CHAPTER 2

SAS Processing

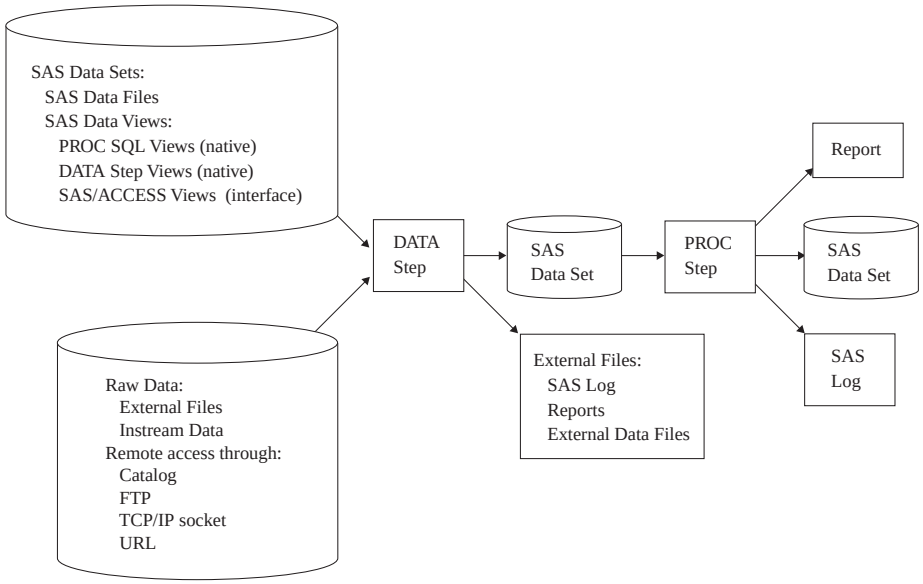
Definition **11**
Input to a SAS Program **12**
The DATA Step **13**
DATA Step Output **13**
The PROC Step **14**
PROC Step Output **14**

Definition

SAS processing is the way that the SAS language reads and transforms input data and generates the kind of output that you request. The DATA step and the procedure (PROC) step are the two steps in the SAS language. Generally, the DATA step manipulates data, and the PROC step analyzes data, produces output, or manages SAS files. These two types of steps, used alone or combined, form the basis of SAS programs.

The following figure shows a high level view of SAS processing using a DATA step and a PROC step. The figure focuses primarily on the DATA step.

Figure 2.1 SAS Processing



You can use different types of data as input to a DATA step. The DATA step is composed of SAS statements that you write, which contain instructions for processing the data. As each DATA step in a SAS program is compiling or executing, SAS generates a log that contains processing messages and error messages. These messages can help you debug a SAS program.

Input to a SAS Program

You can use different sources of input data in your SAS program:

SAS data sets can be one of two types:

- **SAS data files** store actual data values. A SAS data file consists of a descriptor portion that describes the data in the file, and a data portion.
- **SAS data views** contain references to data stored elsewhere. A SAS data view uses descriptor information and data from other files. It allow you to dynamically combine data from various sources, without using storage space to create a new data set. Data views consist of DATA step views, PROC SQL views, and SAS/ACCESS views. In most cases, you can use a SAS data view as if it were a SAS data file.

For more information, see Chapter 28, “SAS Data Files,” on page 411, and Chapter 29, “SAS Data Views,” on page 455.

Raw data specifies unprocessed data that have not been read into a SAS data set. You can read raw data from two sources:

- **External files** contain records comprised of formatted data (data are arranged in columns) or free-formatted data (data that are not arranged in columns).
- **Instream data** is data included in your program. You use the DATALINES statement at the beginning of your data to identify the instream data.

For more information about raw data, see Chapter 22, “Reading Raw Data,” on page 285.

Remote access allows you to read input data from nontraditional sources such as a TCP/IP socket or a URL. SAS treats this data as if it were coming from an external file. SAS allows you to access your input data remotely in the following ways:

- **SAS catalog** specifies the access method that enables you to reference a SAS catalog as an external file.
- **FTP** specifies the access method that enables you to use File Transfer Protocol (FTP) to read from or write to a file from any host machine that is

	connected to a network with an FTP server running.
TCP/IP socket	specifies the access method that enables you to read from or write to a Transmission Control Protocol/Internet Protocol (TCP/IP) socket.
URL	specifies the access method that enables you to use the Universal Resource Locator (URL) to read from and write to a file from any host machine that is connected to a network with a URL server running.

For more information about accessing data remotely, see FILENAME, CATALOG Access Method; FILENAME, FTP Access Method; FILENAME, SOCKET Access Method; and FILENAME, URL Access Method statements in the Statements section of *SAS Language Reference: Dictionary*.

The DATA Step

The DATA step processes input data. In a DATA step, you can create a SAS data set, which can be a SAS data file or a SAS data view. The DATA step uses input from raw data, remote access, assignment statements, or SAS data sets. The DATA step can, for example, compute values, select specific input records for processing, and use conditional logic. The output from the DATA step can be of several types, such as a SAS data set or a report. You can also write data to the SAS log or to an external data file. For more information about DATA step processing, see "DATA Step Processing" in Chapter 21, "DATA Step Processing," on page 259.

DATA Step Output

The output from the DATA step can be a SAS data set or an external file such as the program log, a report, or an external data file. You can also update an existing file in place, without creating a separate data set. Data must be in the form of a SAS data set to be processed by many SAS procedures. You can create the following types of DATA step output:

SAS log	contains a list of processing messages and program errors. The SAS log is produced by default.
SAS data file	is a SAS data set that contains two parts: a data portion and a data descriptor portion.
SAS data view	is a SAS data set that uses descriptor information and data from other files. SAS data views allow you to dynamically combine data from various sources without using disk space to create a new data set. While a SAS data file actually contains data values, SAS data views contain only references to data stored elsewhere. SAS data views are of member type VIEW. In most cases, you can use a SAS data view as though it were a SAS data file.
External data file	contains the results of DATA step processing. These files are data or text files. The data can be records that are formatted or free-formatted.

Report — contains the results of DATA step processing. Although you usually generate a report by using a PROC step, you can generate the following two types of reports from the DATA step:

- Listing file — contains printed results of DATA step processing, and usually contains headers and page breaks.
- HTML file — contains results that you can display on the World Wide Web. This type of output is generated through the Output Delivery System (ODS). For complete information about ODS, see *The Complete Guide to the SAS Output Delivery System*.

The PROC Step

The PROC step consists of a group of SAS statements that call and execute a procedure, usually with a SAS data set as input. Use PROCs to analyze the data in a SAS data set, produce formatted reports or other results, or provide ways to manage SAS files. You can modify PROCs with minimal effort to generate the output you need. PROCs can also perform functions such as displaying information about a SAS data set. For more information about SAS procedures, see the *SAS Procedures Guide*.

PROC Step Output

The output from a PROC step can provide univariate descriptive statistics, frequency tables, cross-tabulation tables, tabular reports consisting of descriptive statistics, charts, plots, and so on. Output can also be in the form of an updated data set. For more information about procedure output, see the *SAS Procedures Guide* and *The Complete Guide to the SAS Output Delivery System*.

The correct bibliographic citation for this manual is as follows: SAS Institute Inc., *SAS Language Reference: Concepts*, Cary, NC: SAS Institute Inc., 1999. 554 pages.

SAS Language Reference: Concepts

ISBN 1–58025–441–1

SAS Institute Inc., SAS Campus Drive, Cary, North Carolina 27513.

1st printing, November 1999