G. DIACONESCU
Department of Economic Cybernetics
Academy of Economic Studies

A Method of Data Files Dynamic Reorganization

The study presents a method of direct files organization according to a key which allows the marked reduction of the reorganizing necessities. The method is based on the introduction of the present effective and dynamic personnel into the file and the construction of access tables.

The method is recommended to be used in case of high volatile file with fixed length records. It is pointed out that the method creates the possibility of an easy use of inverted files.

1. CONCEPT OF RE-ORGANIZATION

It is well known in the literature of speciality that the representation of a data collection on a physical store socket is liable to strict hardware limitations. The modern virtualizing methods used at present aim to relieve the user from the physical representation of data. Nevertheless, the representation of data collection must dynamically correspond to the very collection of data considered as the set in which continuously exist delations, additions and updatings. It results that any access algorithm which is based essentially on the initial (load) representation of the data collection will undergo a slow process of degradation (worsening) of the initial performances.

After a period of existence of the representation of the data collection, the delations and additions tend to entirely change the configuration of representation as compared to the initial one, so that the access algorithm which was performant at the moment of loading may become entirely inoperative. In this case, the solution which is frequently used consists in reorganizing the files, which in fact means to consider the last configuration of representation as loading data for another representation. Since upon its creation the access

algorithm uses to the maximum the particularities of the data (the sequence in the sorting order, the grouping around certain values of HASH-ADDRESS, etc.), the access becomes performant again.

2. DYNAMIC REORGANIZATION

It is known that in classical organizations, a special storage space is allocated for additions and that the space becoming free through delations is not available.

Moreover, due to the chaining of adresses in additions, the access time increases with the growing number of adress chains for additions and with the length of these chains.

Though the space becoming free through delations can be made available through a chain of free addresses (for example, the LIFO technique as in BOMP), however a too marked extension of this space leads to a heavy access.

The dynamic reorganizations must thus correct these shortcomings, namely:

1) to immediately make available the space released through delations;
2) to maintain constant the performances of the access algorithm;
3) to eliminate the necessity for a repeated creation of the file and to treat additions and delations similary to the creation transactions.

Such requirements imposed on a dynamic reorganizing method can be achieved in various ways.

3. THE MUSCEL METHOD OF DYNAMIC REORGANIZATION

In the following, a method is described, based on the use of the following elements:

— the dynamic personnel of the file,
— the present personnel of the file,
— the access table,
— the delation table,
— the presence table.

The main point of the method is the clear-cut separation between the dynamic personnel of the file and its present personnel and the preserving of an access interface between the two categories of personnel through the access tables.

By *the present* personnel of the file, we mean the collection of records at a given moment.

1. It stands to reason that the present personnel varies in time, nevertheless, in most of the applications, a superior limit which remains valid for a long period of time can be established to the number of records.

For example, if we refer to the personnel file of an enterprise, the present personnel may vary as follows:

1967	1968	1969	1970	1971	1972	1973
23 000	26 000	27 000	24 000	24 500	26 000	26 500

The value of 35 000 can be established, for example, as a superior limit of the present personnel, provided that this value be valid for a long period of time.

By *the dynamic* personnel of the file, we define the collection of all records included into the file for an established period of time, as a rule, as long as possible. By resuming the above example with the personnel file, the dynamic personnel for the 1967—1973 period will represent the number of all employments of personnel in this period of time (consider this number to be 120,000 -by admitting a big fluctuation of personnel). The word "personnel" in "dynamic personnel" and "present personnel" does not mean that the file must be always a personnel file.

Essentially, it is supposed that each records of the file has a single key within the dynamic personnel. In our example, the key will be the identity

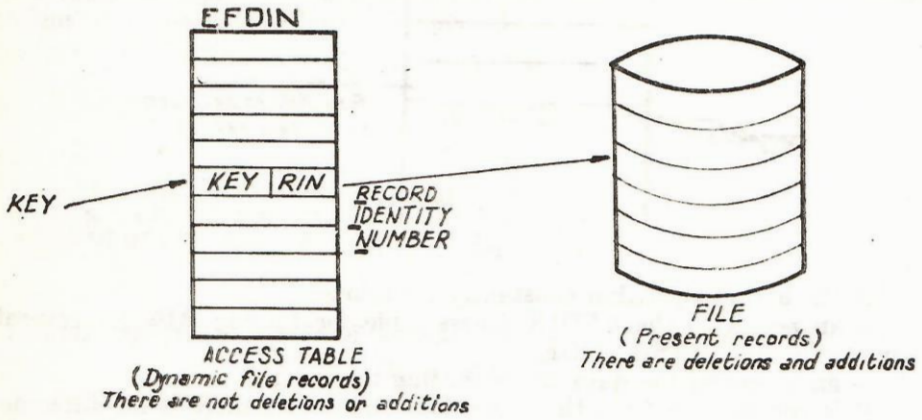


Fig. 1

number which will be single for each of the 120,000 employees in the 1967—1973 period.

The data file proper (in the classical sense) will be the file of the present personnel containing the data of the records present in this file, erases and additions being made continuously.

The access table is a dynamic personnel file, each record containing the single key of a record in the dynamic personnel and the RIN access key in the present personnel (the data file).

The input into the access table is made with a hash-code according to the external key (the identity number in our case).

The RIN RECORD IDENTITY NUMBER key of access to the data file is the serial number of identifying the present records. This key is dynamically granted with the appearance of records and does not imply the user in any way but only the system.

It is very important that on delations the records in the data file, the corresponding RIN is sequentially deposited into the delation table. If after delations additions appear, the RIN granted to an addition is taken from the erase table. The filling in and taking of a RIN from the delation table is made according to the LIFO technique (the list with a fixed end of free RIN) Fig. 2.

Finally, useful is also the presence table which represents, in fact, a bit pattern of the RINs present in the data file the number of bits being equal to the maximum attained present personnel.

The advantages of such an organization are obvious in the case of a file with big fluctuation. All requirements are observed for a dynamic reorganizatio n as they had been previously announced.
Thus:

1) a delation in the data file makes immediately available the space for a new addition through the free RIN table (the delation table);

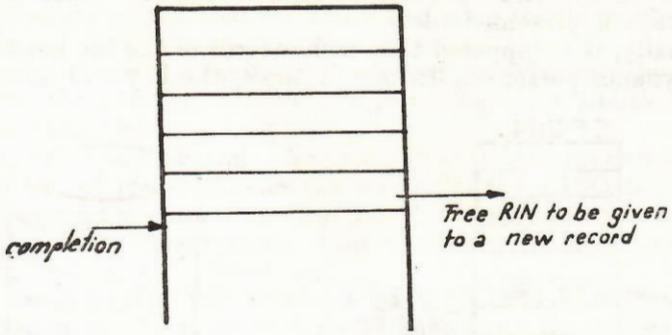


Fig. 2

2) the access algorithm constantly containts:
— an access to the EFDIN access table for finding RIN (or several, depending on the HASH-CODE),
— an access to the data file according to RIN;

3) it can be seen from the execution model that there is no difference between creation, addition or delation. The only creation problem is, in fact, the initialization of the FEDIN table and of the data file as fictitious tables — their formatization and filling in with blanks or zero.

D. CONCLUSIONS

1. The method proposed for a dynamic reorganization of files can be used with great efficiency in the case of files with fixed length records, organized directly and with great volatility.

2. This method requires, as a preliminary, the existence of single keys of records from the dynamic personnel. Otherwise (duplicate keys) the system may take over the construction of single keys by assigning some serial numbers to the multiple keys. This is the most general case and it was developed in the MUSCEL system (THE STABLE UNIFIED MATRICULATION ON THE ELECTRONIC COMPUTER) for a stable descriptive coding of persons. This method will be presented in a subsequent development of the method.

3. For the beginning, in experimenting the method we can use the comparison between the average time of access to this method and that of a method with the natural increase of the number of indexes and the control of free space (the VSAM method).

4. By introducing the RIN key into the present personnel the method facilitates the inverting of the present personnel and correspondingly, the in-

tensive use of the file for retrieval problems.

5. The MUSCEL dynamic reorganizing method finds its applicability in the designing of big files with many additions and delations which are used for a long period of time and in which the problem of uniformizing the average access time for the whole period of exploitation is of particular importance.

6. The advantages of the above method in comparison with other existing methods cannot be made evident outside the above mentioned context (relatively stable present personnel, dynamic personnel with a marked growth, files with direct access according to key, the necessity of a inverted file). A quantitive estimation of the performance of the method was not achieved yet for a concrete implementation, however, the author hopes that he has made evident at a qualitative level, the advantages of the method proposed.

REFERENCES

1. DOS-VS *OPERATING SYSTEM* — IBM, 1973.
2. DOS-VS *DATA MANAGEMENT* — IBM, 1973.
3. SIRIS-III-*OPERATING SYSTEM* — CII, 1973.

UNE METHODE DE REORGANISATION DYNAMIQUE DES FICHIERS DE DONNEES

L'auteur présente une méthode d'organiser les fichiers directs selon une clé qui permet la réduction considérable des nécessités de réorganisation. La méthode consiste dans l'introduction dans le fichier des notions d'effectif présent et d'effectif dynamique et la construction de tableaux d'accès.

On recommande l'utilisation de la méthode au cas d'un fichier à effectif présent stable, d'un effectif dynamique à croissance marquée et des enregistrements de longueur fixe. On met en évidence le fait que la méthode crée la possibilité de l'utilisation facile des fichiers inverses.

МЕТОД ДИНАМИЧЕСКОЙ РЕОРГАНИЗАЦИИ КАРТОТЕК ДАННЫХ

В настоящей статье представляется метод организации прямых картотек по ключу, позволяющему значительно уменьшить потребности в реорганизации. Метод основан на внедрении понятий «настоящее наличие» и «динамическое наличие» в картотеке и построении таблиц для использования.

Применение метода рекомендуется для картотек со стабильным настоящим наличием, значительным ростом динамического наличия и неизменной длиной записей. Подчеркивается что этот метод позволяет легко использовать обратные картотеки.

21.11.1974