

Exploiting Health Care Data Through Hybrid OLAP

Rush Daugette, TEC Associates, Indianapolis, IN

ABSTRACT

Using the Hybrid OLAP architecture to present and analyze health care data can provide several benefits. One is the ability to partition claims of different types into multiple physical MDDBs that can appear as one to the end-user. This type of partitioning is known as racking. Another benefit of Hybrid OLAP (HOLAP) is the ability to use tables within the data model that may have special aggregation rules, such as can occur in enrollment data. This technique is called stacking.

This paper will present a hypothetical health care OLAP data model that uses both racking and stacking. This will include an overview of the model design, definition of the HOLAP Data Group and exploitation of the final proxy MDDB using multidimensional reports created in SAS/webEIS. SAS products used include Base SAS, SAS/EIS, SAS/MDDB Server, SAS OLAP Server and AppDev Studio.

INTRODUCTION

The purpose of this paper is to illustrate some of the capabilities of SAS Version 8 Hybrid OLAP (HOLAP). Namely, the stacking and racking features are shown in the context of a healthcare data warehouse. The use of stacking addresses a particular challenge in healthcare analysis concerning non-additive enrollment data. The use of racking with various types of claims tables shows how these can be maintained separately but be viewed as one. This paper also shows how to create the HOLAP Data Group and the resulting proxy MDDB. Finally, SAS/webEIS is used to exploit the proxy MDDB with two multidimensional reports. The intended audience is intermediate SAS users and users of OLAP applications.

OVERVIEW OF OLAP MODEL

In our hypothetical health care data warehouse, we have claims data residing in three different tables based upon a standard claim categorization scheme, i.e. Institutional Inpatient, Institutional Outpatient and Non-Institutional. For the enrollment data, we have a "member months" table which contains various aggregations of the number of months each member had been exposed in a given time period. These four tables will provide the basis for our OLAP data model.

For the OLAP model, the dimensions and the levels we are interested in are as follows:

Dimensions and Levels	
Time	Incurred Year
	Incurred Quarter
	Incurred Month

Demographic	State
	Provider Network
	Group
Service	Category of Service

Figure 1 Dimensions and Levels

Enrollment Component

What we want in our OLAP model is the ability to view claim utilization per 1000 members and claim cost per member per month (PMPM) at various levels of claim aggregation. For example, a desired aggregation using the Service dimension would be:

Category of Service	Annual Utilization/1000	Cost PMPM
Medical/Surgical/Non-Delivery		
Maternity Delivery		
Well Newborn		
Emergency Room		
...		

Figure 2 Category of Service Level

At this level of aggregation, we use member months values in the denominators of Annual Utilization/1000 and Cost PMPM that will not be additive as we "roll up" to other levels of the Service dimension. These values must remain the same as those used at this Category of Service level. Once we leave the Service dimension, the member months are additive across the other dimensions, Time and Demographic. Given these special rules of aggregation, the HOLAP architecture allows us to create and store the exact subtables needed to effect these aggregations. The key thing is to create all the possible subtables that include any levels of the dimensions that are not additive. For our model, we create the non-additive subtables through an additional data step which generates Category of Service values followed by a Proc Summary using a type statement with all combinations containing Category of Service. We then create the remaining additive subtables using only Proc Summary against the member months warehouse table. Using stacking, these two summary data sets will provide the enrollment component of the OLAP model.

Claims Component

For the claims component of the model, we start with the three different warehouse tables mentioned earlier. These correspond to the three claim categories generally used in our healthcare warehouse design: Institutional Inpatient, Institutional Outpatient and Non-Institutional. In our example, these tables lend themselves naturally in creating three claims MDDBs using Proc MDDB. Through the use of racking, these MDDBs can be

maintained as separate data sources while appearing as a single source when defined in the HOLAP data group.

In creating the necessary subtables for the MDDB's, we start by using the Spiral method to obtain the following ordered list of classes:

Incurred Year
Category of Service
Network
Incurred Quarter
State
Incurred Month
Group

This list is used to generate the initial hierarchy statements used by Proc MDDB in creating the three claims MDDBs. For our multidimensional reporting, we have chosen the following three drill paths:

Time Drill Path – Year>Quarter>Month

Demo Drill Path - State>Network>Group

Service Drill Path - State>Network>Category of Service

With this in mind, we also include the additional hierarchy statements to create subtables for all possible combinations of these three drill paths. Generating these three MDDBs gives us the claims component of our model.

For HOLAP stacking and racking to be used successfully, each data source should have identical columns and statistics. In our model, in addition to the class columns given above, the enrollment summary data sets should contain the claims-related columns with value zero. Likewise, the claims MDDBs should contain the enrollment-related column with value zero (Figure 3).

Data Source	Analysis Columns	Statistic	
Claims MDDBs	Days_Services	Sum	Actual
	Admits Cases	Sum	Actual
	Paid Amount	Sum	Actual
	Member Months	Sum	0
Enrollment Data sets	Days_Services	Sum	0
	Admits Cases	Sum	0
	Paid Amount	Sum	0
	Member Months	Sum	Actual

Figure 3 Analysis Columns

As a result of the above design, the OLAP data model consists of the following data sources and physical structures (Figure 4).

Data Source	Physical Structure
Enrollment Additive	Summary data set
Enrollment Non-Additive	Summary data set
Institutional Inpatient Claims	MDDB
Institutional Outpatient Claims	MDDB
Non-Institutional Claims	MDDB

Figure 4 Data Sources

HOLAP Data Group

Definition

Now that the data sources exist in the desired form, we can create the necessary HOLAP metadata. This is done either by using the MDMDDB command or by selecting Edit > Distributed Multidimensional Metadata from the SAS/EIS Metabase Window. After defining the HOLAP Data Group, we create a proxy MDDB by exporting the metadata to the SAS/EIS Metabase. At this point, we can modify the metadata further to add computed columns as will be required for our multidimensional reporting.

The HOLAP Data Group is named HEALTH_UTIL and the five Data Sources in our model have been added to the group (Figures 5 and 6).

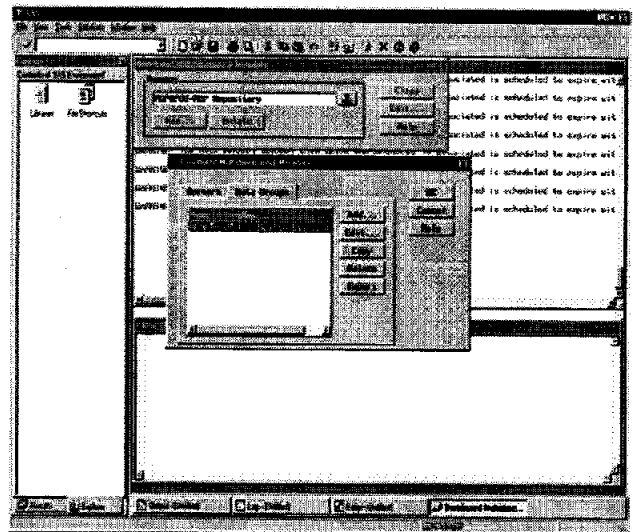


Figure 5 Distributed Multidimensional Metadata Window

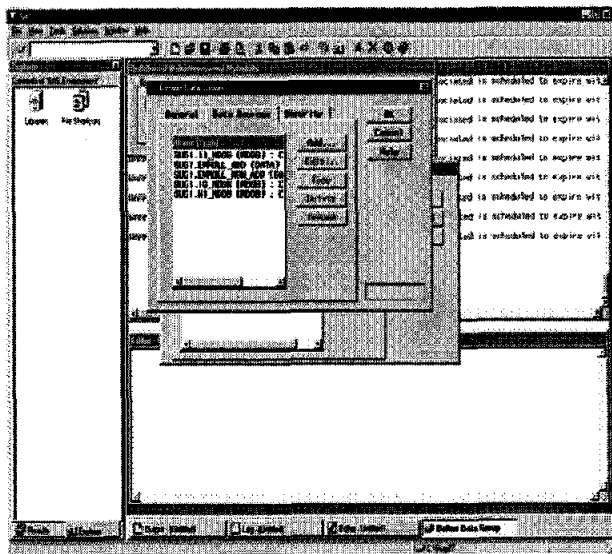


Figure 6 Define Data Group Window

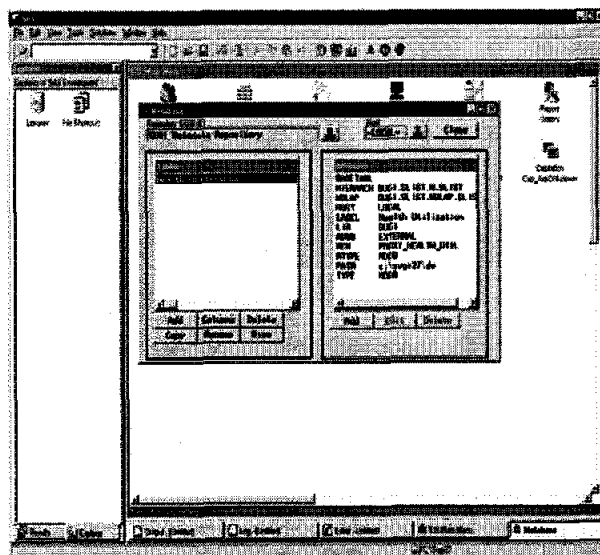


Figure 8 SAS/EIS Metabase Window

Exporting Data Group to the Metabase

After defining this Data Group, we export the HOLAP metadata to the SAS/EIS Metabase. During this process, the proxy MDDB, PROXY_HEALTH_UTIL, is created and registered to the Metabase (Figures 7 and 8). This allows multiple SAS applications to use this proxy MDDB like any other "real" MDDB. In particular, we will be using this proxy MDDB to create our multidimensional reports in SAS/webEIS.

[4] Scerbo, Marge and Levine, Stuart B. (1999), A Health Care Claims Data Mart: Construction and Exploitation, SUGI24 Proceedings

Once the proxy MDDB exists, we can modify and add to its metadata. For our reporting, we add two computed columns to obtain the results shown in Figure 2:

$\text{Utilization per 1000} = \text{Days_Services} / \text{Member Months} * 12,000$

$\text{Cost PMPM} = \text{Paid} / \text{Member Months}$

The proxy MDDB is now complete and ready to be used by webEIS.

EXPLOITING THE MDDB USING webEIS

Our reporting needs include two multidimensional reports: **Demographic Drilldown** and **Category of Service Drilldown**. These reports will use the Demo Drill Path and Service Drill Path, respectively, for the Down direction. The Time Drill Path will be used for the Across direction in both reports. The webEIS software within AppDev Studio provides a quick capability for exploiting MDDBs, proxy or otherwise. For our reports, we are using the Multidimensional Table Component to create the desired results.

An example of the webEIS Demographic Drilldown report is shown in Figure 9. An example of the Category of Service Drilldown report is shown in Figure 10. This second report illustrates both the racking and stacking aspects of our OLAP model. Note in this report that the stacking of the Enrollment data comes into play as the Member Months values remain constant between the Category of Service drilldown level and the Network drilldown level.

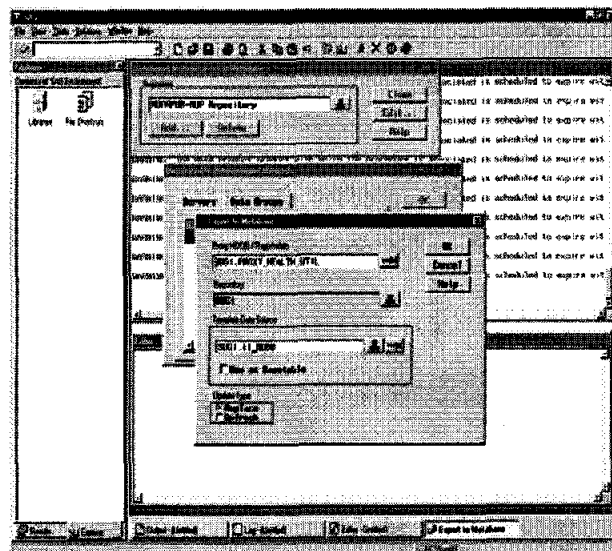


Figure 7 Export to Metabase Window

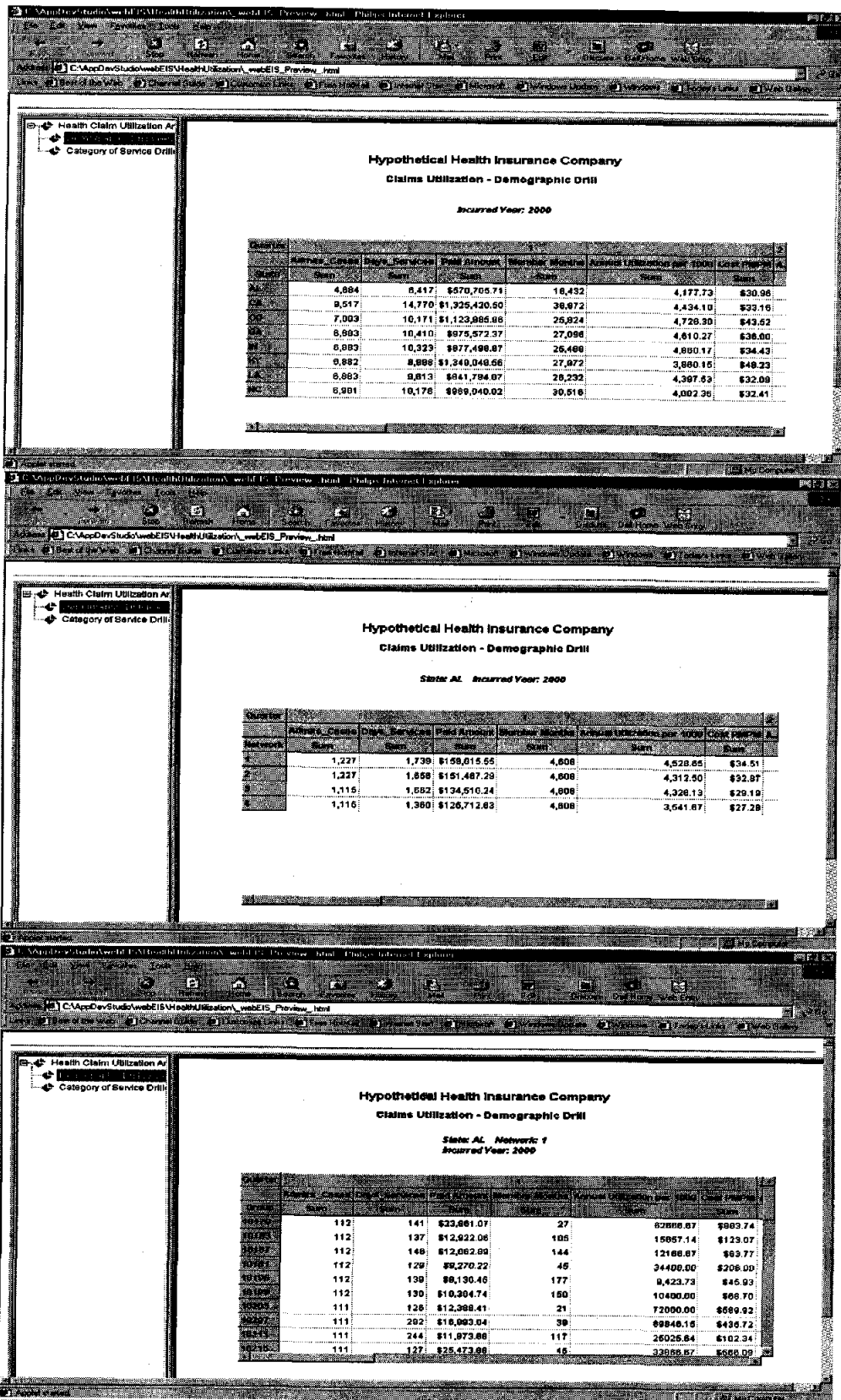


Figure 9 Demographic Drilldown Multidimensional Report

Health Claim Utilization At Demographic Drilldown

Hypothetical Health Insurance Company Claims Utilization - Category of Service Drill

Inclusive Year	2006		2007		2008		2009	
	Admitted Cases	Days Services	Days Services	Days Services	Admitted Cases	Days Services	Admitted Cases	Days Services
2006	18,735	30,137	\$2,442,275.01	73,729	4,903.48	\$33.13		
2007	36,184	51,046	\$6,030,133.32	169,898	4,641.70	\$31.46		
2008	27,713	40,152	\$3,410,188.65	103,298	4,584.50	\$33.01		
2009	27,593	43,720	\$3,590,110.00	108,384	4,641.23	\$33.18		
2010	27,663	41,859	\$3,528,505.13	101,852	4,501.01	\$34.59		
2011	27,664	38,317	\$3,517,605.96	111,888	4,109.60	\$35.01		
2012	27,655	43,539	\$3,652,720.48	104,828	4,079.30	\$36.72		
2013	27,673	42,788	\$3,527,349.09	122,084	4,208.25	\$29.59		

Health Claim Utilization At Demographic Drilldown

Hypothetical Health Insurance Company Claims Utilization - Category of Service Drill

State: AL

Inclusive Year	2006		2007		2008		2009	
	Admitted Cases	Days Services	Days Services	Days Services	Admitted Cases	Days Services	Admitted Cases	Days Services
2006	4,907	7,587	\$940,254.30	18,432	4,928.43	\$45.99		
2007	4,907	8,131	\$529,646.56	18,432	5,293.82	\$26.73		
2008	4,461	6,842	\$564,115.20	18,432	4,454.43	\$30.06		
2009	4,460	7,587	\$518,356.73	18,432	4,839.46	\$26.12		

Health Claim Utilization At Demographic Drilldown

Hypothetical Health Insurance Company Claims Utilization - Category of Service Drill

State: AL Network: 1

Inclusive Year	2006		2007		2008		2009	
	Admitted Cases	Days Services	Days Services	Days Services	Admitted Cases	Days Services	Admitted Cases	Days Services
Category of Service								
H0101 - Medical/Surgical/Obstetrical Delivery Maternity	154	1,194	\$289,428.79	10,432				
H0102 - Maternity Delivery	233	592	\$145,711.78	18,432				
H0403 - Vaginal Anesthesia	54	104	\$5,049.29	18,432				
H0404 - General Anesthesia	10	101	\$0.10	18,432				
H0405 - Unconscious Anesthesia	0	0	\$0.00	18,432				
H0406 - Emergency Room	106	238	\$18,005.24	18,432				
H0407 - Surgery	41	41	\$18,399.44	18,432				
H0203 - Other Outpatient	1,038	2,026	\$58,268.76	18,432				
H0301 - Prescription Drugs	0	0	\$0.00	18,432				
H0302 - Over the Counter Drugs	0	0	\$0.00	18,432				

Figure 10 Category of Service Drilldown Multidimensional Report

CONCLUSION

The Hybrid OLAP architecture provides great capability and flexibility in the exploitation of health care data. Stacking addresses the non-additive nature of enrollment data in the case of our Category of Service reporting requirements. Racking allows claims MDDBs to correspond to their separate warehouse tables while appearing as a single source in the reporting. The final proxy MDDB has an underlying structure that is transparent to the user. It is usable by any SAS application that exploits MDDB's.

REFERENCES

- [1] Henderson, G (2001), What You Need to Consider When Building and Deploying an OLAP Application, SUGI26 Proceedings
- [2] Moorman, Mark (2000), The Art of Designing HOLAP Databases, SUGI25 Proceedings
- [3] Weinberger, Ann and Ender, Matthias (2000), The Power of Hybrid OLAP in a Multidimensional World, SUGI25 Proceedings
- [4] Scerbo, Marge and Levine, Stuart B. (1999), A Health Care Claims Data Mart: Construction and Exploitation, SUGI24 Proceedings
- [5] SAS Institute Inc., SAS OLAP Server Administrator's Guide, Release 8.1, Cary, NC: SAS Institute Inc., 2000

CONTACT INFORMATION

Your comments and questions are valued and encouraged.

Contact the author at:

Rush Daugette
TEC Associates
7351 Shadeland Station, Suite 150
Indianapolis, IN 46256
Ph: (317) 202-0711
Fax: (317) 202-0712
Email: rush_d@tecassociates.com
Url: www.tecassociates.com