Visualising Large Data Volumes with SAS/MDDB® Server, SAS/CONNECT® software, SAS/EIS® Software and the HOLAP Add-ins.

Daniel Morris
Amadeus Software Ltd.

Abstract

A successful Data Warehouse Design for a Major UK Distribution Company resulted in the development of a Data Warehouse using the SAS System. The Data Warehouse is located on an MVS mainframe running SAS® Release 6.09 and this is used as a server to the Windows 95 clients. The Warehouse holds summarised SAS files created from over 200GB of IMS-DL/1 data. This data is then accessed through a Client-Server structure by SAS/EIS objects within a Graphical User Interface running SAS Release 6.12. The use of the HOLAP Add-Ins to SAS/EIS Software allowed the SAS/MDDB Server to provide extremely fast data access times. For a summarised SAS file of 1.5 million observations, the longest drill time was under 5 seconds and in many cases much shorter. This paper documents this Data Warehouse and its associated EIS and OLAP aspects.

Introduction

Amadeus Software Ltd is a SAS Institute Quality Partner based in the UK. We offer SAS development and programming services to clients worldwide. Amadeus Software has completed several major Data Warehousing projects and this paper describes a specific example of a successful SAS Data Warehouse application that we have been involved with. The Data Warehouse has several aspects which make the concepts used and issues raised in this paper very generic and applicable to a wide range of other situations.

The importance to deal with large volumes of data in an efficient and straightforward manner, the enhanced functionality of reporting tools and the fast performance of the Data Warehouse are all important factors in consideration.

Project Description

Overview

The Data Warehouse design consists of two Data Warehouses, one for Sales and one for Purchases of electrical goods for a large multinational distribution company in the UK. The data is held in a variety of sources, including DB2, and IMS-DL/1 databases. The server that is used is an MVS mainframe running batch and TSO. The users use Windows clients to run the application and are not able to access the mainframe directly. The application running on the server accesses data transparently to the user. The reporting requirement is for a large variety of standard and ad-hoc reporting screens. Many different ways of looking at the data are provided. Slice and Dice techniques along with multi-dimensional drilldown screens are available.

Detail

SAS/ACCESS® software to IMS-DL/1 is used to extract data from the IMS-DL/1 database. This database holds 3 years worth of sales and purchasing data for over 120,000 products distributed through 20,000

vendors. The data volumes amount to 224GB and the extraction program takes several hours to create SAS datasets from this data. The initial data load only contains 1 years worth of data at 12 million records, but monthly refreshment of the Data Warehouse eventually builds up to holding 3 years, 36 million records, worth of data.

The data steps involved create monthly files which are appended to the historical data and then summarised to form the basis of the Data Warehouse.

Other data (mainly used for formats within the reporting side) is extracted from DB2 data files and merged in with the main sales and purchasing data.
During a monthly refresh cycle, the new monthly data are extracted from the database into a SAS dataset. The data is then summarised into several different levels of summary datasets. Proc Summary is used to perform this work and takes several hours to create the summary files.

The different levels of summary files reflect the business reporting requirements as defined in the Data Warehouse Design and catered for by the client application.

After the various summary files are created for the current month, the program appends the data to the historical data in the warehouse. By creating small summaries of the current month and appending in this way, we minimise the processing involved and add only in part to the processing required by the client application. The client application sees a "summary" dataset which contains a small number of records that will make up the required data. Totals, averages and counts will still need to be made from the summary data.

Data Provided by the Server

The end result of the summarisation performed on the MVS server is a series of SAS datasets that can be accessed with SAS/CONNECT software from the Windows clients. The summary files contain summary data with a variety of levels, ranging from a dataset holding 12 observations and 2 variables to the "nway" dataset holding 5 million observations and 20 variables.

Client Server Implementation

A Client/Server implementation creates the "best of both worlds" for an application. Data can be held on a server and designed and tuned for efficient access. Users of the application work within a familiar environment of Windows style screens. Any processing required by the application can be performed on with the client or the server, depending upon the complexity and size of the data involved.

SAS/CONNECT software is used to provide the seamless connection between the client and the server. This enables the application to use the data on the server and also execute programs on the server to subset the summary data required by the client screens.

Reporting Screen Prototypes

The prototypes that were developed for the Client used SAS/AF$^{®}$ software and SAS/EIS software. One of the main requirements of the client application is that the screens be easy to maintain and enhance. For this reason there were several prototypes that were assessed for ease of use, functionality and maintainability. The software considered was:

- MOTORE
- EIS Multi-Dimensional Report
- TabGraph
- Custom Built AF screens
- HOLAP Objects

Motore was implemented as the first prototype and used with some success for the initial screens.

The EIS Multi-Dimensional Report object within SAS/EIS was also used to some degree of success. Limitations in this approach were the use of remote library

services to download the data for local processing by the client.

TabGraph was implemented and although the object performs true remote processing on the mainframe, the object deals exclusively with SAS datasets and not MDDB's, which were seen to provide a much faster access time.

Custom Built AF screens were implemented and assessed, however, the issue of ease of maintenance arose in this case. AF and SCL involves a fairly high degree of technical expertise to maintain, and this was seen as contrary to the design requirement of ease of maintainability.

AF screens were sometimes used as a front end to some of the EIS objects, as a sub-setting facility to improve speed of the EIS objects.

The HOLAP objects were seen to be the most appropriate solution in this case as they allow true remote processing of the SAS datasets and return SAS MDDB files to the client application. This mix is ideal in that the use of MDDB's in the client application results in significant performance gains.

Client Application

The client application consists of a series of custom built SAS/AF screens that provide a menu system and act as front ends to the main reporting screens.

The Reporting Screens use the HOLAP objects, and in particular the Multi-Dimensional Report object to create the screens defined by the Business Requirement. The other HOLAP objects were assessed but not used to any great extent.

HOLAP Benefits

HOLAP (Hybrid On-Line Analytical Processing) combines the best features of relational OLAP and the use of multi-dimensional databases. The main benefits of using the HOLAP add-ins to SAS/EIS software are:

- use of compute server for improved performance
- the use of distributed data for increased scalability
- a cache to provide immediate responses for repeated requests
- partitioning of data into separate data sources to improve performance
- reducing the storage space

(Taken from the SAS/EIS Technical Report: HOLAP Extension, Release 6.12).

By using the Multi-Dimensional Report and the HOLAP add-in, the application allows true client server processing and has the increased benefits listed above. The application is also extremely easy to maintain and enhance, since the creation of multi-dimensional reports in SAS/EIS software is a simple process that does not require extensive technical expertise, SCL programming knowledge, or knowledge of database systems. Administrators can add a screen to the client application in an half an hour or less, including testing of the application.

Intelligent Selection of Data

The HOLAP extensions have built in intelligent selection of data files to use for a drill down request. When a user drills through an application, the HOLAP object will look through a list of "associated datasets" that is maintained within the object. The list is searched in order of size of the datasets, smallest first rising to the largest - usually the "nway dataset" - last. This means that the variety of datasets created by the summarisation stage can be listed within the attributes of the HOLAP object and searched through to meet the best match for the drilldown request. When the screen is first shown, this may be a simple 12 observation, 2 variable file that

shows the highest level of drill. When the user drills down into the data, the list is searched again, and the smallest dataset that meets the requirement of the drill request is used. This method allows the user to drill through the datasets in the associated list. The application runs this processing entirely in the background, totally transparent to the user.

This functionality is based on the use of a "proxy" MDDB, which is used by the HOLAP objects attributes to define its data. This "dummy" file attribute works together with the associated data attribute to provide a list of datasets that the object can select from.

Cached Requests

The HOLAP application can be structured to provide a cache. This is a location on the client where the resulting MDDB that is downloaded from the server is placed. A log dataset is maintained by the HOLAP object and when another request for a drill down happens, the dataset is used to determine if the request can be met from the cache area. If so, the cached version is used to fulfill the request and so significantly increase performance.

Summary

The client application allows users to drill into large data files and provides extremely fast performance figures. The seamless integration of the MVS server data and the Windows client application through the use of SAS/CONNECT software, SAS/EIS software, and the HOLAP Add-ins provides a true client server implementation allowing compute processing. It has the added advantage of returning SAS MDDB files to the client application resulting in further increased performance of the client application.

Contact Information:

For further information:

Amadeus Software Ltd.
13 Corn Street, Witney, Oxfordshire, OX8 7DB, England.
Phone: +44 (0) 1993 775454
Fax: +44 (0) 1993 700577
Email: info@amadeus.co.uk
Web: www.amadeus.co.uk

Our web site contains descriptions of other Data Warehousing projects completed by Amadeus Software as well as all our other services in the SAS System such as the provision of contractors, training, and products.
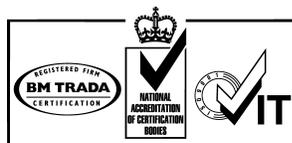
Amadeus Software is a Quality Partner of SAS Institute in the UK for Applications Development

The Quality System of Amadeus Software meets the requirements of ISO9001, the internationally recognized certification for quality services and products.

Trademarks