Appeal to a HIER Authority: Graphing CA-7 Scheduling Dependencies
Brent Turner, The Chase Manhattan Bank, N.A.

Abstract
The Computer Associates' (CA) scheduling product, CA-7, provides a narrative report of antecedent/consequent job dependencies, but in a somewhat congested presentation. The CA-7 output report is used as the input file to a SAS® program which produces two SAS data sets from the report text. The observations in the first data set each contain the job name, its antecedent job name, and, if there are dependencies, a generated reference number. Those in the second set contain the job name, with its dependencies (and the same respective reference number) if any.

By using multiple SET statements in a data step, each observation in the first data set is compared with every other observation in the same data set to determine job sequencing. The SAS supplemental procedure HIER produces a hierarchical diagram showing the jobs' schedule, where dependent job names are concatenated with their respective reference numbers. A cross-referenced requirements/diagnostics table is produced from the second data set, using the PRINT procedure.

Introduction
The Corporate Data Center of Chase Manhattan Bank, N.A. operates 24 hours per day, seven days per week to provide executive management with data that are vital for directing corporate policy. With the exception of preventive maintenance (PM), the center is continuously processing production workloads.

These workloads include online transactions during business hours and production batch jobs at other times. Consequently, batch work must be completed prior to the business day for the system to be able to accommodate the online workloads.

The batch workloads have job dependencies which dictate both when and the order in which they can be processed. Certain jobs produce files used by subsequent jobs; others must await the arrival of tapes from other data centers. The timeframe window for these jobs, which is narrow due to the volume of work, severely restricts any reprocessing if there are scheduling errors.

The Problem
The Computer Associates' (CA) scheduling product, CA-7, provides a narrative report of antecedent/consequent job dependencies, but in a somewhat congested presentation. An excerpt of the report for a specific job scheduling hierarchy is depicted in Figure 1.

Although this report provides detailed information of the requirements for job processing, it does not illustrate the temporal sequence of the schedule in a structural format. In addition, the report contains superfluous information (beyond what is essential for illustrating the schedule structure).

The Solution
SAS can process any readable file as an input source, and is especially well-suited to string manipulation. Therefore, the CA-7 output report is used as the input file to a SAS program which extracts the necessary information for graphical and print reporting. As such, the report is directed to a disk file, instead of a printer.

The pertinent data are located by searching for specific character strings on each line of a CA-7 scheduling report. All of the report lines are processed as mixed record types, which in fact they are. The record "type" is indicated by the presence of specific character strings at designated locations on the respective report line.

Methodology
The program methodology consists of two data steps followed by three procedure steps: SORT, HIER, and PRINT. The first data step reads the CA-7 output report to produce two output SAS data sets. The first data set is denoted as the Sequence data set. Each observation in this data set contains variables for:

- the name of one of the jobs listed in the schedule report;
- the name of the antecedent job (if any) which must complete successfully to "trigger" the currently listed job; and
Figure 1. Excerpt from CA-7 Job Scheduling Report

Figure 2. Schematic Diagram of Comparison Methodology
• if needed, a sequentially generated reference number corresponding to an identically numbered list item which will appear on a subsequent PRINT report. This contains the specific requirements for the listed job. For jobs which do not have specific requirements, this value will be blank.

The other data set consists of observations for jobs that have specific requirements. This data set is denoted as the Requirements data set. Each observation in this data set contains variables for:

• the name of the listed job from the production schedule;
• a sequentially generated reference number which corresponds to an identically numbered hierarchical node in the subsequent graphical report; and
• the job requirements (text data) themselves.

The second data step utilizes the Sequence data set, containing the job names in a particular CA-7 schedule and their respective antecedent jobs, created in the first data step. The data set is used as its own reference file to establish job sequencing for subsequent hierarchical graphing. The lookup comparison and output technique are illustrated in Figure 2.

This technique is accomplished by using two SET statements to read the data set, each of which keeps a separate observation pointer. One SET statement is placed inside a DO UNTIL loop and utilizes the POINT= and NOBS= SET statement options to enable multiple executions of this statement, within the loop, for each data step execution. References to this SET statement (and its associated observations) will hereafter be denoted by the Inside SET.

The other SET statement (again, reading the same data set) precedes (is outside of) the DO UNTIL loop and executes once for each data step execution. References to this SET statement (and its associated observations) will hereafter be denoted by the Outside SET.

The Outside SET statement reads an observation which becomes the reference record for subsequent comparisons in this execution of the data step. It is necessary to assign the variable values from this observation to differently named variables. This is because these values will be overwritten by the values in the observational variables which will be read by the Inside SET statement, since the variables are, of course, the same.

If there is an antecedent job (hereafter denoted as the Trigger) for the job name identified in the Outside SET observation, it is compared to the job name for the Inside SET observation. The two possible results for this comparison are:

1. The Outside SET Trigger matches the Inside SET job name. In this situation three events occur.
   a. The DO UNTIL truth condition is attained and the loop test variable is set to the numerical value 1.
   b. The Inside SET job name is assigned to a storage variable for the Trigger (of the Outside SET job name).
   c. Execution of the data step continues outside the DO UNTIL loop and an output observation is written. The data step then begins its next execution.

2. The Outside SET Trigger does not match the Inside SET job name. The DO UNTIL loop is executed again and the next Inside SET observation is read.

If there is no antecedent job (no Trigger), one of two situations are indicated:

1. There is no Trigger for the Outside SET job name, such that it represents the top node in the scheduling hierarchy.
2. There is no Trigger for the Outside SET job name but it is not the top node in the scheduling hierarchy, and the all observations in the data set have been read by the Inside SET statement but have yielded no match. The Outside SET job name:
   • is not first in the schedule;
   • is not triggered by a prior job completion; but
   • serves as trigger for one or more subsequent jobs in the schedule.

From the matched Outside SET observation data and the Inside SET observation data a new observation is constructed for the output data set. From the Outside SET observation, the job name and its "reference number" (which may be blank) are concatenated into a new variable. From the Inside SET observation, the job name, which corresponds to the Trigger name of the Outside SET observation, and its "reference number" (which may be blank) are concatenated into a new variable.

These two new variables are combined into an observation which attaches the job and Trigger name (and their associated "reference numbers") to each other. The observation is written to the output data set (hereafter denoted as the Attached data set). The Attached data set is subsequently used as input to the HIER procedure.
Figure 3. Code Segment Used In Comparison Methodology

824
The actual code used in the comparison is presented in Figure 3. Note that the values for the subordinate job name and reference number, plus the superior Trigger job name, are stored before entering the DO UNTIL loop, as is the schedule number. These values are overwritten by the SET statement reading the data set inside the loop.

The Requirements data set is sorted using the SORT procedure by reference number and job name. It is then written to the print file using the PRINT procedure with the reference number and job name listed in both the BY and ID statements. This eliminates the printing of duplicate data for jobs which have more than one requirement, and hence, more than one accompanying observation.

Results

The SAS supplemental procedure HIER produces a hierarchical diagram, showing the jobs' schedule, from the concatenated job name-reference number and Trigger-reference number variables in the Attached data set. The Trigger-reference variable is designated as the SUPERIOR variable in the synonymous HIER control statement. The job name-reference variable is used as the SUBORDINATE variable in its corresponding control statement.

The constructed diagram depicts the temporal ordering of the jobs in the particular schedule by their respective placement in the node structure. Dependent job names are lower in the hierarchy than their antecedent jobs (Triggers). A sample scheduling graph is depicted in Figure 4.

A cross-referenced requirements/dependencies table is produced from the data set containing the numbered job requirements (Requirements data set), using the PRINT procedure. Figure 5 provides an excerpt of the Requirements report.

Discussion

The CA-7 Job Scheduling report contains complete information on job sequencing and dependency requirements. However, the information is not presented in a format which easily indicates the structure of the schedule.

By using the CA-7 report as the input source for a SAS reporting system, the temporal ordering of scheduled jobs can be graphically depicted using the HIER procedure. The job requirements, being written to a separate report, provide the needed scheduling instructions (using reference numbers) without obscuring the processing sequence.

The HIER procedure is currently unsupported in policy by the Institute and also unsupported in practice by its author. At the time of this presentation, there are no announcements that it will be functionally adopted under Version 6 of SAS. The principal intentions of this report are:

- to demonstrate the continuing usefulness of this apparent software orphan; and
- to argue for its preservation.

It is the hope of the author of this paper and project that the HIER procedure will not perish through benign neglect or indifference. This paper is intended as an appeal to the HIER authority.

Acknowledgements

The author expresses his appreciation to Mr. Jeff Hayowy for his valuable comments and review of the project which facilitated its development and implementation.

The author further expresses his gratitude to Messrs. Manuel DeGuzman, Adam Hatami, and Sung-Kwong So for their patience in dealing with the vicissitudes of the author during the course of this transcription.

Finally, the author especially thanks Messrs. Frank Suriani and Joseph Williams for their timely help in obtaining authorization to present this report. For further information, contact:

Brent Turner
The Chase Manhattan Bank, N.A.
One New York Plaza
Floor Eleven
New York, NY 10081
212/676-3596

CA-7 is a registered trademark of Computer Associates, Inc., Garden City, NY, USA.

SAS is a registered trademark of SAS Institute Inc., Cary, NC, USA.
Figure 4. Sample Job Scheduling Hierarchical Graph

Figure 5. Sample Job Scheduling Requirements Report