

Fall 2010

A Data Warehouse Solution Emphasizing the Use in Geographic Information Systems

Catherine Carpenter
Regis University

Follow this and additional works at: <https://epublications.regis.edu/theses>



Part of the [Computer Sciences Commons](#)

Recommended Citation

Carpenter, Catherine, "A Data Warehouse Solution Emphasizing the Use in Geographic Information Systems" (2010). *All Regis University Theses*. 336.

<https://epublications.regis.edu/theses/336>

This Thesis - Open Access is brought to you for free and open access by ePublications at Regis University. It has been accepted for inclusion in All Regis University Theses by an authorized administrator of ePublications at Regis University. For more information, please contact epublications@regis.edu.

Regis University
College for Professional Studies Graduate Programs
Final Project/Thesis

Disclaimer

Use of the materials available in the Regis University Thesis Collection ("Collection") is limited and restricted to those users who agree to comply with the following terms of use. Regis University reserves the right to deny access to the Collection to any person who violates these terms of use or who seeks to or does alter, avoid or supersede the functional conditions, restrictions and limitations of the Collection.

The site may be used only for lawful purposes. The user is solely responsible for knowing and adhering to any and all applicable laws, rules, and regulations relating or pertaining to use of the Collection.

All content in this Collection is owned by and subject to the exclusive control of Regis University and the authors of the materials. It is available only for research purposes and may not be used in violation of copyright laws or for unlawful purposes. The materials may not be downloaded in whole or in part without permission of the copyright holder or as otherwise authorized in the "fair use" standards of the U.S. copyright laws and regulations.

Abstract

This paper will explore use of a data warehouse solution for an electric utility that generates and transmits power across four US states. The pseudonym “Utility “X”” will be used to represent the electric utility for security reasons. At Utility “X” several systems store information that is managed in silos that, in some cases, operate on proprietary database formats. This type of usage is identified as a barrier in the development of an enterprise-wide data source. A data warehouse can ultimately help to bridge the Utility “X” information gap, and provide added accessibility for the company to further develop technology, such as Geographic Information Systems (GIS).

The use of a data warehouse can create an environment that increases the usability and accessibility of data. This project will focus on a method of using a data hub; data warehouse; and extract, transform, and load (ETL) processes to compile data from the proprietary equipment maintenance database, the meter system database, and the plant information database to provide one cohesive dataset in a data warehouse, which would ensure easier access to sets of processed electric meter data. The combination of these three sources offers advantages to other departments within the company that need meter information.

A data warehouse does not require alteration of source systems that would cause deviation from their primary functions. For example, when a data warehouse solution is not used, there is pressure for existing databases to conform to new uses, such as GIS applications. This adaptation can include costly new system changes, database changes, and interface changes. However, a system that is set up for one purpose might also contain data that is useful for another purpose. By leveraging a data warehouse solution and developing a new use of the existing data, the company can avoid changing the current system configuration.

This research will define an approach to building a data warehouse based on qualitative action research methods. Research was conducted at the company by using an iterative interview process wherein information from interviews was gathered and classified based on information from primary sources on the topic of data warehousing. An example using electric meter data from the company was identified that showed several solutions to problems by use of a data warehouse environment.

Table of Contents

Title Page	1
Abstract	3
List of Figures and Tables	7
Chapter 1 – Introduction and Background	8
Problems and Objectives	11
Significance	12
Chapter 2 – Review of the Literature and Research	13
Supporting Research	13
Data Warehouse Method Comparison	15
SOA	16
ETL Basics	18
Utility System Basics	21
Chapter 3 – Methodology	23
Action Research	23
The Process	24
Chapter 4 – Results	26
Interview Results	26
Data Warehouse Method	28
The Current Environment	29
Plant Information	29
Meter Database System	30
Equipment Maintenance Database	33

Chapter 5 – Discussion	35
SOA	35
Data Warehouse Solution	35
ETL with Plant Information and Meter Data	36
ETL for Transmission Planning	36
ETL for Billing	38
Creating and Accessible Plant Information Database	40
Creating an Accessible Equipment Maintenance Database	44
Creating an Accessible Meter System Database	48
GIS Component	52
Conclusion	55
References	57
Appendix A – Consent Form	60
Appendix B – Questions for Data Warehouse Project	61
Appendix C – Sales Report	62

List of Figures

Table 1	26
Figure 1: The Existing Data Transfer Between the Meter System Database and the Billing System.	32
Figure 2: Current Configuration.	34
Figure 3: Meter ODS.	38
Figure 4: Rolling Summary Data Example.	42
Figure 5: Linking Data Warehouse to GIS.	44
Figure 6: Meter Data Hub.	47
Figure 7: The Proposed Connection Between the Meter Data and Other Systems.	51
Figure 8: GIS Access to Summarized Meter Data in the Data Warehouse.	54
Figure 9: Proposed Configuration.	56

Chapter 1 – Introduction and Background

This research explored the use of a data warehouse solution for an electric utility that generates and transmits power across four US states. At Utility “X” several systems store information that are managed in silos that are largely inaccessible to the rest of the company. Data silos are isolated sources of information that are not widely accessible for more than one purpose. Frequently task-based systems are created for a single use and do not account for wide use of the information in the database. The use of silo data in an organization is a barrier in the development of enterprise-wide information sources. Without a more organized infrastructure in the organization, information is not available in a reasonable amount of time. It can take much more time to assemble information for upper management to make decisions about the budget or the utility system in general. Electric system decisions or budget considerations based on old information can create an environment where the company is unable to make accurate decisions. Upper management requested more timely delivery of information in order to improve critical business decisions, which is ultimately a call for better data management. Increased availability of timely information is a basis for better decision making.

At Utility “X”, meter data contains information about what is happening in the electrical system. Timely information delivery to more people in the company can increase the accuracy to what is occurring in the electrical system for the company to make better decisions. Meter data is a basic example of information that is stored in a silo environment and contains valuable information, and sometimes critical information that the company can use to make timely decisions. Although the meter information is not real-time, the data that is made available is not

assembled in a way that makes it immediately useful for decision making. This research explored several problems surrounding the specific use of meter data in the company.

The equipment maintenance system keeps track of substation equipment, which also includes information about meter equipment. Meter data is stored in both the plant information database and the equipment maintenance database introducing a problem with data that is not cohesive. Having meter data in several places reduces the availability of a basic source of information.

The sole purpose of equipment maintenance system and database is to keep track of the equipment in the substations for the purposes of scheduling regular maintenance and testing. This system provides information for the company to keep the equipment running in the substations in order to avoid equipment failure. What is valuable about this system that is underutilized, is the actual equipment inventory that the database houses. This data provides information about the power system that can be used for other purposes such as budgeting upgrades, and identifying system limitations. Making this data available to other systems for other uses would increase its value to the organization. However, there are problems that need to be solved before this can occur.

One problem identified with the current system is that it relies on a vendor to provide information on an as-needed basis about what data is stored where in the schema. Utility "X" does not have any documentation of the schema, so Utility "X" relies on the vendor to maintain the information in the equipment maintenance system. This dependence on an outside vendor creates a disadvantage for the company because the vendor has to be involved if any major changes or additions to the database are required, which increases the overall cost of the system. Changes would need to occur to make the data more available to other systems. Whenever there

are changes needed to the database, the vendor needs to be involved in those changes which increases the time and cost of the system. The company has the expertise in-house to make those changes and it would take less time. Utility “X” has the capacity to maintain and utilize their own equipment inventory without relying on an outside vendor if there was a way to take ownership of the database schema. As a company, taking ownership of the schema would provide control so that the data is available for other uses.

Another source of meter data is information that comes from meters in the substations that is stored in the meter system database. This is a third source of meter data accessed using a separate system. This system stores peak load among all meter data that is used by several departments. The data is accessed by using a Microsoft Excel plug-in that requires the user to build a spreadsheet of the information to use. This is an inconvenient way to access information that has the potential to be used in a form that is pre-assembled. To solve this problem, the utility needs to eliminate the task of building a spreadsheet to access the information.

The plant information data is stored in the supervisory control and data acquisition (SCADA) system. The plant data also contains information that has the potential to be used in conjunction with the same data found in the meter system database. The problem with this is that similar information is stored in two different systems and is used for separate purposes. The Utility “X” Transmission Planning Department has a need for peak load information available in both databases. This department does not have a good way to access and use both sources of information easily as one set of data. An improvement in the system would present that data in an environment where the user would not need to go to two different databases to obtain a complete set of data.

GIS (Geographic Information System) data is any data that has a spatial component. In this case, some data is stored in a GIS format natively. Other data will always exist in other systems at Utility “X” where the primary function will not include a spatial component. This presents a problem where data exists in systems that do not already include a spatial component. Some of these systems contain data that has the potential for map display if it included a link to the GIS system. An improvement to the system would provide a database link to the GIS system. The GIS system ultimately relies on linkages to other non-spatial databases to operate at its full potential. GIS sits on top of all the other databases. In order for GIS to be successful, all the other databases need to be working together in a way that is conducive to its use. This research focused on resolving the issues with the non-spatial databases so that those can have potential use for GIS purposes.

Problem and Objectives

The problem facing the GIS Department within Utility “X” is that company data is stored in isolated databases. Without easily accessible data, the GIS cannot deliver spatial information. Once the database accessibility issues are resolved against the equipment maintenance system, the meter data management system, and the plant information system, the GIS can be leveraged as the vehicle for information dissemination so that upper management and the company as a whole can have timely and appropriately organized information to make the best possible decisions. There have been examples where management was given summarized data that was six months old when newer information was available. Decisions that are made by management rely on information that is up to date in order to make an informed decision.

Data sharing across multiple departments for multiple uses can facilitate improvements in information usage. The meter data could be used by other departments within the company for other uses such as transmission planning. The Transmission Planning Department needs to have access to sources of meter data from both the plant information database and the meter system database. If both databases were combined, it would create a picture of the whole transmission system by compiling records for all the meters in the system. Combining these sources would allow the Transmission Planning Department to have easier access to all the data they need.

This research study defined a possible evolution of Utility “X”’s database environment based on the company’s identified need to improve the access and management of information for varied usage across different departments.

This study sought to understand the following questions:

1. What types of problems with data management exist within the company?
2. What can be done to make meter data more accessible and useable to other departments and for other types of uses within the company?
3. How can a data warehouse benefit a GIS and other company departments?

Significance

Once Utility “X” has a good database foundation, doors will open for new technology such as GIS to emerge. The motivation behind this research project was to have an outcome that not only provides good quality data for the GIS that will fuel its progression in the company; but also provides a framework for other companies to grow and develop their database resources.

Without proper enterprise-wide data management, access, and maintenance, GIS capabilities are limited. Not having true ownership of company information and its management keeps

organizations from progressing with new technology. Proprietary databases that are isolated and managed by consultants prevent organizations from having easy and quick information access.

There are many data management solutions such as data warehousing that enable organizations to improve data management, expand data usage and improve decision making ability. Providing input by means such as this research, can help to define the types and availability of data that can be assembled for use in the GIS.

Chapter 2 – Review of the Literature and Research

Supporting Research

Research is used as the basis for answering the questions about the current environment. It is the way to identify how others have answered these same questions. The following section contains the research examples for this project that were used to identify possible solutions and defend them.

What is a data warehouse? The data warehouse is not just a repository of data that is accessed by several departments; it is a place to organize data for other uses. Compared to legacy systems, it can be more flexible to expand for future usage according to Inmon (2005).

“A data warehouse is a subject-oriented, integrated, nonvolatile, and time-variant collection of data in support of management’s decisions. The data warehouse contains granular corporate data. Data in the data warehouse is able to be used for many different purposes, including sitting and waiting for future requirements which are unknown today” (Inmon, 2005).

A data warehouse is a holding place where information is assembled for the user in a timely fashion. Data design projects can last for long periods of time. Systems that

rely on these designs take time to build. The usability of the information in that system is at stake when systems take a long time to migrate when the use of the information changes. A data warehouse can fill a gap of time where systems can catch up to user requirements.

The most important thing to remember about a data warehouse is that the data is integrated and reformatted, so that there is one version of information that everyone accesses rather than having different versions of the same or similar information in several places.

“Data is fed from multiple, disparate sources into the data warehouse. As the data is fed, it is converted, reformatted, resequenced, summarized, and so forth. The result is that data — once it resides in the data warehouse — has a single physical corporate image” (Inmon, 2005).

In addition to that, the data warehouse holds essentially a processed copy of the data in the legacy system or the operational database where it remains the source data.

“Data warehouse data is loaded (usually, but not always, en masse) and accessed, but it is not updated (in the general sense). Instead, when data in the data warehouse is loaded, it is loaded in a snapshot, static format. When subsequent changes occur, a new snapshot record is written. In doing so, a historical record of data is kept in the data warehouse” (Inmon, 2005).

The advantage here is that the data can be formatted and modified in the data warehouse for the new purpose identified, and it gets updated from the original source systems periodically so that the data does not age and become useless. At the same time, the original systems can continue to be used for which they were intended, without inflicting change to those.

One of the major design issues with building a data warehouse is that of granularity.

“Granularity refers to the level of detail or summarization of the units of data in the data warehouse. The more detail there is, the lower the level of granularity. The less detail there is, the higher the level of granularity. For example, a simple transaction would be at a low level of granularity. A summary of all transactions for the month would be at a high level of granularity....The granular data found in the data warehouse is the key to reusability, because it can be used by many people in different ways” (Inmon, 2005) .

There are many uses for this data in aggregate form. The data warehouse in this case would include more aggregated plant data that can be used for other purposes without having to access the proprietary system. The plant information data would be best accessed corporate-wide by developing rolling summary data (Inmon, 2005).

Data Warehouse Method Comparison

There are basically two schools of thought on data warehousing that include the Inmon Corporate Information Factory (CIF) approach and the Kimball method. There are many similarities between the methods. Both methods have the basic premise that the data warehouse should be created enterprise-wide and not in silos by department or by supporting a single purpose or function. They each use extract, transform and load (ETL) processes to migrate data into a staging area. Both have a presentation layer that is used for reporting, business intelligence and analytics (Ross, Kimball, 1994).

The Kimball method differs from the Inmon method in that it uses a bus architecture built with dimensional models based on business processes. The dimensional models are structured using a star schema. Common data across the enterprise is compiled into one model. Common,

conformed dimensions have consistent descriptive attribute names, values, and meanings. The Kimball method also differs in that the architecture at the presentation level identifies and enforces the relationships between business process metrics and descriptive attributes (Ross & Kimball, 1994). The emphasis here is on the business process rather than on departments.

The Inmon CIF method is different because it focuses on departmental data with the concept of a data mart, among other constructs. Data marts are used to incrementally build the data warehouse by focusing on data in each department and building them so that they can be utilized in the data warehouse environment with the enterprise-wide focus. Similar data may or may not exist in other departments, but an enterprise-wide dimensional model is not the goal with this method. This method is more nimble in that changes can be made modularly in the data mart. The Kimball method is more restrictive in this way because of the business process focus it requires more coordination between departments to make changes in the structure of the data. In an interview with Kalman (2004), Kimball says that he does not agree with the concept of a data mart. He says it's a technology looking for a solution, and a way to create a complex architecture to proliferate more machines (Kalman, 2004). It seems like he thinks it's a way for vendors to make money on hardware and supporting complex architecture rather than a way of simplifying the implementation process.

Service Oriented Architecture

Using service oriented architecture (SOA), enables system designs that incorporate data management, agile development of applications, integration of legacy systems, and re-use across the enterprise (Rosen, et. al, 2008).

“SOA can be defined as an architectural style promoting the concept of a business-aligned enterprise service as the fundamental unit of designing, building, and composing enterprise business solutions” (Rosen et. al, 2008).

This method of building enterprise systems helps to alleviate problems surrounding data silos by incorporating an infrastructure that supports multiple usages.

“One of the main goals of architecture is the promotion of *consistency* and reuse. Thus, one difference between software architecture and Service-Oriented Architecture is that of scope. SOA is concerned with providing consistent services throughout the entire enterprise, so that they can be used by many different families of solutions. SOA must promote the development of business capabilities in such a way that they are easily reused by the different business processes.”

(Rosen, et. al, 2008).

SOA is the framework in which data warehousing can develop and be successful. Without an enterprise-wide support structure such as SOA, developing a data warehouse can be difficult. Technology alone cannot support a data warehouse. It requires many aspects of support throughout the organization. The organization needs to have an architectural approach, a business vision, methodology, a corporate-wide vision of reuse and the staff and resources to support it (Rosen et. al, 2008).

Data warehousing in combination with SOA can help tie information systems together and break down the barriers that exist with silo databases, resulting in a better way to manage information and enable the right people to get the right information at the right time.

ETL Basics

ETL is an acronym for Extract, Transform, and Load, which is a key component of the data warehouse environment that allows the processing of data between the legacy systems and the data warehouse.

ETL processes are responsible for:

- (i) The extraction of the appropriate data from the sources.
- (ii) Their transportation to a special-purpose area of the data warehouse where they will be processed.
- (iii) The transformation of the source data and the computation of new values (and, possibly records) in order to obey the structure of the data warehouse relation to which they are targeted.
- (iv) The isolation and cleansing of problematic tuples, in order to guarantee that business rules and database constraints are respected.
- (v) The loading of the cleansed, transformed data to the appropriate relation in the warehouse, along with the refreshment of its accompanying indexes and materialized views.

(Vassiliadis, 2009)

The data in the legacy systems might have similar information across different systems. The common example used is customer data, which might be stored and maintained in several locations across the company. The ETL process, in conjunction with the data warehouse, can assemble a cohesive customer dataset that meets the needs of various departments within the company.

“The data might be extracted and loaded into a separate data warehouse for other reasons. Data in transactional systems is not always checked for validity. Multiple transactional systems can become silos of information and the data can be inconsistently represented in the different systems. So, the warehouse is often the location where data is cleansed of bad information and brought together in a consistent format with consistent meaning.” (Stackowiak et.al., 2007)

Consistent data with consistent meaning is a critical aspect of the data warehouse that adds value to the information that is stored in it. In disparate systems, data of similar type can have similar meaning, but might still be slightly different when evaluated together. Using a simple example, a set of customer data in one system might be used to track customers that are located in one portion of the sales territory that are limited to the sale of one particular set of products. Another system might track all customers throughout the entire sales territory regardless of what product they are purchasing. This information is the same information, but when evaluated together, there might be duplicate records. One customer that buys the one particular product isolated to the one portion of the sales territory might also buy other products that are tracked in the other

system. This customer would be tracked in both systems. Two systems with two different purposes contain two sets of similar data. The purpose of each database is different and the information in them has different meaning as well. The products are isolated to one type in one database, where the other database does not isolate the product type. If a new purpose for the information was identified, such as consolidation of the customer data; an ETL process could be set up with a data warehouse to accomplish the consolidation task without repurposing the first two systems.

The primary advantage of using ETL is that legacy systems can continue to be the primary source of the information and the application source in which the data is updated and the extracted data can be formatted and cleansed.

ETL processes can free the company from dependence on a certain platform or vendor-specific system. The process of changing systems can be costly and time consuming. ETL processes can help to alleviate some of this by providing a way to use the information outside of the currently used platform that might be outdated or in need of an upgrade.

'...the transformation tool should target a wide variety of ODS platforms, with the flexibility to add new ones as they become available. The transformation tool is the first line of defense against vendor lock-in within the ODS, since it can provide portability to new technologies as they become available.'" (Inmon, 1999)

An example of an Operational Data Store (ODS) would include integrating two sets of similar information in one location. As in the customer data example, a new use for customer information was identified that drove the need for a new consistent customer database. An ODS

can provide a set of consistent information needed for the new purpose or operational environment.

“The ODS is a basis for doing integrated operational processing and, in turn, it feeds the data warehouse. There is still a clear line between operational processing and informational processing. The ODS is a separate architectural entity from the data warehouse. The ODS serves the needs of the operational environment while the data warehouse serves the needs of the informational community (Inmon, 1999).”

An ODS is a subset of the data warehouse that provides processed information that is needed for a particular department or specific entity with a specific function in mind. It is a tailored database built from existing data to meet the needs of the users.

Utility Systems Basics

There are several systems that run the electrical transmission system. In order to better understand the research in this document, an introduction to the meter system is necessary.

The electrical transmission system consists of high voltage power lines that one would typically see elevated with wood structures or steel towers. Those power lines are connected by substations that hold various pieces of equipment that manage the electrical current on the lines. One of those pieces of equipment is the electric meter. Similar to the meter at a home, this meter reads the current flowing through the lines at its location. The transmission system works a little differently than the electricity that flows to a house because the power can flow two ways. The flow at any point in time is either read at a deficit that is called “demand” or calculated as a surplus which is called “load.” The difference with a transmission system meter is that it may

have more than one use, one of which is billing customers for their usage. On a transmission system meter, the demand and load over a period of time is calculated to determine whether the customer owes for the additional demand or if the customer receives a refund for the load that was added to the system at that meter location.

Several types of substation equipment, including meters, are used to monitor and control the current through the use of the supervisory control and data acquisition (SCADA) system.

“Operational data, also called supervisory control and data acquisition (SCADA) data, are instantaneous values of power system analog and status points such as volts, amps, MW, MVAR, circuit breaker status, switch position. This data is time critical and is used to monitor and control the power system (e.g., opening circuit breakers, changing tap settings, equipment failure indication, etc.) (McDonald, 2009).”

The main purpose for the SCADA system is to control the flow of power at the substation in the event of a surge or other significant power fluctuation. For billing, the reading is aggregated to some extent and is typically read at an interval of about 15 minutes. For the SCADA system the reading is every few seconds.

For the purposes of this research it is important to understand that those two different readings are stored in two different and separate systems. The meter readings used for SCADA are stored for later retrieval in the plant information database and the meter readings used for billing and other purposes are stored in the meter system database. Both of these readings contain similar information and could be used together. For example, a meter may be used only for billing or only for the SCADA system but a single meter might also serve both purposes. From the interview process and from working with meter information for use in GIS, it was

determined that for some analysis purposes it is important to evaluate all the meters on the system regardless of use. Currently there is not an easy way to do this, since the information is stored in two separate systems.

Chapter 3 – Methodology

Action Research

This research project is a case study that will contain qualitative research and data analysis through an action research approach. “Action research is an established research method used in use in the social and medical sciences since the mid-twentieth century. Toward the end of the 1990s it began growing in popularity for use in scholarly investigations of information systems” (Communications of the AIS, 1999). It combines the study of the social aspects of IT systems with methods or actions with the goal of improving systems in organizations.

Action research to some scholars is not really research, but is “a formalized method of problem solving relevant to a particular organization or setting (Swanson, 2005).” This is valuable research because it is the exercise of solving problems at an organization in a real setting using research to apply changes, and support decisions that are made. More often than not, people in organizations make uninformed decisions. Action research is a method to use to ensure informed decisions. This study focuses on defining a warehouse solution based on research in the company using an interview process with employees, investigating the various systems throughout the company and defining the current issues associated with those. Scholarly

research was used to define methods and options to use in solving the issues that were identified. Some action research projects include the process of actually completing the suggested changes to the IT systems. In this case, the identified solutions would take many years to complete and is not part of the scope of this project. This project only identifies a suggested set of solutions with the hope that the implementation process could be seriously considered as a result of this work.

The Process

I conducted a survey using a questionnaire and met with various employees in different departments in the company to document the information environment (see Appendix A.). The interview questions focused on potential data duplication and perceived difficulty of information retrieval with the goal exposing potential flaws in some of the information systems throughout the company. The interviews were successful in providing those results, which were compiled into a list of problems (see Table 1) that were then analyzed to find a suitable data warehouse example to use for this research. As a new employee to the company, the interview process helped me gain exposure to several systems throughout the organization and an understanding of the challenges that different people face with those systems.

Once the interview process was complete; I looked over the results and compiled a list of problems that were identified. To isolate more complex issues, I separated the list of problems into two categories: ones that could be solved using a data warehouse solution and those that were just basic information management problems. Initially I found that one of the most significant problems was one that introduced an information constraint from one department to another. This was the problem where Transmission Planning Department was limited access to the meter data by the tools that were offered to them. In this case, the department could only use

an Excel spreadsheet to build a query interface to the database. I started to find that other problems in the list also pointed to limitation on the access and use of meter data. The interesting part of the discovery process was finding that the proposed problem to be solved included many aspects of the data warehouse including high volume, high grain data that needed to be transformed into aggregated data and summarized reporting. This series of problems and the many facets of the data warehouse solution it presented made it a good candidate for this thesis project. Once the suitable data warehouse example was recognized, several individuals with specific knowledge of the system were interviewed to gain more information about the existing systems and data. Some of the same people were interviewed again with different questions and new people were identified and interviewed with questions more specific to the topic.

The goal of the second round of interviews was to attempt to understand what systems existed for what purpose and try and identify the flow of information from one department to another or from one use to another. The more I learned about these different systems, a good picture of how meter data was used and maintained began to form. The project started to take focus on meter data as a whole. It was at this point that I decided to keep the focus there and only the systems involved with collecting, maintaining meter data were evaluated, since this type of information seemed to have the most problems in terms of database maintenance and usage.

The interviews were free-form and did not involve preformed questions. Each person interviewed was asked who else would know more about a particular part of the system that I needed to know more about. I had several other referrals that lead to more information about each of the systems and those people were able to explain to me what data was collected and

moved from one system to another. Additional information about the existing systems was submitted for my review that were contained in proprietary internal documents. Other references to external software vendor websites were also suggested that had additional information. The results of this process are contained in the section that includes the description of the current environment for each of the systems evaluated.

Chapter 4 – Results

Interview Results

The interview process revealed a significant amount of work that would have to be done in order to fix all the problems that were identified. However, the purpose of this interview was not to fix the problems, but just to determine what they might be. The list gleaned from the interview process was compiled into a table illustrating two sets of problems as described previously (see Table 1).

Table 1

Identified Problems from Interview Process

A. Problems that could be solved with a Data Warehouse Solution	
1	Line Rating information not kept in one place.
2	Calculated system impedances of lines and transformers are kept in Planning and Protection departments. They are similar calculation but not exactly the same.
3	Resistance and reactance values are kept in Planning and in Protection. They are currently inconsistent and should match. They are working on fixing the inconsistencies now.
4	Relay settings are sourced at the Relay DB, but kept other places like the Planning DB.
5	Two telecom DB's both have similar information kept in it, but the systems have different purposes.
6	Lands department needs information from several departments. Environmental also, does, so they sometimes share information since Environmental also gleans information from different sources (Environmental capital projects folder).
7	Four levels of management (different levels of detail) keep maintenance information for telecom. This will be resolved with Equipment Maintenance DB.

8	Budget and accounting information is difficult to get because of different levels of information that is needed for different purposes. Three different locations for similar information.
9	The Equipment Maintenance DB has information that is also in the Planning DB. There are also individuals in Trans. Maintenance with spreadsheets with similar information. The Equipment Maintenance DB is the source, but there is not a good way to update these other sources with similar information.
10	SCADA data (plant information) is not made easily available to Trans. Planning for them to use in the planning models.
11	Data in the Equipment Maintenance DB and the DOBLE DB are duplicated. Test results get additionally stored in Documentum.
12	There is not a read-only version of the data in the Equipment Maintenance DB so this information is not easily accessed.
	B. Other Basic Data Management Problems
1	Employee information (phone numbers) not kept up to date in HR for telecom and for Dispatch.
2	Telecom and Environmental have problems deciding where to keep pictures.
3	Telecom site GPS coordinates are kept several places.
4	Telecom does not know when things have been built. A DB exists to enter this, but Maintenance doesn't use it properly.
5	Equipment and Equipment Wiring Drawings for telecom are out of date.
6	Paper files in different places are difficult to access and are not filed properly.
7	Inconsistent use of forms (old forms, new forms) in the Lands Dept.
8	People keep information in Microsoft Outlook that needs to be shared like access codes to telecom sites, phone numbers of outside contacts, inside employee cell phone numbers and cell phone account information etc.
9	Other utility information outside of Utility "X" that resides in schematic diagrams or other drawings are difficult to obtain or keep updated because of the competitive environment that FERC has created.
10	Several departments do their own mapping for the same project. Coordination is starting, but not ideal. Map information might be the same but from different sources.
11	Surveyors PLS CAD drawings are too difficult to decipher so structure coordinates are gathered all over again. CAD drawings need to be simplified for different purposes, which is time consuming. Models are not standardized and individual engineers use their own models.

Classifications into the two categories were determined by applying a basic overall understanding of data warehousing as a result of the research effort. Based on the research done

for this project, just over half of all of the problems identified could be solved with a data warehouse solution. The topic for this research was chosen from items 9-12 from list A in Table 1. These items were chosen because they were related in terms of the type of data, and had the most complex issues associated with them. The topic determined from the interview process will cover the application of the data warehouse more specifically using research examples. The next section covers the research supporting the topic.

Data Warehouse Method

For this project, the Inmon CIF method was evaluated for two reasons. One reason is because of its flexibility of focusing on departmental information using a data mart. The terms data hub and data mart are used interchangeably in this project, but the meaning is intended to be the same. At Utility “X”, a lot of company data resides in silos by department. Some departments do not need to, or literally cannot share information for security reasons or marketing reasons, but some do need to share information. To coordinate an enterprise-wide business process driven data warehouse would be cost prohibitive due to the time and effort involved. Building coordinated data marts one at a time based on the need to share information is a realistic option in this environment.

The second reason the Inmon CIF method was evaluated was because a lot of Utility “X” information does not follow the process framework of other companies where data falls into standard categories such as “order”, “customer”, and “widget”. This might lend itself more to the Kimball method where similar data across the enterprise is categorized and assembled in one place. At Utility “X”, there is a lot of very different data across the enterprise that does not necessarily get replicated or duplicated in other departments because the need does

not exist for this to occur. No one is going to try and duplicate meter data that is gathered at a high volume and it updated within minutes or seconds. This type of data is more conducive to remaining in the one department as an ODS and relevant meter information from there can be more easily shared can be assembled and aggregated in a data mart.

The Current Environment

Plant Information

The example for this research was chosen from the list of problems that were identified in the survey. The initial problem that was identified was that the plant information data were not easily accessible to the Transmission Planning Department. The tool the department currently has to use is an Excel plug-in that accesses the data but requires the user to build an Excel spreadsheet to use the data. There is no other tool available for novice users to access information that is stored in a proprietary format. With the Microsoft Excel plug-in, the planner needs to know which “tags” to choose for the program to summarize the data. Tags are the unique identifier for each of the meter readings in the plant information database. If the planner knows that he needs totals on a particular transmission line, then he needs to know what “tags” to choose by using the plug-in. This process forces the user to understand a cryptic unique identifier rather than just referencing something that is easily understood, such as a location name.

This process is additionally restricted by only having meters from the plant information database and it does not include any meters from the meter system database. What the department needs is aggregated data, but the tool they currently have simply provides an ad hoc way to process very granular data, a solution that is both awkward and limiting.

The plant information data, as mentioned previously, is data collected from meters in the substations that are used initially in the SCADA system and are then stored separately in the plant information historian database. The plant information historian database uses data compression to store large amounts of granular data, in this case using swinging door compression.

“...A value is not stored if the time between that value and the last stored value is less than the compression minimum time. The exception to this rule is when a value changes from good to bad status or bad to good status. A value is always stored if the time between that value and the last stored value is greater than the compression maximum time.” (OSI Software, 1991)

It's a very simple concept wherein the system only stores the data it needs based on the rate of change and in time. The value that is stored is within a particular threshold and not every second of data that is coming in is getting written to disk. There is a process in place that gleans out the unnecessary data so anyone who is using this data needs to understand that it is essentially re-sampled. A balancing act is taking place wherein this process lowers the cost of storing large amounts of data, and simultaneously weighs thresholds from the intended use of the data (Inmon, 2005). Since this system is aggregating granular data, this demonstrates that the existing system is using a data warehouse.

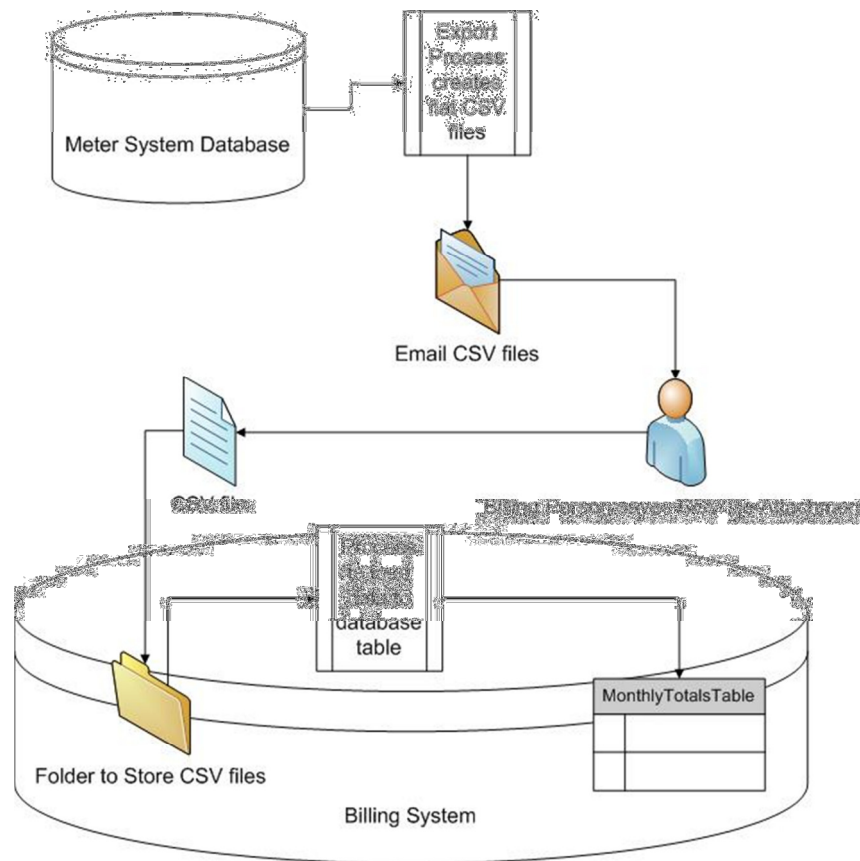
Meter Database System

With more investigation stemming from the initial identified problem, another meter system was identified. The meter database system is another separate system that also collects meter data. This system is not connected to the Plant Information system even though

the data is very similar. This data is read about every 15 minutes as opposed to the SCADA system that requires more frequent reads.

One of the main uses for the data in the meter database system is for billing but the billing system is yet another separate system. It would be logical to assume that there would be a connection between the meter database system and the billing system, but there is not. Currently the information from the meter database system is emailed to the billing department for processing and integration into the billing system.

The system emails a flat CSV file over to the Billing Department. Someone in that department takes that email attachment and saves it in a particular folder. Periodically those files are processed into the billing system. Oracle SQL Data Loader is used to process that CSV file into a table in the billing system. There is another separate table in the billing system that stores the meter identifier with the member it belongs to. Using a relationship with this table allows each of the new records coming into the system to be associated with a member to bill against (see figure 1). This process occurs once every month which limits the availability of information in the billing system to that timeframe.

Figure 1. Existing Data Transfer Between the Meter System Database and the Billing System

There is also an outward facing website that delivers meter data to members who purchase electricity based on this information. This data is also coming from the meter database system and is emailed to another entity that assembles the website in a manner similar to the data transfer workflow of the billing system.

Currently the employees who are responsible for the meter database system are working on a method to import data from the plant information system and import it into the meter database system in order to have all meter readings in one system. There are potential pitfalls

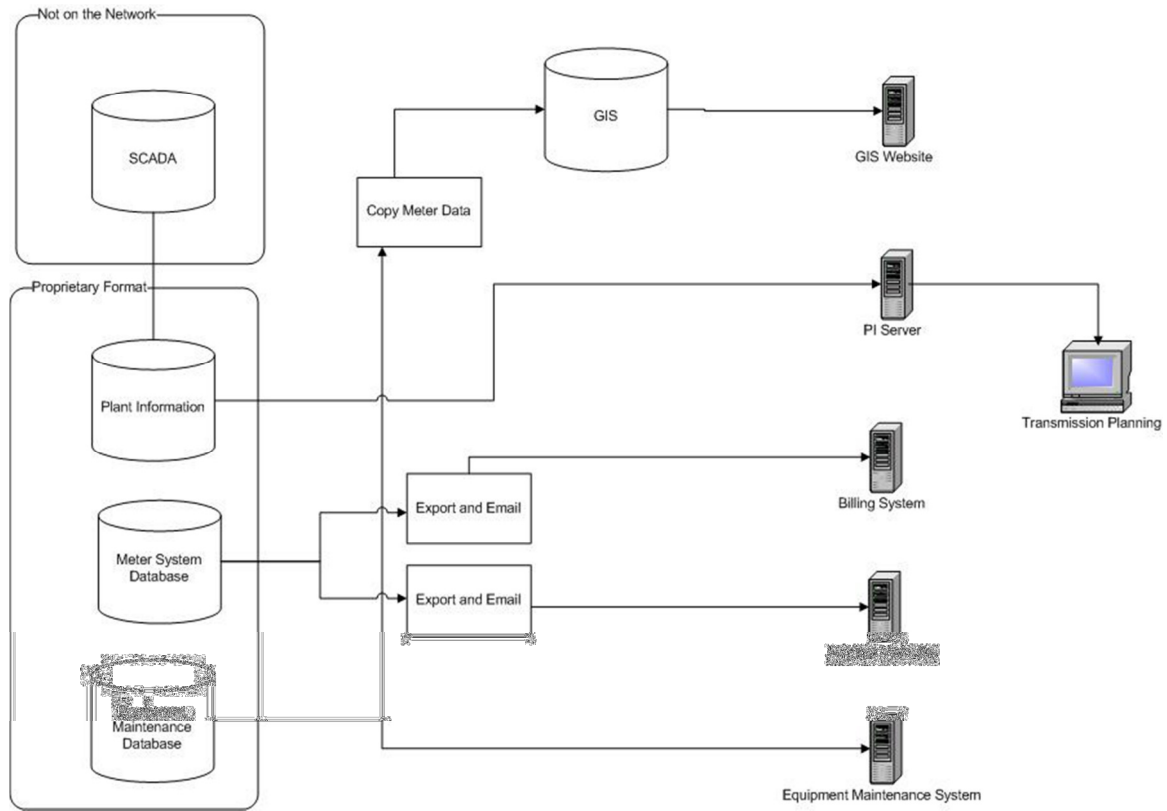
with doing this because the readings are at different intervals and might also store different values with different meanings.

Equipment Maintenance Database

As the research developed, it became apparent that the GIS meter database that was being built also tied into these systems. The equipment maintenance database is used to keep information about different types of transmission system equipment that needs regular maintenance. Like some of the other systems in Utility “X”, this database is in Oracle but it is in a proprietary format. The vendor does not publish the metadata of the schema, so there is not an easy way to be sure what type of information is in the table structure. The software is heavily tied to the database as well, creating a situation where Utility “X” relies on this vendor to keep and maintain the data and software at the same time. This situation makes it nearly impossible for other systems, such as GIS, to access the data in this database and, as a result, the meter data in this system was abandoned and recreated in GIS.

The GIS meter database was to be created from the information in the equipment maintenance database, which would subsequently be used for both maintenance and other mapping purposes in GIS. The schema was reorganized and converted into a GIS format. While the meters in the equipment maintenance database include only meters that Utility “X” is responsible for maintaining, the new meter database in GIS format includes those plus other meters that the company has other interests in.

Figure 2. Current Configuration



Chapter 5 – Discussion

SOA

Utility “X” has had a data warehouse in place, but as a company, it has lacked the enterprise-wide support and infrastructure to continue its growth. Through the interview process there is an indicated belief among those interviewed is that the information technology (IT) department along with upper management has recognized this disconnect and are taking measures to acknowledge the need for SOA to build the infrastructure to support better data management practices and flexible applications. SOA is the infrastructure needed to support a data warehouse solution. The past data warehouse effort has proven to this company that there is more involved in building and maintaining a successful data warehouse. SOA is the foundation that holds the data warehouse together through time. At a company that has been operating in much the same way for 50 years, this idea of shared information across systems is a new paradigm. With past experience with data warehousing, there is hope that a new data warehouse effort would be more successful with the right infrastructure and support systems in place to begin with.

Data Warehouse Solution

In the current environment at Utility “X”, the silo databases are not easily accessible to GIS. A large portion of information is not being fully utilized because it exists in separate, disconnected databases. This data could potentially be spatially enabled using GIS if it were placed in a data warehouse. It would not make sense financially or in terms of the time involved to migrate these systems solely for the purpose of making them accessible to other systems because each of these database systems has a valid purpose and works fine for the primary

designated purpose. As an alternative, a data warehouse can provide a mechanism to allow access to information across different databases. Data that is significant for analysis or reporting can be gleaned out of the current configuration and processed into a format that is more useable to other systems for other purposes.

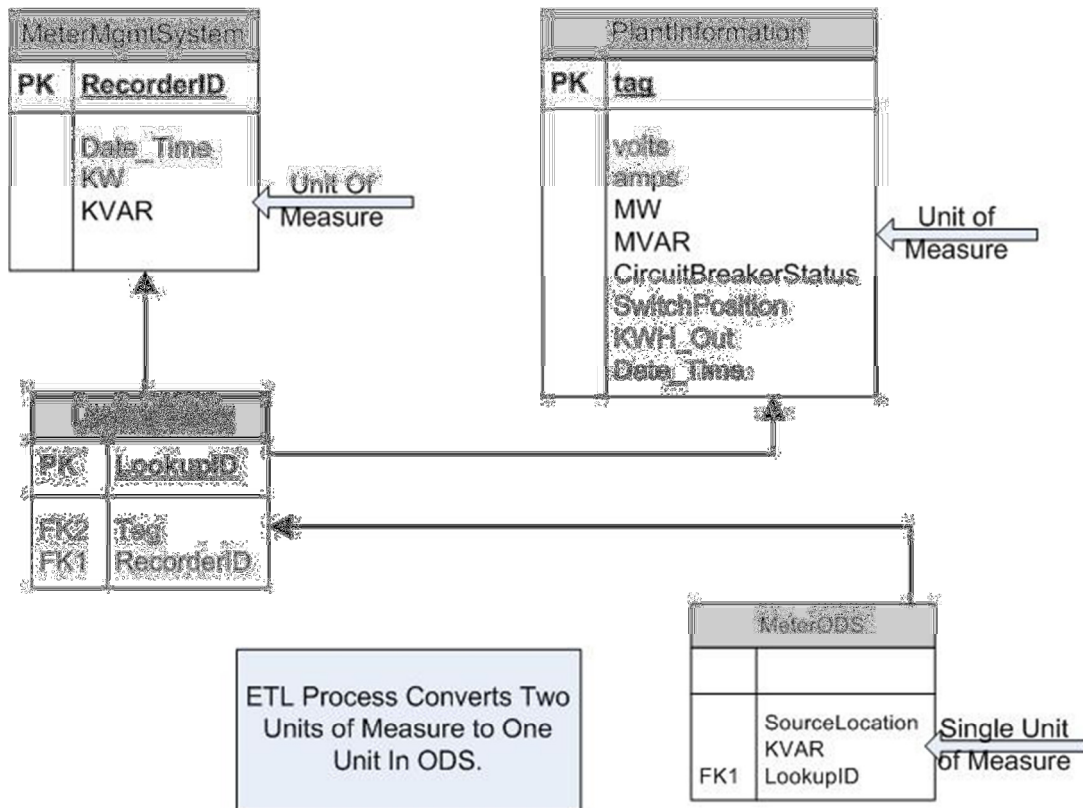
ETL with Plant Information and Meter Data

Whether in an ODS or the data warehouse, use of ETL tools can disassociate the data from a particular vendor or proprietary system by extracting what data is needed from the system, changing it into the format in which it must be used, and storing it for use in another location. ETL tools can be incorporated using a data hub, which is an Oracle concept that consolidates similar information from different systems to capture a single set of the same information “to help create an infrastructure providing a single source of truth across multiple systems (Stackowiak, 2007).” Since Utility “X” is primarily using Oracle for data warehousing, a data hub can be set up to process the granular meter data in the plant data historian and be used in the meter system database. One set of meter data can be accessed from one location that would incorporate the two slightly different sets of information into one comprehensive set of meter information. The advantage is that this data can be used for several more uses rather than just the two initially identified uses.

ETL for Transmission Planning

There are meters tracked in the SCADA system that is not tracked in the meter system. Aggregated meter data using ETL processes and storing the new processed information in an operational data store (ODS) provides a comprehensive list of all the meters in the transmission system that anyone would need. In this example, Transmission Planning needs to see all of these

meters in the transmission system to evaluate load forecasting. The department has to go to two places to get that information since both of those sets of data are not translated into a seamless dataset. By implementing an ODS, and setting up transformation processes, it is possible to conform the two datasets into one so that Transmission Planning only needs to access one location to get all of the data that is needed. The meter system database might have measurements of load stored in different units of measure than what is stored in the plant information database. The transformation process can be set up to calculate one of those values to match the other so that one value can be stored in the ODS (Inmon, 1999).

Figure 3. Meter ODS

To summarize this example, the meter data can be set up in an operational data store (ODS) by using an Oracle data hub framework. The data hub uses Extract, Transform, and Load processes (ETL) to pull data from the legacy systems and process that data into one cohesive meter dataset to store in the ODS (see figure 3).

ETL for Billing

For billing purposes, the meter readings from the meter system data could be similarly aggregated using demand and load values and store tables based on member. Another ETL process could be created to prepare billing tables based on member. Currently this process of

emailing an exported file over to the Billing Department is occurring once a month. There are several reasons why this is not the most reliable way to transfer information.

- 1) In order to email the data, it needs to be exported into a different format, such as a CSV file.
- 2) The email system must be relied upon to deliver data periodically, which adds another variable to the list of potential problems that could occur from relying on a separate system to deliver information. That variable could be eliminated by keeping the data in its original format.
- 3) The data that has been exported to a different format now needs to be imported into another system, which requires someone to interface with the e-mail to capture the file and import it into the other system. This activity could be eliminated by keeping the data in a database format.

In combination with the customer care website, the Billing Department could have continuous access to the information rather than just on a monthly basis. In this case the member might need to see total demands and loads for each month for each meter location. Or the member might need to see a total KWH (kilowatt hours) of all the meters in the member area aggregated for the month. With aggregated data stored in the data warehouse using rolling summary tables, the member can have data updated more frequently, maybe nightly or even hourly. Overall, the billing department and the customer (member) have access to more timely information (see figure 4).

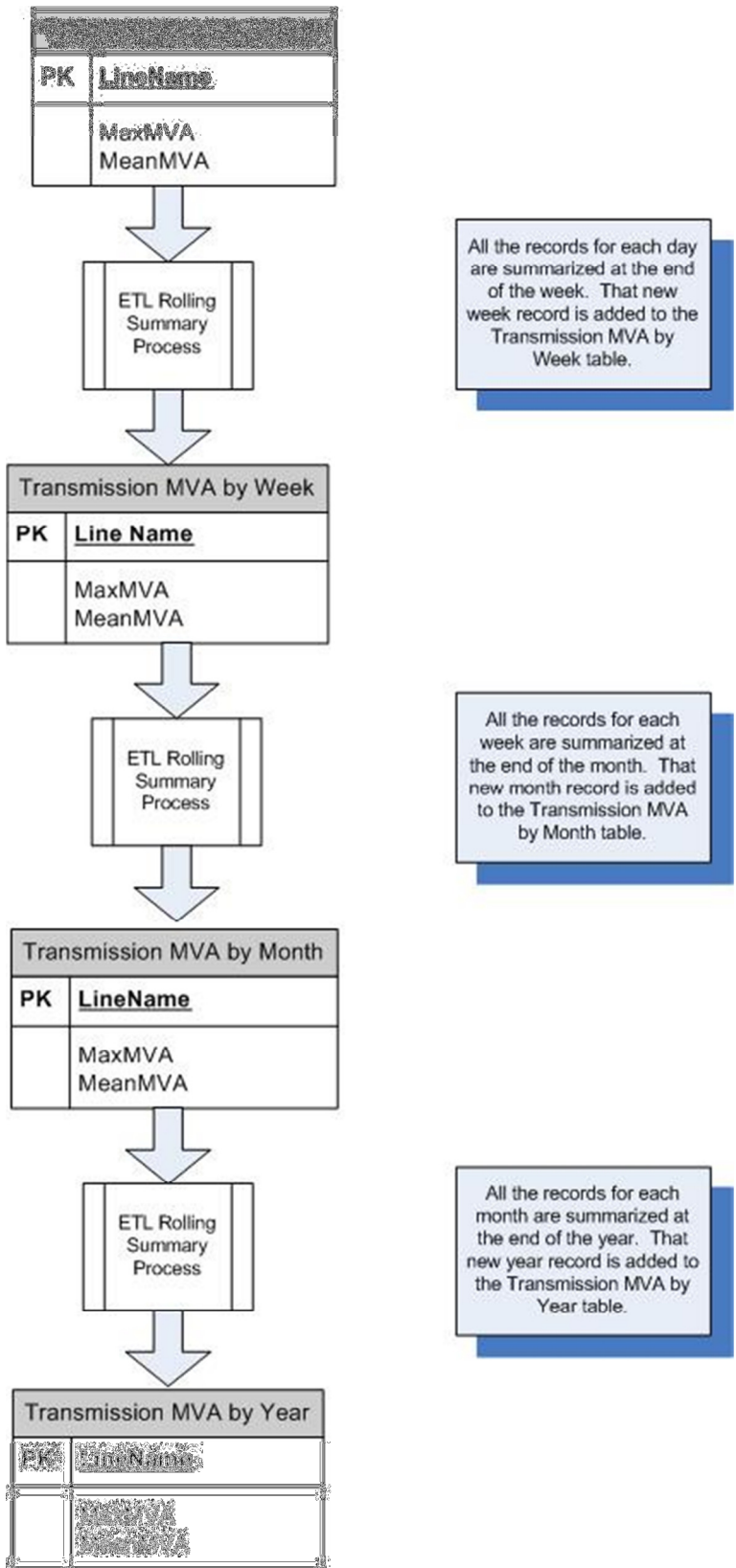
Creating an Accessible Plant Information Database

The plant information a good example to use for data warehousing since the data is essentially stored in a data warehouse. It is highly granular sampled data stored using an archiving system. It is in a proprietary format. The only easy way to access the data is through the products supplied by the vendor. The database is limited to access using ODBC and SQL, or using interface software used for data exchange supplied by the vendor (Hydrocarbon Online, 2010). In this case, the end user who is not a heavy user of the system has access to an Excel plug-in to build a spreadsheet. It is a simple tool that accesses data from the plant information database by entering in a date range, for an identified location or several locations, for example. It will fill in the range of values that meet the criteria that is chosen. This might be adequate for some applications, but a data warehouse might be more useful when there is a need to analyze data or use aggregate information.

In this example, the plant information data has a low level of granular data since it stores individual meter readings at a high rate. This creates a large volume of data that is processed to some degree because it only stores data that it needs to show change over a designated period of time and data that meets a threshold. Referencing back to the research, the operational data (ODS) would be the plant information database and the data there could be extracted and summarized on four levels. In this case the rolling summary data could be aggregated for each day; that data is then aggregated and stored for each week, and that data is then aggregated and stored for each month, and so on. The plant information data could be aggregated this same way but based on peak load values. Another tier of rolling summary data can show those values assembled by region. Because this data is very granular, most applications would require aggregate information to be useful.

In the case of Transmission Planning, this department needs to analyze the peak load data for particular locations over a period of time. In an example supplied by the department, the monthly max and mean MVA (Mega Volt-Amperes) and a total over the entire period which was about two years was shown for a particular transmission line. This was built with the Microsoft Excel add-in. Instead of building this table one transmission line at a time, all of the transmission lines could have this data stored in the data warehouse. Using rolling summary data it could store these values for each line by month, by day, by week, by year or all of those. At any point in time someone could access this information from the data warehouse without having to build something to look at that information on each individual occasion it is needed. Similarly, the rolling summary data could be collected on a regional or member level. Transmission Planning is interested in peak values for a region or a city area. In the data warehouse that data could be identified geographically and then aggregated and stored based on those values. The aggregated data could also be time based as in the previous example. In the table it could show a regional Max MVA for the week, month and year. At any point in time Transmission Planning could access this information from the data warehouse rather than building a spreadsheet.

Figure 4. Rolling Summary Data Example

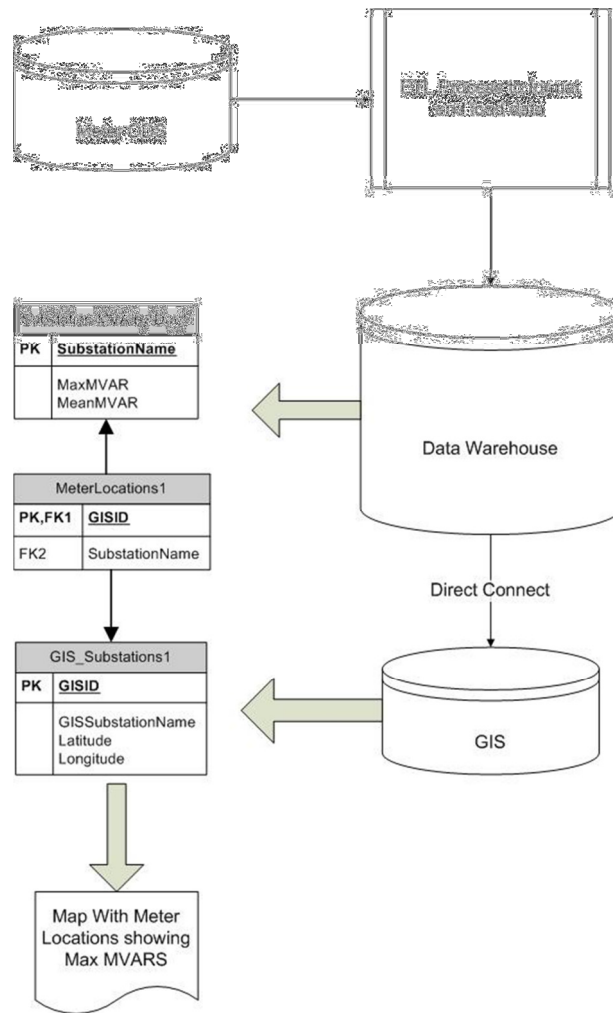


There are three benefits realized from this. One benefit is that the data is removed from the proprietary format making it available to use in other systems for reporting or a dashboard application, for example. The second benefit is the processing is already done for the end user based on a previously determined set of criteria. The data the user is looking for is ready to go and all in one place. And the third benefit is that Transmission Planning Department staff no longer needs to spend time building a spreadsheet to retrieve the information.

Another way the plant information can be used if it were in a data warehouse is for use in GIS. Once the peak load (Max MVA) information is in a data warehouse, it can be accessed further to include the link to the GIS database. In this case, the substation locations are stored in GIS. The substation name is a way to identify the location, but the name can sometimes be duplicated across the dataset. For example, there might be two substations named “ROCK CREEK”. To eliminate this problem, there is a unique identifier on most datasets in the GIS system called GISID.

The plant information database uses a particular named location that does not correspond with the substation location names in the GIS system. That problem can be solved using a data warehouse. The peak load information is a set of aggregated data that would be useful to see on a map. In a data warehouse, the data is stored separately from the proprietary database which provides the flexibility for making schema changes. The ability to change the schema to the way that is needed for the new use is one of the advantages realized using ETL processes. With this additional GISID field, each of the individual spatial locations in the plant information database can be linked to the GIS locations by having this field in both the data warehouse and in the GIS system. This allows the GIS system to query the data warehouse just like the Excel plug-in does, but the difference is the ability to show peak load data on the map (see figure 5).

Figure 5. Linking Data Warehouse to GIS



Creating an Accessible Equipment Maintenance Database

The equipment maintenance database, as explained previously, is also in some form of a proprietary format. At least it's an Oracle database, but the schema is formatted to work with the software that goes with it making it highly difficult to retrieve data without a lot of help from the vendor that supports it. For example, the names of the fields in the database do not identify the type of data that is stored. This is because the software is designed to allow someone to set up a

database without having any database knowledge. The software is driving the database schema and not a database administrator or designer. This allows the user to manage his database without having a lot of database skill. At the same time, it does not allow the data to be easily used outside of its initial purpose. In this case there is valuable information locked up in a system designed only for one purpose. By applying a data warehouse, this data can be accessible for other usage.

The initial purpose of this database is to manage the maintenance of equipment in the substations across the transmission system. There is a set of meter locations in this database that are managed by Utility “X”. The information stored in this database is specific for maintenance purposes. Only that type of data is to be collected in this repository. The metering department does not currently have a meter database to keep track of all the meters that the company has an interest in. This would not only include the meters that Utility “X” maintains, but also other meters on the system that are used for revenue that are not maintained by Utility “X”. The equipment maintenance database is not a good fit for all the meter data that needs to be collected since it is not designed for any other purpose except for maintenance. The decision was made to collect that extra meter data in a GIS format. In the interest of time, a new GIS database was created to collect all the meter data including that data which is currently in the equipment maintenance database. This created a situation where the meter data in the equipment maintenance database was duplicated in the GIS database in order to serve a different purpose.

If a different solution was chosen, Utility “X” could have used data warehousing to avoid copying data from the equipment maintenance system and abandoning it. Copying the data from the equipment maintenance database, bringing it into GIS, and adding other meter data essentially makes the data in the equipment maintenance database useless if it is not updated or

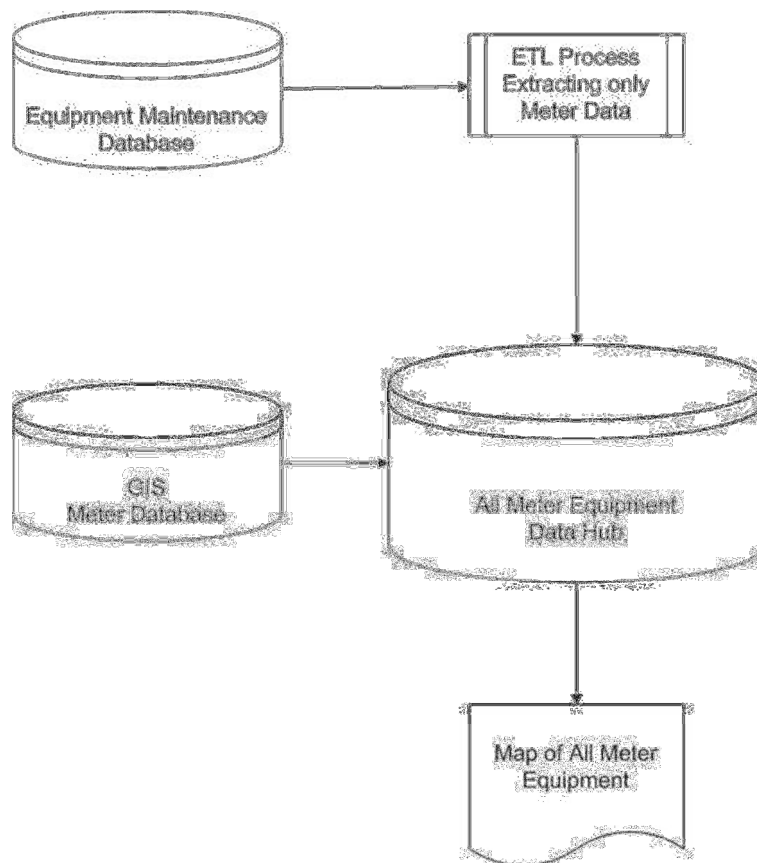
used at that location any longer. If the data were retained, the information would need to be maintained in two locations, thus creating twice the amount of data entry work for the end users and increasing the possibility of disparate data.

A data warehouse solution can be applied to retain the data in the equipment maintenance database and use it in conjunction with the new GIS meter database to avoid the duplication of data.

A data hub could be set up to point to the meter data in the original equipment maintenance database. Only the meter equipment data that was needed would be queried to isolate it from the other equipment. The source data would remain the same and would not need to be altered. The data hub would then be the more flexible location where changes can be made.

“Data hubs are centralized repositories used to reconcile data from multiple source systems. They are often used where companies have deployed multiple vendors' transactional solutions with different data definitions and where it is desired to have a single location where an official definition lives.” ...” a hub is different in that it usually points to data residing in the originating systems without physically moving the data into the hub. (Stackowiak, 2007)”

In this example, the newly identified meter data could be stored here that has nothing to do with equipment maintenance, but it would meet the needs of the metering department. At that point the meter department has seamless access on the application side to all the meter data they need. The data just happens to remain in two places, but since the data hub points to the original source data, the meter department essentially accesses only one location from the data hub. At the same time, the original source data retains its original purpose and the data is not duplicated improperly (see figure 6).

Figure 6. Meter Data Hub

With a data hub, the equipment locations in the equipment maintenance database can be located spatially based on the substation location. Again, like the plant information system, the location field does not correspond to the GIS substation dataset location field that it would be linked to. Also, because of the nature of the database as it is now, it is difficult to maintain an added key field to it without a re-design effort. Adding onto the data hub construction, we can add another GISID field to each meter record to attach it to a substation location in the GIS system. That provides the ability to put all the meter locations on a map. Now there are two new functions with the data hub one meets the needs of the meter department by adding fields to data

that is sourced from the equipment maintenance database, and the other provides the mapping functionality.

Either of these solutions could not have been done as easily if the only option was to add fields to the equipment maintenance system that currently only has one purpose. The purpose of the system would have changed, and the administration of the system would have changed as well. The equipment maintenance system would have turned into something that included the metering database with the option for mapping. Then the interface to the database would have still been restricted to the original interface that is designed around equipment maintenance. This would not make a lot of sense. On the other hand, this would not have been impossible to do, but it could be expensive and not meet the needs of the end user. This data hub option avoids expanding systems that were not originally designed for this type of expansion.

Creating an Accessible Meter System Database

The meter system database is currently a system that is not connected to any other system. Data transfer out of the system occurs via email. A data warehouse solution could create a better data repository for other systems to retrieve data. I was not able to spend enough time to get the required information from this system to create an entirely realistic example. So for the purposes of this research, this section will apply a data warehouse concept to a theoretical example.

In this case an old system is in the process of being migrated into a new system. From the limited amount of information gathered, it is understood that the current configuration was to be migrated first and then new components would be added later. The belief is that the transfer of data from the meter data management system to billing, etc., was one aspect of the system that was to be saved for future adaptations. It is also understood that the new meter database in

combination with the new software has data warehouse capabilities that are not being implemented as part of the initial migration. With that idea in mind, this research covers a solution to that situation that may or may not be valid to apply to this new meter database system due to the limited amount of input into the current system configuration. However, this example applies a data warehouse solution that might serve as a framework for either this new meter database system or can be applied to other similar situations.

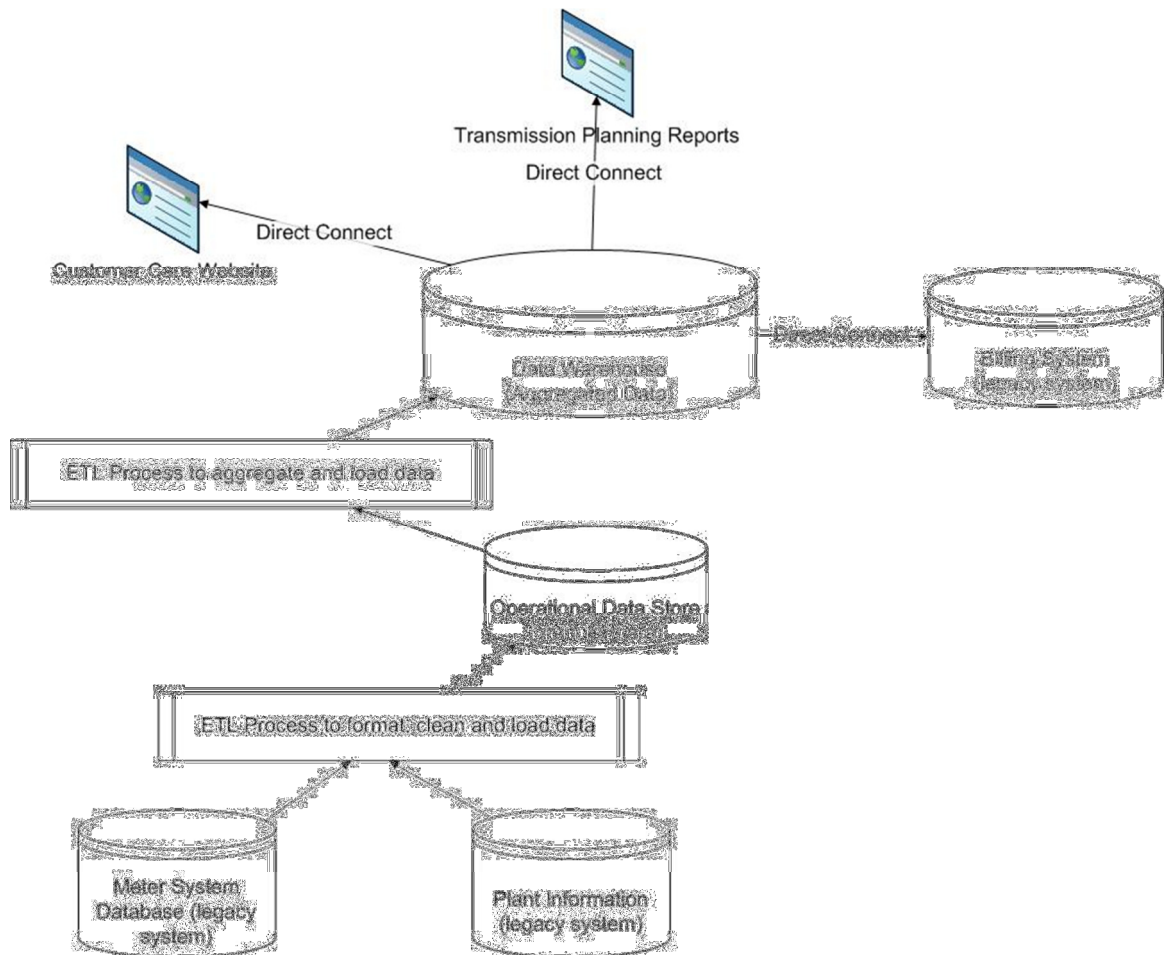
One of the purposes of the data warehouse is to store data that is useful for analysis. Data that is summarized is usually good data to use for analysis. A data warehouse has different levels of summarized data. Some data might be lightly summarized and some might be highly summarized which is sometimes stored in a data mart (Inmon, 2005). This summarized data might be very subject oriented and useful to only one department oriented, so it might be considered a data mart (Inmon, 2005). Data marts typically read from the data warehouse. In this example, summarized meter data is useful for many departments.

One of the advantages of using summarized data is that it can be accessed faster than granular data. “By using the high level of granularity, the data is much more compact and can provide an answer relatively quickly. If it contains sufficient detail, data with a high level of granularity is much more efficient to use. (Inmon, 2005)” When building the data warehouse, it is important to consider that too much summarization might lead to a less useful dataset for queries. “As the level of granularity of data rises, there is a corresponding loss in the ability to answer queries using the data (Inmon, 2005).” So there is a trade-off between too much information where processing speed is affected, and not enough data to support the query. In this example, there are meter readings gathered either every second or every 15 minutes depending on whether the data is used for billing or for SCADA. Over time, that turns out to be large

amount of individual records. Typically, a user of this data is not going to be interested in one reading at a particular moment. What is more useful is the reading of occurrences of peaks and valleys in the values over a period of time. It could be that just the peaks or just only the valleys are useful. When looking at meter readings, these peaks and valleys are the demand and load values. The processed data showing max demand and max load values is valuable information for the Transmission Planning Department. This is the information that they use to make decisions on where upgrades are need for increased demand, or to determine where surplus power areas are.

For the Billing Department, it is the demand and load values that this department is interested in over the period of a month. The billing reports show that they use these values to calculate a total Kilowatt Hour (KWH) of energy to bill each of the members.

The proposed data warehouse environment would allow connectivity to the meter system database so that the billing system can access the data it needs in the format that it needs at any point in time. By eliminating the transfer of data through several conversion processes and email, the data warehouse can be used to summarize the meter readings monthly by using rolling summary data as illustrated in a previous section. The aggregate monthly totals for each meter would be stored in the data warehouse for direct access by the billing system (see figure 7).

Figure 7. Proposed Connection Between the Meter Data and Other Systems

The Transmission Planning Department also uses similarly summarized meter data. The planners in this department need to see peak values for a member over a certain area. They use two different types of peak values. A member peak value is the max system peak for a member area. This is the value that is also used for billing purposes. The second type of peak value is the member non-coincidental peak. This is a value of a smaller area within the member area that is usually a major city. A major city within a member area is going to show a peak sooner than other areas because of the larger population concentration.

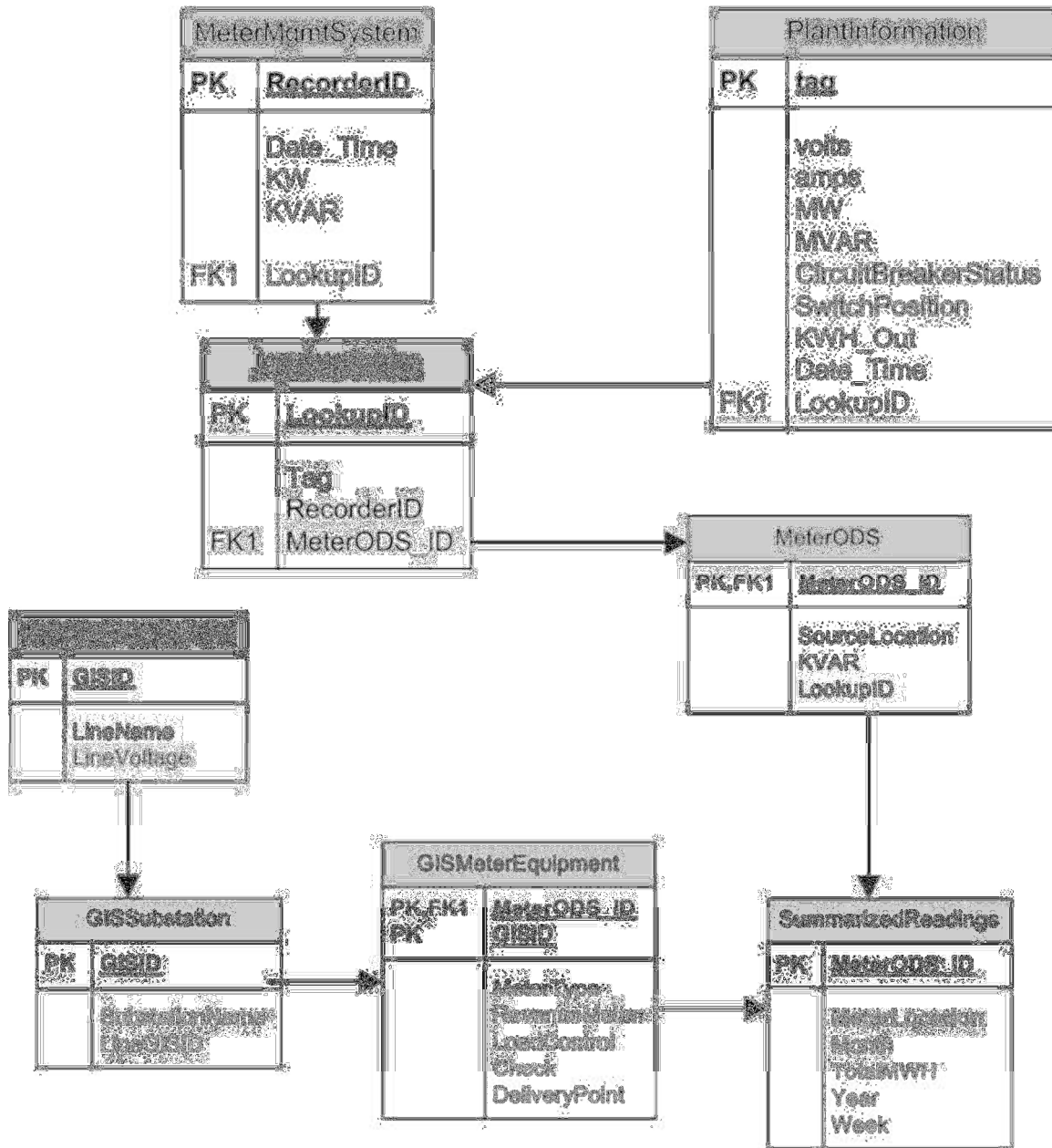
GIS Component

The geographic component of the max system peak and the non-member coincidental peak has not gone unnoticed. There has been a request from the Transmission Planning Department to have the ability to place these peak values on a map. Currently the department is limited by its use of this information since it is stored in its original detailed small-grain form, and the only access to it is by using a Microsoft Excel plug-in, as described previously. An example sales report (See Appendix C) obtained from the department shows the use of aggregate totals by month by substation. Several meter readings along a transmission line are used to show line flow. This is also a geographic component that can be utilized in the GIS. So with the use of the data warehouse to aggregate the totals by month using rolling summary tables, and by incorporating a value to the table that is associated with the meter location in the GIS, we can utilize this combination to put this information on a map. This eliminates the restriction of having only two peak totals by member area or a subset of a population center.

The geographic components described in the last section include the member area, the meter locations and the transmission lines. All of these components exist in GIS, so by adding a link to the summary tables that includes data that is also in GIS, the door is opened to map that information in several combinations. The association between the meter location, the meter location on the transmission line, and the meter location to the member area is retained in GIS. So in the data warehouse, the only item that needs to be stored to link to GIS is the meter location. Having this meter location link allows the Transmission Planning Department to select any combination of meters to illustrate a load in a certain area without being restricted to a geographical construct of a member area or a city population center.

The data warehouse would create a place where all meter locations were stored with their readings. The method should include a way to avoid storing duplicate data because keeping an updated copy of a highly voluminous dataset in such a granular form would create a maintenance issue. As an alternative, the meter data from both the plant information database and the meter system database could be assembled in one place in the data warehouse, and the data can be aggregated by month. This way the Transmission Planning Department would have all the meter locations needed for analysis in one place. Additionally, adding a link to GIS to the summarized meter data, the system holds the information about which meters are on what transmission line. In combination with the summarized and aggregated data in the data warehouse, the planner would only need to know which transmission line was needed to calculate the totals. The planner would no longer be restricted to know what tags were on what line or in what member area since that information would be located and accessible in GIS (see figure 8).

Figure 8. GIS Access to Summarized Meter Data in the Data Warehouse

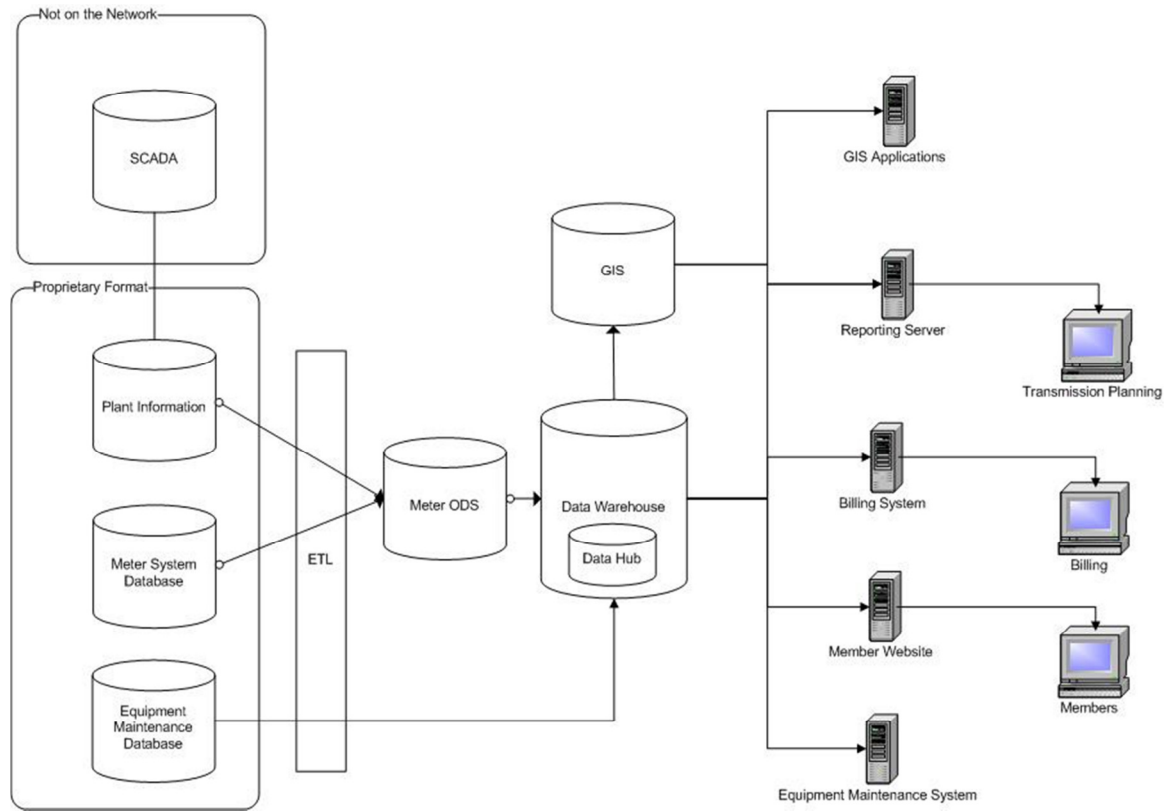


The other additional advantage is that the Transmission Planner can obtain this information in a report form and display it in a map form if needed. The clear advantage is that the information does not have to be map-centric. The spatial aggregation is occurring using GIS without having to use a map to display the information. A report can still be the end product, but without the work of associating meters with a location. This is what the GIS does automatically with the additional link between the GIS and the summarized meter data.

Conclusion

The foundation for the data warehouse is SOA architecture. With support of the enterprise in combination with building a business-wide SOA, the data warehouse can be the tool that makes information accessible to all departments at Utility “X”. Once the architecture is in place and all departments have cooperated in developing their systems to participate in it; different systems can be leveraged by having access to new information. GIS is a system in Utility “X” which use can be expanded by implementing a data warehouse environment. As illustrated in the Proposed Configuration (see Figure 9), systems throughout the company can access the data warehouse and provide an interface to what were previously isolated and largely inaccessible databases (see Figure 2). With support of the company, the data warehouse can alleviate pressure to change legacy systems to accommodate new usage, provide a way to make meter data accessible for multiple uses, and provide a way for other systems such as GIS to expand by providing a spatial interface to data that would not otherwise be spatially enabled.

Figure 9: Proposed Configuration.



References

- Inmon, W. H.. "Chapter 4 - Granularity in the Data Warehouse". Building the Data Warehouse, Fourth Edition. John Wiley & Sons. © 2005. Books24x7.
<http://common.books24x7.com.dml.regis.edu/book/id_12458/book.asp> (accessed January 1, 2010)
- Inmon, W.H.. Building the Operational Data Store, Second Edition. John Wiley & Sons. © 1999. Books24x7.
<http://common.books24x7.com.dml.regis.edu/book/id_2923/book.asp> (accessed July 21, 2009)
- Inmon, W. H., Claudia Imhoff, and Ryan Sousa. Corporate Information Factory, Second Edition. John Wiley & Sons. © 2001. Books24x7.
<http://common.books24x7.com.dml.regis.edu/book/id_6693/book.asp> (accessed July 21, 2009)
- Investigating information systems with action research, Communications of the AIS. Volume 2 , Issue 3es (November 1999), Article No.: 4.
<http://portal.acm.org.dml.regis.edu/toc.cfm?id=J812&type=periodical&coll=ACM&dl=GUIDE&CFID=79061838&CFTOKEN=51832941>
- Kalman, David. "The doctor of DSS. (decision support system expert Ralph Kimball) (Interview)." DBMS 7.n8 (July 1994): 54(5). Computer Database. Gale. BCR Regis University. 7 Sept. 2010
<http://find.galegroup.com.dml.regis.edu/gtx/start.do?prodId=CDB&userGroupName=regis>

McDonald, John. Substation Automation Basics - The Next Generation. Electric Energy Online.

http://www.electricenergyonline.com/?page=show_article&mag=43&article=321.

[accessed December 31,2009]

OSI Software, Inc. Swinging Door Compression. April 1991. Rev. August 1999.

Rosen, Mike, Boris Lublinsky, Kevin T. Smith, and Marc J. Balcer. "Chapter 2 - SOA —

Architecture Fundamentals". Applied SOA: Service-Oriented Architecture and Design Strategies. John Wiley & Sons. © 2008. Books24x7.

<http://common.books24x7.com.dml.regis.edu/book/id_25199/book.asp> (accessed February 16, 2010)

Stackowiak, R., Rayman, J., and Greenwald, R., 2007. Introduction. *In* Stackowiak, Rayman, and Greenwald, *Oracle Data Warehousing and Business Intelligence Solutions*: New York (John Wiley & Sons).

http://common.books24x7.com.dml.regis.edu/book/id_17063/book.asp%20 [accessed October 5, 2009]

Swanson, Richard A., and Elwood F. Holton III (eds). "Chapter 1 - The Challenge of Research in Organizations". *Research in Organizations: Foundations and Methods of Inquiry*. Berrett-Koehler Publishers. © 2005. Books24x7.

<http://common.books24x7.com.dml.regis.edu/book/id_11859/book.asp> (accessed March 7, 2010)

Vassiliadis, Panos. "A survey of extract--transform--load technology." *International*

Journal of Data Warehousing and Mining 5.3 (July-Sept 2009): 1(27). Computer Database. Gale. BCR Regis University. 8 Nov. 2009

<<http://find.galegroup.com.dml.regis.edu/gtx/start.do?prodId=CDB&userGroupName=regis>>.

VertMarkets, Inc. Hydrocarbon Online.

<http://www.hydrocarbononline.com/product.mvc/PI-ODBC-Integration-0001?VNETCOOKIE=NO>. Accessed Feb. 18, 2010.

Appendix A**INTERVIEW CONSENT FORM**

My name is Katy Carpenter. I am working on my thesis project at Regis University for my Masters of Science in Information Technology degree. You have been chosen for an interview regarding the thesis topic which has to do with data warehousing at Tri-State. The purpose of the study is to find ways that Tri-State might develop better information management through the use of a data warehouse environment. The results will be used in the study and the final thesis will be published at the university and possibly used by the Tri-State IT department for reference.

I would like your permission to interview you and use your comments in my research. The interview will last approximately 15 to 20 minutes. The interview questions are in reference to different kinds of information you manage or gather to do your job. There also might be a follow-up questionnaire depending on the results of the first questionnaire. Your name will not be used in the study. All identifiable information collected will remain confidential. You do not need to answer any questions that you do not want to. Any time during the interview you may stop your participation with no questions from me.

If you have questions regarding the study or your participation, you can contact me at #3406. You can also contact my advisor, John Holmes at (303) 419-7837.

Please contact me if you would like to see a copy of the final thesis document.

Consent Statement:

I have read this form. I understand that nothing negative will happen if I choose not to participate. I know that I can stop my participation at any time. I voluntarily agree to participate in this study.

Signature _____

Printed name _____

Date _____

You will be given a copy of this form.

Appendix B

Questions for Data Warehouse Project:

- 1) Do you find that several different people collect the same or similar information?
- 2) If yes, then what type of information do you see duplicated?
- 3) If yes, how many different sources of the same or similar information do you go to, and where are those sources located?
- 4) Do you find that several information sources of similar kind may be inaccurate or out of date?
- 5) Do you have examples of good information that is easy to access that you use?
- 6) Do you have examples of good information that might be difficult to use or access?
- 7) What information do you need to do your job that is not in form that is easy to access?
- 8) Do you share information with other departments? If so, what types of information and what departments does it go to?

