

# BAB 7

## 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 ke dalam memori komputer lalu diproses sort (sortir).
- ✓ Pada metode sort eksternal, record-record yang diproses tidak semuanya dapat dimuat ke dalam memori komputer, karena keterbatasan memori komputer.
- ✓ Metode sort eksternal di dalam penerapannya nanti, menggunakan pula metode sort internal.

### **Contoh :**

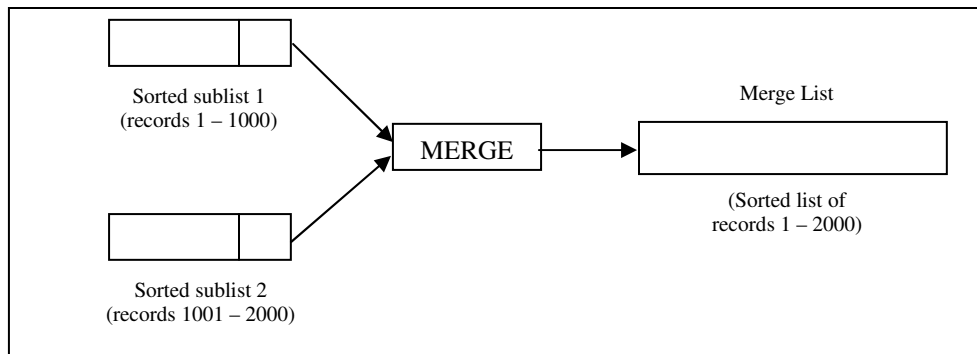
Sebuah file berisi 2000 record harus disortir ke dalam memori yang hanya dapat menampung 1000 record sekaligus. Untuk itu digunakan metode sort eksternal.

*Langkah-langkah penyortiran ini adalah :*

- ❖ Record-record dibagi ke dalam beberapa file agar dapat ditampung sekaligus di memori komputer, lalu masing-masing bagian disortir 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-recordnya telah disortir.



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

- Sort internal, dimana file dibagi menjadi beberapa bagian file, kemudian disortir.
- 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 disortir.
- Output, yang menyalin file gabungan yang telah tersortir ke media 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 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)

*Ada 4 teknik dalam 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 terbagi 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 disortir ke dalam memori komputer yang kapasitasnya 1000 record.

Buatlah dengan menggunakan 2 way natural merge !

### **Lihat gambar 1**

Lihat pula contoh 3 way natural merge, yang ditunjukkan pada :

### **Gambar 2**

Lihat pula contoh 2 way natural merge dengan kapasitas memori 500 record.

### **Gambar 3**

Lihat pula contoh 3 way natural merge dengan kapasitas memori 500 record.

### **Gambar 4**

## *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  $2M$  file.

Pada balanced merge terbagi 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.

**Lihat gambar 5**

## *Polyphase Merge*

Pada  $M$  way polyphase merge digunakan  $2M - 1$  input file dengan 1 output file. Jadi 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 ke input file. Jika kita menggunakan 2 way polyphase merge, berarti 17 run tersebut harus didistribusikan ke dalam 3 input file.

Dari pendistribusian tersebut, maka diperoleh :

**Lihat gambar 6**

## *Cascade Merge*

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

Cascade merge dengan derajat  $M$  menggunakan :

$2^{M-1}$ ,  $2^{M-2}$ ,  $2^{M-3}$ , ... , kemudian 2 input file selama merge

Setiap merge pass dimulai dengan merge dari :

$2^{M-1}$  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.

**Lihat gambar 7**