



USING A DATA MART TO IMPROVE THE MONTH-END CLOSE PROCESS

*By: Kurt D. Duncan
Applications Consulting Specialists, Inc.*

Introduction

The month-end close is a long and arduous process for many companies, oftentimes spanning several days. One very large portion of the close process is running reports, such as Aging Reports, Transaction Reports, and Dimensionalization Reports. These reports are a requirement for any well-run company and must be run within a certain time frame during the close process. The larger the sales of a particular company, the larger the aging and dimensionalization reports will be. Likewise, larger companies will inherently have larger databases to store the data concerning sales, receivables, payables, customers, vendors, etc. Therefore, all other things equal, running complex reports in large companies with large databases and stiff competition for processing time could take a considerable length of time.

One manner in which to reduce the time requirement of running reports is to create a data mart to house relevant Accounts Receivable data. The data mart would primarily include all data necessary to run month-end reports. Three major benefits from using an AR data mart are as follows: 1.) the data mart would reside on a different database/server than the AR module and consequently would not be fighting for system space and processing power; 2.) the data mart would only include data that is relevant to the month-end close process, thus would be a significantly scaled-down version of the AR tables; and 3.) end-users could use the data mart to perform much of their ad-hoc month-end querying and reporting. This practice results in a twofold benefit - queries and reports would run faster in the data mart than in Production environment [due to the reduced size of the tables and limited users], and the strain on the Production database would decrease.

This paper focuses on the many steps/decisions involved in creating a data mart, including: making a decision as to whether or not a centralized location for storing data (such as a data mart) is needed, whether this location should be a data warehouse or a data mart, who will own the data mart, who will design the data mart, who will build the data mart, who will monitor and maintain the data mart, who will be the Project Manager, who will test the data mart, what will the front end reporting tool be, how will the data be updated, how often will the data be updated, what is the time frame for each step in the plan, etc.

In addition to the above topics, this paper will outline the methodology used to form the data mart structure. Key areas include arriving at an appropriate schema for the data mart, choosing the correct tables and columns, building the queries and update statements, and documenting the design.



What is a Data Mart

A data mart is a query and reporting tool incorporating a subset of the enterprise's data, such as Accounts Receivable or Sales. The data in a data mart is focused on a specific function or activity of the enterprise. Data marts have specific business-related purposes, such as measuring and forecasting sales performance, or measuring the impact of marketing promotions.

Data marts are often promoted as being the ultimate decision-making tool, or a must for all high-level decision makers. This assertion masks the true use of most data marts in today's market, which is as a tool for post-decision monitoring of the effects of a decision. Data marts typically are used to answer the following types of questions: Did the sales promotion in the Midwest market lead to increased sales in the 3rd quarter? Did collections on past due amounts from Customer ABC increase after the new dunning letter process was implemented last quarter? In most instances, data marts are used to run queries and reports on historical data, with analysis performed on the output.

Data marts are not necessarily implemented for the needs of decision makers or used in the decision making process. However, using a data mart in the decision making process can indeed lead to greater quality decisions. In the near future, as companies become more accustomed to using data marts and the information within the data marts, we will see the use of data marts expand to not only a query and reporting tool, but also a powerful decision-making tool.

A data warehouse, on the other hand, contains historical data, both summarized and in detail, that is *common to the entire corporation*. This structure differs from that of a data mart due primarily to scope. The scope of a data warehouse is much wider than a data mart, usually encompassing a wide range of corporate activity, whereas data marts are much more narrow in focus. The data that resides in a data warehouse is at a granular level, while the data in a data mart is at a refined level.

Reasons for Developing a Data Mart

If a data warehouse contains only a small amount of data, then it will most likely serve the needs of the different departments within a company, thus making a data mart unnecessary. However, as a rule, data warehouses do not remain small for very long. In a short period of time, the amount and scope of the data will increase greatly. With an ever increasing amount of data on a wide range of subject areas, it becomes very difficult for each department to quickly and efficiently gather the specific information that they need. Consequently, as data warehouses grow larger, the need for data marts increases. Following are four major reasons why companies are driven towards data marts as the size of the data warehouse increases:

- The competition for processing time within the data warehouse grows fierce
- Data becomes more difficult to customize
- The cost of processing in the data warehouse increases as the volume of data and number of users increases



- The software (front-end tool) that is available for the access and analysis of large amounts of data is not nearly as elegant as the software that can process smaller amounts of data

Reasons why data marts are attractive are as follows:

- customization – data can be easily manipulated, aggregated, etc., to best suit the needs of end-users
- historical data – companies are able to include as little or as much historical data as necessary according to the desires of the data mart's users
- resource utilization – processing power in the applications and the data warehouse is increased because more users will run reports/queries in the data mart
- analytical software – sophisticated, relatively easy to use front-end tools are abundant
- cost of processing and storage – due to the smaller scale of the data mart as opposed to the data warehouse, cost of processing and storage decrease

Do I need a Data Mart *and* a Data Warehouse

The above question is hotly debated in the current literature, with the outcome uncertain. Which setup is optimal for a particular company depends on what the company is attempting to achieve, how large the company is, how much funds are available for the project, and, ultimately, which side of the debate can argue the most persuasively.

Exhibit 1 displays a model that includes three different scenarios for setting up a data mart and/or data warehouse. All three scenarios begin with extracting and transforming source data, which could be data that resides within the applications, external data, etc. The first scenario is simply constructing independent data marts from the transformed source data. There is no data warehouse in this design. The next scenario involves creating an enterprise data warehouse from the transformed source data and then constructing dependent data marts from the data within the data warehouse. Lastly, the third design combines the two previous ones into a truly dynamic data environment that includes independent data marts fed from the source data, an enterprise data warehouse fed from the source data, and dependent data marts fed from the data warehouse. Which design is best? None of them, and all of them. Constructing a system that encompasses only independent data marts fed from source data is not an enterprise-wide solution to data needs. On the other hand, perhaps an enterprise-wide solution is not necessary. If this is the case, why allocate scarce funds to a lengthy and complicated project that is not necessary. However, oftentimes a central depository of data from a wide range of sources is critical to a company's reporting and analysis needs. In this instance, a data warehouse that feeds dependent data marts (AR, Sales, Promotion, etc.) would be a better solution. The key is determining how the company will use the data mart and/or data warehouse. Once this is known, then select the structure that will allow you to achieve this goal. In essence, let the *end* justify the *means*.



Constructing a Data Mart

The core value of an AR data mart is the assurance of a single answer to a given question no matter which user is making the inquiry. In day-to-day business operations, several people will research the same matter-of-fact issue and, through various sources, find different answers. Which source is correct? One manner in which to ensure everybody has access to the same data is to implement a data mart. An AR data mart can provide a secure environment for analysis and reporting. Considerations of time and resource often lead companies to data marts rather than a fully integrated data warehouse. Data marts require smaller amounts of source data, fewer data elements to define, and simpler data models. Consequently, a data mart implementation is a rather quick solution to accessing and analyzing AR data.

The design and development of the data mart are the most crucial steps in the life of a data mart. Likewise, the most crucial aspect of the design and development stage is to obtain end-user input from the onset of the project and continue to use end-user input throughout the life of the project. Many well-built data marts with exceptional front-end software are never used because the data they contain is not information end-users want or can use. Nothing is wrong structurally or technically with the data marts - queries run quickly, tables can be joined easily, etc. However, the data marts were designed from a purely technical perspective, leaving out the only people who would eventually be using the system - end-users.

Data is not classified as information unless it is useful. A data mart full of numbers, addresses, and percentages means nothing unless the data is able to be manipulated, summarized, or ordered into a form that is meaningful to the user. For this reason, building a formal data model is advantageous, especially for data marts that will contain a large amount of data and be subject to predictable processing. The data mart model is strongly influenced by the department for which the model is built. The data mart, in essence, should contain whatever data end-users need to satisfy their information needs. Therefore, a diversity of data will exist in the data mart, such as summary data, detailed data, ad hoc data, and prepared data. Generally, data marts have mostly ad hoc summary data and prepared detailed data.

Performance expectations are different for data warehouses and data marts. In a data warehouse where there is an abundance of data and many users are exploring, the issue of performance is somewhat relaxed. This, however, is not the case for data marts. Given the small amounts of data and the nature of the queries run by end-users (mostly predictable), the expectation for performance is high in the data mart environment. Several methods of achieving above-average performance in a data mart are as follows:

- make extensive use of indexes
- use star joins
- limit the volume of data in the data mart (administrators should purge, archive, or condense data when necessary)



- create arrays of data
- create profiles of records (aggregate records)
- create pre-joined tables

The Basics of a Star Schema

A star schema is a manner in which to arrange the components of a database instance, named for the star-like arrangement of the entities. The center of the star is the Fact table, which contains measurable numeric values. [See Exhibit B for an illustration of a star schema centered around transaction information.] Connecting to the Fact table are Dimension tables, which provide descriptive information about each row in the Fact table. For example, a Fact table in a star schema from a retailing company might be populated with information regarding payment *amounts* and *quantities*. The Dimension tables would include the following data: information on customers who made purchases, information on the products that were sold, information on the location sales were made, information on the time-frame of the sale, etc. Fact tables usually have large amounts of rows, while Dimension tables are smaller in nature. This setup enhances performance, as one large table can be joined with a few small tables.

Steps in the Construction of a Data Mart

1. Obtain management approval – As everybody knows, without management approval, even the best planned and designed projects never get off the ground. Convincing key management personnel that constructing a data mart to improve the month-end close process will save the company money and improve reporting capabilities is an absolute must. Focus on how the data mart will help the department and company and avoid technical jargon. Several key benefits of a data mart over a data warehouse include the following: they cost less, they are completed quicker, and they serve a specific function for a specific department.
2. Define the purpose of the data mart in very explicit terms.
 - Which department will use the data mart? AR.
 - How will the AR department use the data mart? Primarily for month-end reporting and querying needs, but not limited to this.
 - How does a data mart assist with the close process? The close process entails running dozens of reports, which restricts processing power and time from all other users in that environment. With a data mart, once the close is complete, the necessary AR data is transferred from Production to the data mart. Now, the Production environment can continue with normal activities. All AR reporting needs will be taken care of in the data mart. Processing in Production will not be affected by processing in the data mart.
 - What other uses will the data mart have? Users will be able to use the data mart to run ad-hoc and as-of queries.
3. Name key personnel, both technical and functional – A strong team of knowledgeable people is necessary to complete the project in a timely and cost efficient manner. One essential



requirement of the data mart team is to have both technical and functional representatives. Without functional input throughout the entire project, the data mart will not be a success. Demand that feedback be given by all parties throughout the entire data mart project. In addition, ensure DBAs are involved and that space requirements (memory) are adequate.

4. Define a decision-making process and project plan – Will a data model be used? If so, will the data model be defined in house or will a standard data model be used? Once the decision whether or not to use a data model is made, team members should develop a project plan, focusing on the items in the data mart that will be used on a regular basis. The plan should include interviewing end-users, analysts, management, and technical personnel (including DBAs) to determine their issues and needs.
5. Require Leadership, Direction, and Responsibility – Who owns the project? Unless someone with authority owns the project and is responsible for its completion, the project will drag on for much too long. The design process should have an owner, the build process should have an owner, and the maintain process should have an owner. All owners should be held accountable for the completion of the project. Devise a time table/schedule and stick with it, unless major unforeseen problems occur.
6. Define the data mart structure – What schema will be used? How many data sources will there be? Will the data be normalized? How often will the data be updated? Will external data be utilized? Internal data is often enhanced by the addition of external data. Consequently, many companies have significant data integration challenges. Also, the data mart structure must be able to adapt and extend to support changing business needs, user requirements, and available technologies.
7. Write queries and update statements – Technical personnel should first use the interviews from Step 4 above to get an idea of how to structure the queries. Next, they should again consult with end-users to determine the purpose of the data mart and refine the data mart queries. All team members must keep in mind that the data mart has no business value if end-users are not satisfied.
8. Decide which front-end tool best suites the needs of end-users – Will end-users use the data mart more for reporting or more for ad-hoc querying? Are graphics necessary? How sophisticated and knowledgeable is the average end-user? Have end-users been exposed to other front-end tools?
9. Build the data mart prototype and test with end-users – Extract production data from the data sources and load it into the data mart. Allow end-users, analysts, and management to use the front-end tool chosen in Step 8 and run queries and reports. Verify that the output is correct. Feedback is very important at this stage of development. Request feedback from all those who test the prototype.
10. Make changes to ensure the data mart meets the needs of end-users – Use the feedback garnered from Step 9 above and make the necessary alterations to the data mart design. If feedback has been given by all parties involved throughout the project and, in turn, incorporated into each step of the process, then only minor changes should be needed at this point in time.



11. Test the data mart prototype again with end-users – Team members should define a set number of iterations that will take place. Too few iterations allowed will translate into a unsatisfactory result to end-users, while too many iterations will lead to frustration from all parties. A minimum of two iterations is suggested.
12. Fine tune the data mart – Use the feedback from each iteration to fine tune the data mart.
13. Ensure the data mart meets the goals defined in Steps 2 and 4 above and that end-users are satisfied with the results. Ensure each team member is satisfied with the results.
14. Make go/no-go decision – The owner of the project will make a go/no-go decision whether or not to place the data mart into the Production environment. All team members should provide input to this decision.
15. Document everything – Documentation is one of the most important aspects of the project. Unfortunately, it is also one of the least enjoyable and least liked steps in the data mart process. When the data mart structure needs to be changed in the future, the original documentation will be essential. Likewise, employees that were not part of the project team and new hires will need to be able to understand the process. Undoubtedly, original team members will leave the company, be transferred, or retire. Thus, the only method of passing down this key information concerning the data mart development is through documentation.
16. Conduct training – Training classes and/or seminars should be offered for several levels of users. A basic how-to course should be mandatory for all end-users, managers, and analysts. Then, depending on time and personnel available, an intermediate course should be available for the more frequent users and an advanced course should be available for the super-users.
17. Provide support – Both technical and functional support personnel will be necessary. Perhaps team members can implement a help desk or FAQ web site to assist users with common problems. If the data mart is a success, then more and more users will be added each month and the need for a support group will only increase.

Summary

The end-users need to be involved from the very beginning of the project when the data mart is being designed, through the development of the schemas and business areas, to the testing and final implementation of the data mart. The focus point throughout the entire project should remain on the subject matter of the data mart, i.e. the AR data that is necessary to meet the needs of the users. A few other important points to remember are as follows:

- Whereas a data warehouse is data driven, a data mart is a response to real business needs
- Data marts provide analysis on a narrower range of data than provided by data warehouses
- User requirements should dictate what type of data mart is built; the technology and design techniques should be based on these requirements – not the other way around
- Use real legacy data when building prototypes; build prototypes early and often
- The keys to data mart success are CLARITY, FOCUS, and FEEDBACK



About the Author

Kurt D. Duncan is a Consultant for Applications Consulting Specialists, Inc., located in St. Petersburg, FL. ACS specializes in the implementation of relational database packages. Kurt can be reached at kduncan@acsi-usa.com or via the company's website: www.acsi-usa.com.

References

'Data Warehouses and Data Marts: A Dynamic View', by Joseph M. Firestone, 1997.

'What is Data Marts', by Bill Inmon, 1996.

'Knowledge Through Data Warehousing: Financial Data Marts', by Stephen Cranford, 1998.

Further Information:

Web Sites = www.datawarehouse.com
www.dwinfocenter.org
www.knowledgecenters.org

Books = *90 Days to the Data Mart*
By Alan R. Simon

Understanding and Implementing Successful Data Marts
By Douglas Hackney



EXHIBIT A

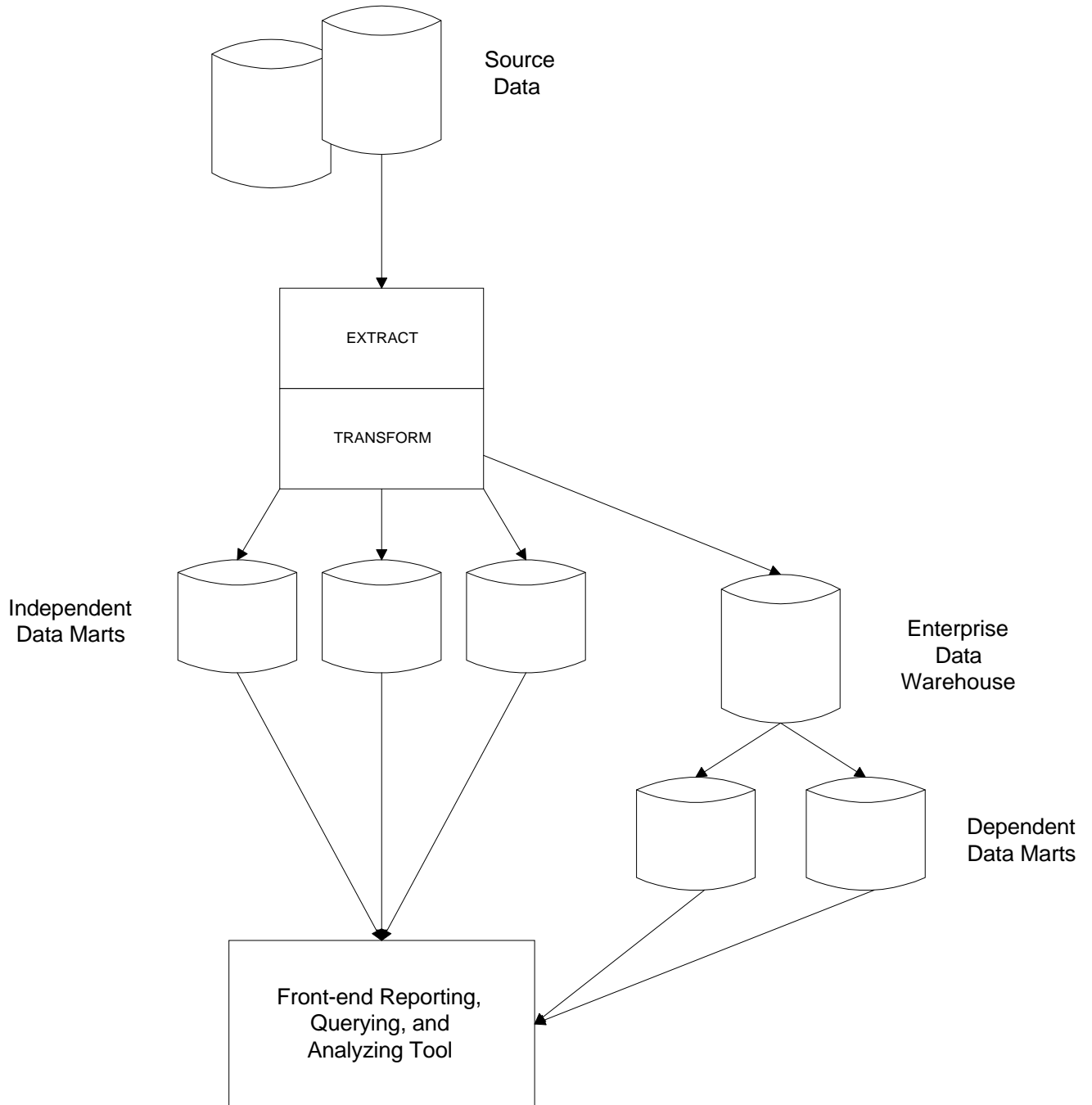




EXHIBIT B

