Algoritmos e Estruturas de Dados III

Aula 3.2 - Ordenação Externa

Prof. Felipe Lara



Roteiro do Conteúdo





3.1 Arquivos

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

3.2 Ordenação de Arquivos

- Ordenação Externa
 - Intercalação Balanceada
 - Intercalação Polifásica

Roteiro do Conteúdo





3.1 Arquivos

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

3.2 Ordenação de Arquivos

- Ordenação Externa
 - Intercalação Balanceada
 - Intercalação Polifásica

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)

Etapas:

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

Ordenação Externa - Intercalação Balanceada

(de m caminhos) ordenação em se única fonte Arquivo memória principal Fonte 1 ordenado bloco Fonte 2 Intercalação Fonte m Arquivo se única fonte original Fonte m+1 Fonte m+2 Intercalação Fonte 2m

Execução do algoritmo



 Considere um arquivo armazenado em uma fita de entrada:

INTERCALACAOBALANCEADA

- Objetivo:
 - Ordenar os 22 registros e colocá-los em uma fita de saída.
- Os registros são lidos um após o outro.
- Considere uma memória interna com capacidade para para três registros.
- Considere que esteja disponível seis unidades de fita magnética.

Fase de Distribuição

Criação de blocos ordenados

fita 1: INT ACO ADE

fita 2: CER ABL A

fita 3: AAL ACN

Fase de intercalação - Primeira passada:

- 1.O primeiro registro de cada fita é lido.
- 2. Retire o registro contendo a menor chave.
- 3. Armazene-o em uma fita de saída.
- 4.Leia um novo registro da fita de onde o registro retirado é proveniente.
- 5. Ao ler o terceiro registro de um dos blocos, sua fita fica inativa.
- 6.A fita é reativada quando o terceiro registro das outras fitas forem lidos.
- 7. Neste instante um bloco de nove registros ordenados foi formado na fita de saída.
- 8. Repita o processo para os blocos restantes.

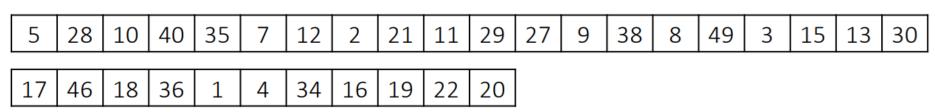
Resultado da Primeira passada

fita 4: AACEILNRT

fita 5: *A A A B C C L N O*

fita 6: AADE

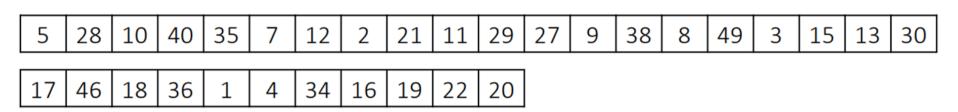
Arquivo original



Observação:

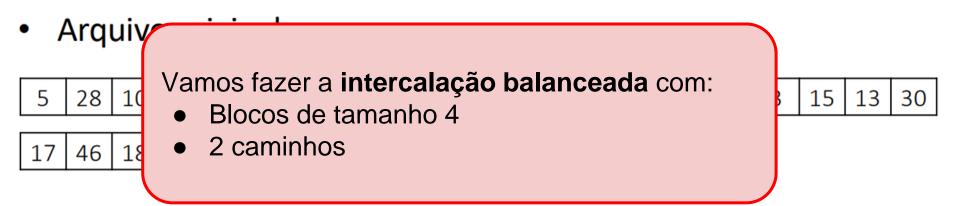
Nesse exemplo, apenas os IDs dos registros estão sendo apresentados, mas os registros são completos, isto é, contém diversos outros atributos.

Arquivo original



Etapa 1: Distribuição

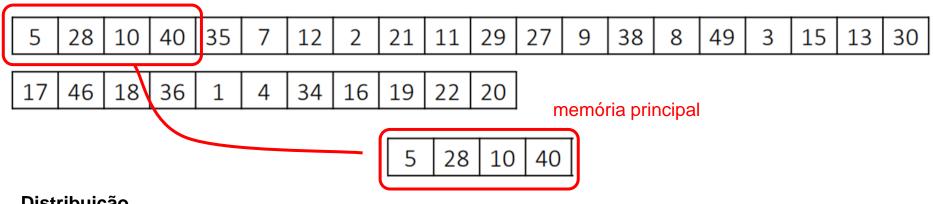
Os registros serão lidos em blocos de grande tamanho (**4 neste exemplo**) para a memória principal, ordenados em memória principal (Quicksort, Mergesort, ...) e escritos nos primeiros M caminhos (**2 neste exemplo**)



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

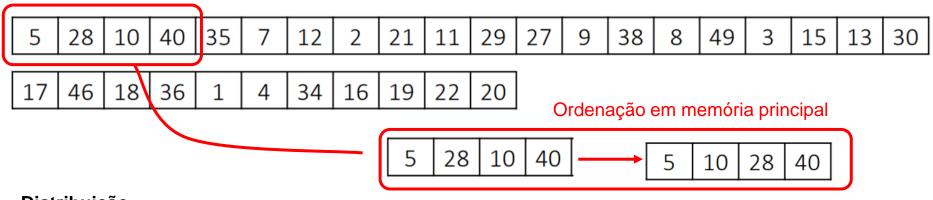


Distribuição

Arq.Temp 1:

Arq.Temp 2:

Arquivo original

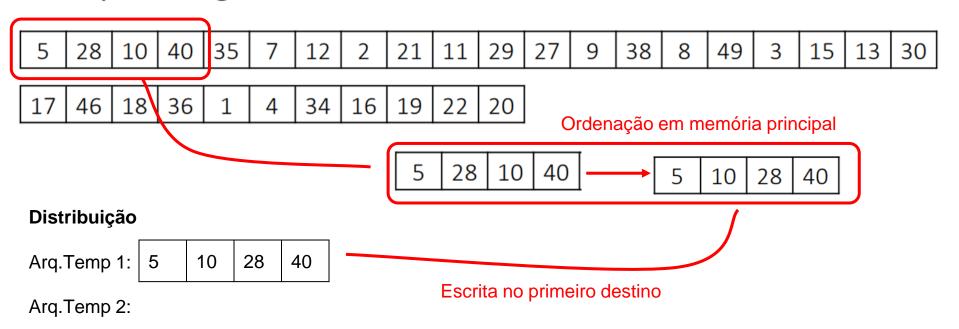


Distribuição

Arq.Temp 1:

Arq.Temp 2:

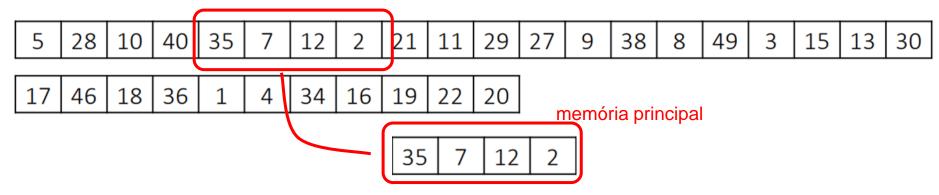
Arquivo original



Felipe Soares

Algoritmos e Estruturas de Dados III

Arquivo original

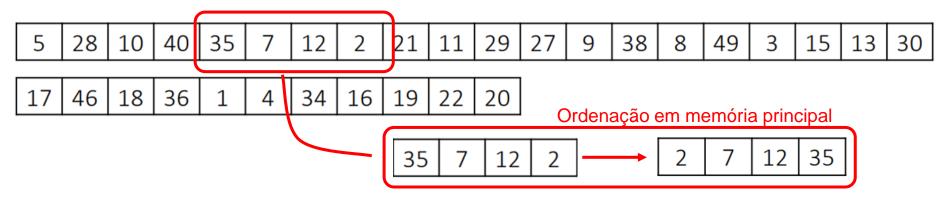


Distribuição

Arq.Temp 1: 5 10 28 40

Arq.Temp 2:

Arquivo original

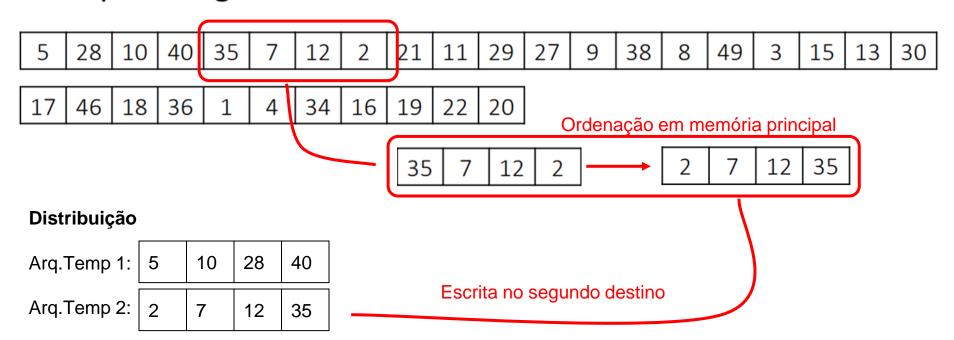


Distribuição

28 Arq.Temp 1: 10 40

Arq.Temp 2:

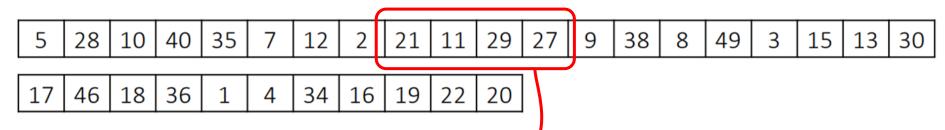
Arquivo original



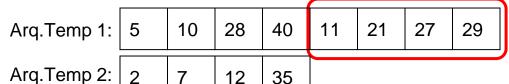
Felipe Soares

Algoritmos e Estruturas de Dados III

Arquivo original

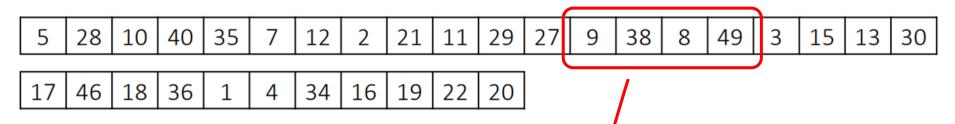


Distribuição

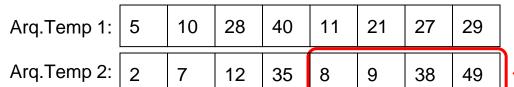


Ordenação em memória principal e escrita no primeiro destino (todos os destinos já receberam um segmento ordenado)

Arquivo original



Distribuição



Ordenação em memória principal e escrita no segundo destino

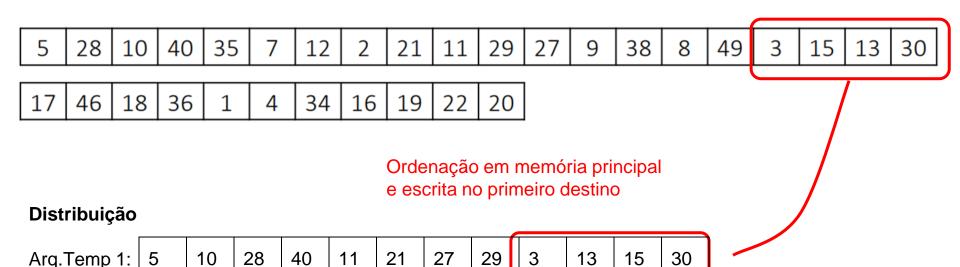
35

12

9

38

Arquivo original



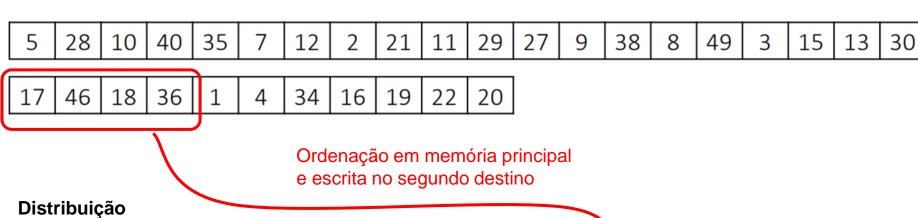
49

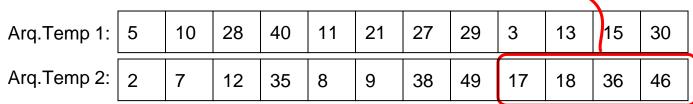
Felipe Soares

Arq.Temp 2:

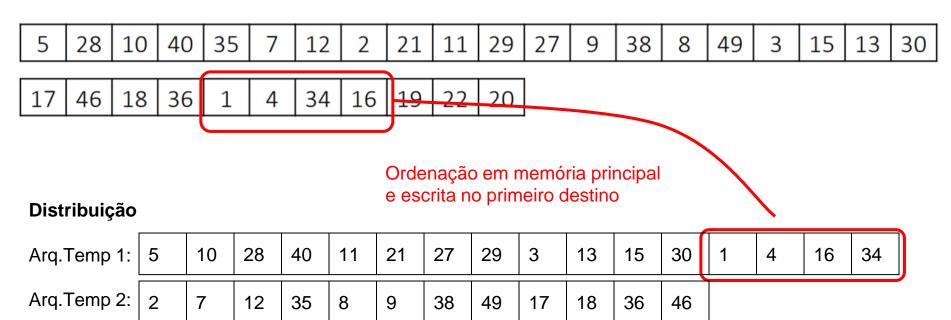
Algoritmos e Estruturas de Dados III

Arquivo original





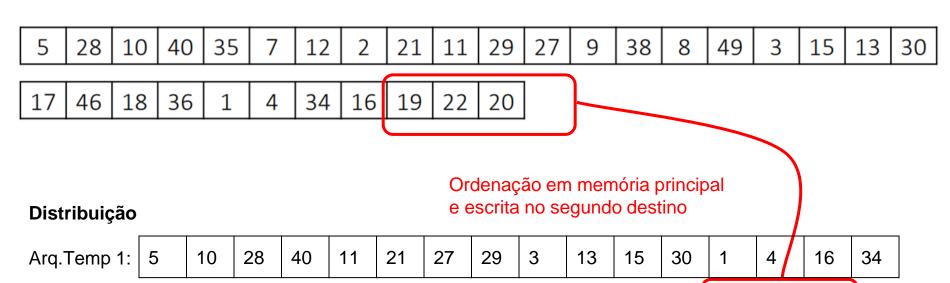
Arquivo original



Felipe Soares

Algoritmos e Estruturas de Dados III

Arquivo original



Felipe Soares

Arq.Temp 2:

Algoritmos e Estruturas de Dados III

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									

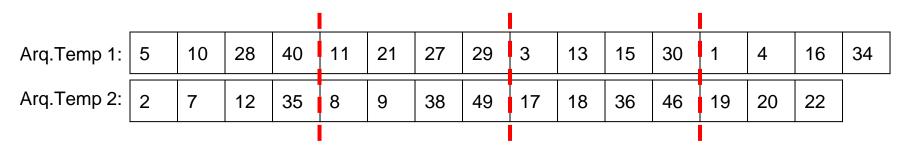
Fim da Etapa 1: Distribuição

Distribuição

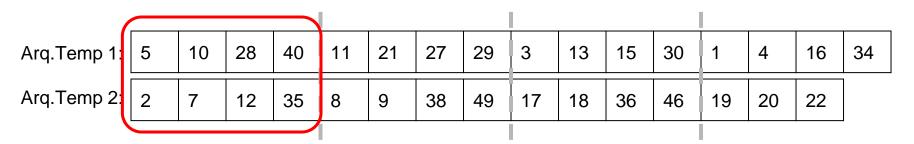
Arq.Temp 1:	5	10	28	40	11	21	27	29	3	13	15	30	1	4	16	34
Arq.Temp 2:	2	7	12	35	8	9	38	49	17	18	36	46	19	20	22	

Felipe Soares

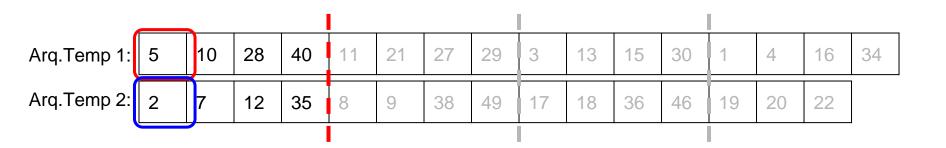
Algoritmos e Estruturas de Dados III



Etapa 2: Primeira intercalação



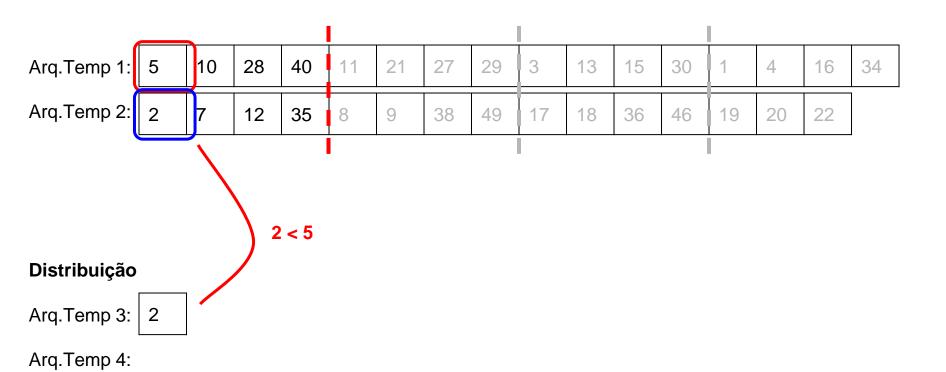
Etapa 2: Primeira intercalação

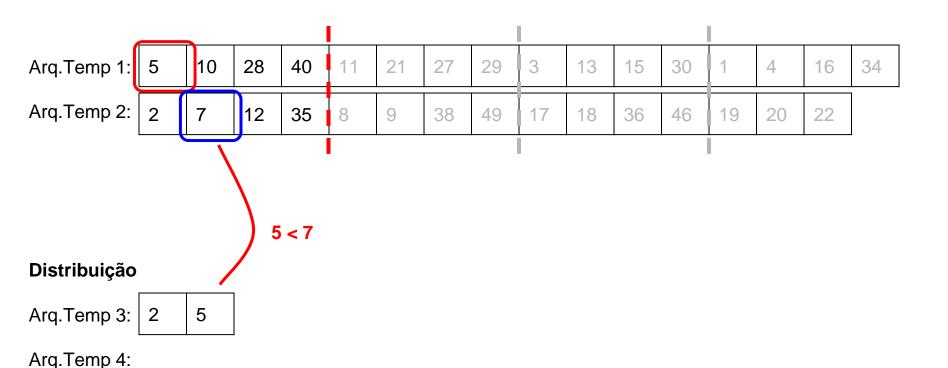


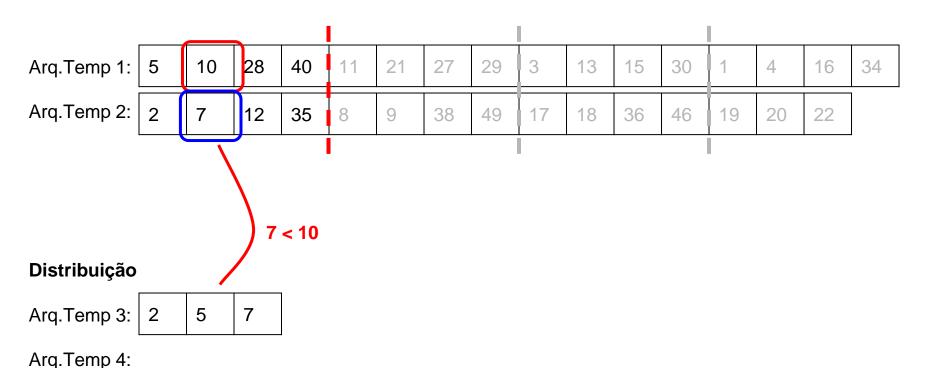
Distribuição

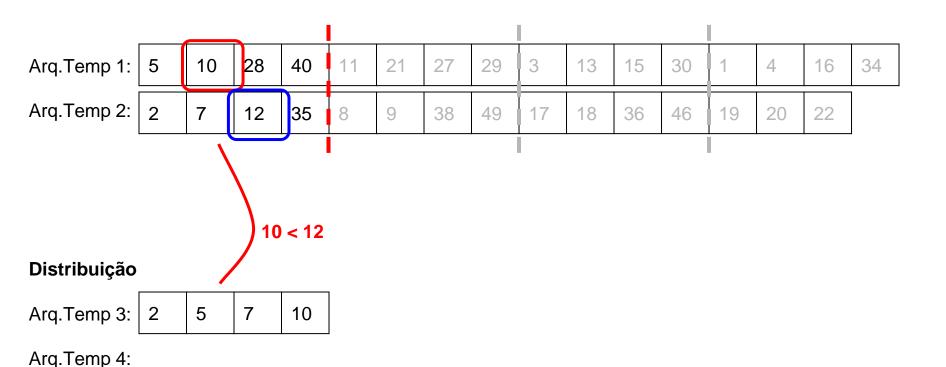
Arq.Temp 3:

Arq.Temp 4:

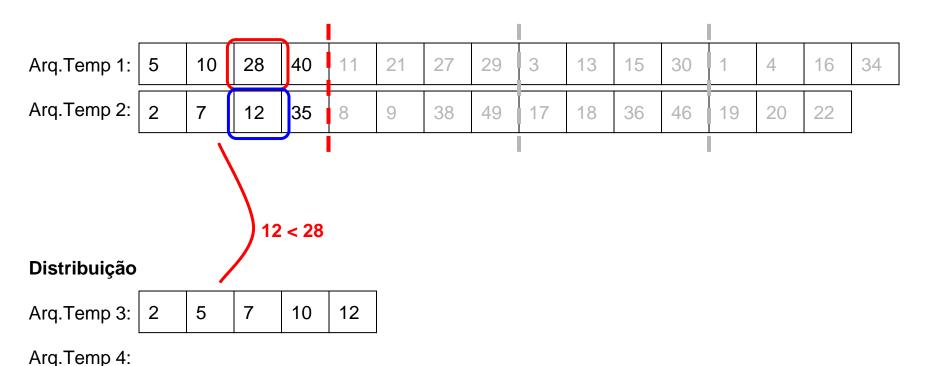




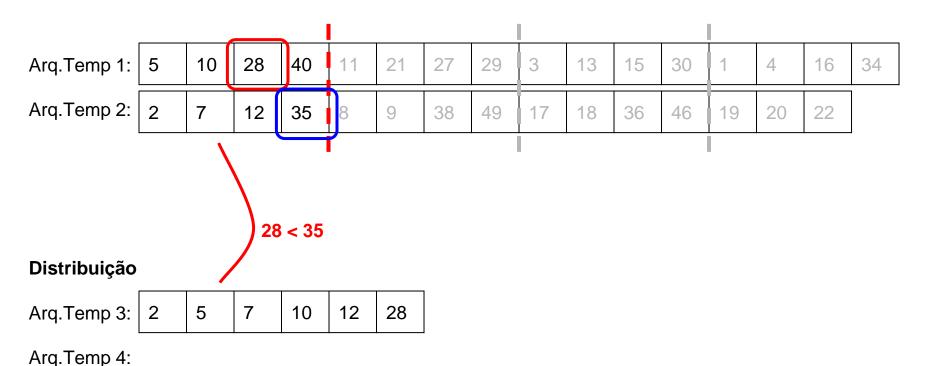




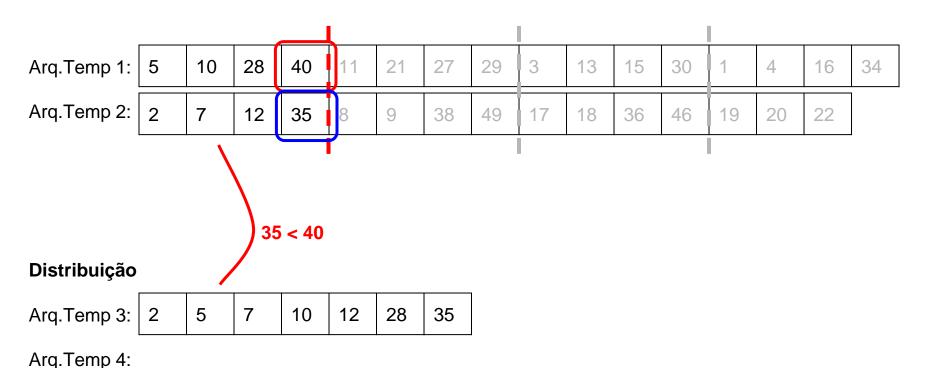
Felipe Soares



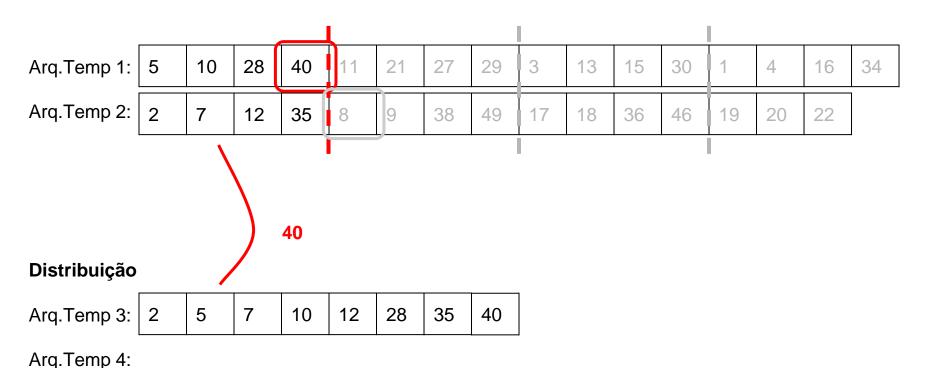
Felipe Soares



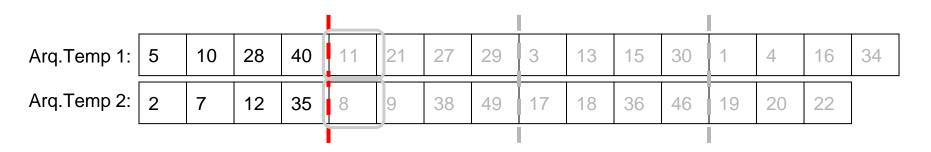
Felipe Soares



Felipe Soares



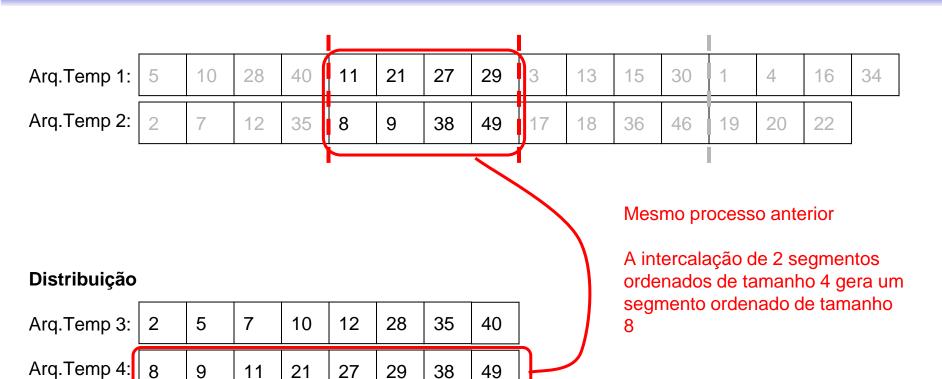
Felipe Soares

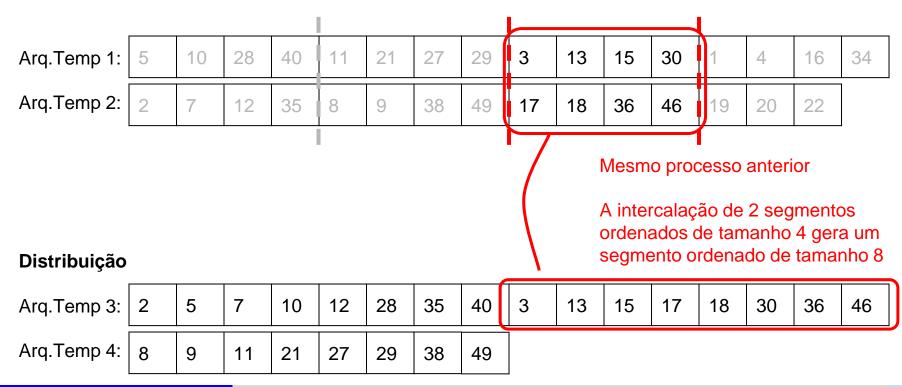


A intercalação de 2 segmentos ordenados de tamanho 4 gera um segmento ordenado de tamanho 8

Distribuição

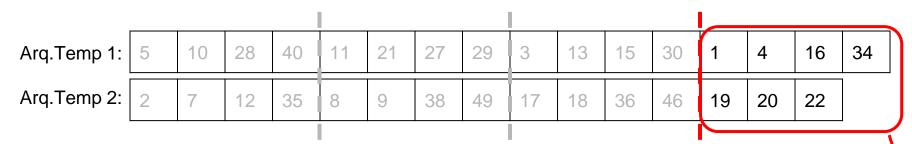
Arq.Temp 3: 2 5 7 10 12 28 35 40





Felipe Soares

Algoritmos e Estruturas de Dados III

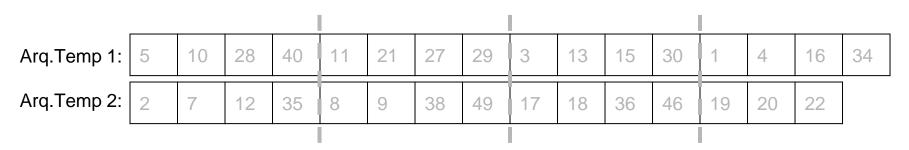


Mesmo processo anterior

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

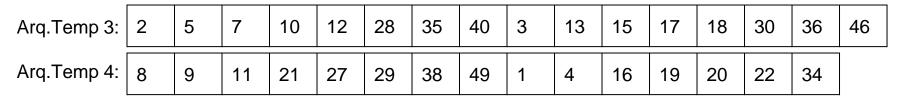
Distribuição

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



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

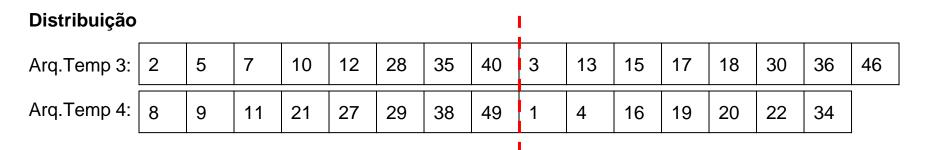
Distribuição

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	

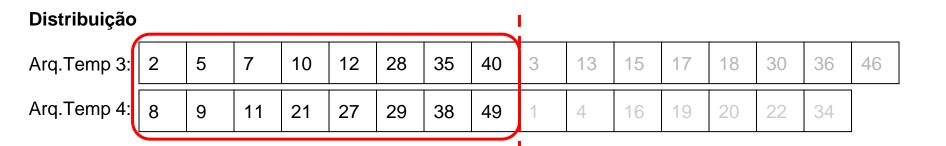
Etapa 3: Segunda intercalação

Os arquivos temporários 1 e 2 podem ser reaproveitados

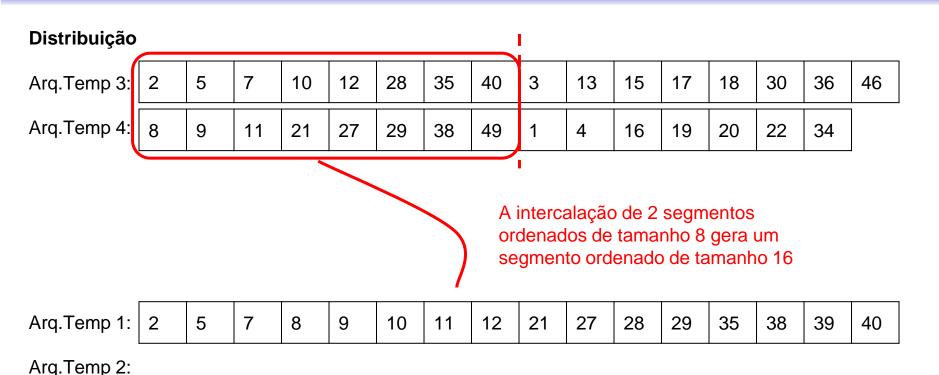
Arq.Temp 1:



Arq.Temp 1:

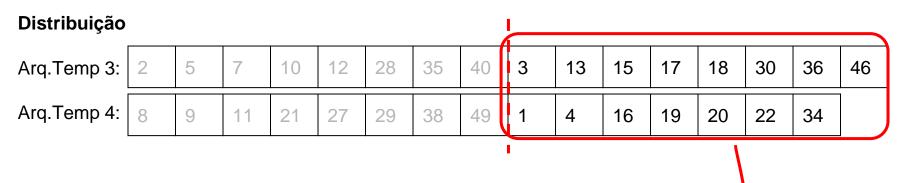


Arq.Temp 1:



Felipe Soares

Algoritmos e Estruturas de Dados III



A intercalação de 2 segmentos ordenados de tamanho 8 gera um segmento ordenado de tamanho 16

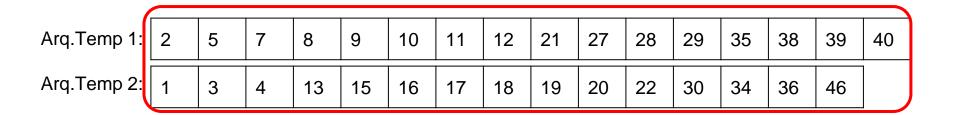
Arq.Temp 1:	2	5	7	8	9	10	11	12	21	27	28	29	35	38	39	40
Arq.Temp 2:	1	3	4	13	15	16	17	18	19	20	22	30	34	36	46	

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	

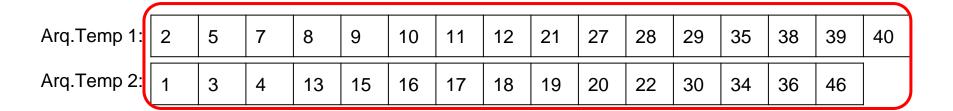
Etapa 4: Terceira intercalação



Etapa 4: Terceira intercalação

Os arquivos temporários 3 e 4 podem ser reaproveitados

Arq.Temp 3:



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

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	

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	

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	

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.



- 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)

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

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

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

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

$$passadas = 1 + 6$$

$$passadas = 7$$

•
$$N = 50.000$$

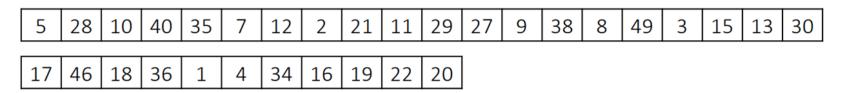
•
$$b = 150$$

• m = 3 (mas serão usados 6 arquivos temporários)

log10 (333,33) / log10 3

Vamos testar

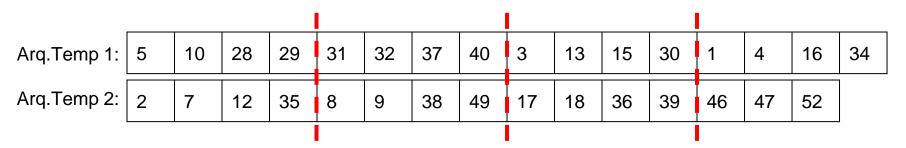
Arquivo original



- Blocos de tamanho 4
- 2 caminhos
- Quantas passadas são necessárias?

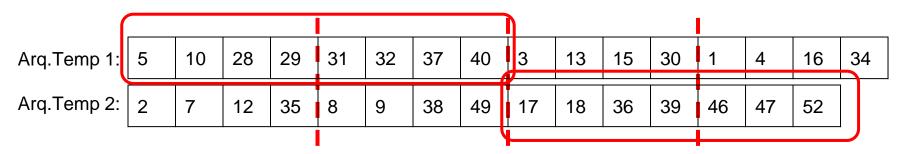
Estratégia 1 Segmento de Tamanho Variável



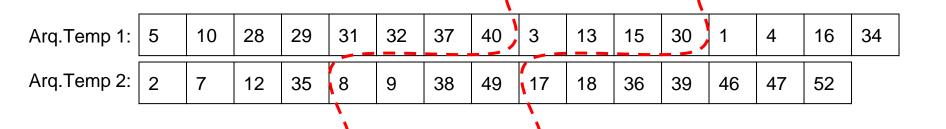


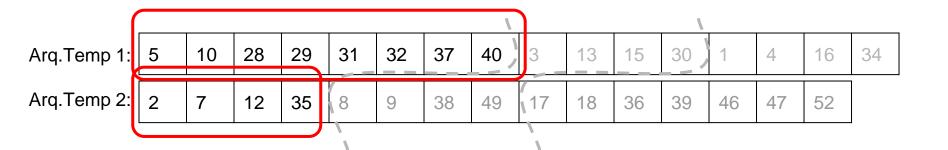
A Etapa 1 (ordenação em memória principal) já foi feita

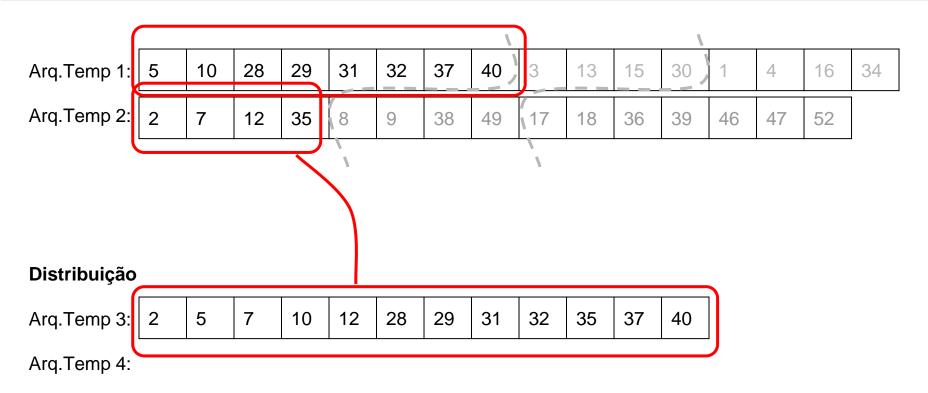
Estamos na Etapa 2: Primeira intercalação

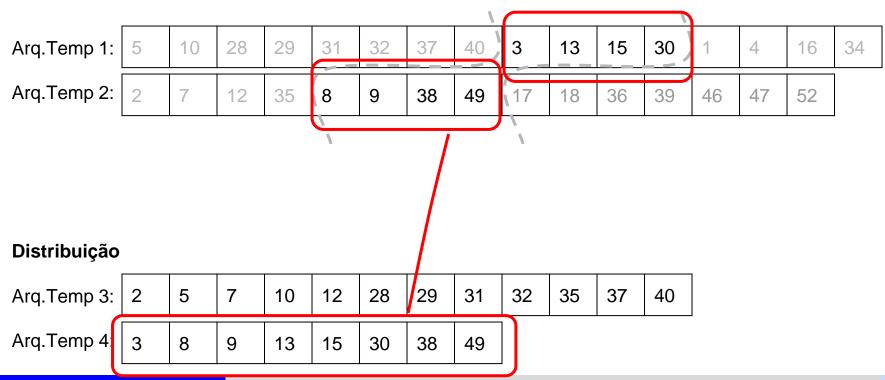


Uma primeira estratégia de otimização é tentar aproveitar a eventual ordenação entre os blocos:



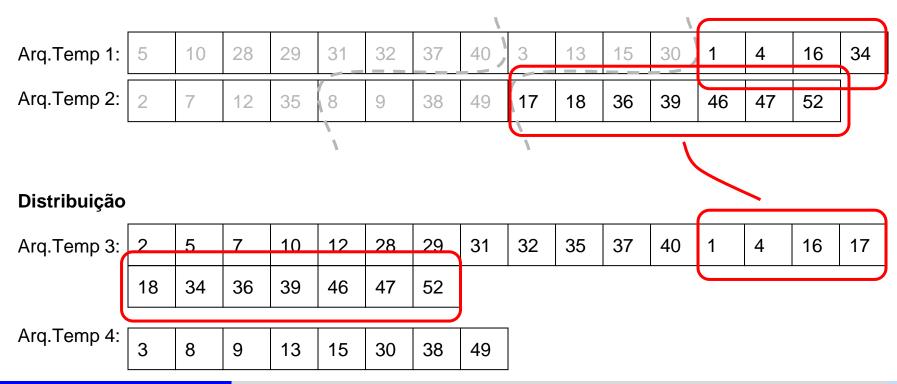






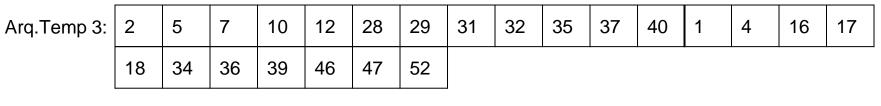
Felipe Soares

Algoritmos e Estruturas de Dados III



Felipe Soares

Algoritmos e Estruturas de Dados III



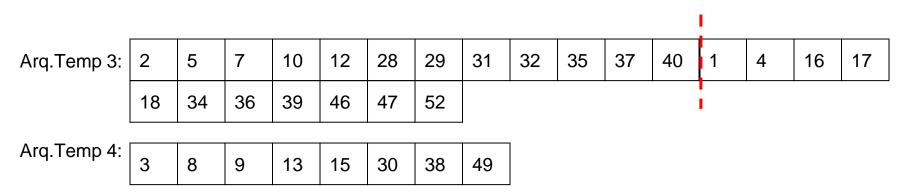
Arq.Temp 4:

3 8 9 13	15	30	38	49
----------	----	----	----	----

Fim da primeira intercalação

Início da segunda intercalação

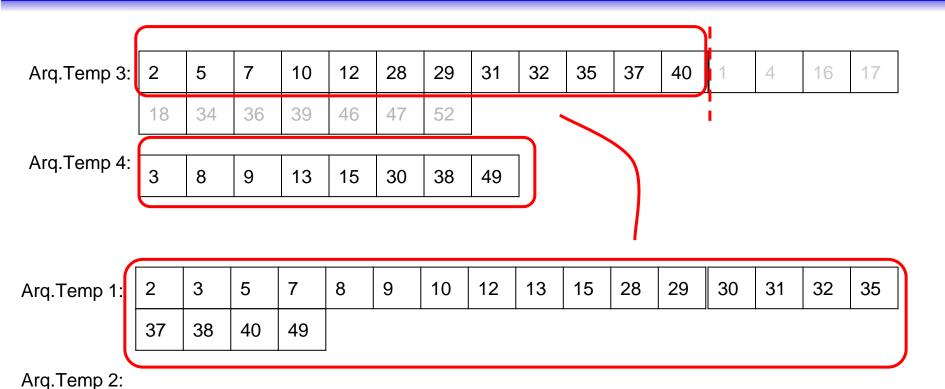
Arq.Temp 1:



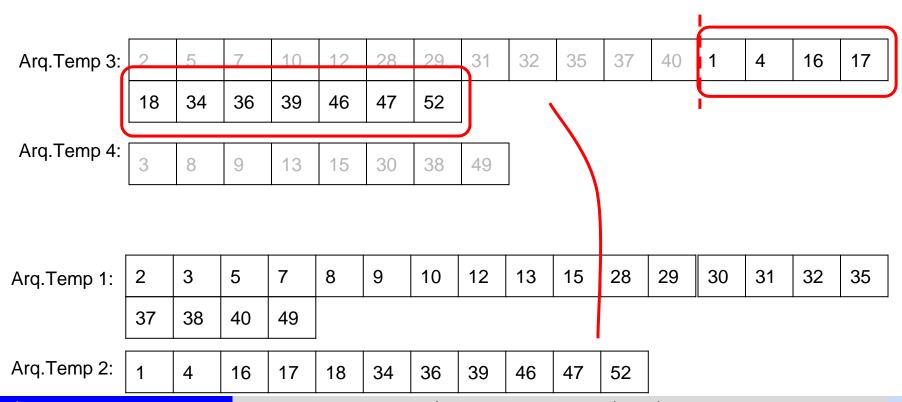
Fim da primeira intercalação

Início da segunda intercalação

Arq.Temp 1:



Felipe Soares



Felipe Soares



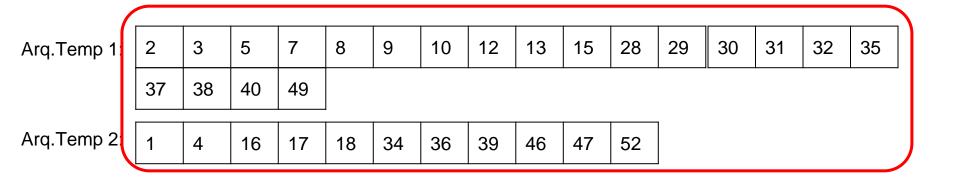
2	3	5	7
37	38	40	49

Arq.Temp 2:

1	4	16	17	18	34	36	39	46	47	52

Fim da segunda intercalação

Início da terceira intercalação



Arq.Temp 3: