

Configuring and Optimizing Scalable Performance Data Server on HP-UX— a Performance Achievement

Udo Schmitt—Hewlett-Packard Company

Abstract

In recent months, Hewlett-Packard and SAS Institute have jointly embarked on performance tuning and optimization techniques addressing real customer needs in large-scale data warehousing environments.

This paper describes the scalability of the SAS Scalable Performance Data Server® software on HP-UX and illustrates the performance improvements available when running the HP-UX 11 64-bit operating system. It focuses on resource utilization, system architecture, new Scalable Performance Data Server 64-bit enhancements, and the results to meet and exceed the requirements for a large telecommunications customer, Mannesmann Arcor.

Two configurations of SAS Scalable Performance Data Server on HP-UX systems are described, and some hints for optimizing this application on HP-UX are offered.

The Problem

Mannesmann Arcor, a European telecommunications company with some 4,000 employees, had a problem: Too many existing and potential customers were being lured away by competitive offerings. Mannesmann Arcor's managers had a full suite of products and a data warehouse that held a large amount of customer data. However, the data, although extensive, lacked the granularity and precision that would allow Mannesmann Arcor to tailor services for particular customers and to match customers and products precisely. Mannesmann Arcor also expected business growth, which would cause the data volume to grow in a linear manner.

Mannesmann Arcor's Original Database

The database itself was structured using standard SAS relational tables and views, with a raw data size of 400 GB. The database was linked to servers and back-up systems via SCSI technology. This provided sequential and indexed access to the data but did not support any parallel access.

The data warehouse was set up as a few large tables that stored both detail and summary records. The detail records were updated nightly from multiple operational database systems via a high-

speed network. Because of analytical needs and legal requirements, detail records were kept in the warehouse for several months and were deleted only when they were at least six months old. Summary records were also kept in the warehouse. After the warehouse was loaded each night, Mannesmann Arcor's IT staff performed database backup via the high-speed network connection.

At Last, a Proven Data Warehouse Solution ...

In today's highly competitive business environment, you need all the tools you can leverage to gain that critical edge. Moreover, as your business grows, you want to ensure that those tools—and the technology and partners you choose—can grow with you. At the same time, your choices must be cost-effective for your organization. Data has emerged as a primary technological tool in the competitive struggle. Suddenly, there are diamonds in those growing volumes of detailed customer transactions—and even more riches if you can unearth patterns from the history. Data warehousing and its sister, data mining, are keys to exploiting all this data. But what technology can reliably store the massive data volumes required for a useful data warehouse and, more importantly, allow you to retrieve and process that information in a timely manner? The answer is found in a unique partnership between Hewlett-Packard and SAS Institute, a partnership that has created just such a data storage solution. Constructed first for a specific customer, the solution is detailed in this paper—and now it also is available to you. By leveraging the technological advances of HP's newest V-Class servers and SAS Institute's Scalable Performance Data Server (SPD Server) software, your investment in data warehouse technology can deliver maximum benefits. Your valuable data assets are stored securely in an environment that is ready to grow and was built with the special demands of data warehousing in mind.

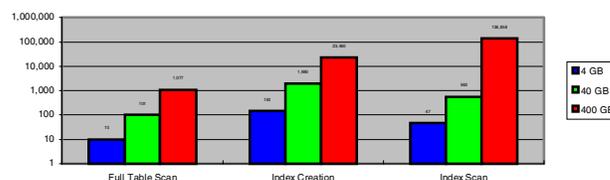


Figure 1: SPD Server scalability on HP-UX

Problems with the Original Database

Mannesmann Arcor had determined that its original data warehouse was too small to contain the customer data necessary for Mannesmann Arcor to be competitive. The problem was not so much that the number of customer records needed to grow—though that certainly was a goal—but rather that the information for each customer was insufficient. Mannesmann Arcor wanted to increase the granularity of data, allowing much more information about each customer's likes, dislikes, desires, and purchase habits to be stored. Why not simply increase the database size? Under normal circumstances, when data increases, absolute performance of an installation remains the same. The result is that relative performance per data record *decreases*. The time to access each record slows to a crawl. Yet for Mannesmann Arcor, both data size and data performance were business-critical issues.

The Goals

Mannesmann Arcor outlined its primary goals for its data warehouse operation:

- A fourfold increase in database size
- No loss of performance

The company planned a benchmark test. Mannesmann Arcor also had some ideas for how to achieve its business goals. The desired solution was to include the following:

- Scaling upward
- 64-bit operating system
- Fibre Channel technology
- Fast, direct support from the software and hardware vendor

The Solution

Mannesmann Arcor opted for benchmark tests using SAS SPD Server software running on an HP 9000 server with the HP-UX operating system. The company hoped that this heretofore untested configuration would meet its business requirements and performance goals for the data warehouse.

Mannesmann Arcor tested two benchmark configurations:

- Basic configuration—an HP 9000 V2250 server running SAS SPD Server software with two EMC 3430 disk arrays
- Enhanced configuration—an HP 9000 V2500 server running SAS SPD Server with nine HP SureStore E Disk System FC10 arrays

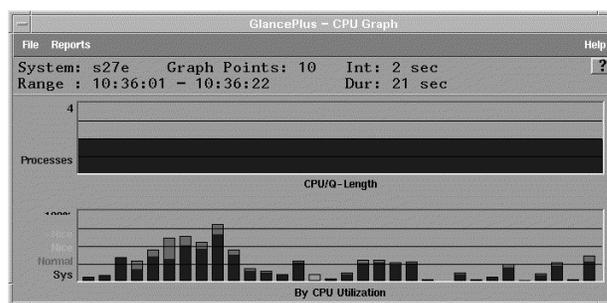


Figure 2: CPU utilization during a full table scan

Together, HP and SAS Institute have ensured that a marriage of hardware and software has already been performed. Your precious IT resources can focus on delivering business value from your data warehousing projects, not fighting technological integration battles with tools from disparate vendors. Naturally, you have pressures to justify your IT allocations—to prove that they are delivering business value. And this data storage solution is constructed to maximize return on investment. For instance, the SAS SPD Server software allows for extremely lean storage and special features such as SAS formats in order to minimize the need for expensive index overhead—all without jeopardizing performance. The HP V-Class technology is completely scalable, too, permitting you to ration your IT budget by adding processors and storage when you need to grow. What's more, HP and SAS Institute consultants practice an iterative approach to implementation that is driven by business priorities, not merely amassing bigger and bigger piles of data.

What about security? As you contemplate opening your data warehouse to a wider base of users—including customers—you can relax, knowing that HP and the SAS Scalable Performance Data Server are protecting your corporate assets with advanced security concepts. So let those intranet, extranet, and Internet users in! This data warehouse solution is ready.

Contact Information

Hewlett-Packard
 Udo Schmitt
 Herrenbergerstrasse 130
 71034 Boeblingen
 Germany
 Phone: +49-7031-14-8229
 E-mail: udo_schmitt@hp.com

Reference Paper

URL:
http://www.enterprisecomputing.hp.com/unixpartners/pdf/spds_whtpr.pdf

Presentation

URL:
http://www.enterprisecomputing.hp.com/unixpartners/pdf/hp_s_uqi25_293.pdf