

BÀI TẬP EXTERNAL SORT – NHÓM 3

Cho file F0 gồm các phần tử ($n = 8$):

5 2 -1 6 5 0 9 -2

1. Thực hiện thao tác trộn run trên File (*). Trả lời những câu hỏi:

Số file con là 2, sắp xếp tăng dần

a. Tổng số lần trộn:

- $m = 1 < 8 \rightarrow$ Số Run = 8 \rightarrow Tổng số lần trộn: $8/2 + 8\%2 = 4$
- $m = 2 < 8 \rightarrow$ Số Run = 4 \rightarrow Tổng số lần trộn: $4/2 + 4\%2 = 2$
- $m = 4 < 8 \rightarrow$ Số Run = 2 \rightarrow Tổng số lần trộn: $2/2 + 2\%2 = 1$
- $m = 8 = 8 \rightarrow$ Dừng thuật toán.

\Rightarrow Kết luận: Tổng số lần trộn là $4 + 2 + 1 = 7$

b. Chi tiết file f1 sau lần trộn thứ 1:

- $m = 1$:

F0:	5	2	-1	6	5	0	9	-2
f1:	[5]	[-1]	[5]	[9]				
f2:	[2]	[6]	[0]	[-2]				

F0:	2	5	-1	6	0	5	-2	9
-----	---	---	----	---	---	---	----	---

- $m = 2$:

F0:	2	5	-1	6	0	5	-2	9
-----	---	---	----	---	---	---	----	---

f1:	[2 5]	[0 5]						
-----	-------	-------	--	--	--	--	--	--

f2:	[-1 6]	[-2 9]						
-----	--------	--------	--	--	--	--	--	--

F0:	-1	2	5	6	-2	0	5	9
-----	----	---	---	---	----	---	---	---

\Rightarrow Kết luận: Vậy file f1 sau lần trộn thứ nhất sẽ là: **2 5 0 5**

2. Thực hiện trộn tự nhiên trên File (*). Trả lời những câu hỏi:

Số file con là 2, sắp xếp tăng dần

Các phần tử trong các file thực chất là liên kề và liên tiếp nhau, dùng Tab và “[”, “]” để dễ dàng phân biệt các Run với nhau.

- Lần 1:

F0:	5	2	-1	6	5	0	9	-2
-----	---	---	----	---	---	---	---	----

BÀI TẬP EXTERNAL SORT – NHÓM 3

f1: [5] [-1 6] [0 9]

f2: [2] [5] [-2]

F0: 2 5 -1 5 6 -2 0 9

- Lần 2:

F0: 2 5 -1 5 6 -2 0 9

f1: [2 5] [-2 0 9]

f2: [-1 5 6]

F0: -1 2 5 5 6 -2 0 9

- Lần 3:

F0: -1 2 5 5 6 -2 0 9

f1: [-1 2 5 5 6]

f2: [-2 0 9]

F0: -2 -1 0 2 5 5 6 9

- Lần 4:

F0: -2 -1 0 2 5 5 6 9

f1: [-2 -1 0 2 5 5 6 9]

f2: *NULL*

Dừng thuật toán.

⇒ Số lần trộn: $3 + 2 + 1 = 6$

a. Nguyên nhân hình thành phương pháp trộn tự nhiên:

Nguyên nhân đơn giản là do phương pháp trộn Run chưa tận dụng tối đa được độ dài có thứ tự của một Run. Phương pháp trộn tự nhiên sẽ chọn và đưa các Run có độ dài tối đa vào các file con ngay từ ban đầu, từ đó có khả năng giảm thiểu được số lần trộn các Run giữa các file con.

Một ví dụ đơn giản:

A: 1 2 3 4 -2 6 7 8

Với phương pháp trộn Run thì sẽ tốn 7 lần trộn với các kích thước Run lần lượt là $m = 1, 2, 4, 8$. Nhưng với phương pháp trộn tự nhiên thì chỉ với 1 lần trộn là đã hoàn thành với 2 Run là:

[1 2 3 4] và [-2 6 7 8]

Trường hợp tệ nhất của phương pháp trộn tự nhiên sẽ có độ phức tạp ngang bằng với phương pháp trộn Run: $O(N\log(N))$.

Trường hợp tệ nhất là Mảng đầu vào là một dãy số có thứ tự ngược với yêu cầu đầu ra:

- Yêu cầu **sắp xếp tăng dần** nhưng đầu vào là **mảng giảm dần**.
- Yêu cầu **sắp xếp giảm dần** nhưng đầu vào là **mảng tăng dần**.

b. Chọn 3 phần tử riêng tùy ý để thêm vào file ban đầu (điều kiện là số lần trộn tăng đúng 1 lần):

BÀI TẬP EXTERNAL SORT – NHÓM 3

Chọn 3 phần tử: 9 8 10 thêm vào cuối file F0

- Lần 1:
Run: [5] [2] [-1 6] [5] [0 9] [-2 9] [8 10]
- Lần 2:
Run: [2 5] [-1 5 6] [-2 0 9 9] [8 10]
- Lần 3:
Run: [-1 2 5 5 6] [-2 0 8 9 9 10]
- Lần 4:
Run: [-2 -1 2 0 5 5 6 8 9 9 10]

c. Hãy cho biết nguyên nhân hình thành phương pháp trộn đa lối cân bằng:

Phương pháp trộn tự nhiên đã tận dụng tối đa độ dài vốn có của 1 Run nhưng mỗi lần thao tác, ta phải copy các phần tử từ file nguồn ra các file con. Để giảm thiểu thời gian lãng phí vào việc copy tài nguyên từ file nguồn ra các file con, ta sẽ vận dụng trực tiếp kết quả sau khi trộn cho lần trộn tiếp theo của file mà không cần phải mỗi lần trộn xong ta đưa kết quả vào file nguồn rồi lại copy lại từ file nguồn.

Hiệu quả của phương pháp trộn đa lối cân bằng này là ta hình thành 2 tập file nguồn và đích:

- File nguồn: chứa các Run chờ để trộn.
- File đích: chứa các kết quả sau khi trộn các Run của các file nguồn.

Sau mỗi lần trộn, file nguồn và file đích sẽ hoán đổi vai trò cho nhau và tiếp tục thực hiện trộn. Các file đích và nguồn luân phiên hoán đổi vai trò cho nhau. Thuật toán dừng khi mà chỉ tồn tại duy nhất 1 file nguồn của Run (cụ thể là chỉ chứa 1 Run).

VD:

- Lần 1:
Sour: F1, F2, F3
Des: F4, F5, F6
- Lần 2:
Sour: F4, F5, F6
Des: F1, F2, F3
...

Phương pháp trộn đa lối cân bằng giảm thiểu số lần copy dữ liệu nguồn thông qua số lượng file của tập file nguồn và file đích.

BÀI TẬP EXTERNAL SORT – NHÓM 3

THÀNH VIÊN NHÓM:

1. Phạm Bùi Nhật Huy – 20521410
2. Nguyễn Thị Như Vân – 20520855
3. Nguyễn Minh Trí – 19522388