

MDDB Support in Scalable Performance Data Server(2.5

Advanced Server Development
 SAS Institute Inc
 Cary, Nc 27513

On-line Analytical Processing (OLAP) allows individuals to gain insights to large amounts of data through fast analysis techniques that are supported by organizing the data into dimensions of the enterprise that are understood by the user. Typically, OLAP is performed on very large volumes of data maintained in a data warehouse or datamart.

Standard database methods to query/analyze the enterprise are not efficient due to the magnitude of the data resulting in an unacceptable time to respond to the request. A key requirement of a successful OLAP solution is to minimize the time of the query and analysis of the data to provide realtime results to the user. This can be accomplished by building a smaller multi-dimensional database (MDDB) from the enterprise data that is specifically organized to support the OLAP requirements of the user.

A dimension contains attributes of a hierarchical nature representing a similar concept in the user's view of the data. For example a time dimension could consist of year, month, and day attributes while a geography dimension could consist of continent, country, region, and state attributes. Associated with the dimensions are measures that are of interest to the user. For example a measure could be sales. Associated with a measure are calculations such as sum, min, max, standard deviation, etc. An MDDB is used to pre-calculate and store the measure calculations across dimensions, organized in such a way as to support OLAP. In the above example the MDDB would maintain the calculation of sales across the time and geography dimension. The hierarchical structure of a dimension allows for drilling up or down the levels of the dimension to get the more general aggregates (up) or more specific aggregates (down) of the calculations across the hierarchies. For example one could display the sum of sales for North America in 1994, then drill down to the sum of sales for the countries in North America in 1994, then drill down to the sum of sales for the states of a particular country in 1994, etc.

The addition of the Microsoft OLE DB for OLAP support in SPDS is a natural evolution for SPDS. SPDS was designed to provide solutions for large volumes of data. The multi-processing design of SPDS provides support for time sensitive solutions of complex queries. These are key attributes for providing OLAP MDDB solutions. An SPDS MDDB can be created from an SPDS table using the "SPDS" option to PROC TMDDDB. Using PROC TMDDDB a user can specify the base SPDS table to create an SPDS MDDB containing the specified dimensions, hierarchies, and measures. SPDS will read the base table, do aggregations across the hierarchies to pre-calculate results, and store them in the MDDB. .

The SPDS MDDB presents a logical view to the user of a single table. However, under this logical view, the SPDS MDDB is actually a set of several SPDS tables that contain manageable pre-aggregated

results across the hierarchies in the dimensions of the MDDB. The aggregation tables are constructed from the base table or from already created aggregation tables in the SPDS MDDB (whatever is more efficient) using SPDS parallel group by technology. This technology breaks the problem into smaller parts that are done in parallel. The result is then aggregated from the parts. The aggregation tables themselves are also created in parallel, further minimizing the create time for the MDDDB. SPDS hybrid bitmap technology is utilized to maintain indexes on the tables to support efficient slicing and dicing of pre-aggregated hierarchies.

An SPDS OLE DB for OLAP provider and an SPDS MDDB Server are provided to query the SPDS MDDB. The SPDS OLE DB for OLAP provider provides a standard interface to Microsoft compliant OLE DB for OLAP clients running on Windows/Windows NT. Microsoft OLE DB for OLAP defines a query language known as MDX (multi-dimension expressions) that provides a full set of capabilities to query MDDBs. The SPDS OLE DB for OLAP provider acquires meta-data from the SPDS server that allows the client to construct MDX queries for the MDDB. The client driven queries are given to the SPDS OLE DB for OLAP provider, who sends the query to the SPDS MDDB server and awaits the the answer set. When the answer set is retrieved, it is made available to the client through provider supported OLE DB for OLAP query methods.

The SPDS MDDB Server is a multi-threaded server that can support connections to multiple SPDS OLE DB for OLAP clients. The SPDS MDDB Server will parse, plan, and execute MDX queries and send the results back to the SPDS MDDB for OLAP provider where they can then be obtained by the OLAP client. The SPDS MDDB Server uses knowledge obtained from previous SPDS efforts to utilize parallel processing technology to more efficiently respond to MDX queries on multi-processor platforms. Specifically, the SPDS MDDB Server will plan the execution of the MDX statement as a set of smaller queries based on the p-e-aggregated SPDS MDDB tables that will be involved in the query. A thread is created for each of the smaller statement parts that will execute a query on the appropriate SPDS MDDB sub-table to retrieve it's part of the answer set. The thread will send its partial response back to the MDDB OLE DB for OLAP provider, which will accumulate the parts into the complete solution. The combination of the SPDS MDDB Server parallel query technology coupled with the organization of the SPDS MDDB data to efficiently handle parallel queries will result in excellent performance of MDX queries generated by OLAP clients against an SPDS MDDB.

Refer to the SAS SPDS web gage, <http://www.sas.com/software/components/spds.html> for more information.