DW + DM = $avings

Robert J. Scharl, LTV Steel
Paul M. Centric, LTV Steel

Abstract

Data Warehouse. Data Mart. Data Mining. These are all recently created terms that apply to data extraction and decision support based on a solid understanding of the data. Companies have experienced the methodology of Deeming for Statistical Process Control and Continuos Product Improvement. The recent improvements in computer hardware and improved data access has lead to a revolution in the ease of data analysis for decision support.

History

The word market place has been shrinking since the Industrial Revolution. Industry used to think only in terms of its community, country or continent for supplying services or products. Improvements in communication and transportation have made it possible for a company to compete in a world market. The world market has also made customers more demanding of the product for price and quality. Some governments may allow their basic industry to have 'price supports' or they may operate in a totally 'free' market place where a company will determine its own destiny and only the strongest and best will survive.

Since 1945, a lot of basic manufacturing in Europe and Japan has been rebuilt because of the destruction during World War II. Japan started its Statistical Process Control (SPC) with Mr. Deeming. In the 1970's, United States basic manufacturing, (i.e., automobiles, steel, farm and construction machinery) was loosing its global market share to foreign competition. The United States had to very quickly reduce the manufacturing cost and increase the quality level.

Steel customers were demanding 'Just in Time' delivery of the product. The supplier/manufacturer was required to have the product in 'his' inventory to supply the product upon demand if a continuing customer/supplied relation was to be maintained. The Customer did not want to be charged with the inventory tax charges. The manufacturer also does not want to pay an increased inventory tax either. Therefore, procedures had to be found to decrease the time required from order placement to product deliver to the customer.

The steel customers are demanding a product to their specifications with an increased quality level. Reject levels between 3.0% and 5.0% at the customer plant were acceptable prior to the 1980's. Today, customers want a 0.0% internal reject level in their plant. The actual level maintained seems to be around 0.3% to 0.5%.

LTV Steel Corporation began a serious manufacturing rebuild in the late 1970's early 1980's. The processes that used to be controlled by humans were now controlled by computers. By the mid 1980's there was an information overload of data. The Process Control Computers were 'controlling' the process, but the engineer responsible for the delivery and quality of the product found it very difficult to access the data. The data seemed to reside for the exclusive use of the process computer or for limited use of the Information Systems (IS). The IS department had the data stored in sequential and VSAM file formats and basically captive in the IS department. A Method had to found to enable the Quality Engineer to readily access the data to enable continuous product and/or process improvement.

The Quality Assurance Department and Process & Product Quality Departments went on a mission to make the data available to the end user analytical user. In 1990 when SAS® released Version 6, they incorporated into their system a quasi relational data base structure. The best update to their data structure was incorporation of indexes or keys. This made the SAS® data files behave similar to the larger more established data base systems.

Our departments were actively pursuing a more established data base system until statistics were reviewed that were conducted by a outside consultant comparing data retrieval from data files ranging in size from a 1,000 to a 1,000,000 observations. In every instance, the SAS® data was retrieved at a 90% improved rate than were the conventional true data base system files.

The purchase cost for a conventional data base system could be in excess of $0.5 million plus a data base administrator to maintain the system and files. The SAS® System was already purchased with a strong user base that had a intimate knowledge of their local data. A data base administrator therefore, would not be necessary.

DATA WARE HOUSING AND GETTING STARTED

There are many considerations that need to be addressed in the development of a Data Warehouse. Some questions that must be explored and answered for the establishment and maintenance of a Data Warehouse are:

1.) Is the Data Warehouse for Decision Support or Real Time Business Support?

A Data Warehouse designed for Real Time Business Support requires a robust full featured Database Engine like DB2/Oracle/etc... which can perform record locking while records are being posted or manipulated. If yesterday's information is sufficient to answer business questions and trends, then a with static, well indexed data is adequate. This simply means that, do I buy a full feature Database or utilize SAS® System with indexes to support the Warehouse. Each path has its merits and costs.

2.) Is the Information dynamic and if so how often does it change?

The information can be gathered from many sources and can be in need of updating from time to time. For example, if the information is entered into a Legacy system once a day the Warehouse will only require, at best, a daily revision. Consequently, does information that changes every minute required a minute to minute Data Warehouse update.

3.) Is the information historical in nature or time based?

Information may need to be kept by KEY and not Transactions (i.e. Social Security system versus a checking account). This question will help inform you on the type of file manipulation techniques to use in handling the information.

File types can be defined as one of the following:

STATIC- Static files are files that can be loaded in their entirety possibly nightly, and this process is conducted each and every night with total replacement of the file.

OPEN- Open file system allows for data to be appended to the data set and routinely verified that there is adequate space for future postings. This eventually will require system maintenance and purging of perhaps the oldest data.

TIME BASED-Time Based file system operates on a time system. Data is maintained through a definite time/date format. The data may be in multiple files and divided into time units of month, quarter or year.

MOVING TIME WINDOW- The data is maintained in a drop and add fashion. For instance, there is a conscious decision to have data available for 180 days back in history from today. Records 181 days or older are deleted and current records are added to the data base.

HISTORICAL- These files can be less comprehensive because of new additional variables that are continually added. They may also be incomplete or simply used as *BENCHMARK* information for deviation or reference checking. This data could be summarized.

SUMMARY-Summary files are basically self descriptive. They will be defined as needed (i.e. Annual/Monthly/Weekly/Daily). These files will save the drudgery of building routines cook the data down into groups and terms users need and want. These files are the main stay for drill down systems with all their elaborate front end push button user friendly menus.

STATISTICAL-Files of this nature can be utilized help in the analysis of the information. Meta Statistical files etc...i.e. how particular variables correlate to other data elements/files/areas/systems and the ability to

link list your way through the various systems for analysis.

4.) Can the information be used or accessed from the original source?

If information can be accessed from the original source data, will a special tool or programming be required? Will the source support SQL and will records be accessed directly with pass through code?

5.) What resources are available to complete project(MAN HOURS/MONEY/MACHINES)?

Is there a proverbial deep pocket or is there a strict budget? If their is defined maximum amount, be aware of what you can reasonably afford.

6.) Who are your users and what is their training and expectations?

People who have a very technical job, tend to embrace new technology with more enthusiasm. The upper management team and support staff tend to require a different type interface to Data Warehousing than the technical staff. This will require different training and possibly a smoother more intuitive interface metaphor for management and clerical staff.

7.) Will data be created and stored specifically for the Data Warehouse?

Any data required to operate a business needs to be backed up and entrusted for users. Information garnered from Legacy systems, this data probably has a back up scenario. Data entered directly into a Data Warehouse and is pertinent in the operation of business, also requires backing up.

8.) What is the scope and can it be defined sufficiently to ascertain a cost and is expansion possible?

The scope on most Data Warehousing endeavors can be very difficult to define. First define if the system is Local, Plant wide, Corporate, or International. Define the purpose the system is to fulfill and how comprehensive will it be in comparison to the current functioning. Including everything, even the kitchen sink, is a common solution. Don't accept all data elements as a possible scenario.

Decision Support

When material rejects were/are at a high internal rate, it can be a fairly easy task to reduce the rate. However, when the rates are quite low, (less than 5%) an improvement strategy is not all that obvious. The 'Mining' of huge amounts of data in what may seem like totally unrelated sources may be normal instead of the exception. Such is the case for the examples to follow.

EXAMPLE 1

Problem: The process Control Computer for the Hot Mill takes readings of 15 vital attributes to define the hot strip being produced. Each of the 15 attributes are controlled in micro increments and then summarized into 240 to 440 incremental readings. The engineer could view only a limited number of coils 'stored' on the process computer. It was desirable to be able to view an individual coil trace that may be up to a year since production because of a customer problem. The data was not in a format that could be used for statistical analization for process or product improvement.

Solution: Provision was made to have the process data transferred immediately after the coil is produced from the process control computer to the mainframe. The IS Department has the data stored in a sequential format. This data was transformed into the Data Warehouse where upon demand an engineer can request a coil, series of coils or random coils for analysis. The data is stored in 22 separate files with average 1020 bytes per file, and up to 450 variables per file. Production rate is 600 coils per day for 362 days production per year.

The Engineer using this information in combination with other files for the hot mill and historic reject trends, correlation analysis and ad hoc statistical queries against the Data Warehouse, there has been an improvement in the rate for internal rejections for one (1) of the rejection classifications has decreased from an average of 0.58% to 0.22%. The 0.36% improvement in the rejection rate equates to a savings of $6.3 million per year.

A long term trend analysis file has summary weekly and monthly for production and rejects by producing unit with defects summarized into major groups.

Example 2

Problem: A production foreman was spending approximately 4 hours per day manually researching 'Roll Mark' rejects for the previous day for processing history using the Legacy 'Corporate Coil Tracking System'. The legacy system could only perform simple 'look-up tasks'. The foreman would then do a double transpose of the information first by hand and then to an Excel Spread Sheet. The rejected coil in conjunction with the previous 5 and post 5 coils where high lighted as potential problems. Potential 'problem' coils were being missed by the current production/inspection feed forward process.

Solution. All the production, rejection and text files were already in the Data Warehouse. The Engineer assigned to reduce 'Roll Marks' requested a daily summary report listing. The Foreman was not a computer literate person so a local SAS® programmer was consulted. A simple SQL query was written to access and assimilate the data. Programmer time spent was 4 hours total with the implementation of a automatic print going to several remote printers. The foreman's time was better utilized for other production problems. $32,500 could be projected to better time management savings plus intangible benefits of potential liable product not having value added processing and shipped to the customer.

Example 3

Problem One of the most persistent and damaging problems in the productions of steel is surface 'Rust'. As LTV produces steel, internal stresses are formed during rolling. These stresses need to be relieved. LTV does this by heating up the recently rolled steel and slowly cooling (annealing). When the steel is in the annealed condition, it is highly susceptible to rust formation due to the lack of any protective coating.

Solution To assist in the reduction of rust, a **Dead Soft Report of Conformance** report was developed. This conformance report measures the lag time between Breakout (leaving the Furnace) and the current day. The greater the time between breakout and the temperature humidity index, the likelihood of rust forming is increased. The need to move the older coils into the next production operation before the formation of rust has occurred is critical. The daily report serves as type of report card of conformance and an action list of coils to be immediately processed. The Data Warehouse is used to trace these coils and calculate the lag time. A graphical Dead Soft Report is created and automatically distributed to mill laser printers for immediate action by Production Control and the turn foreman.

Example 4

Problem It has long been thought that the hardness of the rolls used in making flat rolled steel directly relates to 'Roll Mark' defects. Various rolling and support departments generate and gather data relating to rolling practices, roll chemistry, roll grinding practices and subsequent roll hardness.

Solution The condition of rolls become increasingly important to the ultimate quality outcome with the desire to increase production and reduce rejections for roll related defects. There are many facets to determine roll performance and quality producing levels. One aspect is the hardness level and the defect occurrence within a roll hardness range. The graphical **Finishing Hardness Range** chart was created to demonstrate the various defect ratios and hardness levels at which the defects These defects are all roll related and having the correct roll hardness level is extremely important. This report revealed any perceived patterns of hardness to defect level. Most of the harder surface rolls are more expensive and harder to service and keeping the hardness at lower levels will save millions of dollars when you consider the vast number of rolls involved.

All the reporting on roll performance is derived from our roll tracking repository. This warehouse consists of ten files which can identify the rolls that were in service as every coil was produced and any roll related defects that occurred as the coils were produced. There are five (5) production stands. Each stand has a top and bottom work roll and a back up pressure roll. Depending upon production requirements, some of the work rolls in the last two (2) production stands are changed are changed after every two produced coils. Roll logistic

usage was difficult to track with any manual system.

Since the inception of the roll hardness program with relation to defect tracking, it is estimated that $9.48 million savings per year can be attributed to a reduction in roll related defects. Statistics are not yet available for potential increases roll costs and roll repair sequencing.

THINGS NOT TO DO!

1.) Don't Include the world!
2.) Don't rely on committees and experts to define the entire system. Have lots of End User input and direction. The best person to define areas is the user not the bureaucrat in charge.
3.) Don't skip areas i.e.(labeling etc...)
4.) Don't think cost more is better.
5.) Don't allow power people to bully the outcome of your warehouse.
6.) Don't over commit startup resources.
7.) Terabytes of data does not constitute a better Warehouse.
8.) Don't ignore training. If training appears to be to time consuming or expensive, ask yourself, can I afford to do this and if the answer is yes then these busy people will benefit the most from a warehouse. They obviously are spending a lot of time gathering information which a trained person can easily get from a well designed Warehouse.
9.) Don't build single user files. Coordinate for usage to the general user population or reprioritize. The pay back may be insufficient.
10.) Do Not restrict information unless it is sensitive in nature (i.e. payroll, personnel).
11.) Do Not assume that all data must be in one file/one database format/one system. A Data Warehouse can consist a combination of STATIC DATA/RELATIONAL DATA/PC vs. MAINFRAME FILES etc...In other words all your eggs don't have to be in the same basket.
12.) Don't insist on single data storage (**BE REDUNDANT WHERE IT IS PRUDENT**)
13.) Don't incorporate private user files without adhering to the Data Warehousing rules that have been established in your company.

THINGS TO DO!

1.) Format data both in and out.
2.) Keep data sets of format and definitions
3.) Define variable definitions and labels and store into a data set.
4.) Data Glossary-Data dictionary to define files, data sets, variables, labels, addresses, structure, location.
5.) Common key variables between data sets and files and indexes superkey (build var.).
6.) Provide for backup solutions.
7.) Understand your problem before you try to define system.
8.) Start with Pilot and allow for expansion.
9.) Don't limit your vision. This process is extremely evolutionary.
10.) Have trained personnel assist in all designs not untrained committees.
11.) Have data available to everyone. This may require duplicate storage but in comparison DASD is cheap and people are not.
12.) Secured Information i.e.(personnel) may require special treatment.
13.) Front end your system where ever possible for ease of use.
14.) TRAIN TRAIN TRAIN and then retrain!!!!!! Training is the best support tool money can buy.
15.) Start a communications bulletin either weekly monthly or at least bimonthly.
16.) Get solid written commitments/feedback from users(deliverables etc...). End users know the information/process the best.
17.) Have system maintained by a single group with end user input.
18.) Build and support ongoing documentation to the point that a manual can be delivered to new users as needed and incorporate into TRAINING.
19.) Coordinate requests and commonize.
20.) Develop common naming convention and file naming. This may be difficult.
21). Build a collage reports and output graphics.
22.) Load clean data where possible or validate as much as possible.
23.) Try to assign a local GURU and include into documentation.
24.) Document user problems What/Who/When/Why/Outcome and most important was this person adequately trained.
25.) Try to use a group of software products compatible with each other and that support ANSI STANDARD SQL Pass through technology.
26.) Utilize SQL code wherever possible. This will eventually become the default processing technique and appears to have more support.

27.) Establish User feedback feed forward mechanism for improvements and enhancements.

Conclusion

The Technology (software) used in the development for the Data Warehouse and Information Delivery System is purchased from SAS® Institute. The methodology is an evolutionary process for all companies. It is the responsibility of the Data Warehouse designers to have a working familiarity with the data. The Information Delivery requires personnel with the desire to improve the process with the use of statistical analysis. Creative thinkers unhindered by the 'old timer syndrome' is not acceptable. People must be free to think unencumbered of what happened in the past. The most important desire is for maximum improvement at minimal cost.

Since 1994, the topic of Data Warehousing and Data Mining have become the 'Vogue' terms. For LTV Steel, having the data readily available in an easily understood format is the most important requirement. The IS Department policy has been to furnish to the user community a sequential flat file either on disk or tape. A record layout would also be provided. Minimal training would be given for the extraction of the data from the flat file into a useable format. Consequently, the potential existed either the whole file or extracts of the file could be contained in many different user ID's. The realization was soon made that ready access to a uniform data format was critical to the survival of the company. The data is now available to any one in the company in a consistent format.

Attending various conferences has lead to a variety of terms being spawned to define the fast emerging technologies. The underlying need is to made the data available for information delivery. The data may be called 'a large data set (file)', or may say 'a Data Warehouse', or subset the data warehouse into a 'Data Mart' for a specialized portion of the master data residing on a remote server for the use a an intended end user.

Data Warehouse and Data Mining are new terms applied to old ideas and practices. The '**NEW'** is making the data available to the process engineer or analyst who understands the interactions between cause and effect to solve product or process problems. The ownership of the data has been removed from the IS Departments and placed in the hands of the end user. The Information Age has made data and data analysis easier to accomplish. The interaction of data variables has long been understood in the statistical word. The ready access for the data that defines the variables is easier today than yesterday. The computing power of the personal computer today exceeds the power of a mainframe of a few years ago. The ability to merge large data files on unlike variable names without transformation is a success to data mining.

LTV Steel has computer generated files from order inception to product shipment and customer claims plus all the department process control computers. The production of one coil of steel can create at least 70,000 bytes of information in 60 plus files.
