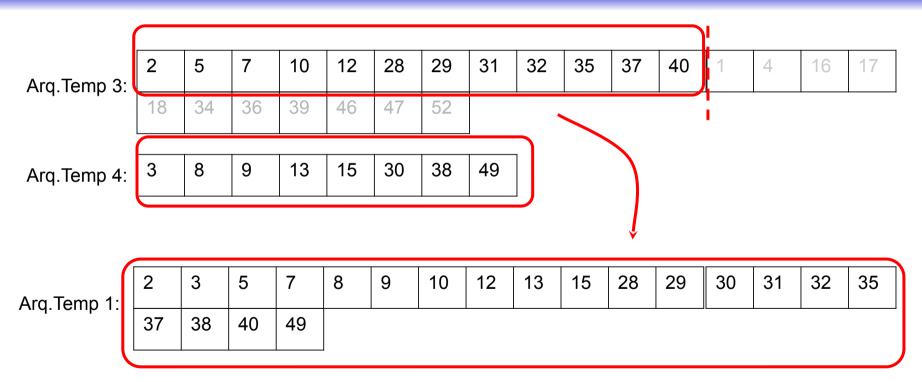
Arq.Temp 1:



Fim da primeira intercalação

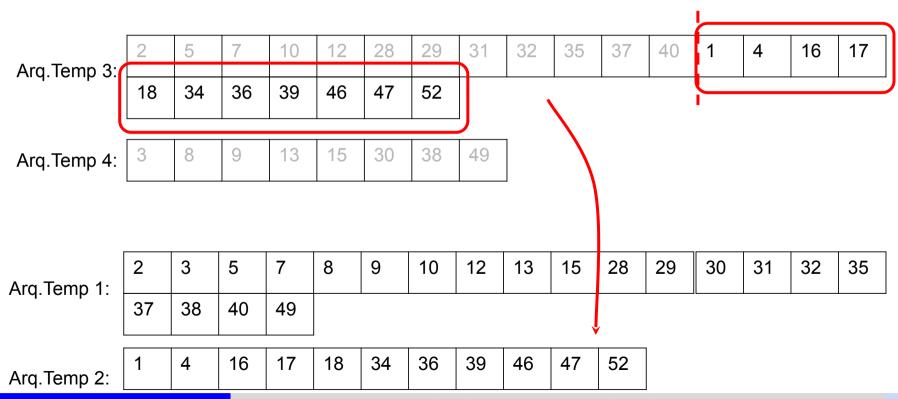
Início da segunda intercalação

Arq.Temp 1:



Arq.Temp 2:

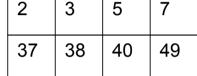
Hayala Curto



Hayala Curto

Algoritmos e Estruturas de Dados III



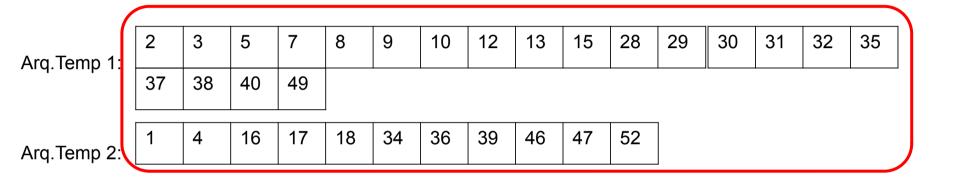


Arq.Temp 2:

	1	4	16	17	18	34	36	39	46	47	52
- 1											

Fim da segunda intercalação

Início da terceira intercalação



Arq.

. Tomp 2:	1	2	3	4	5	7	8	9	10	12	13	15	16	17	18	28	
լ.Temp 3:	29	30	31	32	34	35	36	37	38	39	40	46	47	49	52		

Arg.Temp 4:

Estratégia 2 Intercalação por Substituição

