

Using Desktop OLAP Tool in Easy Interface for SAS

Zhuan (John) Xu

Big Creek Software, L.L.C, Polk City, Iowa

Abstract

Online Analytic Processing (OLAP) is a popular front end for data warehousing. It provides an easy way to 'slice and dice' and 'drill down'. Therefore, it is a very important tool in data exploration. A SAS OLAP application can be built with SAS /MDDDB and SAS/EIS.

In the desktop environment, a more portable OLAP is desired. This paper describes the concept of OLAP and demonstrates Easy Interface for SAS (EI), a desktop OLAP application. It uses SAS/BASE to generate summary data and has an easy to use GUI for users to point and click through different data dimensions and drill down to more detail.

EI is a Windows 95/98/NT application developed as a front end to SAS. EI is available free of charge on the Internet at <http://www.bigcreek.com/sas>.

Introduction

Today, more and more companies are building Data Warehouses to fulfill their decision support needs. The SAS Institute has played an active role in this area by developing tools such as SAS/Warehouse Administrator, SAS/MDDDB, SAS/EIS, etc. Among many methods for exploring the data warehouse, OLAP is a popular way for users to look at their data to answer business problems. It provides an easy way to do 'slicing and dicing' and 'drilling down' of the data.

Typically, data warehousing requires many resources and is a significant business endeavor. Oftentimes, the OLAP functionality is developed in the later phase of the project. As a result, OLAP is only available to a small percentage of SAS users. Yet the OLAP itself is quite useful for users wanting to explore their data, even without a formal data warehouse.

This paper presents a simple desktop version of an OLAP tool built into Easy Interface for SAS. EI is a Windows application developed using Borland Delphi with SAS as an OLE automation server. It adds a graphic user interface (GUI) for many procedures in SAS, and is designed exclusively for SAS users who work with SAS data sets on Windows 95/98/NT platform.

What is OLAP?

According to a SAS Institute's white paper on OLAP¹, "Online Analytic Processing is a hardware/software configuration that provides a fast way of viewing and analyzing data from any perspective without having to specify the perspective and the exact level of detail required in advance. In other words, OLAP is a technology that offers fast access to multidimensional information for exploration in a wider business intelligence contents."

For example, suppose we are interested in the sales performance of a company, and we want to see the sales by region, product, and time. Furthermore, if we find something interesting, say the trend shows that overall sales for the current period is better than previous periods, then we want to 'drill down' to see more details. We may want to know if certain regions were doing much better than other regions or if a new product made the difference. This is a dynamic process called 'slicing and dicing'. The finding from one step will help us to decide what to look into next.

In general, OLAP is a discovery and query tool for uncovering information and a support tool for analysis and confirmation at the overview level.

OLAP does have its limitations. It is often necessary to do more detailed analysis to meet business needs. If the dimensions are unknown, you need to identify these dimensions first. On the other hand, if there are too many dimensions, you need to conduct analysis to identify the relevant dimensions.

SAS and OLAP

As mentioned earlier, SAS institute has products specialized in OLAP. SAS/MDDDB is a tool for creating multi-dimensional database. SAS/EIS includes some multi-dimensional objects specifically designed for OLAP. One can also use SAS/AF to develop an OLAP front end. These tools make SAS a good choice for OLAP-type analysis. Details of SAS OLAP tools and their application are beyond this paper and can be found in SAS documentation and on the SAS Institute's Web

site. You may also find a good discussion in Peter R. Welbrock's book² on data warehousing.

It is worth to mention that PROC MEAN and PROC SUMMARY, two very similar procedures for simple statistics in SAS/BASE product, provide us with a tool to generate results closely related to OLAP. Using the sales example in the previous section, suppose we have a SAS data set SALES containing sales transaction data with three dimensional variables: REGION, PRODUCT, and TIME as well as an analysis variable AMOUNT. The following simple SAS program will produce all the data we need for OLAP:

```
/* summarize data */
proc means noprint data=SALES;
  var AMOUNT;
  class REGION PRODUCT TIME;
  output out=RESULT n=N sum=SUM
         mean=MEAN std=STD;
run;

/* output result */
data _null_;
  set RESULT;
  file 'C:\RESULT.TXT';
  LEVEL=put(_type_,binary3.);
  /* write header */
  if _n_=1 then
    put "LEVEL"    '09'x
        "REGION"  '09'x
        "PRODUCT" '09'x
        "TIME"    '09'x
        "N"       '09'x
        "MEAN"    '09'x
        "SUM"     '09'x
        "STD"     ;
  put  LEVEL '09'x
      REGION '09'x
      PRODUCT '09'x
      TIME   '09'x
      N      '09'x
      MEAN   '09'x
      SUM    '09'x
      STD    ;
run;
```

The resulting output is a tab-delimited text file like the following:

LEVEL	REGION	PRODUCT	TIME	N	MEAN	SUM	STD
000				100	50	5000	40
001			97	40	45	1800	30
010		1		20	50	1000	40
100	East			30	40	1200	45
011		1	97	10	30	300	33
101	East		97	10	40	400	46
110	East	1		10	45	450	45
111	East	1	97	3	44	132	33

Note that the use of the CLASS statement is the key in generating such a multi-dimensional summary. Also note that the LEVEL represents the dimensions and it is the SAS generated variable _TYPE_ in binary format.

To create an OLAP tool, we first need a graphic user interface to select the desired analysis variable and dimensional variables. Next, the corresponding SAS program should be dynamically generated and submitted to produce the result file. Finally, the most important part is providing a user-friendly front end for 'slicing and dicing' and 'drilling down' the multi-dimensional summary statistics. Easy Interface for SAS provides such a front end.

Overview of Easy Interface for SAS

Since version 6.11, SAS has been an OLE Automation server. Through OLE Automation, SAS exposes itself to another application as a programmable object. This means that we can use any application that can act as an Automation controller to create a SAS session and control it using the methods and properties that the SAS System makes available. Details of such technology can be found in the SAS on-line help file under OLE automation of SAS system. My MWSUG 97 paper³ and SUGI23 paper⁴ have some information and examples as well.

Easy Interface for SAS is a Windows 95/98/NT GUI (graphic user interface) application developed using Borland Delphi. EI uses SAS as a database and analytic engine through the SAS OLE automation server. Users are provided with a familiar main screen styled after Windows Explorer. EI allows users to summarize data, calculate correlations, and perform sample size estimation, random sampling, T-tests, ANOVA and regression. In addition, EI supports viewing, subsetting, sorting, merging, importing and exporting SAS data sets. Another nice feature is EI's Ad Hoc Application, which allows most SAS programs to be integrated into Easy Interface for SAS by creating a dynamic GUI for the SAS programs.

Its Desktop OLAP function is powerful yet quite easy to use for 'slicing and dicing' and 'drilling down'. The details of DOLAP will be presented in the next sections.

Design and Implementation of OLAP in EI

As discussed earlier, SAS/BASE can be used to generate a multi-dimensional summary. The goal is to create a

simple yet powerful front end that creates and accesses the summary data.

OLAP tools are often used by a knowledge worker, who is interested in extracting information and may not be skilled in programming or using software. It is desirable to have a simple graphic user interface support with 'point and click' capabilities. In addition, since the user will decide the next step of analysis based on the findings from the current step, it would be nice to integrate the user input with output so the user can focus on both simultaneously. Furthermore, the result should be displayed both numerically as well as graphically.

From a programming point of view, building the above functionality from ground up is a big task. Fortunately, most modern programming tools provide ready-to-use components. For example, Borland Delphi comes with database, report, and graphic components. These three major pieces were used in OLAP part of EI.

The first part of OLAP in EI was designed to dynamically create a multi-dimensional summary based on user selection. The second part is for conducting OLAP. Screen snapshots are presented in the Appendix.

Using the OLAP functionality in EI

Using the OLAP functionality in EI is quite easy. You first start EI and select a SAS data set to be analyzed with the mouse, as illustrated in Figure 1 of the Appendix. Then choose Analysis, OLAP, New... from the menu. This will start a new window similar to Figure 2. Now you choose an Analysis (Factual) Variable and Dimensional (Descriptive) Variables by 'Point and Click' or 'Drag and Drop'. Click the OK button to start summarization. When it is done, EI will display a new window with a chart and a table of the summarized results (see Figure 3). This is the place you actually conduct analysis. You may:

- Slice and dice to another dimension.
- Drill down to next level.
- Select different statistics for analysis.
- Sort the table and chart.
- Save data or graph.
- Print/Preview report/graph.
- Copy data/graph to other application through Windows' Clipboard.
- Change between Bar, Pie and Line Charts for display.
- Make comparisons.
- Change options such as titles, footnotes, and labels.

Mainly, you 'point and click' to do the above tasks. For example, clicking on a blank cell in another dimension at the same level will switch to that dimension. This is called 'slicing and dicing'. Clicking on a blank cell at a lower level will drill down. Clicking on a cell of statistics (Count, Average, Total, STD) will change the chart to display the selected statistics. Clicking on the header will sort the data and chart by that column. For example, clicking the header TOTAL will sort the data and chart by the total of the analysis variable.

Clicking on buttons in the toolbar will accomplish other tasks. For example, clicking on the button with a printer then selecting Report from the pop-up menu will display a report preview window as shown in Figure 5. You can then print the report.

If you have saved the summary data, you can simply choose Analysis, OLAP, Open... from menu to bypass data summary step. EI also automatically saves the last OLAP summary data in c:\temp\olap.dat. So you can always continue with the OLAP analysis you did last time. Furthermore, the summary file can be forwarded to another user of EI. In this case, since the summary has been done, the new user can perform OLAP with EI alone, without SAS on the computer.

Conclusion

Desktop OLAP is a powerful tool for use in exploring data. It enables us to perform OLAP without a formal data warehouse. Easy Interface for SAS provides such tool to SAS users in Windows. Its user-friendliness and other features enable us to focus on business rather than programming and other technical issues.

References

- ¹ SAS Institute Inc., *A formula for OLAP Success*,
- ² Peter R. Welbrock, (1998), *Strategic Warehousing Principles Using SAS Software*, Cary, NC: SAS Institute Inc.
- ³ Zhuan (John) Xu, (1997), SAS Application Development Using Windows RAD Software for Front End, *Proceedings of the 8th Annual Conference of the Midwest SAS Users Group*, Chicago, Illinois.
- ⁴ Zhuan (John) Xu, (1998), Using SAS as an Automation Server in Windows Application Development,

Proceedings of the 23rd Annual SAS Users Group International Conference, Nashville, Tennessee.

Acknowledgments

SAS, SAS/AF, SAS/BASE, SAS/MDDDB, and SAS/EIS are registered trademark of SAS Institute Inc. in the USA and other countries. Other brand and product names are registered trademark of their respective companies.

The author may be contacted at:

Zhuan (John) Xu
Big Creek Software L.L.C.
400 53rd PL
W. Des Moines, Iowa 50266
(515) 984-6243
johnxu@bigcreek.com

Appendix: Selected Screens of OLAP Part of EI

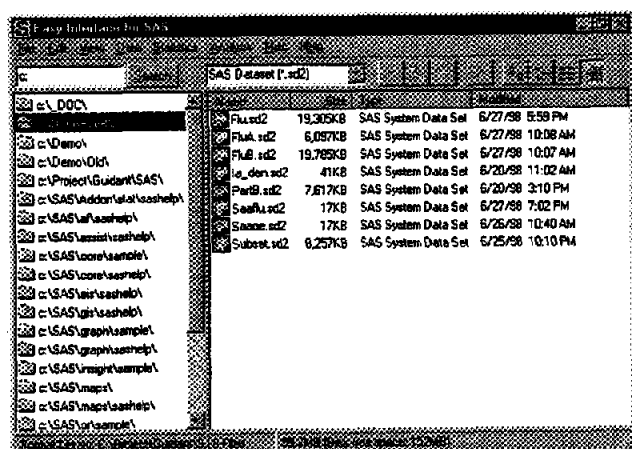


Figure 1, Windows Explorer Style Main Screen

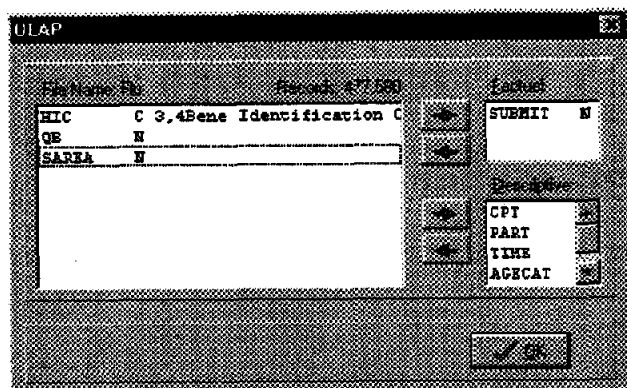


Figure 2, OLAP Summarization Selection

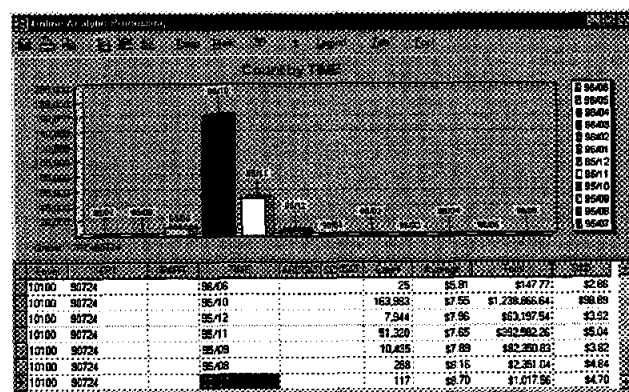


Figure 3, OLAP Analysis Screen 1

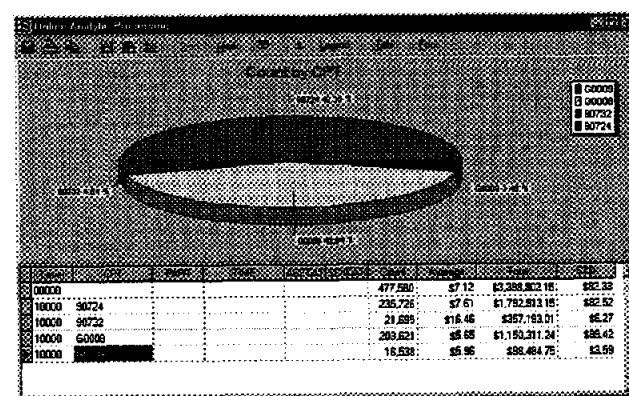


Figure 4, OLAP Analysis Screen 2

Average By CPT

Level	CPT	Count	Average	Total	STD
00000		477,580	\$7.12	\$3,398,902.15	\$92.33
10000	90724	235,726	\$7.61	\$1,792,913.15	\$82.52
10000	90732	21,895	\$16.46	\$357,193.01	\$6.27
10000	00008	203,621	\$5.85	\$1,150,311.24	\$89.42
10000	00009	16,538	\$5.96	\$98,494.75	\$3.98

Figure 5, OLAP Report Print Preview