

## Measuring the Success of a Data Warehouse



It is difficult to determine success metrics for a data warehouse. Certainly there is evidence that return on investment (ROI) is frequently used to determine warehouse success, but there are many potential metrics that may be used. For example, it is possible to define financial measurements, usage measurements, customer satisfaction, availability, performance, response time, etc., as reasonable warehouse success indicators. However, many of the possible measurements can be interpreted in different ways. For example, a decrease in queries against a data mart may indicate that use is decreasing, or it may indicate that the users are so happy with the data that they find it more efficient to download it to a cube on their desktop for easier access, thus increasing usage. Taking action based simply on query volume may lead to the wrong actions being taken. In order to obtain as objective a view of warehouse success as possible, the Balanced Scorecard, as described by Robert Kaplan and David Norton, can be modified to measure the data warehouse.

The metrics presented in this article are not the only ones that can be used to measure the warehouse. Rather, these should be seen as a set of example measurements that can be refined over time as the organization's perspective on the data warehouse changes and matures.

### **The Balanced Scorecard**

Robert Kaplan and David Norton proposed the Balanced Scorecard in a series of articles in the Harvard Business Journal. They took the position that traditional financial metrics which had been applied to measure business success did not give a complete picture of the true state of the business. Instead, they proposed the Balanced Scorecard as a way to solidify an organization's focus on future success by setting objectives and measuring performance from four viewpoints or perspectives: financial, customer, internal and organizational learning. These perspectives are diagrammed in Figure 1.

<b>The Balanced Scorecard</b>	
<b>Financial Perspective</b> Measures the ultimate results that the business provides to its shareholders.	<b>Customer Perspective</b> Focuses on customer needs and satisfaction as well as market share.
<b>Internal Perspective</b> Focuses attention on the performance of the key internal processes which drive the business.	<b>Organization Learning</b> Directs attention to the basis of all future success - the organization's people and infrastructure.
Figure 1	

In order to achieve the goal of measuring the data warehouse, the measures described in Figure 1 have been modified and have made the assumption that a standard architecture and methodology for warehouse construction have been defined. If this is not true, other internal perspective metrics should be defined. Given that, the financial perspective will focus on the costs to build the warehouse and the financial benefits gained from using it. The customer perspective will measure user satisfaction, usage of the warehouse and adherence to service level agreements. The internal perspective will focus on the amount of reuse of warehouse processes and procedures and adherence to the architecture and methodology. The organizational learning perspective will measure warehouse communication and the training of IT people in the tools and techniques required for warehouse construction and maintenance. Given these definitions, the graphic for the Data Warehouse Scorecard can be defined as shown in Figure 2.

<b>The Data Warehouse Scorecard</b>	
<b>Financial Perspective</b> Measures costs to build the warehouse and the benefits accrued from its use.	<b>Customer Perspective</b> Focuses on warehouse usage and user satisfaction.
<b>Internal Perspective</b> Determines the amount of process/procedure reuse and the adherence to the warehouse architecture.	<b>Organization Learning</b> Measures IT awareness of the warehouse, and the readiness to build and maintain it.
Figure 2	

### Financial Perspective

The financial perspective focuses on two basic measurements: construction/maintenance cost and financial benefit. Of these two

measurements, only the construction/maintenance cost can – and should – be collected at least partially by IT. A second part of the financial cost, provided by the users of the system, is the business cost to help build, maintain and use the system. Although the users will gain the benefit for the system, it may be necessary to have user experts assist in developing and maintaining end-user packages for less experienced users. This is a part of the warehouse “total cost of ownership” and should be included in the financial perspective. The financial benefit measurement must be completely projected and collected by the users of the data warehouse application.

Many warehouses are measured by ROI. This is a valid measurement of warehouses. However, using ROI as the primary measurement of the warehouse is sometimes seen to imply that a single organization can be responsible for producing the measurement. Therefore, while ROI can be used for warehouse measurement, the cost and benefit measurements should be collected separately and the ROI calculated from the results.

### Construction/Maintenance Cost

Construction costs should include all the costs involved in building and maintaining a data warehouse. These include:

- Personnel (internal and consulting)
- Servers
- Network
- Tools
- Maintenance Licenses

Obviously, some of these costs are easier to collect than others. It is fairly simple to track how many people (or fractions of people) it takes to both build and maintain a system. Hardware and software costs are harder to collect. Many warehouses are built incrementally, using hardware and software already purchased. If these costs are allocated to the early warehouse projects, they may become cost-prohibitive. Therefore, the hardware and software costs may be apportioned to warehouse projects as they use them.

The other way of allocating the hardware and software costs is to take them as a “cost of doing business,” and just allocate them to the IT budget, only charging people costs back to each project. This method makes it easier to cost-justify warehouse projects, but has the disadvantage that the true cost of the application is not calculated.

### Financial Benefit

Although traditionally IT has been asked to determine ROI on a project, in fact IT can only realistically estimate and collect part of the cost side of the equation.

The financial benefit to the users should be estimated and collected by them. This benefit can take different forms: it may be estimated in terms of greater revenue, decreased costs or some combination of both. However, in order to determine the success (or otherwise) of the warehouse, it is necessary to have both a preliminary estimate of the financial benefit and a means of measuring the benefit in place.

## **Customer Perspective**

The customer perspective can be measured by three metrics: usage, service level agreement (SLA) adherence, and customer satisfaction. IT can collect all these measurements, with the participation of the users of the implemented applications.

### Usage

As discussed earlier, measuring usage of the warehouse can produce results that can be interpreted in opposing ways. Therefore, instead of trying to measure query volume or the number of people accessing the warehouse at any one time, the usage measurement should be defined in as simple a manner as possible. To that end, measurements to be collected for usage can be defined as: total number of enabled users and number of enabled users who access the warehouse at least once per update cycle.

Although these measurements do not give an indication of how busy the warehouse is, they are less subject to interpretation than other usage numbers. Note that these measures may be made for the warehouse as a whole or for individual warehouse components, such as servers or data marts.

### Service Level Agreement Adherence

SLAs should be negotiated for each warehouse implementation. They should cover (at least): response time, population frequency and availability. Obviously, these agreements define the user requirements for the warehouse. In order to be able to view the warehouse from the customer perspective, it is necessary to put mechanisms in place to measure the adherence to SLAs. These procedures should include both automated capture of information such as response time or system availability, and manual “anecdotal” information from the users which may give an indication of potential problems not captured by the automated tools. This anecdotal information should be collected during regularly scheduled meetings with warehouse customers as well as keeping records of user complaints about SLA adherence.

### Customer Satisfaction

Surveys are a proven method of measuring customer satisfaction. Depending on the function of the implemented warehouse system, there are many questions that could be asked. However, at a minimum, the following areas should be covered:

- ❑ Overall satisfaction. (It is always important to measure overall satisfaction, as the users may be satisfied with the system even if they see room for improvement.)
- ❑ Data quality satisfaction.
- ❑ Number of times the user uses the warehouse data (not how often does the user access the warehouse).
- ❑ Does the application meet its goals?

The survey may also be used as one of the vehicles to measure the SLA adherence, but it should not be seen as a replacement for regular user meetings.

### **Internal Perspective**

The internal measurements are those that IT uses to measure its own success in warehouse implementation. These may differ from one IT organization to another. However, making the assumptions on methodology and architecture defined earlier, the measurements of the internal perspective may be defined as reuse and methodology/architecture adherence. They are largely IT measurements and should be produced and collected by IT.

### Reuse

The ability to promote and institutionalize reuse of warehouse components and techniques is often cited as a goal of a warehouse implementation organization. The only way to determine if the organization is meeting its goal is to measure the amount of reuse that is being realized. Reuse can be measured for (at least) the following components:

- ❑ Population Modules: extract routines, transformation rules, load routines and scheduling components
- ❑ Business Intelligence Modules: screens, reports, Web-publishing components and on-the-fly calculations
- ❑ Database Designs: data warehouse, data marts and OLAP “cubes”

### Methodology Adherence

To some degree, methodology adherence may be seen as a subjective measure. However, with any methodology, there will be some concrete deliverables

required, and they may be used to give a more objective measure of methodology adherence. For example, these components may include:

- Data and business information models
- Service level agreements
- Statements of work
- Performance engineering models
- Development plans

### **Organizational Learning Perspective**

Because this Balanced Scorecard is being directed at the data warehouse, the organizational learning perspective should be focused on information specifically about the Warehouse. Therefore, the required measurements are defined and collected by the warehouse management team. These measures are communication and training.

#### Communication

Communication, seen from this perspective, is the number of people in the IT organization who have been communicated to on any given warehouse subject. These subjects may include strategy, architecture, methodology, standards and guidelines, tools and techniques, etc. For each defined subject, the number of people communicated to should be tracked.

#### Training

The training referred to here is not training the users in the delivered applications, rather the ability of IT people to be active participants in the construction of the warehouse. Therefore the specific measures collected for the training portion of the organizational learning perspective are:

- Number of people trained in population tools and techniques
- Number of people trained in business intelligence and techniques
- Number of people trained to support the warehouse DBMS
- Number of people trained in architecture/methodology

These measures may be enhanced by adding the number of people in the organization who have experience in warehouse construction and/or maintenance projects.

#### Targets

It is necessary to set targets for data warehouse metrics because if they are not set, there is no way to tell if a response time or availability figure is good or bad. For example, if customer satisfaction is 80 percent, that in itself doesn't tell us

anything. If the target satisfaction Index is 98 percent, then 80 percent is probably unacceptable, if it were 60 percent, then 80 percent is very good. A figure of 25 people trained in population tools is probably good if you need 20, but bad if you need 40. These types of arguments are valid for all metrics proposed in this document.

Therefore, if targets cannot be set for a particular measure, then there is little point in measuring the category. This is true of all proposed metrics, including those that are the user's responsibility to define and collect. Note that failure to collect one measure may affect the usefulness of other measures, particularly those in the same perspective. If the users are unable to define a benefit, then ROI will be impossible to calculate.

## **Conclusion**

Metrics provide a means of measuring data warehouse success. Without the ability to measure objectively and consistently, it is impossible to define whether or not the data warehouse is adding value.

The Balanced Scorecard provides us with a proven means for measuring the success of an organization. While there is little empirical evidence that this is translatable to a data warehouse measurement system, the "perspectives" quadrants give us the ability to measure success factors in more than simply a financial area. Historically, efforts to provide data warehouse measurements have proved difficult to implement largely due to the difficulty in defining financial measures. The greater balance afforded by the scorecard suggests that this may be a better measurement mechanism for the data warehouse.

The specific measures defined in this article are not meant to be an exclusive definition of all measures that will ever be appropriate for measuring the data warehouse. Rather, they should be perceived as a starting group of measures that will be refined over time, both as the warehouse and sophistication of the organization increases.

**About Threshold**

Threshold Consulting Services is a leader in Data Warehousing (DW) and Business Intelligence (BI) solutions. Since 1995 Threshold consultants have delivered pragmatic solutions that leverage corporate data to deliver insights and improve business results. Threshold understands how to acquire information, edit it, organize it, store it and make it easily accessible to those who need it – when they need it. We utilize a tools-centric, cost-effective approach which leverages industry leading solutions and minimizes labor intensive programming.



Visit [www.thresholdcs.com](http://www.thresholdcs.com) for more information about Threshold Consulting Services