Paper 169-2009

Using SAS to Build the Managed Care Data Warehouse at Community Care Behavioral Health Organization

Mary Doyle, Qingfeng Liang, Gretchen Ciocco, Michele Mesiano, Shouqiang Wang
Community Care Behavioral Health Organization, Pittsburgh, PA

ABSTRACT

Community Care Behavioral Health Organization (Community Care) needed to increase the sophistication of its ability to manage, analyze and report data in order to support organizational improvement and to accommodate its rapid growth. Improving efficiencies in summarizing and displaying decision support data to company management was also vital. To accomplish these tasks, Community Care needed to create a data warehouse that integrated large volumes of data from a variety of divergent sources. Community Care decided to implement SAS strategies to accomplish these goals. The company partnered with SAS in a consulting engagement to apply SAS best practices in their data warehouse work. This paper discusses the strategies successfully implemented using the SAS Data Integration platform to build the Community Care data warehouse. The SAS solution is briefly described in the following areas: business needs of the data warehouse, data warehouse architecture, the SAS platform and the DI Studio tool, data warehouse access, a case study illustrating the benefits gained by creating the data warehouse, conclusions, future enhancements and lessons learned.

INTRODUCTION

Community Care Behavioral Health Organization (Community Care) is a member of the Insurance Services Division of the University of Pittsburgh Medical Center (UPMC). Community Care is the only Pennsylvania-based, not-for-profit, licensed behavioral health managed care organization (BH-MCO) created to serve HealthChoices programs throughout Pennsylvania. Community Care holds both risk-bearing and Administrative Services Only (ASO) contracts. The organization has seven offices across the Commonwealth of Pennsylvania and currently has contractual arrangements with 35 of the 67 Pennsylvania counties. With nearly 500 employees and by utilizing a network of approximately 2,000 behavioral health providers, Community Care manages behavioral healthcare for nearly 600,000 Medicaid managed care enrollees. Community Care was designed primarily to serve the needs of public sector consumers, their families, and their communities. The mission of Community Care is to improve the health and well-being of the community through the delivery of clinically effective, cost-efficient, and accessible behavioral health services.

To support the mission of Community Care and its rapid growth over the past decade, the company needed to increase the sophistication of analyses in the clinical, financial and outcomes areas. To accomplish these goals, the IT department was charged with the development of an in-house data warehouse. Community Care decided to implement SAS strategies and partnered with SAS in a consulting engagement to apply SAS best practices in their data warehouse work. The SAS solution used is described in this paper.

Business needs of the data warehouse

Business needs

Community Care relies on its data to make informed business decisions in a variety of areas ranging from financial to operational to clinical. Data are received from numerous sources including both external and internal sources. Internal data sources are generally from the databases supporting software applications used in daily operations. External sources are generally from government offices related to the various county and state contracts held by Community Care as well as Community Care's claims processing vendor. Both internal and external data are complex in both structure and in timing which presents numerous reporting challenges. The diagram below is a simplified illustration of the data sources used in Community Care's daily operations.

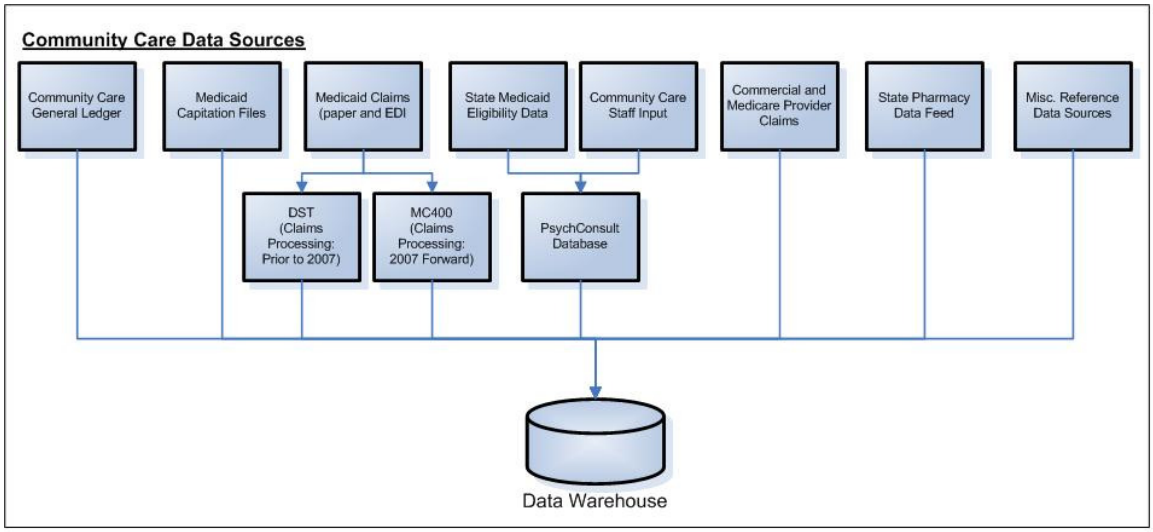

Figure 1: Community Care Data Sources

Community Care's rapid growth necessitated the construction of a comprehensive reporting environment to support reporting and analysis. Prior to the construction of a centralized data warehouse, the time and effort required creating reports and analyses was substantially higher. By creating a centralized store of data, Community Care is more nimble to respond to the unique requests received from its various county and state government contract representatives as well as from its internal management.

Business requirements

Community Care defined several business requirements at the onset of the data warehouse project. First, it was clear that the warehouse must create an integrated decision support environment to support reporting and analysis, as well as hold cleaned and standardized data in an integrated fashion. Next, designing, creating and populating data marts and similar stores was necessary to support access and flexibility, since the intention was for multiple types of users with different business needs to access the data warehouse.

In addition to the requirements above, the establishment of useable metadata documenting the warehouse, associated data marts and standard reports was also defined as a project requirement. A process for developing standards for accepting data into the warehouse, approving transformations of data, and approving reports to ensure the accuracy of system outputs and their associated documentation was another key requirement of the data warehouse project. The standardization of the logic and algorithms used to analyze the data as well as a process to support the correction of data were also vital to this project. Finally, creating a process to ensure that replacement reports and/or report modifications are completed as the modification of data stores occurred was identified as a project requirement.

Mission

The mission of the data warehouse project was to create a data and technology foundation for a decision support environment. Such an environment fosters an integrated body of knowledge about Community Care decision support, operations reporting, and outcomes analysis requirements. This body of knowledge and comprehensive data analysis is leveraged to drive organizational improvement, support organizational growth, and to provide higher quality services to our members.

The data warehouse provides a common, standardized, open data repository for reporting and analysis. In addition, it simplifies information access for all users by improving access tools and data documentation and by establishing standards for common metrics. This warehouse environment also supports access by non-technical users and senior management to further streamline the delivery of quality data.

DATA WAREHOUSE ARCHITECTURE

After gaining a better understanding of our data sources, our data consumers, and the informational support our various business areas needed, we proceeded to lay out our data warehouse architecture. Our goal was to take disparate data sources, apply our business rules through transformations and combine these data in a useful way.

We needed an end product of meaningful information so a data warehouse user could easily query the data. Obtaining the proper data sources and ensuring quality of the data were critical parts of the data integration in building our central data warehouse (CDW). Setting up cross data source data definition standards was also a key success factor in our warehousing. The architecture we implemented for our data warehouse displays the complexity of the data sources and the route we took to integrate these sources together in our CDW database. The SAS Data Integration platform was successfully utilized in building our CDW database.

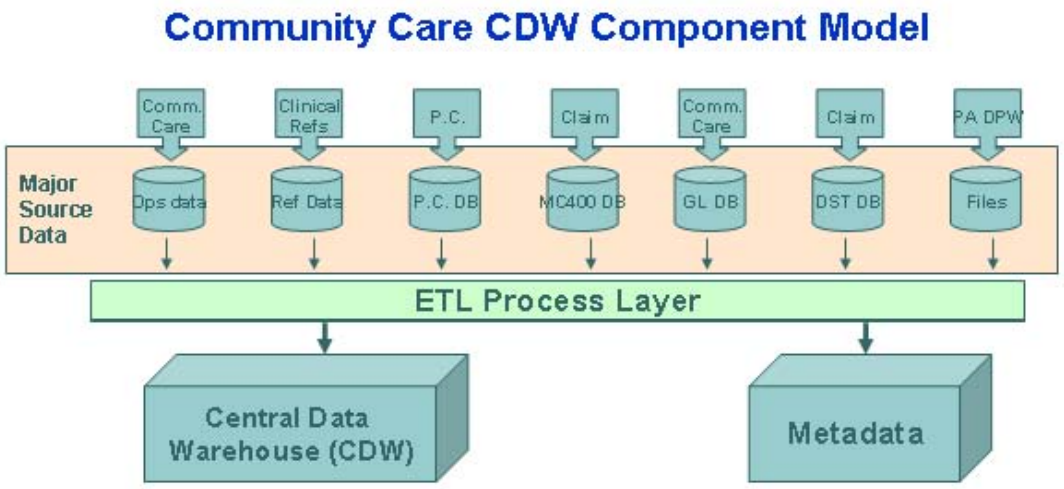

Figure 2: Community Care CDW Component Model

The next step after building our CDW database was to build the enterprise central data marts (CDM). With a better understanding of the informational support needed by our various business areas, the challenge was to organize data in a fashion that would marry subsets of the data in a manner customizable to each business unit. Defining clear business rules and constructing data constraints in the CDM by utilizing the SAS platform was vital to successfully constructing these customized marts.

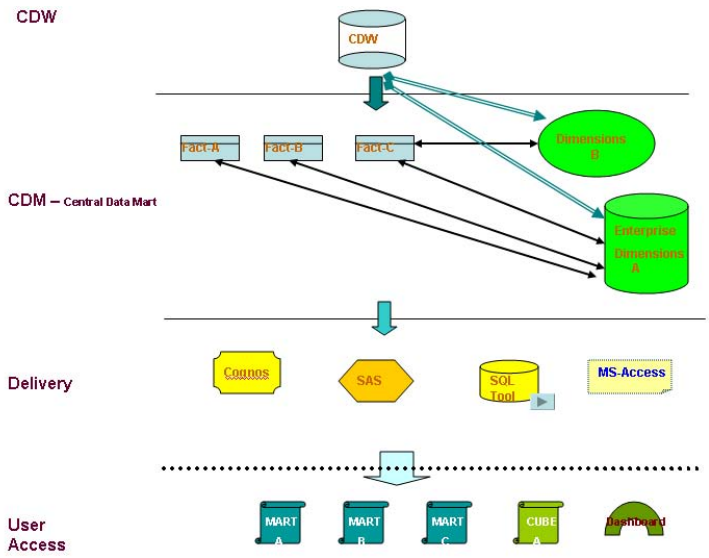

Figure 3: General Warehouse Architecture

Based on the architecture described above, we built our data marts utilizing dimensional tables. The fundamental design principle in designing our CDM was to provide an accurate and flexible approach to our business users to query the data at any time, in any format and from many angles. By using our data marts, a user can query and interpret behavioral healthcare data from any one of the dimensions presented in the CDM. To further improve query efficiencies and to set up the standard mechanisms for analyzing major pieces of data, our CDM also stores

aggregated forms of some dimensional attributes.

The SAS DI tool was a perfect tool to support the construction of our CDM based on dimensional table structures. The design below is a sample entity-relationship diagram (ERD) for the Member Enrollment mart.

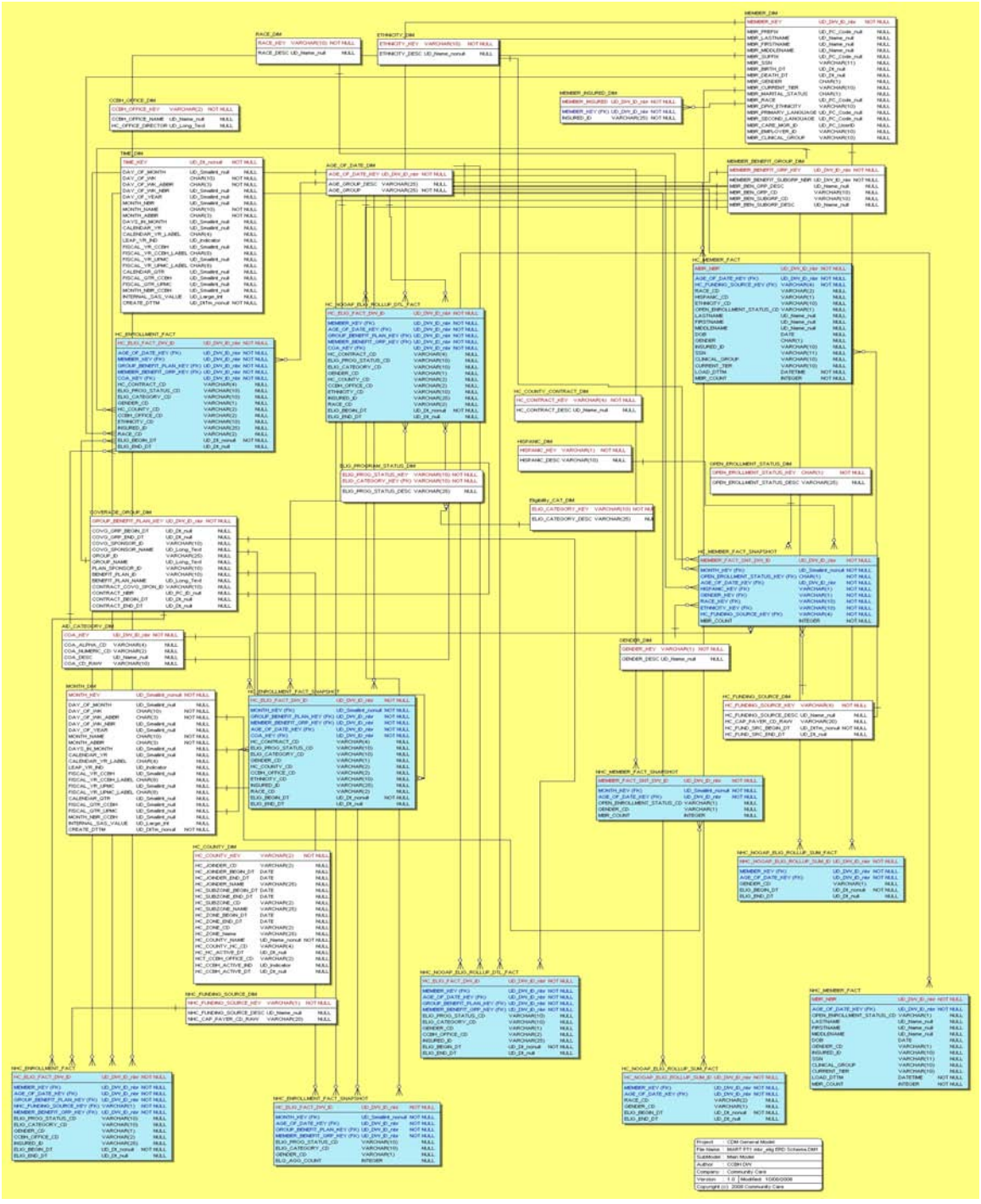

Figure 4: Member Enrollment Mart Design

The SAS platform and the Data Integration Studio tool:

In deciding upon a tool to use for building our warehouse, SAS Data Integration (DI) Studio offered some important

features which were really attractive to us. First, the extract, transform, load (ETL) process design in DI Studio is an intuitive, point-and-click tool which makes it easy to build logical process workflows and to quickly identify input and output data flows. Second, this tool provides a powerful yet easy-to-use transformation language that supports collaboration and the reuse of processes and ETL jobs. Transformations can run on any platform with any data source. SAS programmers can also create their own custom transformations which incorporate specific business logic. Third, it includes an embedded data quality feature which allows data quality processing to be imbedded within ETL jobs. This data quality feature is robust and accurate. Choosing SAS DI studio as our data integration (DI) tool was our logical choice because of these features. In addition to these DI Studio features, SAS also provides unparalleled technical support and professional services, giving our organization the flexibility to go beyond the robust prepackaged transformations and build unique routines without fear of losing vendor support for customized work.

We fully utilized the features and functions provided by SAS DI studio to create nearly a hundred ETL jobs which serve to build and maintain our enterprise data warehouse and data marts. We not only used pre-built transformations, but also wrote custom transformations containing specialized business logic to integrate data from different system sources into our CDW housed in an Oracle database. In addition, metadata was captured and documented throughout the data integration and transformation processes. ETL developers share metadata and therefore are able to utilize consistent definitions across data sources, simplify their designs and reduce maintenance costs.

We started the design of the necessary data marts from an analysis of user needs and emphasized ease of access and usability throughout the design process. We interviewed users from different departments to understand their current and future data needs as well as the types of decisions they typically base on data analysis. We designed data marts to incorporate various analytical and operational needs and to serve the needs of functional departments such as finance, clinical/utilization, network, IT, outcomes research and claims. The CDM was built using the graphical user interface provided by SAS DI studio dimensional models organized as star-schema made of fact tables and multiple associated dimension tables. The picture below shows one of the most simple of ETL jobs that we created.

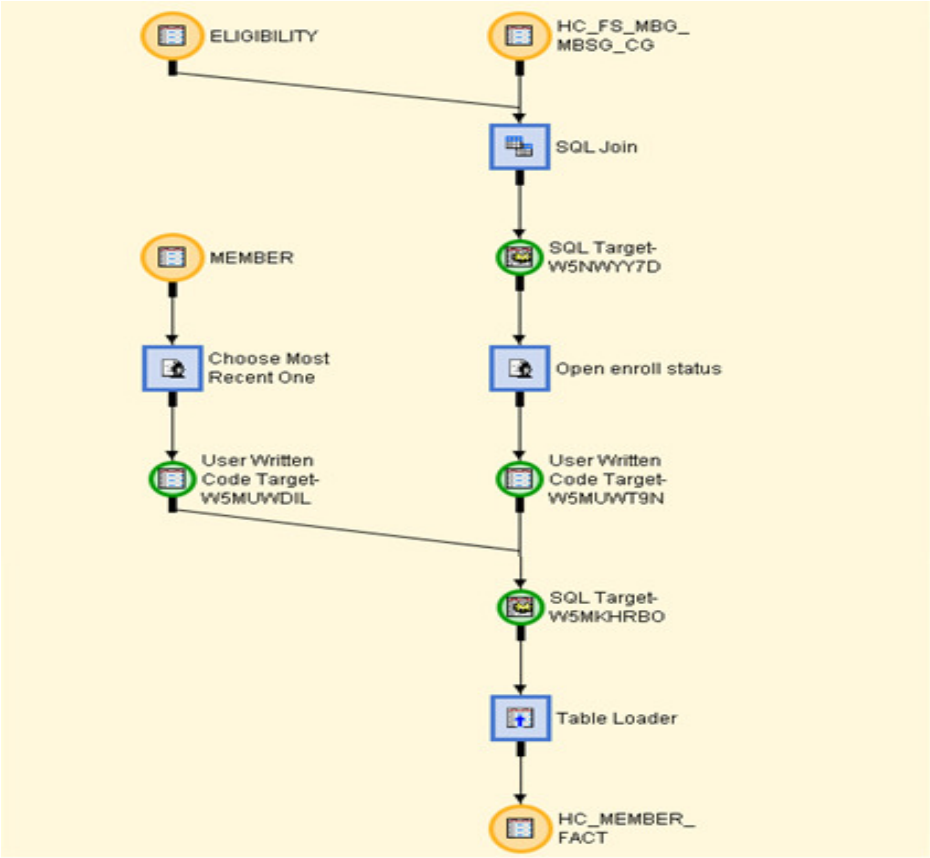

Figure 5: Sample ETL Job

Providing clean, credible and correct information is one of the IT department's foremost charges. Bad data quality can lead to poor decision making and very high costs. We utilized the SAS data quality solution (which is a combination of dfPower studio, SAS data quality server and SAS DI studio) to ensure data quality. This solution provided us with many quick and effective ways to clean and enrich data. First, we focused on profiling data to indentify the consistency, accuracy and validity of the data. Identified issues were then documented and business rules were applied to correct the issues if applicable. For example, if a member's insurance coverage period start date is later than the coverage end date, the record is not valid according to our business rules. Second, we improved data quality by applying processes such as de-duplication, business rule creation, and cleansing of data where required. Third, we set up processes which automatically detected and alerted ETL developers when data violates business rules or standards. For example, follow-up letters are sent to our members to ensure they receive timely and appropriate services subsequent to other types of services. In instances when a member has an invalid address or multiple mailing addresses with different spellings associated with the same address, the SAS data quality solution provides us with the ability to standardize the member address and apply fuzzy logic to clean errors associated with the address field. The following picture depicts one of the ETL jobs designed to deal with data quality.

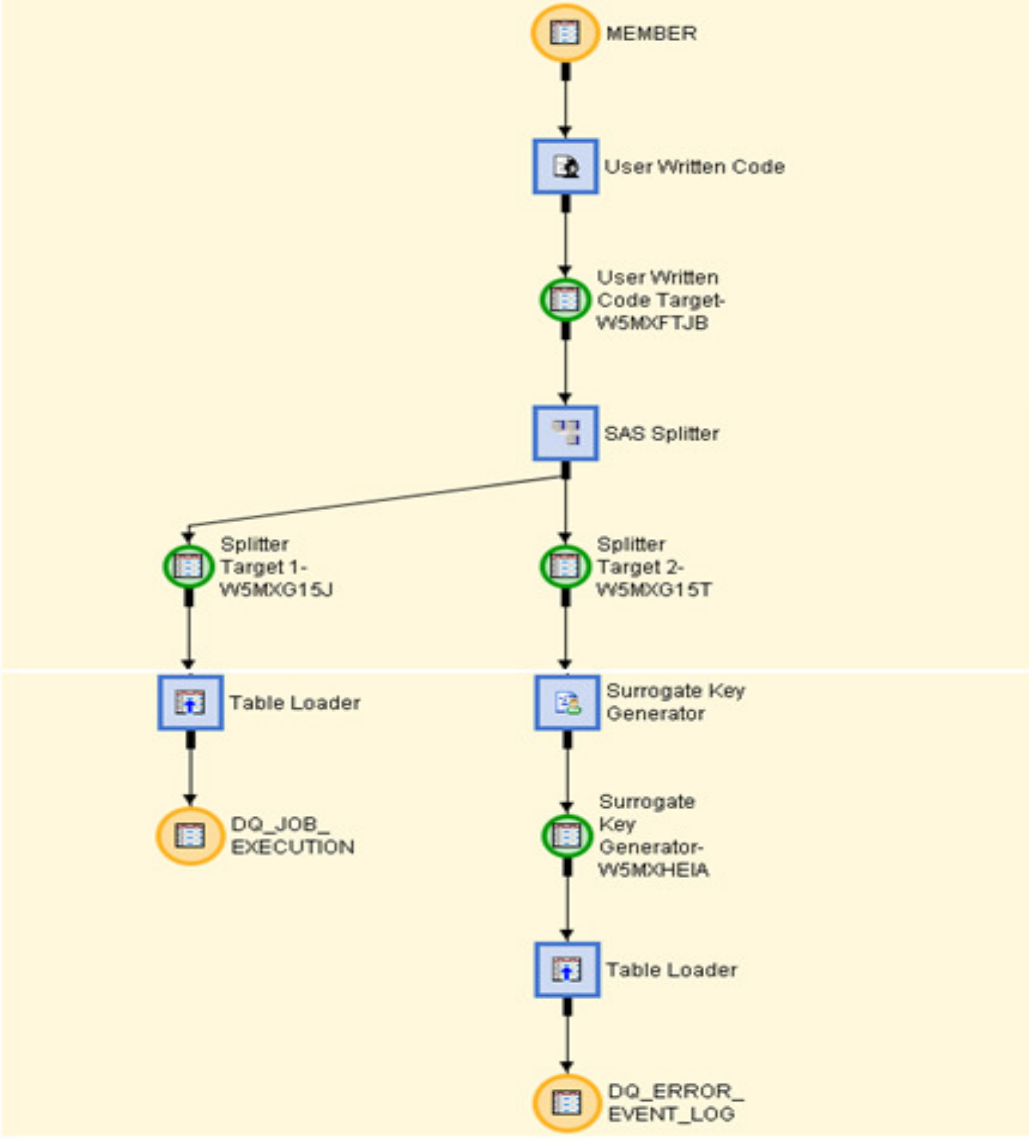

Figure 6: Sample Data Quality ETL Job

Data warehouse access

Base SAS, Cognos, SQL and Microsoft Access are available for end users as tools to access the CDW and CDM. For security reasons, we only allow highly skilled "power" users to access the CDW through either base SAS or SQL tools. According to users' different analytical needs, we allow general users access to data through specific data marts. By granting analysts access data through marts, we ensure consistent and standard data analysis. It is important that we provide flexibility in the tools available to access the marts because of the varied skill sets and experience of our end users. Similar to the CDW and CDM, our data marts constructed for specific business purposes can also be accessed using Base SAS, Cognos, SQL or Microsoft Access. In addition, Cognos frameworks are built on top of data marts and report writers can easily create analytical reports and construct dashboards and scorecards in the Cognos reporting environment.

Case study

Under our current business intelligence platform, analysts benefit by consistently obtaining accurate and timely data when and where it is needed. This saves analysts programming time and effort. For example, if an analyst wants to know if a claim submitted for a member's inpatient stay occurred while the patient was eligible for insurance coverage, s/he previously had to access both the claims detail table and the eligibility table to apply complex aggregation logic and convert multiple records into one row. These two aggregated tables then had to be matched to identify whether the inpatient stay was covered for payment. Because both the claim detail table and the eligibility table are tables containing transaction details, one inpatient stay could have several claims associated with it in the claim detail table, while a member could be eligible for coverage during multiple spans of time which appear as multiple transaction lines in the eligibility table. By using the eligibility and service event marts we created, an analyst can simply join two tables to get the correct results for this example because those marts transform the transactional records into streamlined aggregate records that allow for more precision and ease of use. This new functionality is one example of how our business intelligence platform substantially decreases programming errors and increases productivity and accuracy.

Conclusions, future enhancements and lessons learned

Community Care needed to increase the sophistication of its data management as well as analysis and reporting to accommodate its rapid growth, to support organizational improvement, and to continue its mission of improving the health and well-being of the community through the delivery of clinically effective, cost-efficient, and accessible behavioral health services. To accomplish this, Community Care created an enterprise data warehouse that integrated large volumes of data from a wide variety of divergent sources. Community Care approaches data warehousing as an ongoing process. As such, the intent is to constantly reshape and enhance our data warehouse environment to meet the changing informational needs of the company.

Community Care learned that data warehousing success can be greatly influenced by the choice of the tool used for its creation and maintenance. The decision to implement SAS strategies in building our warehouse had a direct impact upon its success. We have benefited greatly from our consulting engagement with SAS. By partnering with SAS, we were able to apply SAS best practices in our data warehouse work, thus reducing the "trial and error" that we might have otherwise endured.

Consequently, by creating a data warehouse that integrated large volumes of data from a variety of divergent sources, Community Care has positioned itself to provide accurate and sophisticated analyses with much less difficulty and in less time. This ability will drive organizational improvement, support continued organizational growth, and will provide a higher quality of information and services to our management, our oversight agencies, our providers, and most importantly, to our members and their families.

REFERENCES

SAS Institute Inc. 2007. SAS Data Integration Studio 3.4: User's Guide. Cary, NC 27513: SAS Institute Inc.

SAS Institute Inc. 2006. Using the SAS Data Quality Solution to Cleanse Your Data Course Notes. Cary, NC 27513: SAS Institute Inc.

SAS Institute Inc. 2007. The New Data Integration Landscape, White Paper. Cary, NC 27513: SAS Institute Inc.

ACKNOWLEDGEMENTS

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the authors at:

Name: Mary Doyle, Associate Chief Information Officer
Enterprise: Community Care Behavioral Health Organization
Address: Suite 700, Chatham One Center
City, State ZIP: Pittsburgh, PA 15241
Work Phone: 412-454-2156
Fax: 412-454-2177
E-mail: doylems@ccbh.com
Web: www.ccbh.com

Name: Qingfeng Liang, Lead Programmer
Enterprise: Community Care Behavioral Health Organization
Address: Suite 700, Chatham One Center
City, State ZIP: Pittsburgh, PA 15241
Work Phone: 412-402-8748
Fax: 412-454-2177
E-mail: liangq@ccbh.com
Web: www.ccbh.com

Name: Gretchen Ciocco, Operations Coordinator
Enterprise: Community Care Behavioral Health Organization
Address: Suite 700, Chatham One Center
City, State ZIP: Pittsburgh, PA 15241
Work Phone: 412-454-7155
Fax: 412-454-2177
E-mail: denkgm@ccbh.com
Web: www.ccbh.com

Name: Michele Mesiano, Director, Decision Support & Application Services
Enterprise: Community Care Behavioral Health Organization
Address: Suite 700, Chatham One Center
City, State ZIP: Pittsburgh, PA 15241
Work Phone: 412-402-7519
Fax: 412-454-2177
E-mail: mesianomm@ccbh.com
Web: www.ccbh.com

Name: Shouqiang Wang, Manager, Data Warehousing
Enterprise: Community Care Behavioral Health Organization
Address: Suite 700, Chatham One Center
City, State ZIP: Pittsburgh, PA 15241
Work Phone: 412-402-7522
Fax: 412-454-2177
E-mail: wangs2@ccbh.com
Web: www.ccbh.com