



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?

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

Why is SAS® Scalable Performance Data Server® important?

The capabilities of SAS Scalable Performance Data Server ensure that reporting and analytic applications maintain consistent performance and that data integration processes do not exceed the time windows available as 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 who are interested in significantly speeding up their reporting and analytic applications or shortening extraction, transformation and loading times for their data warehouses, irrespective of the amount of data held across the enterprise.

Decision makers who rely on business and analytic applications are looking for one thing: accurate answers when they need them. Unfortunately, the subset of 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. Because transactional systems were not designed for large-scale data retrieval, reporting and analysis, such actions can reduce system performance, consume additional 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 fast loading and fast retrieval of data subsets from large stores of data. Large-scale reporting and analytic applications will benefit from the swift access to detail data stored in this server. It ensures that applications perform consistently and that data integration 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 the elimination of 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 reporting and analytic applications require only a subset of enterprise data. SAS Scalable Performance Data Server uses parallel storage technologies and hardware to reduce overhead, such as the two-phase commit, which is incurred with transactional systems. This significantly shortens search and delivery times.
- **Optimize performance for reporting and analytic applications.** By separating the transactional data store from the reporting and analytic application 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 data integration processing times.** An integrated extraction, transformation and loading process moves data from transactional systems to the SAS Scalable Performance Data Server using parallel storage and efficient indexing. This reduces processing time, 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 vast quantities 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 launches 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 High Performance

Scalable I/O

SAS Scalable Performance Data 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 of the 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 Scalable Performance Data Server deployments contain more than 8 terabytes of data with single tables exceeding 500 gigabytes. SAS Scalable Performance Data 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.

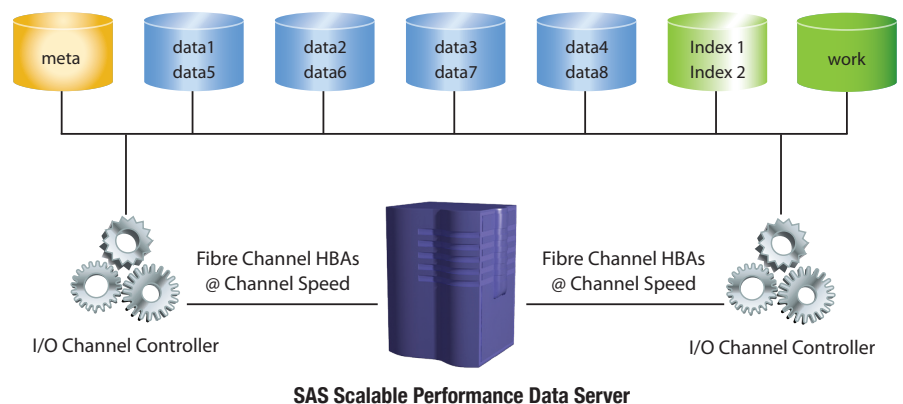
Parallel Group By enhancements

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 data 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.

The WHERE clause planner costing feature evaluates the effectiveness of different strategies for query processing and determines the least resource-intensive option. It also determines whether to use an index and which one to use if choices 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 be retrieved and processed so that 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 multiple-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 Scalable Performance Data 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

SQL Planner optimizations improve performance for the types of queries most often found in reporting and analytic 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 2 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 that are 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.

Dynamic cluster tables

SAS Scalable Performance Data Server uses a virtual table called a dynamic cluster that facilitates the loading of vast amounts of data in parallel, reducing the time needed for this task. At the same time, it enhances the handling of large data sets by breaking them into manageable pieces, e.g., for effective aging of old data. The speed of data retrieval also can benefit significantly from dynamic cluster tables by easily excluding entire areas of the virtual table from the query.

64-bit wide-mode support

64-bit support enables SAS to allocate and address more than 2 gigabytes of memory at a time, which improves efficiency and performance when processing large amounts of data.

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. The audit logs captured allow IT organizations to tune and comply with regulatory obligations.

Key Features

Partitioning and parallel processing

- Partitions data across multiple CPUs, I/O channels and disks.
- Supports dynamic clustering of tables with identical structure.
- Leverages parallel CPUs and I/O for summarizations.
- Supports use of table aliases.
- Integrates into the SAS SQL engine.
- Provides multithreaded WHERE clause processing.
- Evaluates subsetting criteria via WHERE clause planning.
- Supports full table scans where no indexes are defined.
- Launches multiple threads to process different portions of a table simultaneously.
- Evaluates strategies for query processing to determine the least resource-intensive option using the WHERE planner costing feature
- Uses memory lookups in phase two of a star join instead of index points to a table for enhanced performance.

Unique indexing technology

- Hybrid bitmap index combines best features of B-tree indexes and bitmap indexes.
- Analyzes each table and determines the best indexing option.
- Builds and updates multiple indexes with a single read of a data table.
- Maintains a table of statistics about indexes.
- A new SQL reset option increases the maximum number of hash joins that are available.

Query optimization

- Provides star schema optimization and parallel and optimized join methods for data models other than star schemas.
- 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 2GB to reduce the number of bins needed to complete the sort.
- Provides star join optimizations for use in more unique query instances.

SQL Pass-Through support

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

Management and security

- Handles large numbers of users and increasing amounts of data.
- Offers dynamic cluster tables for enhanced manageability and scalability of large tables.
- 64-bit wide-mode support allocates and addresses more than 2GB of memory at a time.
- Provides sophisticated security implementation that gives administrators read, write and modification access to tables on a per-user or per-group basis.
- Ensures data integrity and consistency throughout the reporting and analytic application life cycle via integration with the SAS metadata framework.
- Offers a central point of control and administration via the SAS Management Console.
- Allows user names to begin with either a character or number.
- Provides row-level security using WHERE constraints.

SAS® Scalable Performance Data Server® System Requirements

To learn more about SAS Scalable Performance Data Server system requirements, download white papers, view screenshots and see other related material, please visit www.sas.com/spds.