Introduction to Saspsort

CAUTION:
Saspsort is an experimental sort available with Version 7. This sort is not supported by SAS Institute Technical Support and might not be available in future releases of the SAS System.

Version 7 includes an experimental sorting procedure tailored for symmetric multi-processing (SMP) machines. This sort, named saspsort, implements a parallel version of the quick sort algorithm. Saspsort sorts a data set by dividing it up into a number of smaller collections. Each collection can be sorted independently on one of the SMP machine’s processors, so the time needed to perform the sort can be greatly reduced.

Saspsort runs best on a platform with multiple processors and an operating system that uses threads. However, depending on the characteristics of the data set being sorted, performance on single-processor machines might also be improved.

If your data set is already roughly ordered on the sort key, the default SAS sort will probably sort the data set faster than saspsort.

Using Saspsort

Invoke parallel quick sort just as you would any other host sort. Specify the SORTPGM and SORTNAME system options, and when you run PROC SORT, the procedure uses the sort that you specified:

```
options sortpgm = host;
options sortname = saspsort;
```

Setting the SORTSIZE System Option

To make the best use of your machine, the SORTSIZE option should be set to a value large enough so that the entire data set, plus quick sort’s overhead, can be held in memory. This value is given by the following expression:
sortsize = n * (s + k + 4)

where
n is the number of rows in the data set.
s is the size of a row in bytes.
k is the size of the sort key in bytes.

For example, consider a data set with 100,000 eighty-byte rows that is being sorted on an eight-byte numeric variable. The sort runs fastest if the value of the SORTSIZE option is set to the value 100000 * (80 + 8 + 4).

    options sortsize = 9200000;

If the entire problem cannot be held in memory, the quick sort procedure performs a number of individual sort steps, each one as large as the value of the SORTSIZE option permits, and saves the results on disk. After all data has been sorted, the individual sorted collections are merged to produce the final sorted data set. In this case, parallel quick sort may offer no performance advantage.
Your Turn

If you have comments or suggestions about SAS Companion for UNIX Environments, First Edition, please send them to us on a photocopy of this page or send us electronic mail.

Send comments about this book to
SAS Institute Inc.
Publications Division
SAS Campus Drive
Cary, NC 27513
email: yourturn@unx.sas.com

Send suggestions about the software to
SAS Institute Inc.
Technical Support Division
SAS Campus Drive
Cary, NC 27513
email: suggest@unx.sas.com