

Paper 116-29

8 Characteristics of a Successful Data Warehouse

Marty Brown, Lucid Analytics Corp

Abstract

Most organizations are well aware that a solid data warehouse serves as the foundation from which to build meaningful business and analytical intelligence. Unfortunately the gulf that exists between being aware of the need and actually building a data warehouse seems too large to cross. Instead of building a data warehouse, many organizations build their analytical solutions from spreadsheets and query tools.

There is a better way. This paper serves as a roadmap from which to bridge the gap between concept and reality. Included are brief discussions in the following areas:

- The executive sponsor
- The project team
- The business
- The data model
- The metadata
- Extract, transform, and load (ETL)^Q
- Marketing and change management
- The users

Once complete, the data warehouse allows organizations to unleash the power of their data in the form of analytical intelligence. It is at this point that they can begin to achieve a true competitive advantage.

Introduction

As organizations seek to move up the intelligence value chain, it becomes clearer everyday that their previous strategies for delivering information are becoming inadequate. What was once “good enough” for reporting and analytics has now been legislated into a more robust and controlled environment. Businesses simply cannot be run on spreadsheets anymore. Official reporting data that touches many hands is quickly becoming a thing of the past.

Enter (or re-enter as the case may be) the data warehouse. What was once primarily considered an IT centered project, now has the eye of the business. With the passing of legislation like the Sarbanes-Oxley Act, corporate executives must feel comfortable that the financial (and other public) reports that their teams prepare and that they sign are accurate. Many of the failed data warehouse projects of the past lacked true commitment on the part of the business. The fact is it is nearly impossible for a data warehouse to be successful without active participation from the data owners, i.e. the business, on a daily basis.

The focus of this paper is getting from here to there. It includes the roles the business plays as well as some suggestions on how the warehouse, once implemented, remains the firm foundation from which organizations build out the rest of the intelligence value chain.

Characteristic 1 - The Executive Sponsor

At the heart of the data warehouse is the executive sponsor. This is the person who believes in the value of the warehouse and is willing to use his/her political clout to fund the project. The executive sponsor is the one individual that must clearly understand the ROI of the project and be able to articulate to his/her peers.

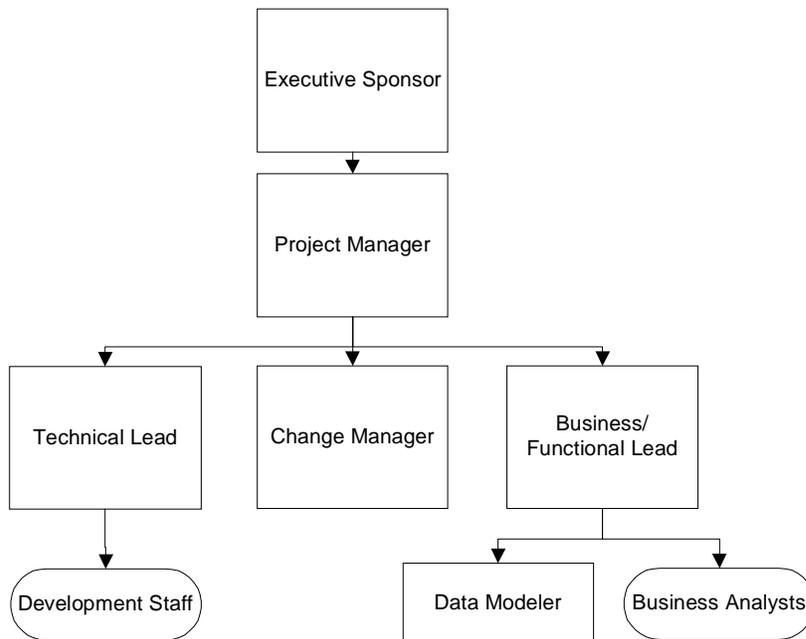
So what level of an organization does the executive sponsor represent? Generally it's the “C” level (CFO, CMO, CRO, etc) of an organization. By definition a data warehouse crosses over functional boundaries within an organization so it's extremely important that the executive sponsor carry enough clout and have a high level of buy in from his/her peers. It is important that there be an individual as the executive sponsor rather than a group of executives that collectively make up the “sponsorship”. If the scope of the data warehouse is going to cross all functional areas, then many times the CFO makes the best sponsor. This is

because this position is responsible for the financial stability of the entire company. Many would say that the CIO is an adequate executive sponsor since his/her responsibility already crosses the functional areas of the business. While the crossover is true, the CIO's job is to support the business rather than run the business and thus usually does not carry enough clout to effect the changes in business processes necessary to support the data warehouse. Also, having the CIO as the executive sponsor continues down the path of an IT centric view of the warehouse and does not clearly address the issues of business process and change management.

One important note is that the executive should be an active member of the data warehouse project from concept through implementation. It is very difficult to progress through the politics of the project if the sponsor's only involvement is to supply funding and receive monthly status reports.

Characteristic 2 – The Project Team

Equally important as the executive sponsor is the organization of the project team. Although the project team is made up of resources from the business and technical side of the house, for the purposes of the project, they cannot report up separate chains of command outside of the project manager. The following diagram depicts one way to organize the team for optimum performance.



The Project Manager

The project manager is responsible for the overall project. This person, in conjunction with the executive sponsor, manages the direction of the project and directs the project team. The project manager also removes obstacles for the project team and relays important information to non-project members of the company.

The Technical Lead

The technical lead is responsible for executing on the requirements defined by the business. The technical lead is also responsible for ensuring that the warehouse, once implemented, is supportable within the operational standards of the existing IT organization. The development staff reports to the technical lead and resolves technical issues and requirement clarification through the technical lead.

The Change Manager

The change manager is responsible for 2 things:

- Identifying and communicating business process changes throughout the business
- Marketing the new system to the user base

On the surface, these tasks may seem like something that can be done as the project closes and the users get engaged. In reality, however, this may be the single most important role on the project team in terms of preparing the business environment to both support the data warehouse through business process change and get the user base accustomed to getting their information from the warehouse versus their previous method.

The Business Lead

The business lead is responsible for a large number of things:

- Defining requirements for relaying to the technical team
- Ensuring the data model is of good form and supports all of the defined requirements
- Identifying business process changes in conjunction with the change manager
- Mapping the source data to the data model
- Defining the transformations (and business rules) to be applied to the source data prior to loading the data model
- Creating and maintaining all reference or cross reference data to be used in the ETL and/or reporting process

This position is key to the success of the data warehouse as well as its sustainability. A senior individual that has a working understanding of multiple functional areas of the business must staff it. Also, communication is key with this position.

The Data Modeler

Although this position technically reports to the business lead, it is mentioned separately in this paper because of its importance to the success of the data warehouse. Many data warehousing experts will readily agree that the data model is the center of the universe when it comes to measuring the success of the data warehouse. This person must have a remarkable understanding of business data as well as the ability to generalize and organize it to support the business requirements of the data warehouse.

Characteristic 3 – The Business

It is really the business that represents the sustaining factor in data warehouse success. As mentioned above, the business lead's team has an extremely important role in the development of the data warehouse. But.... What about sustaining its effectiveness after the warehouse is in production? That is where the business comes in. The responsibility for the ongoing accuracy of the data lies squarely on the business. After all, who will use the warehouse if its quality is in question? Many organizations actually create a department who is solely responsible for maintaining data quality. The lead position of this department is called a data steward (or data dictator in some cases). It is the job of this department to be obsessive about data quality. This department must have the continued blessing from the highest levels of the organization.

Characteristic 4 – The Data Model

As stated before the data model is the center of the universe for the data warehouse to be effective. Once a data model is defined, it must be enforced with rigor. Any changes to the model should be made with extreme caution. Generally once a data model is implemented, it can only be extended. Changes to the model usually have a major impact on existing process both business and technical. In reality, business process change should drive data model change.

While the data model in the cross business function warehouse is typically designed in third normal form, it is not uncommon for several smaller functional-area specific data marts to be created at the same time to facilitate query and reporting for the user community. These smaller data marts usually take the form of a star or snowflake schema.

Characteristic 5 – The Metadata

Metadata is defined as data about data. While most in the industry know about metadata, it is probably the most overlooked facet of data warehousing. Metadata is extremely important. It serves as a means of

communicating the ins and outs of the data warehouse to the users as well as the support staff. For example, if a user needs to know how net present value is calculated in the warehouse, he should be able to go to the metadata and find out.

For the technical support staff, the metadata serves as the documentation source for how the warehouse works. It defines names of data servers, application servers, etc. This allows for troubleshooting the warehousing process (i.e. ETL) by even the most novice members of the support staff.

Metadata, if maintained well by the data quality team, serves as an invaluable resource for the users in reporting and analyzing the business as well as impact analysis for those responsible for extending the data warehouse.

Characteristic 6 – Extract, Transform, and Load (ETL)Q

At the heart of the technical side of the data warehousing process is extract, transform, and load (ETL). This process is responsible for connecting to and extracting data from one or more transactional systems (source systems), transforming it according to the business rules defined through the business lead's organizations, and loading it into the all important data model. It is at this point where data quality can be either gained or lost. Of the many responsibilities of the data warehouse, the ETL process represents a significant portion of all of the moving parts of the warehousing process.

Although it is possible to code the ETL process in many languages, it is much better to obtain and use an ETL tool to aid in the management of the process. When selecting a tool there are several discreet criteria to measure against. The following list outlines what the authors deem as among the most important:

- Ability to read from and write to an unlimited number of data source architectures (flat files, VSAM, DBMS, etc.)
- Automatic capturing and delivery of metadata
- A history of conforming to open standards
- An easy to use interface for the developer as well as the functional user

While this list is not all-inclusive, it does represent some of the most important criteria when selecting a tool.

SAS Warehouse Administrator and Informatica represent the leading tools in the market today. Debuting with the release of SAS 9 (version 9.1) is SAS ETL Server and SAS ETL Studio. SAS ETL Studio represents a giant leap forward among all tools in the ETL space in terms of flexibility and usability. SAS ETL Studio, when combined with sound business process and a focus on data quality, exponentially increases the effectiveness of the data warehouse. This can be described by the formula (ETL)^Q.

Characteristic 7 - Marketing and Change Management

The next characteristic of a successful data warehouse is one of public relations. In large organizations it is imperative that the user base be made aware of the progress of the data warehouse project as well as which functional areas are online and ready for use. Even the best built and maintained data warehouses can fail without a high level of use by the business.

One of the areas that keep users from turning to the data warehouse for information is a perceived lack of data quality. Whether or not there are data quality issues, any sign from a user that issues exist put the data warehouse at risk. One way to mitigate this risk is to publish data quality metrics on a regular basis. These reports should show marked improvement in data quality from the previous report.

It is the job of the change manager to keep the communication flowing and be ready to address issues as they crop up long before they are blown out of proportion. This can best be done, at least at first, by having regular meetings with representatives from the user base from each functional area. Publishing minutes of these meetings to the entire user community helps to facilitate communication as well.

Characteristic 8 – The Users

Is a data warehouse successful if it measures high on loading it with high quality data on a regular basis? The answer is a resounding NO. A data warehouse is useless if the users can't get the data out. Therefore it is extremely important that this issue be addressed in the planning phase of the project.

There should be multiple ways to get data out of the warehouse each of which is tailored to a specific grouping of users. The executive is less likely to want to do detailed analysis so electronic delivery of static reports or a few OLAP cubes should suffice. The business user, however, will likely want more sophisticated OLAP capabilities as well as the ability to design and run some queries. The analyst will require even more advanced data acquisition tools such as SAS Enterprise Miner to aid in predictive analysis through statistical modeling.

In order to obtain continued support for the data warehouse, the level of usage should also be measured and communicated on a regular basis.

Conclusion

A successful data warehouse can be a difficult thing to achieve – overcoming the sins of the past as well as avoiding the ever-present silver bullet mentality. It is hard work grounded in thorough business understanding of the value of data, data quality and data organized in ways that facilitate reporting and analysis. No matter the motivation – legislation, economic stress or true business vision – the foundation of a successful data warehouse is business understanding.

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. [] in the USA and other countries. ® indicates USA registration. Other brand and product names are trademarks of their [] respective companies.