

BERKAS SORT DAN MERGE

PENGERTIAN BERKAS SORT DAN MERGE

Dalam sistem penyortiran dikenal 2 metode, yaitu:

- Metode sort internal
- Metode sort eksternal

Perbedaannya:

- Pada metode sort internal, semua record yang akan diproses dimuat kedalam memori komputer lalu diproses sort (*sortir*).
- Pada metode sort eksternal, record-record yang diproses tidak semuanya dapat dimuat kedalam memori komputer, karena keterbatasan memori komputer.
- Metode sort eksternal didalam penerapannya nanti, menggunakan pula metode sort internal.

Contoh:

Sebuah file berisi 2000 record harus di *sortir* kedalam memori yang hanya dapat manampung 1000 record sekaligus. Untuk itu digunakan metode sort eksternal.

Langkah-langkah dalam penyortiran ini adalah:

- Record-record dibagi kedalam beberapa file agar dapat ditampung sekaligus di memori komputer, lalu masing-masing bagian di *sortir* internal. Bagian-bagian file yang telah tersortir ini disebut *sorted sublist*.

Maka didapat:

- Sorted sublist 1 (record 1 – 1000) dan
- Sorted sublist 2 (record 1001 – 2000)
- Setelah itu kedua *sorted sublist* ini (RUN) digabung (*merge*), sehingga didapat berkas gabungan (*merge file*) yang record-record nya telah di *sortir*.

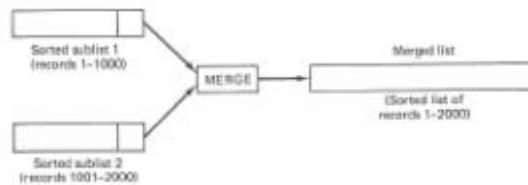


Figure 12-4 Natural merge of two sorted lists.

Gambar 1. Natural Merge of two sorted list

Maka dapat disimpulkan langkah-langkah untuk metode sort eksternal ini adalah:

- Sort eksternal, dimana file dibagi menjadi beberapa bagian file, kemudian di sortir.
- Merge, dimana bagian-bagian file ini (*sorted sublist*) digabung menjadi satu atau lebih file gabungan. File-file gabungan kemudian digabung lagi sampai akhirnya didapatkan sebuah file gabungan yang berisi semua record-record yang telah di sortir.
- Output, yang menyalin file gabungan yang telah ter sortir ke media storage terakhir.

Faktor-faktor yang mempengaruhi metode sort eksternal:

- Jumlah record yang akan di sortir
- Ukuran record (panjang record)
- Jumlah storage yang digunakan
- Kapasitas internal memori
- Distribusi nilai key dalam input file.

Teknik sort/merge file ini berbeda satu dengan yang lainnya dalam hal:

- Metode sort internal yang digunakan
- Jumlah main memori yang disediakan untuk sort internal.
- Distribusi dari *sorted sublist* di *secondary storage* menjadi satu atau lebih file gabungan dalam satu langkah gabungan (*merge pass*).

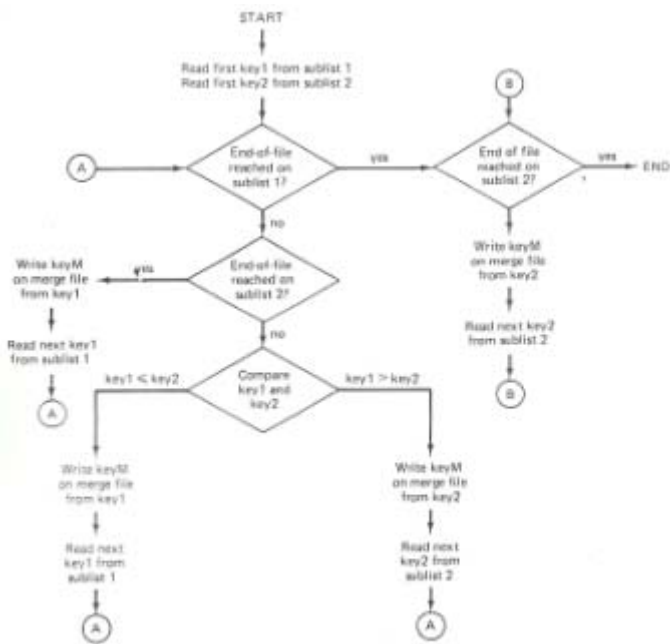


Figure 12-2 Merge logic for two sorted lists.

Gambar 2. Merge logic for two sorted list

Ada 4 teknik sort/merge file, yaitu:

- Natural Merge
- Balanced Merge
- Polyphase Merge
- Cascade Merge

NATURAL MERGE

Merge yang menangani 2 input file sekaligus disebut 2 way natural merge. Merge yang menangani M input file sekaligus disebut M way natural merge. M menunjukkan derajat merge.

Pada natural merge, dibagi lagi menjadi:

- 2 way natural merge
- 3 way natural merge
- ⋮
- ⋮
- M way natural merge

Pada M way natural merge, dapat didefinisikan sebagai merge dengan:

M input file dan hanya 1 output file

Contoh:

Sebuah file yang terdiri dari 6000 record hendak di sortir kedalam memori komputer yang kapasitasnya 1000 record.
Buatlah dengan menggunakan 2 way natural merge!

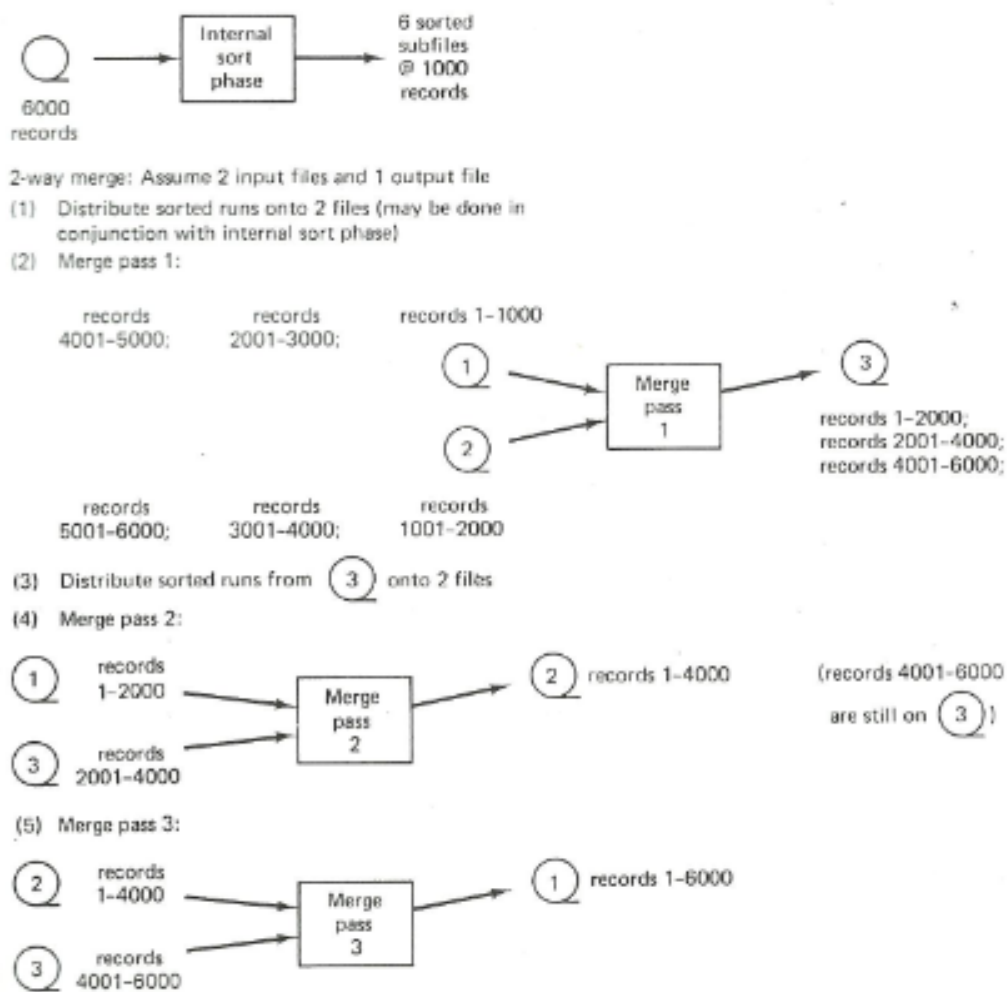


Figure 12-3 2-way natural merge example.

Gambar 3. Contoh 2-way Natural Merge

Contoh 3 way Natural Merge:

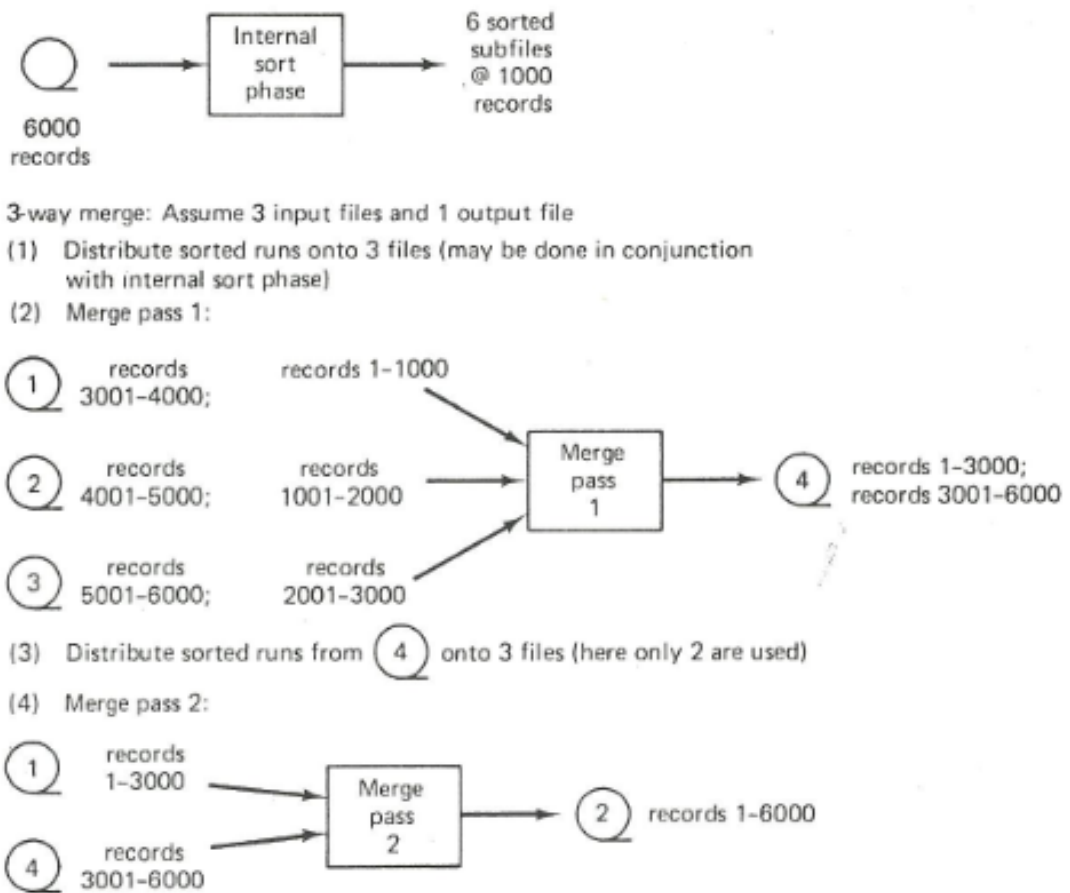


Figure 12-4 3-way natural merge example.

Gambar 4. Contoh 3-way Natural Merge

Contoh 2 way Natural Merge, yang kapasitas memorinya 500 record:

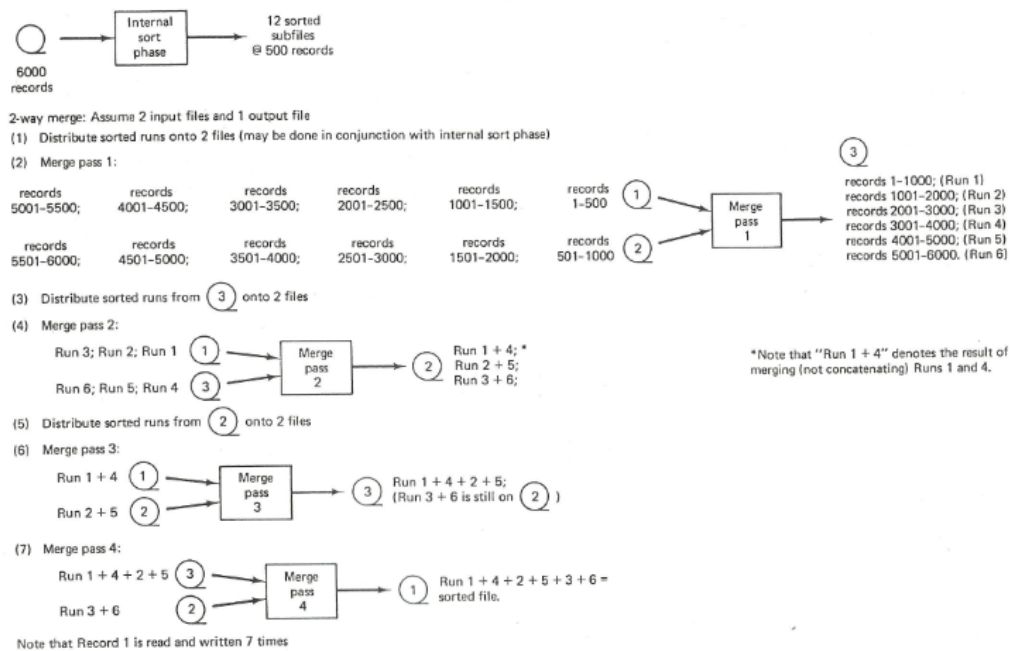
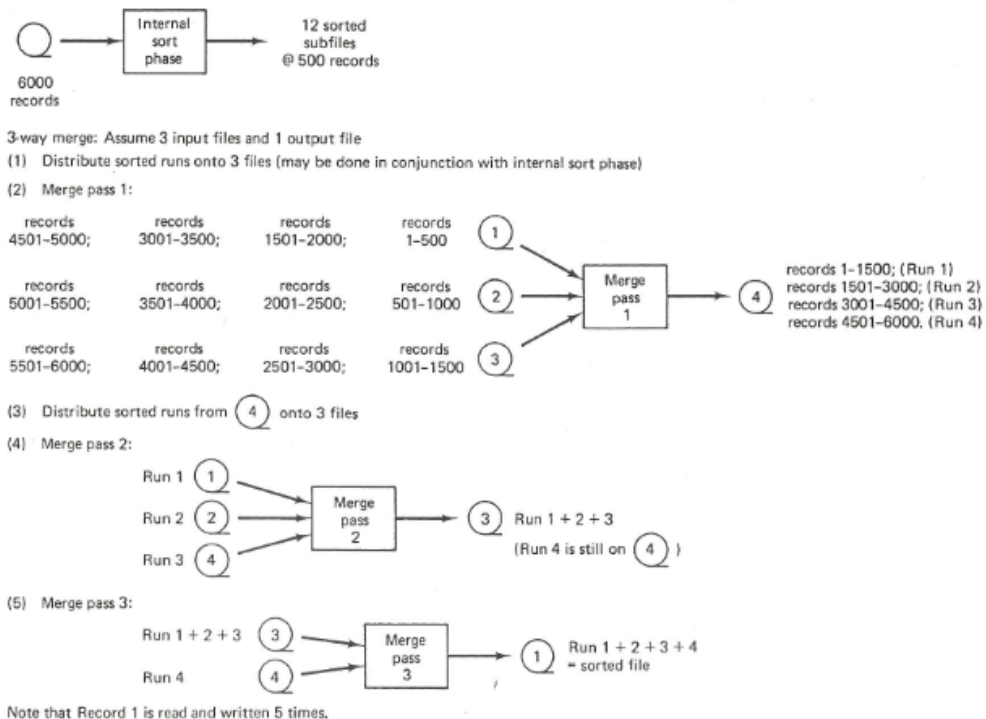


Figure 12-5 2-way natural merge example.

Gambar 5. Contoh 2-way Natural Merge

Contoh 3 way Natural Merge, yang kapasitas memorinya 500 record:



Gambar 6. Contoh 3-way Natural Merge

BALANCED MERGE

Dari metode natural merge kita lihat bahwa, jika kita gunakan M input file, maka file seluruhnya yang kita gunakan adalah $M + 1$ file.

Sedangkan pada balanced merge, jika kita gunakan M input file, maka file seluruhnya yang dipakai adalah $2 M$ file.

Pada balanced merge dibagi lagi menjadi:

- 2 way balanced merge
- 3 way balanced merge
- ⋮
- ⋮
- ⋮
- M way balanced merge

Pada balanced merge, jumlah input file sama dengan jumlah output file, walaupun pada akhirnya tak ada lagi keseimbangan antara input dan output file.

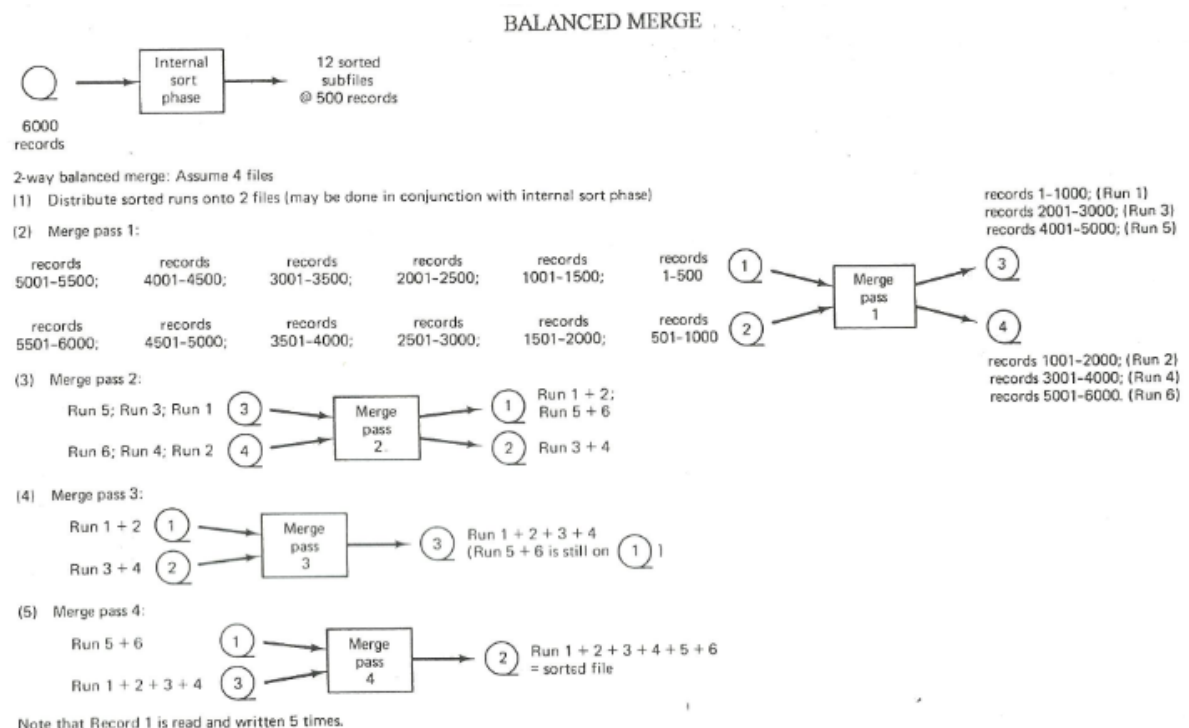


Figure 12-7 Example 2-way balanced merge.

Gambar 7. Contoh 2-way Balanced Merge

POLYPHASE MERGE

Pada M way polyphase merge digunakan $2M-1$ input file dengan 1 output file. Jadi jika kita menggunakan 2 way polyphase Merge, maka banyaknya input file yang digunakan ada 3 input file.

Contoh:

Setelah phase sort internal, misalkan kita mempunyai 17 subfile atau 17 run yang akan didistribusikan kedalam 3 input file. Jika kita menggunakan 2 way polyphase merge, berarti 17 run tersebut harus didistribusikan kedalam 3 input file.

Dari pendistribusian tersebut, maka diperoleh:

- input file 1 berisi 7 run
- input file 2 berisi 6 run
- input file 3 berisi 5 run

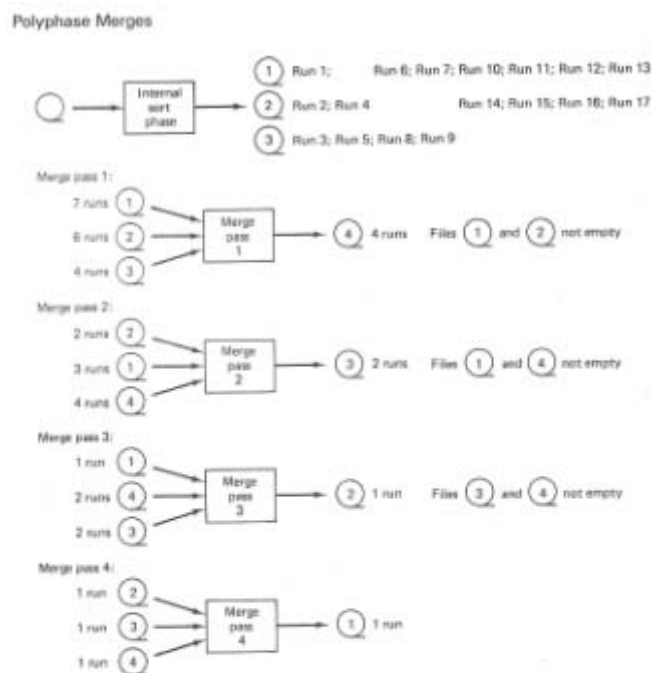


TABLE 12-1 DETERMINING RUN DISTRIBUTIONS FOR A POLYPHASE MERGE

	# runs/file			
	Cycle 1	Cycle 2	Cycle 3	Cycle 4
file 1	1	1	3	7
2	1	2	2	6
3	1	2	4	4
Total	3	5	9	17

Gambar 8. Contoh Polyphase Merge

CASCADE MERGE

Jenis lain dari *unbalanced merge* yang berusaha mengurangi penyalinan/copy dari record-record disebut *cascade merge*.

Cascade merge dengan derajat M menggunakan:

$2M-1, 2M-2, 2M-3, \dots$, kemudian 2 input file selama merge

Setiap merge pass dimulai dengan merge dari:

$2M-1$ input file ke 1 output file

Pada cascade merge pendistribusian run-nya sama dengan pendistribusian run pada polyphase merge, hanya berbeda pada phase merge-nya.

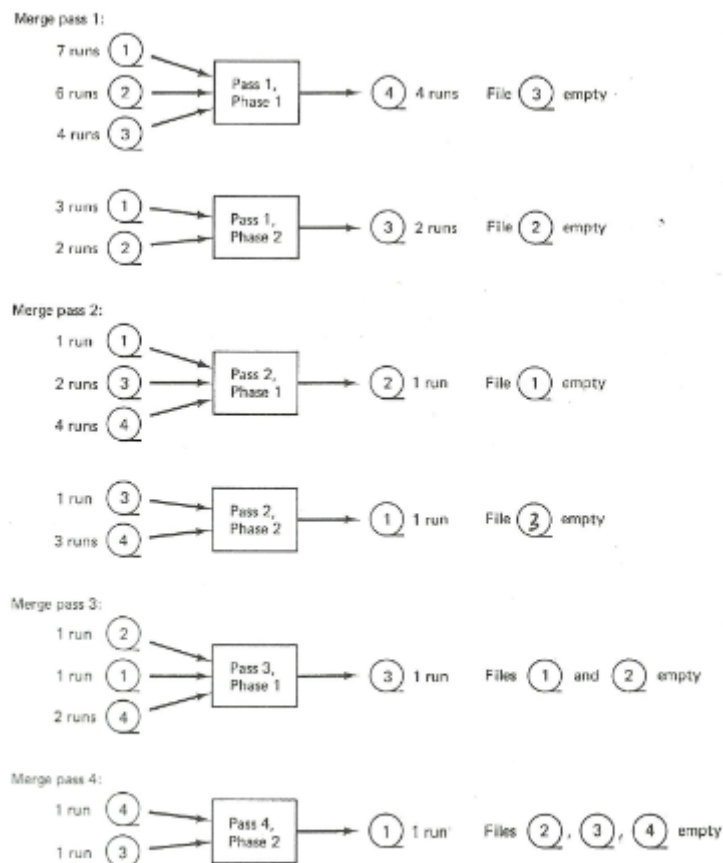


Figure 12-9 Example cascade merge.

Gambar 9. Contoh Cascade Merge