

# SORT DAN MERGE FILE

Dalam sistem penyortiran dikenal 2 metode, yaitu:

## 1. Metode Sort Internal

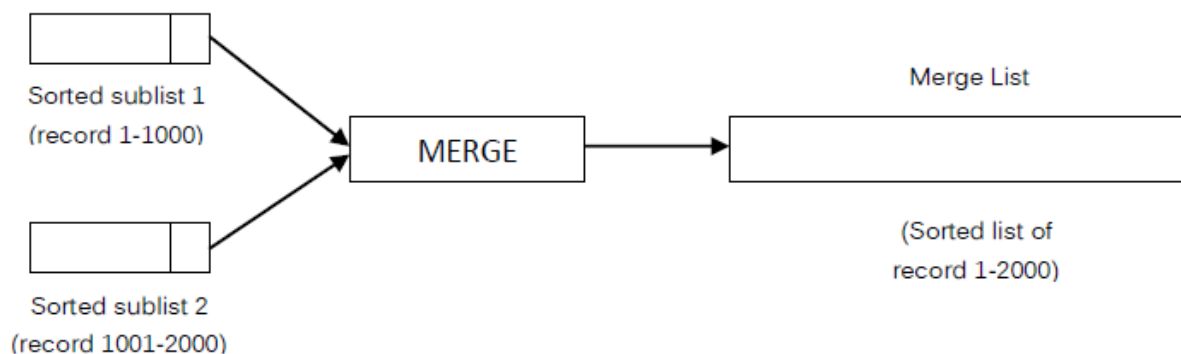
Semua record yang akan diproses dimuat ke dalam memori komputer lalu diproses sort (sortir)

## 2. Metode Sort Eksternal

Record-record diproses tidak semuanya dapat dimuat ke dalam memori komputer, karena keterbatasan memori komputer

Contoh: Sebuah file berisi 2000 record harus disortir ke dalam memori yang hanya dapat menampung 5000 record sekaligus. Untuk itu digunakan metode sort eksternal. Langkah-langkah penyortiran ini adalah:

1. Record-record dibagi ke dalam 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:
  - a. Sorted sublist 1 (record 1-1000) dan
  - b. Sorted sublist 2 (record 1001-2000)
2. Setelah itu kedua sorted sublist ini (RUN) digabung (merge), sehingga didapat berkas gabungan (merge file) yang record-recordnya telah disortir.



- Kita simpulkan langkah-langkah metode sort eksternal ini adalah:
- Internal sort, 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 ini kemudian digabung lagi sampai akhirnya didapatkan sebuah file gabungan yang berisi semua record-record yang telah tersortir.
- Output, dimana menyalin file gabungan yang telah disortir kemudian storage terakhir.

Faktor-faktor yang mempengaruhi metode sort eksternal:

- Jumlah record yang akan disortir
- 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 lainnya dalam hal:

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

## **A. NATURAL MERGE**

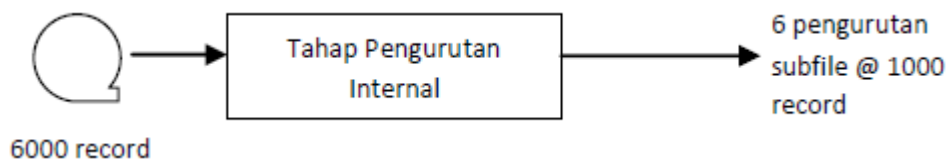
Merge yang menangani dua input file sekaligus disebut dua way natural merge, merge yang menangani M input file sekaligus disebut M way natural merge. M menunjuk derajat merge. M way natural merge dapat didefinisikan sebagai merge dengan M input file dan hanya satu output file. Contoh, sebuah file yang terdiri dari 6000 record hendak disortir ke dalam memori komputer yang kapasitasnya 1000 record. Pada contoh ini, sebagai eksternal storage digunakan tiga tape. Tape satu berisi tiga sorted sublist yaitu record 1-1000, record 2001-3000 dan record 4001-5000. Sedangkan tape 2 berisi tiga sorted sublist yang terdiri dari record 1001-2000, record 3001-4000 dan record 5001-6000. Dengan 2 way natural merge, dimana dua input file sekaligus digabung menjadi sebuah output file di tape tiga.

Record 1-1000 }  
Record 1001-2000 } Menjadi record 1-2000

Record 2001-3000 }  
Record 3001-4000 } Menjadi record 2001-4000

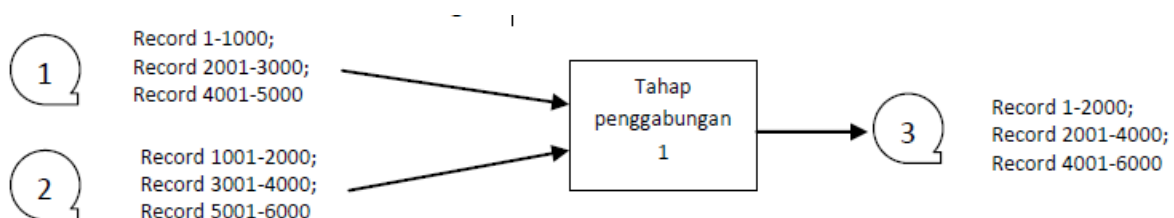
Record 4001-5000 }  
Record 5001-6000 } Menjadi record 4001-6000

Langkah selanjutnya, 1 dari 3 file digabung (record 1 – 2000) pada tape 3 ini disalin ke tape 1. Pengerjaan seterusnya dapat dilihat pada gambar dibawah ini. Hasil akhirnya adalah sebuah file gabungan berisi 6000 record yang telah terseortir di tape 3.

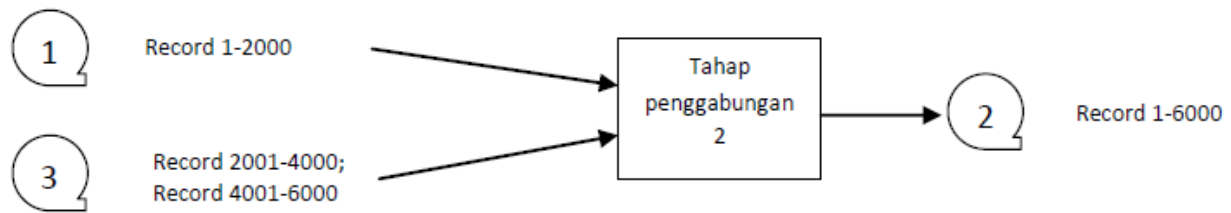


2 way merge. Diasumsikan 2 input file dan 1 output file

1. Pengurutan terdistribusi terhadap 2 file (bisa diselesaikan dalam konjungsi dengan tahap pengurutan internal).
2. Tahap Penggabungan/Merge 1

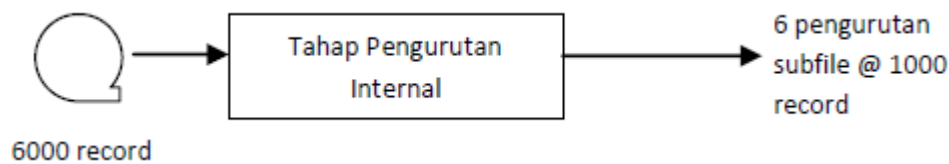


3. Salah satu dari 3 hasil penggabungan sementara pada file 3 dipindah ke file 1
4. Tahap penggabungan kedua



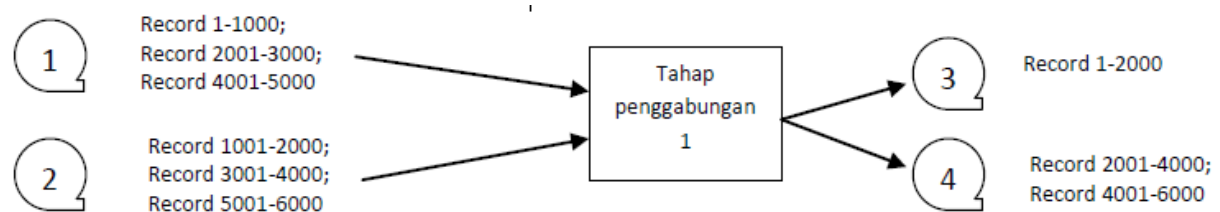
## B. BALANCED MERGE

Keperluan penggunaan tape pada natural merge dapat dikurangi dengan menggunakan balanced merge. Pada balanced merge, tidak ada lagi langkah pendistribusian hasil merge ke dalam beberapa tape. Berbeda dengan natural merge, balanced merge pada awalnya ada keseimbangan antara input file dan output file, walaupun pada akhirnya tidak ada lagi keseimbangan antara input dan output file. Seperti halnya natural merge, balance merge juga ada beberapa cara yaitu 2 way balance merge, 3 way balance merge, dan M way balance merge. 2 way balance merge berarti merge dengan 2 input file sekaligus dan hasilnya 2 output file. M way balance merge berarti M input file dengan M output file

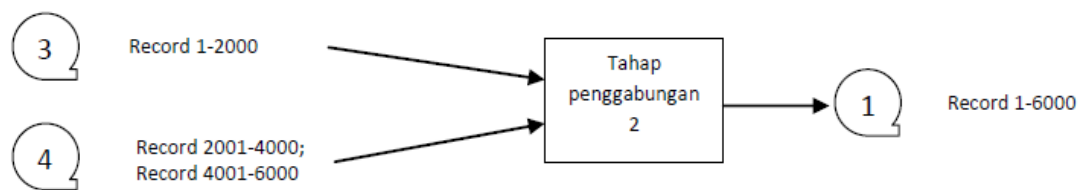


2 way merge. Diasumsikan 2 input file dan 1 output file

1. Pengurutan terdistribusi terhadap 2 file (bisa diselesaikan dalam konjungsi dengan tahap pengurutan internal).
2. Tahap Penggabungan/Merge 1
- 3.

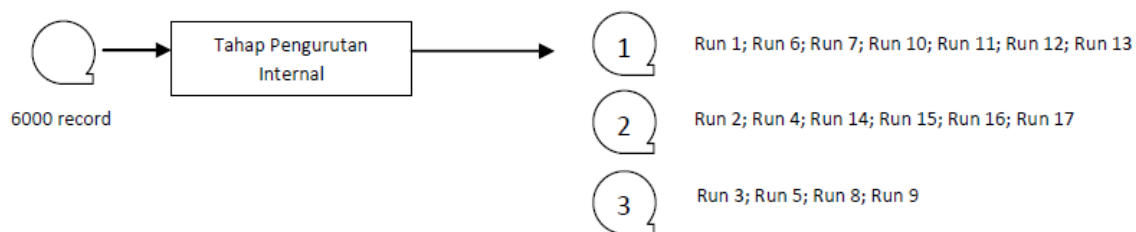


4. Salah satu dari 3 hasil penggabungan sementara pada file 3 dipindah ke file 1
5. Tahap penggabungan kedua

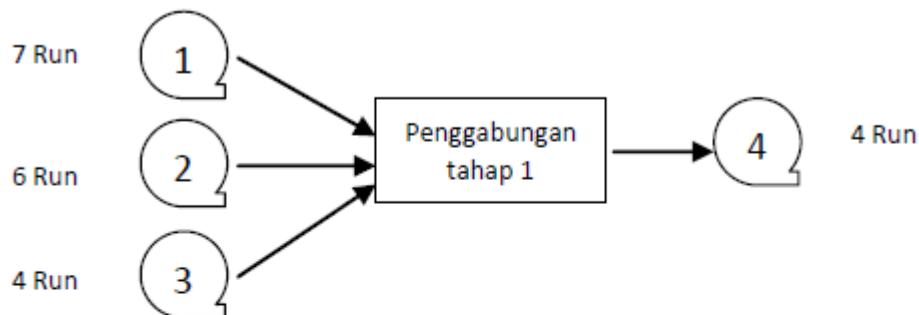


### C. POLYPHASE MERGE

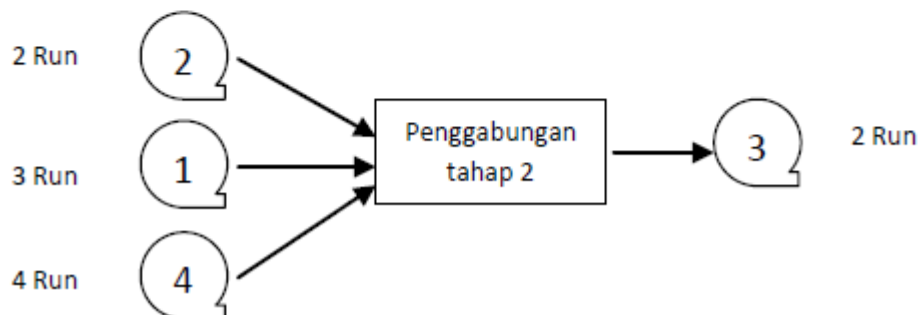
Kita telah lihat bahwa M way balance merge menggunakan 2 M file (M input file dan M output file). Karena yang digunakan setiap saat hanya M file sebagai input dan direkam ke sebuah file maka ada M-1 file yang nganggur (idle). Untuk itulah perbaikan dari kelemahan ini diambil oleh poplyphase merge. Pada M way polyphase merge digunakan 2M-1 input file dengan 1 output file sekaligus.



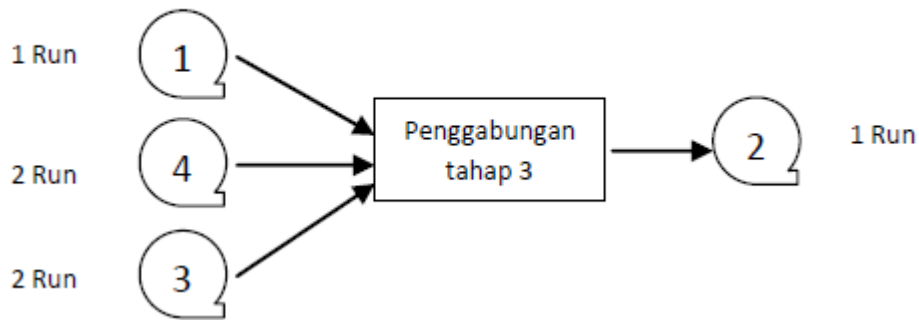
#### Penggabungan tahap 1:



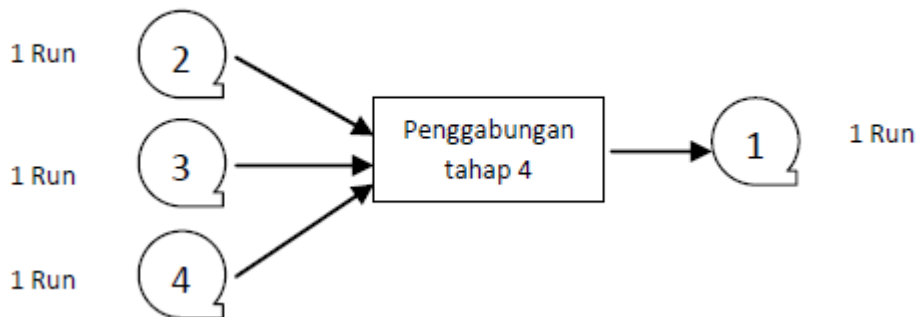
#### Penggabungan tahap 2:



### Penggabungan tahap 3:



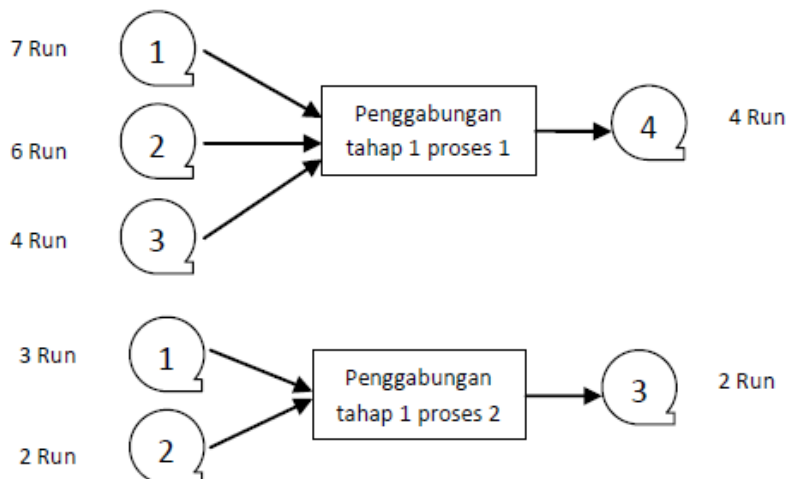
### Penggabungan tahap 4:



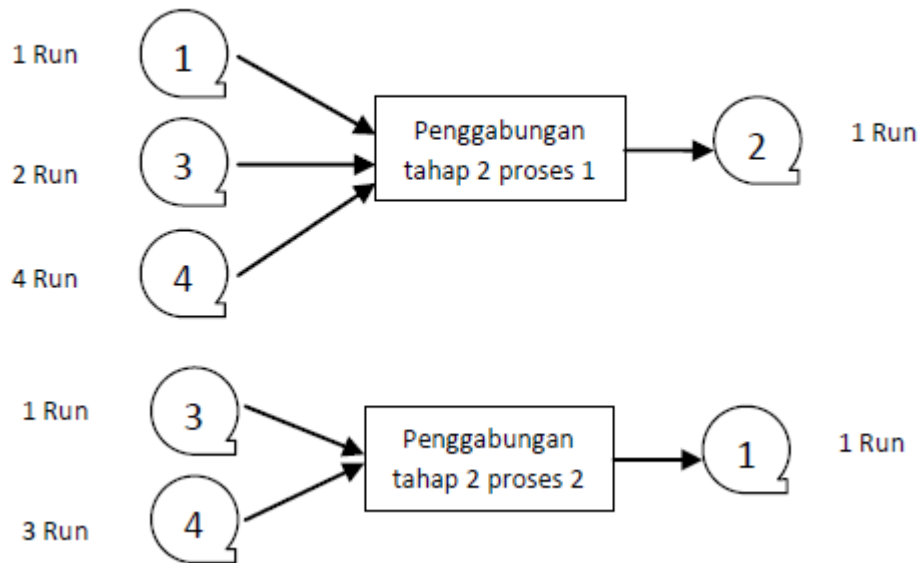
## D. CASCADE MERGE

Jenis lain dari unbalanced merge yang berusaha mengurangi penyalinan dan pembacaan record-record disebut cascade merge. Cascade merge dengan derajat  $M$  menggunakan input file  $2M-1$ , kemudian  $2M-2$  dan  $2M-3$ , ..., 2 input file selama tiap tahap merge. 3 way cascade menggunakan 3 dan 2 input file selama tiap tahap merge.

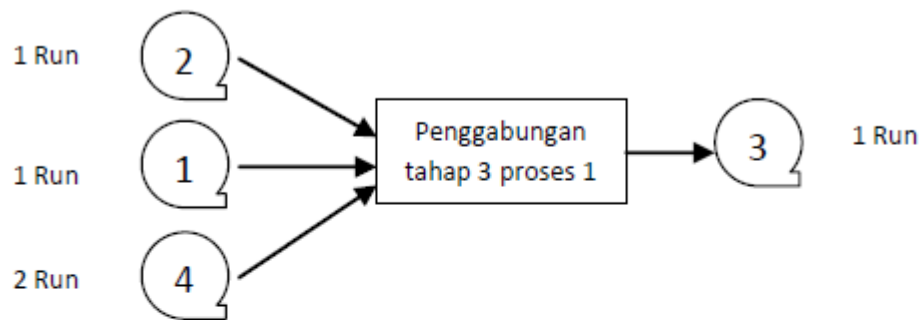
### Penggabungan tahap 1:



### Penggabungan tahap 2:



### Penggabungan tahap 3:



### Penggabungan tahap 4:

