SORT EFFICIENCY USING SAS
UNDER VM/CMS

By Bob Mahan
Syncsort Inc.

Abstract

SAS® applications heavily consume system resources. Collection and analysis of large data files is CPU and I/O intensive. Using SyncSort CMS with SAS applications however, can substantially reduce CPU time, SIO and elapsed time.

Further, with new SAS interfaces, SyncSort CMS can decrease resource consumption up to 80%. The new SAS interfaces are:

- PROC SYNCSORT replaces PROC SORT within SAS applications.
- SYNCSORT SAS, (a CMS command), can sort the SAS data prior to invoking the SAS system.

These features of SyncSort CMS are here delineated focusing upon user performance benefits.

Typical SAS Application

Following is a description of a typical SAS application, the sorts within it, and the consumed computer resources.

The application is a production job that runs quarterly and performs the following functions:

- Reads three tapes with approximately 350,000 records, selects 316,000 of those records, and creates a SAS data set
- Sorts the data set by zip code
- Provides 16 reports requiring 12 sorts
- Prints one set of mailing labels
- Generates summary data.

The job is typically run in a 4 megabyte virtual machine, and keeps 38 variables out of 316,000 records which totals slightly more than 100 million bytes of data.

As a result, considerably more resources are consumed sorting the data using PROC SORT, as compared to sorting the data in an external sort as illustrated in Figure 2.

PROC SYNCSORT

PROC SORTT is sometimes used in SAS for small sorts with a limit of 64,000 observations. According to Dan Squillace of SAS Institute (Share 68, San Francisco, 3/87), PROC SORTT is most appropriate for small sorts with less than 300 observations.

PROC SORT, (as distinct from PROC SORTT), uses the system sort, in our case, SyncSort CMS, and can sort an unlimited number of observations. The standard interface for PROC SORT passes records to the sort, one record at a time.

PROC SYNCSORT, a new SAS sort interface, permits SyncSort CMS to read and sort records more efficiently by replacing PROC SORT or PROC SORTT in any SAS application. PROC SYNCSORT, which may be used only from within an SAS job stream, offers substantial savings in system resources and performance benefits over other methods.

SYNCSORT SAS

SYNCSORT SAS is a CMS command that allows sorting of SAS data sets outside of the SAS system. The SYNCSORT SAS command fully utilizes Syncsort's I/O and optimization techniques, providing the most efficient method available for sorting and copying SAS data sets. SAS files can be presorted before processing by a sequence of SAS applications.

Additionally, SYNCSORT SAS provides resource reductions as well as syntax similarity with a CMS command. The format that you would code is:

SYNCSORT SAS DATA=sas.dataset
OUT=sas.dataset [REVERSE]
[EQUALS] BY [DESCENDING] variab1
[DESCENDING] variab2...

Figure 1: SYNCSORT SAS Command Syntax
Installation for Performance

SyncSort CMS should be installed as the default sort procedure to provide system-wide performance benefits. SAS users can then take advantage of both PROC SYNCSORT and SYNCSORT SAS for even greater performance. The following options must be set to take advantage of these features:

Change SyncSort for SAS Compatibility

Sort options may be modified by turning the SAS Compatibility to ON. This is done by the SYNCDEF interactive utility program provided with the sort. First, you would access the SyncSort CMS library disk in write mode. Second, you would invoke the SYNCDEF interactive utility and change the SAS Compatibility to ON at screen two.

Change SAS Options to use SyncSort CMS

You would access the SAS library disk in write mode and run the SETOPTS procedure to change SORTPGM = SAS to SORTPGM = SYNCSORT and SORTLIB = ' ' to SORTLIB = SYNCSORT. This instructs SAS to invoke SyncSort by default.

Conclusion

In summary, SAS jobs consume much of our system's resources. SyncSort CMS can reduce that resource expenditure. SyncSort's two new interfaces to the SAS system - PROC SYNCSORT and SYNCSORT SAS - can provide even greater reductions in resource consumption - up to 60% savings in elapsed time, 70% savings in CPU time, and up to 90% savings in Start I/Os.

Resources Consumed for a 100 Megabyte Sort Using PROC SORT in a SAS Application

<table>
<thead>
<tr>
<th>RESOURCE</th>
<th>CONNECT TIME</th>
<th>VTIME</th>
<th>TTIME</th>
<th>Start I/O's</th>
</tr>
</thead>
<tbody>
<tr>
<td>PROC SORT</td>
<td>14:08.00</td>
<td>02:08.36</td>
<td>02:36.36</td>
<td>34,313</td>
</tr>
<tr>
<td>PROC SYNCSORT</td>
<td>08:56.00</td>
<td>00:29.03</td>
<td>01:10.56</td>
<td>3,882</td>
</tr>
</tbody>
</table>

REDUCTION %

<table>
<thead>
<tr>
<th>RESOURCE</th>
<th>CONNECT TIME</th>
<th>VTIME</th>
<th>TTIME</th>
<th>Start I/O's</th>
</tr>
</thead>
<tbody>
<tr>
<td>PROC SORT</td>
<td>14:08.00</td>
<td>02:08.36</td>
<td>02:36.36</td>
<td>34,313</td>
</tr>
<tr>
<td>PROC SYNCSORT</td>
<td>07:06.00</td>
<td>00:28.41</td>
<td>01:06.17</td>
<td>2,356</td>
</tr>
</tbody>
</table>

REDUCTION %

<table>
<thead>
<tr>
<th>RESOURCE</th>
<th>PROC SORT</th>
<th>SYNCSORT SAS</th>
</tr>
</thead>
<tbody>
<tr>
<td>CONNECT TIME</td>
<td>14:08.00</td>
<td>07:06.00</td>
</tr>
<tr>
<td>VTIME</td>
<td>02:08.36</td>
<td>00:28.41</td>
</tr>
<tr>
<td>TTIME</td>
<td>02:36.36</td>
<td>01:06.17</td>
</tr>
<tr>
<td>Start I/Os</td>
<td>34,313</td>
<td>2,356</td>
</tr>
</tbody>
</table>

REDUCTION %

Figure 2: Comparative Analysis of Consumed Resources
References:


The author may be contacted at:

Syncsort, Inc.
50 Tice Boulevard, CN18
Woodcliff Lake, NJ 07675
(201) 930-8200

*SAS is a registered trademark of SAS Institute Inc., Cary, NC, USA.*