STRIP AND LOAD DATA

by

Richard H. Jones, Computer Systems Analyst
Business Data Systems Division
Mail Stop 179, Langley Research Center
FTS 928-2721

The Strip and Load process refers to a method of taking batch data files and loading these files into the ADABAS Data Base Management System (DBMS) for subsequent on-line query. ADABAS requires a sequential data set as input to the Loader Utility. For purposes of this document, we assume an IBM/MVS operating environment using COM-PLETE, ADABAS, and NATURAL. Other tele-processing monitors, such as CICS or TSO, are valid substitutes for COM-PLETE.

You just received your new ADABAS DBMS and the primary objective is to quickly become productive. The initial steps to becoming productive are to identify a prototype system and put it on-line. The prototype system should be small in scale but contain the complexity of larger scale systems. Once the prototype system has been successfully implemented, implement next the system with the most immediate benefit to the organization.

The process of defining the ADABAS data base is very simple. Most new ADABAS users simply define the new ADABAS file, field for field, to look like their old file. This procedure will work and is sometimes appropriate; however, with a little effort, you can avoid the potential pitfalls of poor disk utilization and poor performance. If a field is a derivative of other fields in the record, you may not want to store the field. Fields such as City and State may be replaced by Zip Code, Department Name may become a Table File, etc. ADABAS requires fields defined as numeric to contain valid numeric data. It is very important to look at the data values of each field you include in the data base. To assume fields have certain characteristics is not enough; you should run frequency distributions on each field.

One of the primary functions of a Data Base Management System is to provide access to the data. ADABAS provides access through fields defined as descriptors. Descriptors are like index catalogs in the library. Interview the proposed users of the system to determine how they will access the data.

The best descriptors are unique (e.g., Social Security Number) and the worst descriptors are non-unique (e.g., Sex Code 'M' or 'F'. The frequency distribution of the data values will tell you if the field is a good descriptor field or not so good. If a user accesses the data based on values from several fields, you can define a compound descriptor/"superdescriptor." If a part of a field is to be used in a search, for example, purchase order number positions 3 and 4 represent Directorate and Division, they can be made into a sub-descriptor. The phonetic descriptor can be used where the exact spelling of an alphabetic field is uncertain. In your evaluation of descriptors you should compute the on-line disk storage required and justify each descriptor. For Strip and Load Files, liberal use of descriptors is recommended. The only cost is on-line disk storage.

When you have defined the fields to be included in the data base and which fields will be descriptors, it should be documented and reviewed with the user for final concurrence and signoff. Many misunderstandings and erroneous assumptions have been corrected by doing a final documented review with the user prior to implementation.

Now, you are ready to define the data fields and files to the ADABAS DBMS. This is accomplished by using the on-line interactive data dictionary facility called "PREDICT." You must define each field and its attributes, whether it is a descriptor field, and what type of data compression. PREDICT will generate the loader definition statements required by the jobstream used to load a file to ADABAS.

The physical field sequence of the data file to be loaded to ADABAS must be in the same sequence as it was defined to ADABAS. Generally, this requires a COBOL program to reformat your non-DBMS file to conform to the ADABAS file definition. The ADABAS loader utility provides a user exit which gives control to the programer after the data is read and before the data is written and can be used to format data, thus saving one pass of the file.

Once the files are loaded, you can write NATURAL programs to query the data. NATURAL is an interactive development system designed for use with the ADABAS DBMS. If the user has not been trained in coding NATURAL, their basic requirement would be met with canned programs until such time as they are trained. Because these files are Strip and Load, no updating occurs.