APPENDIX

4

Sorting Data Sets on SMP Machines

Introduction to Saspsort **325**
Using Saspsort **325**
Setting the SORTSIZE System Option **325**

Introduction to Saspsort

CAUTION:
Saspsort is an experimental sort available with Version 8. This sort is not supported by SAS Institute Technical Support and might not be available in future releases of the SAS System. △

Version 8 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.

The correct bibliographic citation for this manual is as follows: SAS Institute Inc., *SAS® Companion for UNIX Environments, Version 8*, Cary, NC: SAS Institute Inc., 1999.

SAS® Companion for UNIX Environments, Version 8

Copyright © 1999 by SAS Institute Inc., Cary, NC, USA.

ISBN 1–58025–502–7

All rights reserved. Printed in the United States of America. No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, or otherwise, without the prior written permission of the publisher, SAS Institute Inc.

U.S. Government Restricted Rights Notice. Use, duplication, or disclosure of the software by the government is subject to restrictions as set forth in FAR 52.227–19 Commercial Computer Software-Restricted Rights (June 1987).

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

1st printing, October 1999

SAS® and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries.® indicates USA registration.

Other brand and product names are registered trademarks or trademarks of their respective companies.

The Institute is a private company devoted to the support and further development of its software and related services.