Final Project - BI/DW Solution

Proposal for CareMore Claims Data Mart and BI Solution

Present to:

Dr. Turel

ISDS 556 (Spring 2016)

May 8th, 2016

By:

Ameya Agavekar

Amogh Newgi

Kanda Ann Chaiyakit

Matthew Dawood

Saurin Mehta

Table of Contents

I. Overview

This project will focus on improving upon and creating new analytical capabilities for the Claims department at CareMore, a Healthcare company, so that management will have quicker access to critical data in order to make more timely and effective business decisions. CareMore is a company that currently utilizes business analytics in the areas of strategic planning and resource allocation. In order to compete in healthcare marketplace, it is crucial that CareMore looks into a solutions that minimize the time it takes for data to make it to the decision makers (goal is real time) in order for decisions to be made quicker and improve business efficiency.

II. Description of the Organization and Need for BI

CareMore is a health care delivery system and health plan headquartered in California. The company operates in eight states and has annual revenues of $1.2 Billion serving more than 150,000 patients. CareMore incorporates both payment and delivery of healthcare. Company mainly focuses on management of protracted disease, frail and high-risk patients. It has been regarded by United States health policy experts as one of the most pioneering models to reform Medicare. Company makes decisions and innovations based on medical evidence data.

CareMore requires Business Intelligence (BI) reporting and analytics in order for management to be able to make strategic decisions on where to allocate resources. Claims department currently lacks BI dashboards that are informative and easy to access. Mostly information is available in tabular reports from BI Center application sourcing data from operational data stores. Substantial amount of labor hours are required just to perform analyses, summarize and present recommendations to management, and other tasks that are mostly repetitive in nature.

Data visualization and interactive dashboards will uncover surprising patterns and observations that would not be obvious from looking at operational reports alone. For example, upon looking at dynamic claims dashboard and visualizing the aggregate data for average cost to care for patients across city/region, management can better decide how to distribute existing resources to maximize efficiency and reduce costs as well discover any additional needs.

High level Enterprise Data Warehouse Bus Matrix

	Date	Trading Partners	Members	Provider Contract	Providers	Denial Reasons	ICD Diagnosis
Claim Adjudication	X	X	X		X	X	X

III. Description of the functionality and features of the required/desired BI system.

Dynamic Dashboard with real time data would be one of the key functionalities that is required. A large amount of data used in decision making is currently hosted on disparate databases that are complex in nature and not user friendly. This does not provide non-technical decision makers the best presentation of the data in order to aid in their decision making. An ideal goal is an interactive dashboard with clear visualization of the data. The primary end users of this data will be front line managers and analysts. Currently, the department lacks the proper insight into how claims are processed, what the weekly turnaround time is, the average cost to care for each specialty, etc. The data is complex thereby requiring that it be compiled by individuals with a strong technical background in order to be presented to management. A goal would be to minimize this dependency on highly technical staff, and have it so that data can easily be accessed and analyzed by business team members.

The key performance indicators (KPIs) being considered are:

1) Number of claims received and processed
2) Number of claims adjudicated by Claim Status
3) Average turnaround period
4) Number of claims paid vs. capitated.
5) Cost of care per member per month (PMPM)

The proposal is to create a claims data mart with facts tables on the defined KPI's along with dimensions table mentioned in the above bus matrix. Data should be refreshed once nightly using Microsoft SQL Server Integration Services ETL package. As for the frontend, an interactive claims

dashboard should be developed in Tableau which will source data from the backend claims data mart. The claims dashboard should be interactive, user friendly and accessible online. It should be developed keeping in mind not only the super users but also the management and other end users of the application.

IV. Data Model

The dimensional data model above is a combination of a star schema and snowflake schema. Dim_Member, Dim_Address and Dim_PlanStructure are dimension tables that are of snowflake schema whereas the remaining tables are a star schema. Most of dimension tables are of slow changing Type 2 dimension except dimension tables Dim_ClaimStatus, Dim_ICDCode, Dim_Date and Dim_Time as they are static in nature. Above dimensional design depicts two Fact tables, one with grain at claim header level and another with grain at claim line detail level.

V. Data Analysis:

CareMore is collecting data on various business processes, but specific to the data mart in question and the defined KPI's the following sources are deemed appropriate. The data is available from multiple sources and can be classified as internal or external:

1. **Internal (primary):** The internal data sources are primarily operational systems within the organizations that collect data on. One of the key source system is Facets claim processing application and it has data on following areas:

   o Members: Facets includes data that has been collected on all the members that have been subject to enrollment and describes them according to utilization based on the date of service.

   o Providers: It contains data about Primary Care Physicians (PCP) and Specialists including the services they offer, specialty, and their serviceable region.

   o Diagnosis & Procedures: The data from this area provides information about the different medical conditions that a member can be diagnosed for and the procedures related to the medical conditions that can be used as treatment for members.

   o Claims: The claims data source subject area contains information about claims that are received every day at a transactional level. It also includes data about the cost and time factors associated with each claim in terms of the different channels the claim is received through. Moreover, the status of the claim is also described in the data after it has been adjudicated as adjusted, finalized, void and pending. The status also provides details about whether the claim is pended, adjudicated, void, on manual hold or has been paid for. It also contains data about referring and servicing providers.

   o Product offerings: It provides data about the different products that the organization offers. When a claim is received this data can be used to check what services and benefits a member is entitled to contingent upon the type of enrollment.

2. **External (secondary):** The external data sources provide data on attributes CareMore does not have direct access to. Such data is collected from the following sources:

   - CMS: The Centers for Medicare & Medicaid Services provides data about coverage determinations, policy related documents and other related guidance documents.

- NPD (National Provider Directory): This data is provided by a government body about the various providers that are registered that a member can have access to. Each provider is uniquely identified by the NPI code which describes the region they operate in.
- USPS (usps - zip code, county, state): Data in regards to geography is collected from USPS. This data includes all the zip codes and the associated geographic areas in terms of the city, county and state they belong to.

Unlike the external source systems, the internal source systems are not subject to much cleansing as the data is already collected within the organization in an operational system that is updated every day based on transactions that occur. However, both sources will need to be formatted in identical fashion to make sure the data being loaded into the data mart has consistent formatting. Data fields from different sources need to be defined to have the same data type across the system. Moreover, acronyms from the external sources will have to be redefined to be consistent with the defined internal operational sources. In addition, availability and the time lag on the data varies for each sources. For instance Member Enrollment data from CMS is updated daily whereas update on product offering is made once every six months.

VI. Extraction, Transformation, and Loading (ETL) Issues

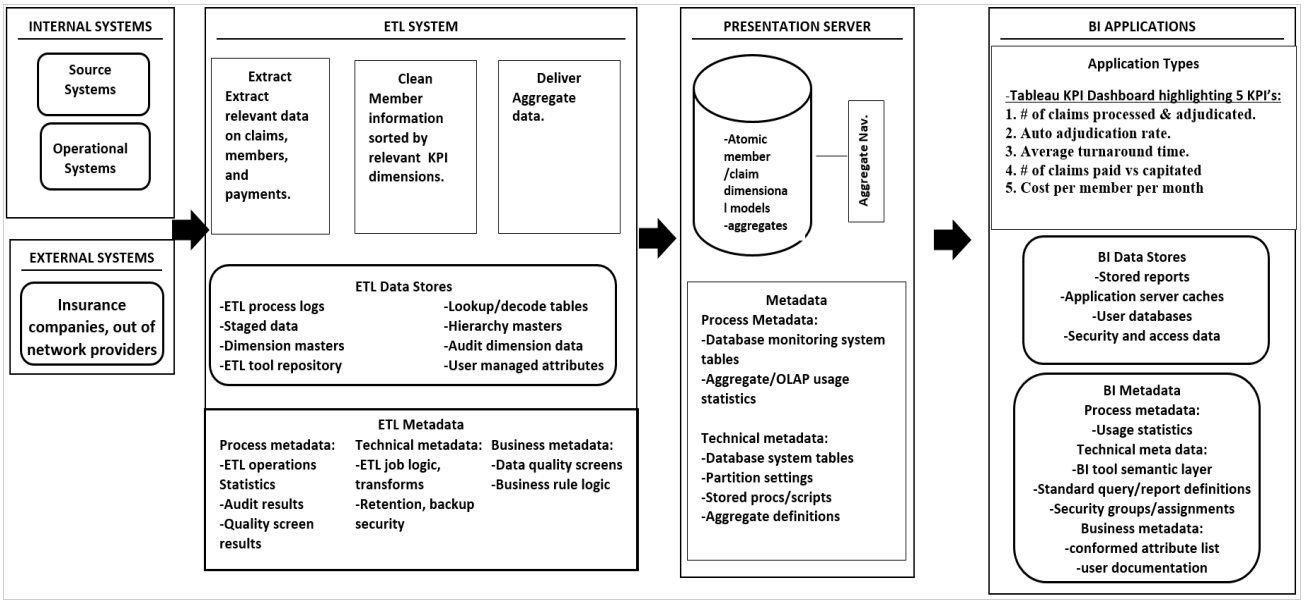
The Extract, Transform, Load (ETL) process is the most important and time-consuming component in the data warehouse. To build efficient business intelligence applications and avoid "garbage in, garbage out" situations, one must ensure the ETL is capable of pulling out the data from the source systems and delivering it to the data warehouse. A number of issues may arise that need to be addressed when developing the ETL package for the claims data mart.

1. Data Formatting: The claims data normally comes from disparate sources and contains data in different formats leading to inconsistency. ETL should be compatible with all data formats and resolve such inconsistency. This problem should be dealt with in the cleansing process before tables are joined to avoid consequences. A validity Check is also needed for some types of data such as addresses and phone numbers which are uploaded into the system in different format. An integration process must be defined to avoid data duplication when loaded into the same table. For example, the claims are submitted by healthcare providers through different trading

partners. Integration is required to consolidate multiple sources of data into a singular multi-dimensional table.

2. Data Quality: A challenge to be dealt with during the ETL process is maintaining the quality of data as non-discrete data can lead to inaccurate reports. Moreover, the ETL system architects must ensure that there is appropriate data mapping and language processing to convert raw data into the meaningful information.
3. Changing Dimensions: The ability to handle slowly changing dimension (SCD) is an important element in the ETL architecture. Most of the dimensions considered for claims data mart are SCDs. The recommendation is to use the famous SCD technique called Type 2. In this methodology all history of dimension changes is kept in the database. With Type 2 SCD, new dimensional rows are added to identify lifespan of a record and whether a record is current. With this practice, the old value can be seen even though a new dimension record is created.
4. Scheduling: Data in a Data Warehouse is not real-time mainly due to lag in data refresh and can create a conflict with BI application that are looking to provide real-time solutions. If the users need latest data, Business Intelligence application should consider pulling data directly from source systems. However, scheduled extract and upload services can be performed with predefined frequencies depending on types of data and data sources. To avoid time conflicts with users during office hours, loading/extracting processes should be scheduled to be done overnight or early mornings.

VII. The Recommended Architecture



Back Room:

The data provided comes from both, internal systems (e.g. operational systems) and external systems (e.g. insurance companies), and needs to be cleaned in order to be useful. The tool used to Extract, Transform and Load (ETL) the data is Microsoft SQL Server Integration Services (SSIS). This system serves as a foundation for enterprise level data manipulation and can work with various types of data including data from relational data sources and flat files. After being cleansed, the data is migrated to the presentation server. The presentation server technology is Microsoft SQL Server. The Data Warehouse including Claims data mart will reside in the presentation server.

Front Room:

A critical element to effective decision making are the Business Intelligence (BI) applications that business end users utilize to analyze the data after it has been extracted, transformed, and loaded from the various data sources into the data warehouse. In the case of CareMore, a healthcare company that competes on analytics, it is crucial that business analysts, managers, and other decision makers have access to an Business Intelligence system designed to provide real time data in order to be able to obtain answers to questions in real time, and be able to make more timely

decisions. The front room BI solution to consider is Tableau, a data visualization/business intelligence software solution. Tableau allows for instantaneous insight by transforming data into visually appealing, interactive visualizations within dashboards.

BI Application:

Team proposes to develop a Claims KPI dashboard. The key components that allow a dashboard to be leveraged as an effective decision making tool are the metrics obtained and the visualization tools used to display the key metrics. The five key performance indicators (KPI's)/metrics that should be displayed interactively are:

1. The measure of the number of claims broken down by number of claims processed (i.e. completed) and ones that are being adjudicated (i.e. in the queue to be determined). This is a numeric value that can be visually represented.
2. The rate of automatic adjudication (i.e. streamlined/automatically processed) sorted by submission type. This rate is a percentage.
3. The measure of average turnaround time for the claims. Grain for this measure would be in minutes. It would help management understand what factors are delaying the processes.
4. The measure of the number of claims paid sorted by those that are paid and those that are capitated (i.e. fixed payments at regular interval made to medical provider by a managed care organization for an enrolled patient). This is a numeric value assigned to the two categories.
5. The cost of care per member per month (PMPM). This is a dollar value.

The Claims KPI dashboard will be made up of the five components above with the primary component of the dashboard being KPI #5, the cost of care per member per month (PMPM). This will be visually represented as a dashboard gauge with three zones: green, yellow, and red zone (from left to right) with the needle on the dashboard indicating the actual average dollar cost PMPM. The goal would be to minimize this cost (i.e. keep it to the left green zone). The break off points for transition from green to yellow and yellow to red will be determined by the executive team based upon average healthcare market trends for average cost PMPM. KPI #1, the number of claims, will be represented on an area chart with the number of claims processed being one color and the number of claims being adjudicated being a different color. The horizontal access will be represent time at a low level grain, in hours. KPI #2, the automatic adjudication rate will

be visually represented by stacked bars where the horizontal access will identify the submission type and the vertical stacked bars will have two colors, one for the automatically adjudicated and the other for those that are not automatically adjudicated. This will give us a visual comparison of the automatic adjudication rate in addition to what the actual percentage is. KPI #3, the average turnaround time of claims will be represented on the dashboard by a days/hours/minutes breakdown. The goal is to minimize this number. KPI #4, number of claims, will be visually presented on an area chart with the horizontal access being time (days) and the two dimensions will be those that are paid and those that are capitated.

VIII. Preliminary Budget Plan

CareMore has an existing data warehouse and resources that can be leveraged to implement the claims mart project. To understand the overall cost to implement the proposed claims data mart, all aspects of the project, including the labor, hardware, and software cost are considered in the preliminary budget plan.

A special project team will need to be formed that will consist of a Project Manager, Business Analyst, ETL Developer, BI Developer, Quality Assurance Tester, and Production Deployment Support personnel. The claims data mart that will be developed will be integrated with different business analytics tools used to deliver fast and dynamic reporting. This project will utilize existing Microsoft SQL Server Integration Service and Quality Center QA environment. Management should consider procuring a separate database server and Tableau data visualization tool for this project.

To implement the data warehousing and business intelligence systems, CareMore must consider one-time costs as well as the annual operating costs. The one-time cost consists of the cost to purchase the hardware to run the server and the labor costs for the project team. The ongoing annual costs will consist of annual maintenance fees for different software tools such as Microsoft SQL Server, Tableau, Visual Studio, ETL and OLAP packages. The estimated breakdown for one-time costs are as below:

Hardware and Software Cost

Item	Cost
Database Server & Storage	$40,000
Tableau	$20,000
Total	$60,000

Resource Plan (in Hours)

(Labor-Hours)	Business Analyst	PM	ETL Developer	BI Developer	QA Tester	Prod Support	Total Hours
Requ Analysis	120	40	0	0	0	0	**160**
Design	8	8	160	40	0	0	**216**
Development	16	60	560	220	0	0	**1,356**
Unit Testing	4	4	110	120	80	0	**368**
UAT	40	40	40	80	40	0	**320**
Prod Deployment	8	20	0	0	8	40	**76**
Post ProdSupport	20	20	0	40	0	40	**180**
Total Hours	***216***	***192***	***870***	***500***	***128***	***80***	***1,986***

The estimated labor hours required to productionalize claims data mart is 1,986 hours. Considering labor cost at $75/hr, the estimated project labor cost is $148,950.

The return on investment is believed to be much higher than the project cost. The new claims data mart will help reduce labor cost and improve operational performance. First, senior managers and executive officers could make strategic and operational decision more timely and efficiently based

on dynamic reports generated by the new system. Initial estimate is that it will add to claims department ops gains of $115,000 per year. Second, it will also increase labor hours required to perform analysis, summarize and present recommendation to management. With implementing claims dashboard and operational reports, claims department can consider reducing workforce equivalent to two full-time employees leading to additional $210,000.

The estimated total investment for the project is estimated to be $208,950. The total savings are expected to be $325,000 per year. The net savings after implementing the claims data mart is believed to be $116,050.

IX. Conclusion and Recommendation

CareMore is a Medicare company that is involved in providing better health care to frail and high-risk patients. The company is actively involved in collecting data and uses analytics to improve services rendered. However, one area the organization would like to gain better insight into is the claims data. Claims data refers to the requests that are received from healthcare providers to make payouts based on members utilizing their product services. This data is crucial because it can save the company a significant amount of time and money.

The project is to design the Claim Star Schema Data Mart for the non-technical users easily generate visualizing reports and use those reports to improve operational efficiency and reduce costs. Multidimensional analysis of relevant data in real time will provide CareMore with more of a competitive advantage.

Building a claims data mart would allow CareMore to analyze aspects such as average turnaround time while processing claims data, number of users registered and the specific services they are authorized to use and the costs experienced with providing care per patient.

Upon implementation of the BI system, the KPIs dashboard created from the data mart provides the better understanding of the KPIs and the possible solutions to improve KPIs. Using Tableau in the BI solution, non-technical decision makers will have access to visualizations that will provide insights that are easier to interpret.

Data modeling is created based on business and information system objective. The data model serves faster and accessible to business user while reduce workload on report generation for IT users. With the well-design ETL system, it ensure the data extracted and loaded into data warehouse timely to support the need of BI application. Data architecture provides the better picture of the Data Warehouse data flow. This would be useful for the future planning and for a non-tech users training.

According to the claims data mart, the front end users can easily access the claims data without having to rely upon highly technical users every time they want to perform an ad-hoc analysis. As a result, this self-service report generation delimits frontend users' creativity to come up with any kind of new inspirational work.

Successful implementation of the claims data mart would provide non technical users access to information without having to rely on technical know-how of the data model. This will save considerable time and money when users such as managers and analysts look to gain access to information quickly on the go thus improving the efficiency of operations.

Clearly, a goal of a business is to be profitable and ensure projects will increase their bottom line. This project guarantees CareMore will break even after implementing the data mart in the short run, and increase their profit margins the long-run.