

# Ordenação externa

Slides do livro Database Management Systems  
3ed, R. Ramakrishnan and J. Gehrke,  
McGrow-Hill, 2003

# Atividade A9

1) Ordene o seguinte arquivo usando um algoritmo de ordenação externa

Pag size=2 (cada página armazena 2 valores)

Buffer size=3 (a memória pode armazenar até 3 páginas)

20,19,18,17, até o número 5

2) No exemplo acima, calcule o número de operações no caso de Buffer=5 de I/O.

# Ordenamento

- Um problema clássico em CC
- Ordenar representa o primeiro passo para Bulk Loading B++;
- Importante para eliminar duplicatas;
- Problema: ordenar 1 TB de dados em 1 GB de ram

# Merge-sort em duas vias

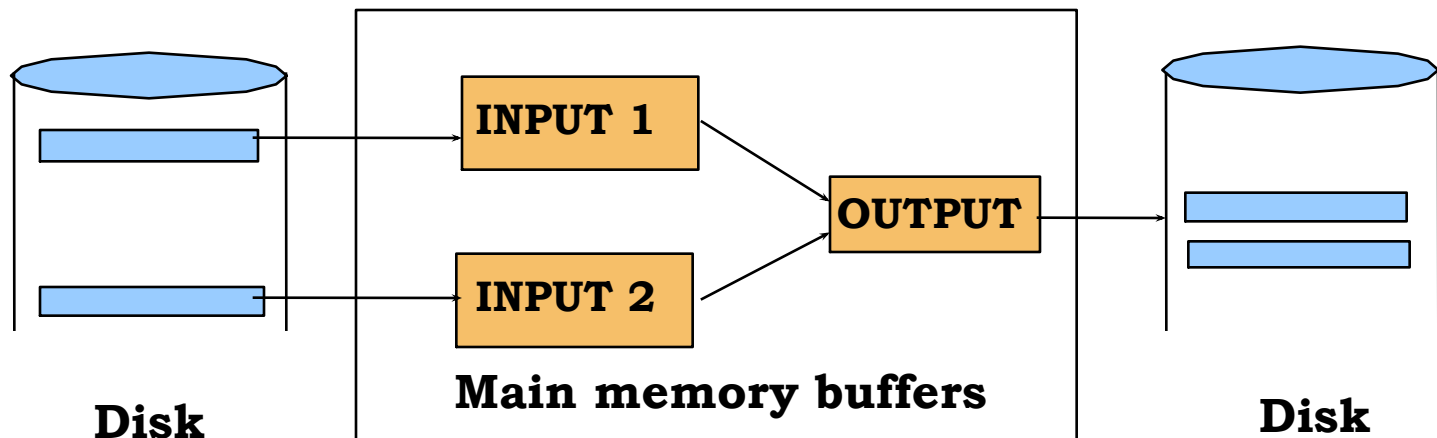
- Passada 0: leia a página, ordene, e a escreva.

Somente um buffer é usado

- Passada 1, 2, ..., etc.:

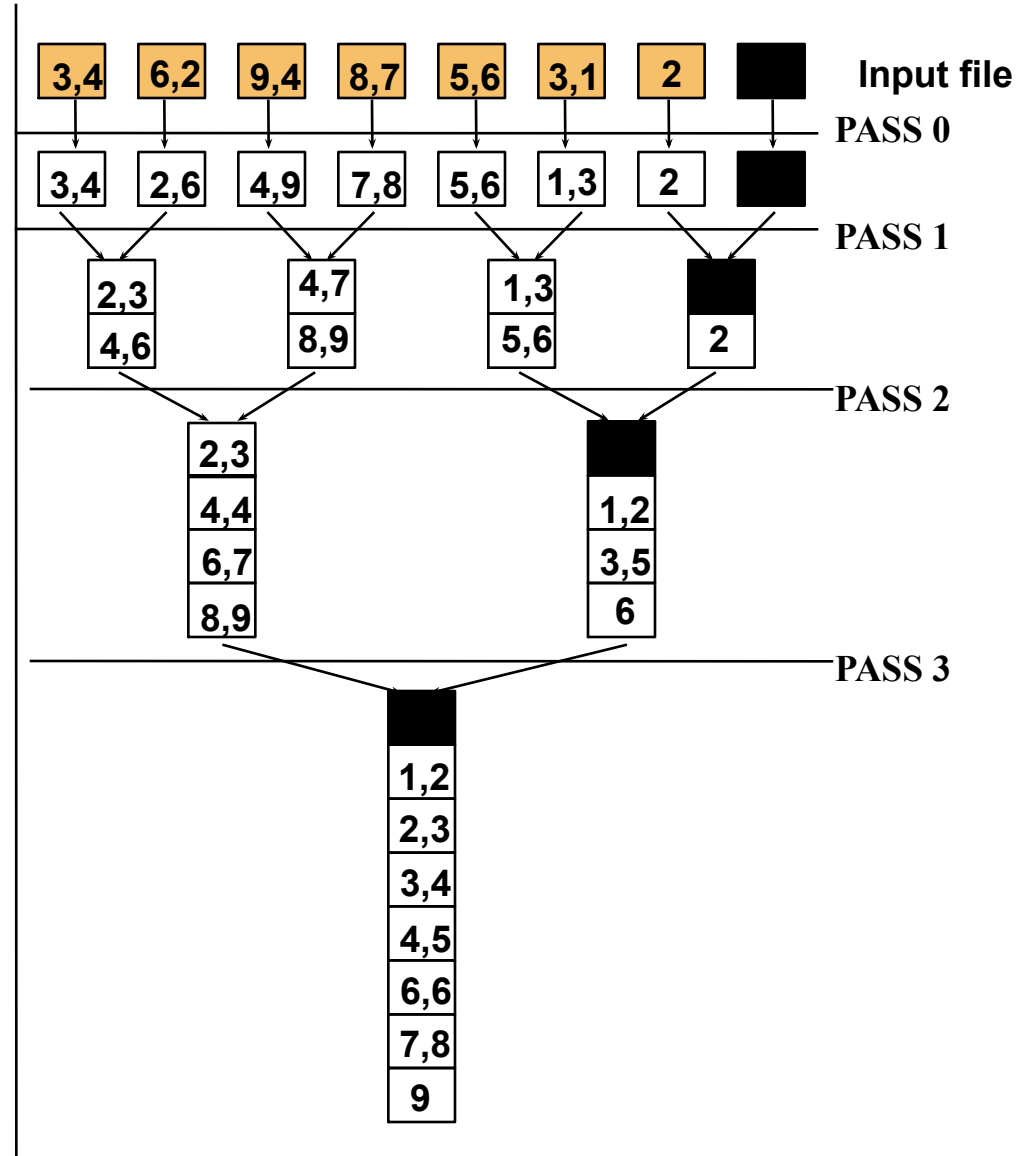
Demanda de 3 buffers

Junta os pares



# Merge-sort em duas vias

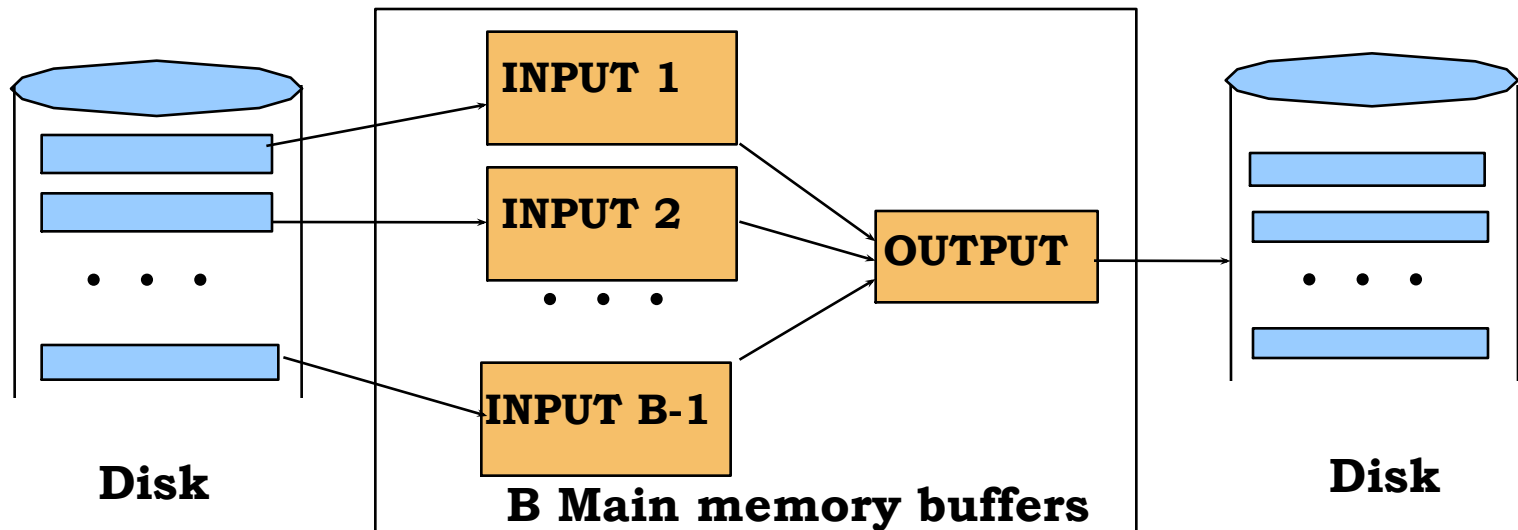
- N número de páginas
- Cada passada lê e escreve uma vez ( $2N$ )
- Número passagens =  $(\log_2 N) + 1$
- Custo final:  
 $2N(\log_2 N + 1)$



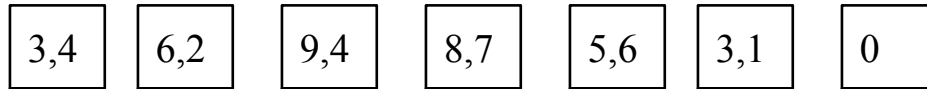
# Merge sort B buffers

- *Mais que 3 buffers de página*

- Para ordenar um arquivo com  $N$  páginas usando  $B$  buffer pages
  - Passo 0: usar  $B$  buffer pages. Produce  $N/B$  páginas ordenadas;
  - Passo 1,2...: merge de  $B-1$  execuções



# Merge-sort em 4 vias



# Custo

- Número de passes:  $1 + \lceil \log_{B-1} N/B \rceil$
- **Custo =  $2N * (\# \text{ of passes})$**
- E.g., com 5 páginas de buffer, e um arquivo de 108 páginas:
  - Pass 0:  $108/5 = 22$  páginas ordenadas

Agora, fazemos o merge com  $B-1$  páginas:

- Pass 1:  $22/4$
- Pass 2:  $6/4$
- Pass 3: ordenado o arquivo

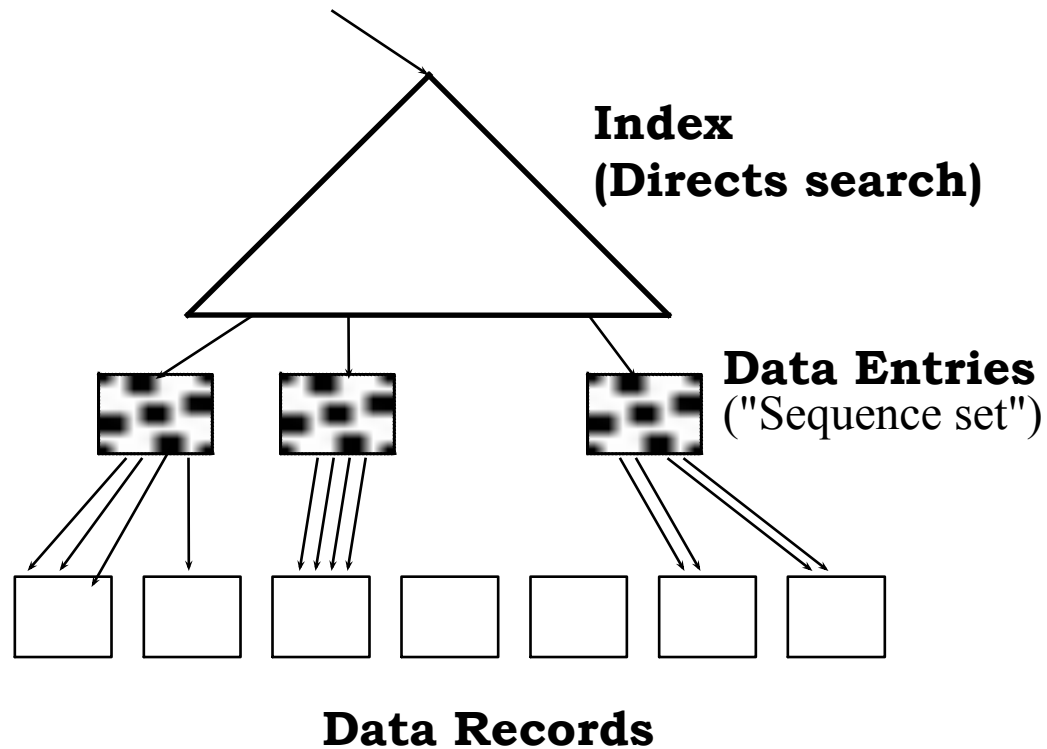


# Number of Passes of External Sort

( I/O cost is  $2N$  times number of passes)

N	B=3	B=5	B=9	B=17	B=129	B=257
100	7	4	3	2	1	1
1,000	10	5	4	3	2	2
10,000	13	7	5	4	2	2
100,000	17	9	6	5	3	3
1,000,000	20	10	7	5	3	3
10,000,000	23	12	8	6	4	3
100,000,000	26	14	9	7	4	4
1,000,000,000	30	15	10	8	5	4

# B+ agrupada para ordenamento



# B+ não-agrupada para ordenamento

one I/O per data record!

