

## File & Database Design

### ❖ Basic terminology

**Data item** : Individual items of data are called data items or fields or simply items. Data items can comprise sub items or sub fields.

**Record** : The complete set of related data pertaining to an entity. When a number of size of data items in a record is constant then such a record is called fixed length record. If the number of data items varies from one record to another record then such record is called the variable length record.

**Record Key** : To distinguish one specific record from another, systems analysts select one data item in the record that is likely to be unique in all records of a file and use it for identification purposes. This item called the record key or key attribute or simply key. Basically it is a part of the record.

**Entity** : An entity is any person, place, thing or event of interest to the organization and about which data are captured, stored or processed.

**File** : A file is collection of related records. Each record in a file is included because it pertains to the same entity.

**Database** : A database is an integrated collection of data stored in different type of records, and in a way that makes them accessible for multiple applications.

### ❖ Types of files

**Master file** : A master file is collection of records about an important aspect of an organization's activities. It may contain the data describes the current status of specific events or business indicators. For example, the master file in accounts payable system shows the balance owed to every vendor or suppliers.

A second type of master file reflects the history of events affecting a particular entity. Suppose in any organization any employee leaves the organization at that time all the details of the employee must be stored. In computer such type of data or information are stored as history file.

Master file are useful only so long as they are kept accurate and up to date.

**Transaction file** : A transaction file is a temporary file with two purpose: Accumulating the data about events as they occurs and updating master files to reflect the results of current transactions. The term transactions refers to any business event that affects the organization and about which data are captured. Each transaction file contains only the records that pertain to the particular entities that are subject to the file. The details are accumulated as a record at a time in transaction file. The transaction and master file are read at the same time by a program and master file is revised depending upon the transactions.

Master files are permanent and exist as long as the system exists. However, the contents of the files change as a result of processing and updating. Transaction files,

on the other hand, are temporary. At some time they are no longer needed and are erased or destroyed, depending upon the method used for time durations.

**Table file** : A special type of master file is included in many systems to meet special processing requirements involving data that must be referenced repeatedly. Table files contain reference data used in processing transactions, updating master files, or producing output.

Analyst often specify the use of table files to store data that otherwise would be included in master files or embedded in computer program. Table file conserve storage space and ease program maintenance by storing in a file data that otherwise would be included in programs or master file.

**Report file** : The central processing unit of computer very often produces data for output at a faster rate than the printer can possibly keep up with. Following the normal sequence of events, processing would have to be delayed while the result are printed. To prevent such inefficient use of the cpu, the analyst generates the output in a text file called the report file. Report files are temporary files used when printing time is not available for all the reports produced, a situation that frequently arises in overlapped processing. The computer writes the output to a file on secondary storage. This file will be used whenever it is required and will be deleted when the use of the file is over.

**Other files** : Other kind of files, as well as special uses of the file types previously discussed, play a role in information system. For example a backup file is a copy of master file , transaction file or table file made to ensure that duplicate is available if anything happens to the original file.

### ❖ **Methods of file organization**

Records are stored in files using a file organization that determines how the records will be stored, located and retrieved. There are basically three methods to store the records are as follow :

**Sequential organizations** : It is the simplest way to store and retrieve records in a file. In a sequential file records are stored one after the other without concern for the actual value of the data in the records. The first record is stored at the beginning of the file. The second record is stored right after the first record and so on. This order never change in sequential file organization. A characteristic of sequential file is that all records are stored by position: first record, second record and so on. There are no addresses or location assignments in sequential file.

**Direct access organization** : This method requires the program to tell the system where a record is stored before it can access the record. In contrast to sequential organization, processing a direct access file does not require the system to start at the first record in the file.

Using the record key as the storage address is called direct addressing. Direct addressing have the a data set with the following characteristics :

- The key set is in a dense ascending order with few unused values. Therefore few open gaps in key values are wanted.
- The record keys correspond to the numbers of the storage addresses: there is a storage address for each actual or possible key value in the file and there are no duplicate key values.

**Indexed organization :** A third way of accessing records is through an index. The basic form of index includes a record key and the storage address for a record. To find a record when the storage address is unknown it is necessary to scan the records. However the search will be faster if an index is used, since it takes less time to search an index than an entire file.

An index is a separate file from the master file to which it pertains. Each record in the index contains only two items of data : a record key and a storage address. To find a specific record when the file is stored under an indexed organization, the index is first searched to find the key of the record wanted. When it is found the corresponding storage address is noted and then the program accesses the record directly. This method is uses a sequential scan of the index, followed by direct access to the appropriate record. The index helps speed the search compared with a sequential file, but it is slower than direct addressing.