SAS® Scalable Performance Data Server®

A data storage system optimized to deliver data subsets quickly and on demand

What does SAS Scalable Performance Data Server do?

The SAS Scalable Performance Data Server is optimized to deliver subsets of information that need to be harvested from large enterprise data mountains, quickly and upon demand.

Why is SAS Scalable Performance Data Server important?

The capabilities of the SAS Scalable Performance Data Server ensure that business intelligence and analytic intelligence applications maintain consistent performance and that extraction, transformation and loading processes do not exceed the time windows available as the amount of your enterprise data continues to grow.

For whom is SAS Scalable Performance Data Server designed?

SAS Scalable Performance Data Server is designed for IT managers with large amounts of data who are interested in significantly speeding up their analytic intelligence and business intelligence applications or shortening extraction, transformation and loading times for their data warehouses, irrespective of the amount of data held within the enterprise.

Decision makers who rely on business and analytic intelligence solutions are looking for one thing: accurate answers when they need them. Unfortunately, the data that drives these answers is often locked within mountains of data.

Transaction-oriented relational database systems have long focused on collecting rather than delivering data for reporting and analytics, and because transactional systems were not designed for large-scale data retrieval reporting and analysis such actions can reduce system performance, consume resources and interfere with business operations. To accommodate users' needs, subsets of data often must be built on disparate machines so, in addition to slow reporting times, it is difficult to achieve a cohesive view of the enterprise.

With SAS Scalable Performance Data Server, you get a data storage system that is optimized for both fast loading and fast retrieval of data subsets from large data mountains. Business reporting and large-scale analytic applications will benefit from the swift access to detail data stored in this server. It ensures that applications perform consistently and that ETL processes do not exceed the available time windows, even as data continues to grow.

SAS Scalable Performance Data Server achieves its scalability through parallel processing and partitioning, hybrid bitmap and B-tree indexing, query optimizations, and eliminating transactional overhead. Data extraction, transformation and loading times are shorter, reports are generated faster and decision makers can quickly get the answers they need.

Key benefits

- **Significantly speed up the gathering of subset information.** Most analytic intelligence and business intelligence applications require only a subset of the enterprise data mountain. SAS Scalable Performance Data Server uses parallel storage technologies and hardware to reduce overhead, such as the two-phase commit, that is incurred with transactional systems. This significantly shortens search and delivery times.
- **Optimize performance for business and analytic intelligence applications.** By separating the transactional data store from the business and analytic intelligence data stores, each can be tuned for appropriate task requirements. You get better performance for reporting and analyses while reducing the load on operational systems.
- **Reduce extraction, transformation and loading times.** An integrated ETL process moves data from transactional systems to the SAS Scalable Performance Data Server, leveraging parallel storage and efficient indexing. This reduces extraction, transformation and load times, allowing data and indexes to be refreshed rapidly, irrespective of data size. The use of integrated metadata ensures data consistency and reduces the time spent maintaining your mountains of data.

Product overview

With SAS Scalable Performance Data Server, IT departments can provide numerous concurrent users with fast, secure access to subsets of information drawn from large enterprise data stores. The server's ability to scale existing host hardware (according to problem size) to refresh large amounts of data within a defined time window results in a maximum return on investment.

SAS Scalable Performance Data Server is best suited for large volumes of data. The size of the data warehouse and the time spent processing data are determined mostly by the available hardware. The server is designed to scale up with the hardware as additional components, such as CPUs, main memory and I/O systems, are added.

To support the fastest access to large volumes of data, SAS Scalable Performance Data Server features a multithreaded I/O engine that will launch multiple, independent lightweight processes to search for rows containing specific information in large tables. This may scale the time needed to find information down to a fraction of what would be needed in a single-threaded sequential search.

Partitioning and parallel processing ensure scalability and performance

Scalable I/O

SAS Scalable Performance Data (SPD) Server speeds the processing of large amounts of data by partitioning the data across multiple disks and I/O channels. This enables parallelization of many SAS I/O functions over multiple data partitions. It is designed to use all resources available on a machine and maximum benefits are gained on machines with multiple CPUs, I/O channels and disks where there are large amounts of data to be manipulated.

Some existing SAS SPD Server deployments contain more than eight terabytes of data with single tables exceeding 500 gigabytes. The server has demonstrated scalability to databases and tables containing billions of rows and was designed with a petabyte-sized address space to support massive data warehouses.

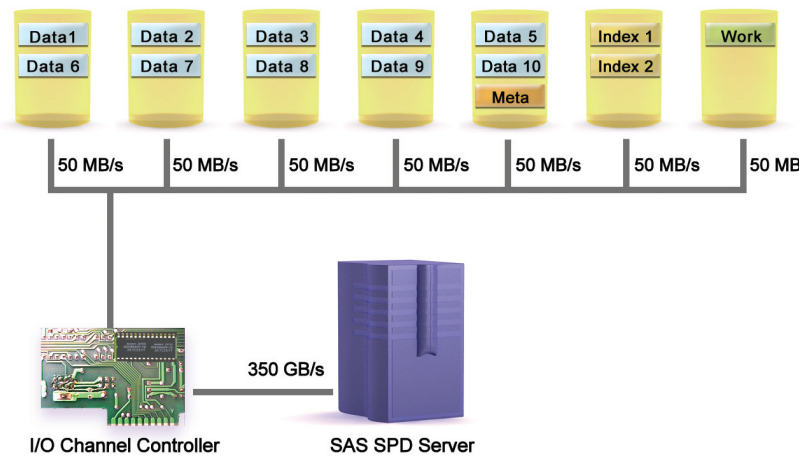
Parallel Group By enhancements

Parallel Group By enhancements leverage parallel CPU and I/O paths for certain kinds of summarizations, including COUNT, FREQ, N, AVG, MEAN, MAX, MIN, NMISS, STD, SUM, VAR and DISTINCT functions. The use of table aliases is also supported. Parallel Group By enhancements have been directly integrated with the SAS SQL engine, expanding the range of situations where these enhancements can be applied.

Multithreaded WHERE clause processing

SAS Scalable Performance Data Server's intelligent WHERE clause planner finds the fastest strategy for evaluating subsetting criteria, independent of whether any (or how many) of the contributing variables are indexed. The largest gains in scalability are achieved when evaluating a complex predicate that contains multiple conditions, or when the server is forced to perform a full table scan such as when indexes are not defined on variables in the subsetting criteria. In these cases, multiple threads are launched to process different portions of the predicate or table simultaneously, ensuring high I/O throughput and scalability.

WHERE planner costing evaluates the effectiveness of different strategies for query processing to determine the least resource-intensive option. It also determines whether to use an index and which one to use if multiples are available based on the density and distribution of values within the index.



SAS Scalable Performance Data Server uses multiple threads across multiple I/O channels to increase the speed at which subsets of data can retrieved and processed so queries are completed much faster.

Unique indexing technology reduces storage and accelerates queries

Fast table joining with hybrid bitmap index

SAS Scalable Performance Data Server uses a hybrid indexing scheme that combines the best features of B-tree and bitmap indexes. The server analyzes each table and automatically determines the best indexing option to use for each segment of data, resulting in better index performance, faster table joins and reduced index storage requirements.

Parallel multi-index builds and updates

When creating a Base SAS data set with indexes, the data processing portion must be finished before the indexes are created sequentially. With SAS SPD Server, you can create or refresh multiple indexes that belong to the same table concurrently, even while data is loaded using the APPEND procedure. Multiple indexes can be built with a single read of a data table, greatly speeding index creation time.

Index statistics table

SAS Scalable Performance Data Server maintains a table of statistics to help determine when to use an index, including which SQL join strategy to use (index join, merge join or hash join), based on cost.

Optimized query performance

SQL Planner optimizations

New SQL Planner optimizations improve performance for the types of queries most often found in business and analytic intelligence applications. Optimized correlated queries transparently restructure and recode deeply nested SQL code to create temporary tables "on the fly," accelerating processing while preserving business logic.

Fast sorting on BY statements

Sorting a SAS data set is a very common occurrence. SAS Scalable Performance Data Server improves sorting performance in two ways. If a WHERE clause or KEEP list has been applied to a subset of data, only the appropriate subset is sorted so that sort time is a function of the subset size and not the entire data set. Secondly, the 64-bit wide-mode version of SAS Scalable Performance Data Server can create single sort bins larger than two gigabytes, depending on the memory available. This reduces the number of bins needed to complete the sort and improves sort time significantly.

SQL Pass-Through support

The SAS SQL procedure's Pass-Through facility enables you to send SQL statements specific to SAS Scalable Performance Data Server directly to the server for execution. This is most beneficial when new SQL statements specific to the SAS Scalable Performance Data Server are used. Multiple independent SQL statements can be triggered for parallel execution.

Time-based partioning

This capability facilitates loading and removing data from very large tables using time-based units of data. SAS Scalable Performance Data Server uses a virtual table called a cluster with a number of slots, each holding a time period of data. When the slots are full and new data is added, the oldest time period is automatically deleted from its slot to accommodate the most recent one.

64-bit wide-mode support

64-bit support enables SAS to allocate and address more than two gigabytes of memory at a time, which improves efficiency and performance when processing large amounts of data.

SPDSSNET consolidation

In previous releases, SPDSSNET has been the path for a Windows desktop application to connect to the power of SAS Scalable Performance Data Server. With SAS SPD Server 9.1, ODBC connectivity goes through the standard server ports defined on the server box.

Sophisticated security implementation

Security features enable administrators to control read, write and modification access to tables, rows, columns and cells on a per user or per group basis.

SAS® Scalable Performance Data Server Technical Requirements

Supported platforms

- AIX (64-bit), Release 5.1+
- HP/UX (64-bit), Release 11i
- HP/UX Itanium (64-bit), Release 11i+
- Linux for Itanium (64-bit): Red Hat RHEL 3.0
- Solaris (64-bit), Version 8, 9, or 10 on SPARC
- Tru64 UNIX (64-bit), Version 5.1A or 5.1B
- Windows (x86-32): Windows 2000 Professional, Windows XP Professional, Windows NT 4 Server, Windows Server 2003

Required software

Base SAS

Key Features

Partitioning and parallel processing

- Partitions data across multiple CPUs, I/O channels and disks.
- Leverages parallel CPUs and I/O for summarizations.
- Supports use of table aliases.
- Integrates into the SAS SQL engine.
- Provides multithreaded WHERE clause processing.
- WHERE clause planner evaluates subsetting criteria.
- Supports full table scans where no indexes are defined.
- Launches multiple threads to process different portions of a table simultaneously.
- WHERE planner costing evaluates strategies for query processing to determine least resource-intensive option.

Unique indexing technology

- Hybrid bitmap index combines best features of B-tree indexes and bitmap indexes.
- Analyzes each table and determines best indexing option.
- Builds and updates multiple indexes with a single read of data table.
- Maintains a table of statistics about indexes.

Query optimization

- Optimizes correlated queries.
- Uses transparent restructuring and recoding of nested SQL code to create temporary tables “on the fly” for accelerated processing.
- Performs fast sorting on BY statements.
- Sorts on subsets of data identified with WHERE clause or KEEP list.
- Creates single sort bins larger than 2 GB to reduce the number of bins needed to complete sort.

SQL Pass-Through support

- Sends server-specific SQL statements directly to server for execution.
- Triggers multiple independent SQL statements for parallel execution.
- SQL Pass-Through can be used with a SAS library (LIBNAME statement).

Management and security

- Handles large numbers of users and increasing amounts of data.
- Time-based partioning loads and removes data from very large tables using time-based units.
- 64-bit wide-mode support allocates and addresses more than 2 GB of memory at a time.
- Server consolidation connects Windows desktop applications to the power of SAS SPD Server.
- Sophisticated security implementation gives administrators read, write and modification access to tables on per user or per group basis.
- Integrated wtih metdata framework of the SAS Intelligence Platform to ensure data integrity and consistency throughout the BI lifecycle.
- Offers a central point of control and administration via the SAS Management Console.

World Headquarters
and SAS Americas
SAS Campus Drive
Cary, NC 27513 USA
Tel: (1) 919 677 8000
Fax: (1) 919 677 4444
U.S. & Canada sales:
(1) 800 727 0025

SAS International
PO Box 10 53 40
Neuenheimer Landsr. 28-30
D-69043 Heidelberg, Germany
Tel: (49) 6221 4160
Fax: (49) 6221 474850

www.sas.com

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 trademarks of their respective companies. Copyright © 2005, SAS Institute Inc. All rights reserved.

101254US_362964.1205