

Scalable Performance Data Server ® 2.1: Tuning and performance

Daniel M. Sargent
Advanced Server Development
Sugi 24 Miami, FL Paper 291
SAS Institute Inc., Cary, NC 27513



Abstract

This paper will discuss the methodology of tuning the SPD server for maximum performance. Factors like disk configuration, striping hardware/logical, thread count, data organization, index types, and memory and their effect to the overall performance of the SAS/SPD Server software will be discussed. Examples and actual performance benchmark results will be presented.

Introduction

The Scalable Performance Data Server™ (SPD Server) is a data base engine designed to scale to the performance needs of an organization. The environment is secure, easy to administer, handles large files, operates 24 hours a day 7 days a week, accessible to end users, and is full SAS language compatible. The software is written to take advantage of Symmetric Multi-Processing (SMP). Symmetric Multi-Processing (SMP), also known as parallel processing takes a given task and divides it into many smaller parts known as threads. As each thread completes its share of the overall task the results are combined with the results of the other threads. Threading allows the system to fully utilize resources such as CPU cycles, I/O, and memory. The SAS/SPD Server software uses SMP technology to optimize where clause processing, load data, build indexes, and sort data. SMP technology is well suited for multi-user computing environments such as warehousing and data mart applications.

Measurements

The focus of the tests will be to determine how to tune the software to minimize run time of jobs used to retrieve data from the SAS/SPD Server software. The unit of measurement will be the average runtime for a job(s). The first test will create a baseline. The baseline will consist of single threaded job(s). Each subsequent test will modify one of the following: threading, disk read/write, I/O connections, and data organization. Each test will remove 1 of the bottlenecks that hinders performance. Test 1 will increase the thread count to enhance performance. Test 2 will increase the number of disks the data is spread across increasing read access to the data. Test 3 will increase I/O connections allowing the server better access to the disk storage devices. The last test will modify the data organization to demonstrate how design changes can take advantage of the SAS/SPD Server software.

Tuning and Scaling the System

Performance Improvements with Threading

This section will show the effects of threading by varying the number of threads a process can launch. Step 1 will run job(s) using a single thread as a baseline. Step 2 will increase the thread count until the run time of the job(s) has been minimized. Using the thread count from step 2 we will then increase the number of jobs (job count) until the average run time is equal to the average run time for step 1. For this test a single disk array with a single directory path will be used.

Increasing Disk Space

Using a single disk array this test will increase the number of directory paths used to show the effects that increasing disk space can have on performance. Although the single I/O connection to the array will be a bottleneck spreading the data across more disks allows better read/write access. The final results from the test above will be used as a baseline for this test. The number of directory paths will be increased until 4 directory paths have been used increasing read/write access by a factor of 4. The results will show how the amount of disks used can effect performance.

Increasing I/O Connections

Using the results from above the number of I/O connections will be increased until there are 4 I/O connections from the server to the disk array(s). The final results from the test above will be used as a baseline for this test. The number of directory paths will remain at 4 but I/O connections from the server to the disk array(s) will be increased until 4 I/O connections are available from the server to the disk. This will eliminate I/O bottlenecks and increase performance.

Data Organization

The last test will modify the order of the data to increase performance. Data is stored on disk in clusters. Cluster size can vary depending on the system setup but typically a cluster is 4096 bytes. Using a where clause on an indexed variable will allow the system to select only clusters that contain desired values. The SAS/SPD Server software minimizes read operations required to complete a task. When data are sorted and indexed the number of clusters a data value resides on will decrease in comparison to a uniform distribution.

Plan carefully, sorting data to achieve performance gains can have tradeoffs. Different users select data based on different criteria. Sorting data to increase performance for a one user will change the distribution of the data and could impact the performance for other users.

Indexing with the SAS/SPD Server software

These tests will examine performance differences when using the three different index types the SAS/SPD Server software supports: hybrid bitmap, global b-tree, and segmented b-tree. Tests will include comparing the performance of the different index types across different cardinality (number of discrete values) levels of data.

Sort options

The SAS/SPD Server software contains different sort options, which allow more flexibility for the users. The options include a spdsort for large files, default heap sort options, and a parallel quick sort. The different uses and performances of these sorts will be explored.

Parameter Files

The software has 2 parameter files that control the resources of the server. How the administrator can use the libname.parm file will be discussed. How the spdserv.parm file can effect performance will be explained.

Other Considerations

SQL Pass-Through

The SAS/SPD Server software includes SQL Pass-through syntax. This allows increased performance when data is moved across networks. Examples of the syntax will be shown and explained.

Security

The SAS/SPD Server software creates a very secure environment for data. It contains features that allow better control over access to the data. Included is a user password feature for libnames. Sample libnames will be shown.

Access to Server data

The SAS/SPD Server software allows access to the data several ways. A list of supported access methods will be provided.

HTML Page

For the complete results of this paper and other papers pertaining to the SAS/SPD Server software please visit our web site and select the button "white papers".
<http://www.sas.com/software/components/spds.html>

References

"SAS/SPD Server Documentation for Release 2.1", HTML available on SAS/SPD Server 2.1 installation media, SAS Institute, Inc., Cary, NC

SAS® and SAS/SPD Server® are registered trademarks of SAS Institute, Inc., Cary, NC