

Measuring Data Warehouse Return on Investment

A Whitepaper by Sid Adelman

[Executive Summary](#)

[Introduction](#)

[Section 1: Data Warehouse Costs](#)

[1.1 Cost Estimates as Part of Cost Justification](#)

[1.2 Costs](#)

[1.3 Total Cost of Ownership](#)

[1.4 Post-Implementation Cost Measurement](#)

[Section 2: Benefits](#)

[2.1 Estimating Tangible Benefits](#)

[2.2 Estimating Intangible Benefits](#)

[2.3 Post Implementation Benefits Measurement](#)

[Section 3: Full Cycle Assessment of Value](#)

[Appendix A – ROI Calculation Process](#)

[ROI Example](#)

[Cost of Capital](#)

[Risk](#)

[Net Present Value](#)

[Internal Rate of Return](#)

[Payback Period](#)

[Appendix B - Cost Calculation Template](#)

[Appendix C - Benefits Calculation Templates](#)

[Appendix D – Glossary](#)

[References](#)

Executive Summary

Management wants to know the business value of the data warehouse, measured in quantitative terms, with the metrics in the form of return on investment (ROI). ROI refers to a few different ways of calculating the value of an investment which can be calculated by payback period (break-even analysis), net present value, or rate of return (yield).

Cost justification - Many data warehouse projects were implemented without knowing how much they would cost. Many were authorized without even estimating the potential benefits. However, ROI should become part of normal cost justification and should also be part of evaluating the system once its up and running. Without an ROI process, the organization may lose business value because the wrong project was implemented.

Data warehouse costs – The costs for the data warehouse can be substantial including hardware, software, internal personnel costs, consultants and contractors, network, training for both IT and the business users, help desk, and operations and system administration.

The total cost of ownership (TCO) – TCO extends beyond initial implementation; the yearly cost for the maintenance of the data warehouse often runs between 40% to 60% of the initial cost. The data warehouse must be maintained for every year it delivers service to your users. It's important to know what the total costs will be, not just the original delivery price tag.

Benefits - The data warehouse can deliver substantial benefits, both tangible and intangible. A project should not begin without the expectations of these benefits. Measuring them is a challenge, but there are best practices being used by leading practitioners from which others can learn. Data warehouse benefits take the form of cost savings through increased operational efficiency and data mart consolidation, as well as business improvements such as fraud reduction, increased customer retention, and the implementation of new business initiatives that increase revenue opportunities. This paper has concrete examples in each of these categories.

Measuring intangible benefits is far more difficult and while the metrics are illusive, they are nonetheless important and real. These intangibles include improved shareholder and Wall Street analyst relations, better and faster decisions, improved customer service, employee empowerment, and getting the jump on the competition.

Post-implementation measurement – Very few organizations measure the ROI of their projects after they are implemented, but by documenting the estimated costs and anticipated benefits, and then measuring the actuals against them, we have a solid understanding of success as the ROI is realized. Post-implementation cost and benefits measurement is crucial to determine the ROI from the project and also helps with future goal-setting, benefits estimates, cost estimates, and plans for proposed expansion and enhancements. The measurement should provide guidance for the organization's process of future estimating and budgeting.

ROI calculation - The ROI calculations are fairly straightforward and should include the cost of capital and the risk associated with the project. The specific ROI metric used (payback, net present value, internal rate of return) should focus on the calculation favored by the CFO and should be the method the organization uses to prioritize its other projects. However, it is often useful to use more than one method because one by itself provides a limited picture. For example, payback period gives no credit to the benefits achieved after the payback has been satisfied. The Appendix gives an example and provides the methods to calculate ROI. The Appendix also has a cost calculation template, benefits calculation templates, a glossary, and a set of references.

Introduction

Measuring return on investment (ROI) is relevant as an organization cost justifies potential data warehouse applications. It is also relevant for post-implementation evaluation. Many of the same procedures and equations can be used for both purposes. This whitepaper will provide an approach to cost justify a data warehouse project and to determine ROI post implementation.

Expected ROI should be a major determinant in evaluating options and making decisions about which architectures to employ, which vendors to choose, and which projects to pursue and how they should be prioritized. Additionally, some companies measure their managers on the success of their projects and bonuses are often tied to the ROI of those projects. These managers want their projects to be fairly and fully represented and measured.

The project to be measured could be a new project, or a re-implementation of an unsuccessful or partially successful system, or simply an existing project whose ROI has never been established. It's important to establish expectations for both costs and benefits *before* a new project is launched. By establishing these expectations, organizations will be able to determine if the project was successful. Measures should *never* be re-cast in an attempt to show success regardless of the actual outcome.

ROI can be demonstrated through a variety of means. This paper will present three of the most commonly used techniques: payback period (or break-even analysis), net present value, and rate of return (or yield).

Section 1: Data Warehouse Costs

The expenses for the data warehouse will vary widely. The cost will be dependent on the size of the database, the number of users, the complexity and quality of the source data, the software tools employed, the need for consultants and contractors, the capabilities of the team, and how well the system is supported and maintained.

It's necessary to understand how costs will be accounted for. Some costs will be expensed immediately, and others will be depreciated over the expected life of the system. Costs will appear in different accounts and all these factors will become important when the actual total costs are tabulated and the actual ROI is calculated.

When there is uncertainty about costs, or to validate your cost estimates, call references to find out what costs they experienced in each of the cost categories. This will not only help you estimate your costs but will also help explain the costs to management.

1.1 Cost Estimates as Part of Cost Justification

It's important to fairly estimate costs of each project since a significant underestimation could launch a project that will result in loss to the organization. Alternatively, an overestimation could kill a project that would return major value. Since cost justification will influence governance, incorrect estimates will affect project prioritization. We'll look at both technology costs and personnel costs.

1.2 Costs

1.2.1 Hardware

For the data warehouse, you will need CPUs, disks, networks, and workstations. The hardware vendors can help size the machines and disks. Some hardware vendors have benchmark capabilities that will help you estimate your hardware requirements. Be aware that unanticipated growth of the data, increased number of users and increased usage will explode the hardware costs. Some vendors, such as Teradata usually bundle the hardware along with the RDBMS.

The amount of the disk required will depend on the raw data and will also depend on the number and size of the indexes required to satisfy performance requirements. It will also depend on the

need and usage of summary tables and on the need for working disk storage. The multiplier can vary up to six times the amount of raw data. The need for indexes and summary tables is highly dependent on the RDBMS as some have no performance requirement for either.

If existing desktops and laptops are adequate to support end users, no additional costs should be charged, but if upgrades or new machines are required, the additional costs should be assigned and depreciated over the expected life of the system. Three years is often used as the expected life, even though the system will probably last longer. The calculation is the cost to purchase or upgrade times the number of anticipated users.

Do not ask the query tool vendor for the minimum desktop configuration; ask for the *recommended* configuration.

1.2.2 Software

The data warehouse always needs an RDBMS. The way the RDBMS is priced, i.e., by node or by the number of users can influence the cost of the RDBMS. End user access and analysis tools such as Business Objects, Brio, Cognos, Crystal Decisions, and MicroStrategy usually price by the number of registered users or price by server. Whereas a new tool may not be required, an upgrade or add-on to user software may be necessary.

Most products have base prices and many add-ons. Many installations choose an extract, transform, and load (ETL) tool such as Informatica or DataStage rather than writing their own ETL code. Add-ons with the ETL tools could include additional costs for each different type of source file or target database. These tools are often priced based on the operating system and size of the machine. Additional tools are often needed for data cleansing and performance monitoring. An organization may decide to forego buying one of these tools but should consider the additional cost of developing and maintaining their own code.

1.2.3 Network Usage

Introducing the data warehouse usually requires network bandwidth, which may already be in place but additional network capability may be required especially as the number of users increases. The architecture of the system can have a major impact on network usage; a federated data warehouse will require more network capacity than will be required for a centralized data warehouse. For a fair cost allocation, all data warehouse use of the network should be included in the cost. The network people are usually in a position to estimate network usage and to assign costs. Use their input for estimating network usage expense.

1.2.4 Personnel Costs

Internal Staff

The fully-burdened rate (salary plus taxes, benefits, support costs, etc.) for the IT folks associated with the project should be included in the project cost. Business personnel are usually not included in calculations for personnel costs, but any help desk staff in the business organization should be. Management will be spending significant time devoted to the data warehouse projects but management time is rarely included in the cost calculation.

Internal people are getting paid regardless of whether or not we use them on this project - why should we include their costs in our budget? There is an opportunity cost of assigning them to this project vs. another project. Include the fully-burdened costs of the people on your project. Your project should, at minimum, include the costs for a data modeler (data administrator), DBAs, ETL people (data warehouse application developers), a business analyst, an access and analysis person, and a project manager.

Consultants and Contractors

Consultants are engaged to help determine requirements, help plan the project, create the scope agreement, cost justify the project, help select the software, and establish the initial and long-term architectures. Consultants are typically more expensive than contractors but usually don't remain on projects as long. Contractors are brought in to supplement technical skills, specifically for software such as the RDBMS, the ETL tools, and the access and analysis tools. A primary role of the contractors should be to transfer their skills to the organization's employees. The cost for contractors will be dependent on how deficient the organization is in the required skills, how fast the organization needs the system implemented, and how long it will take to transfer skills once the implementation is complete.

1.2.5 Training

IT training

IT training will be required for access and analysis tools, ETL tools, and the RDBMS. Your organization may already have DBAs trained and experienced in OLTP usage, but you may require data warehouse-specific training such as the data structures, security, data models, data placement, RDBMS specific tuning, and the different types of user data access. IT personnel will need enough time to work with the products to become proficient in their use. The vendor cost of training is sometimes included in the price of the tool. The more knowledge transfer you receive from the vendor, the lower your ramp-up costs.

The cost for IT training should include:

1. Registration fees
2. Travel and living expenses
3. Fully burdened personnel time including class time, preparation time, travel time, and time to become proficient.

User training

End user training should not only include class time on tool usage, but the additional training that will make the difference between effective and incompetent end users. The additional training should primarily focus on *understanding the data to be accessed*. It should also cover performance issues, security, meta data usage, how and when to call the help desk, and the use of the query and report libraries. This additional training will normally require one more day beyond the training for the tool. While training on the tool itself could be outsourced, internal personnel should teach the company-specific modules. Workshops using a subset of the users' data rather than unrelated sample data will keep learners focused.

User training is usually done on-site. Costs include:

1. Registration fees, or training development costs for internally developed courses
2. Training delivery costs including trainer costs, facilities and equipment costs, and travel costs.
3. The time users spend away from the job and the time it takes them to become proficient.

Help desk/support

Most organizations have found it necessary to upgrade and train their help desks to be able to support the data warehouse efforts. The help desk personnel must know the supported tools, they need to know the data, they need to be familiar with the query and report libraries, and they should know something about the users and their level of training. Data warehouse user support will be an ongoing expense. Depending on the user activity, the effectiveness of the training, the ease of use of the system, and the need for hand-holding, the ratio of help desk to active users usually ranges from 1-to-20 to 1-to-75.

Operations and system administration

This is a grab-bag of roles and costs including monitoring the system performance, executing backups, running the ETL programs, administering security, managing the query and report libraries, administering the meta data repository, dealing with the vendors, and assigning charge-backs. These categories of costs are higher in the initial implementation but exist for the life of the data warehouse. Consultants can provide suggestions on how to estimate this cost category, but organizations similar to your own can give you the best idea of what these costs will be.

1.3 Total Cost of Ownership

The total cost of ownership (TCO) of the data warehouse goes beyond initial implementation. Most organizations focus on the cost to implement the initial application but give little thought to ongoing expense. Over a period of years, the continuing cost will very likely exceed the cost of the initial implementation. The annual cost for the maintenance of the data warehouse often runs between 40% to 60% of the implementation cost. Maintenance must be sustained for every year the data warehouse delivers service to your users. The cumulative TCO is particularly onerous for installations with multiple siloed data marts. CFOs usually only allow benefits for data warehouses to be calculated for three years but the TCO extends for the life of that data mart or data warehouse. It's like a mortgage you must pay until you abandon your house.

Maintenance includes the contractual maintenance cost of the hardware and software (usually 15% to 20% of the software retail or purchase price). If the hardware was purchased, the depreciation of the hardware should be included. For leased hardware, the leasing costs must be included. New or upgraded software and hardware may be required if the system does not perform as expected or if the usage and complexity go beyond initial estimates. Technical personnel will always be required establish and run backup and recovery procedures, to monitor and tune the system, and, with the normal turnover, there will be an additional cost with the introduction of new technical people. There will always be requirements for, and costs associated with, assimilating new data, new capabilities, and new users into any successful data warehouse.

The database will not remain static. Anticipate growing your environment. The growth will be in the number of users (web delivery could significantly increase the number of users), the requirement to perform more complex queries, an extension to users beyond your enterprise (for example, customers and suppliers through an internet capability). New data will be added, sometimes more than for the initial implementation. The design most likely will change, and the database will need to be tuned. Additional historical data will increase CPU and disk requirements. New software will be introduced, new releases will be installed and some interfaces will have to be rewritten. As the data warehouse grows, the hardware and network will have to be upgraded.

At some point, you may need to migrate to a more robust and better performing platform. Migration is always costly. The lost time and opportunity cost of not having the system deployed and available could far exceed the money wasted on a prototype or “phase one” implementation. Migration should never be considered as an alternative to fully understanding how much capacity and capability will eventually be needed or as an alternative to implementing a scalable configuration in the first place. All these factors must be considered in the TCO. These costs may sound onerous, but if implemented well, a data warehouse can provide financial benefits that dwarf the costs, even in the first year; there are examples in production in the marketplace today. Balancing the cost and value so that the resulting ROI is significant is key to achieving a successful data warehouse.

1.4 Post-Implementation Cost Measurement

After the initial project is implemented, a cost assessment should be performed. As part of cost justification, the organization *estimates* costs, but post implementation, you will want to measure *actual* costs incurred. This measurement will be an important factor in determining if each project phase was a success, which means that a cost assessment should be performed for each major phase. This measurement is crucial to determine the ROI from the project and to highlight areas where costs have been excessive. This will help with future cost estimates.

You may have needed more hardware than the vendor had estimated. Rework may have ballooned personnel and contractor expenses. The network may have had to be expanded to facilitate unanticipated line traffic. Data volumes typically grow faster than planned, and the cost to clean up the data is almost always underestimated.

Post-implementation cost measurement is not a one-time activity. If the project takes more than three months, the spending run-rate should be monitored, especially for the consultant and contractor costs and the internal personnel costs. These costs can quickly exceed budgets if the project is not properly managed.

Section 2: Benefits

If the right project is chosen and the implementation is designed to meet the business needs, it's likely that the benefits of the data warehouse will significantly outweigh the costs. However, to be successful the project must be well planned and executed and the benefits must be identified to be sure that the investment will truly pay off. Data warehouse benefits, whenever possible, should be related to improvements in the effectiveness or productivity of the business analysts. Clearly, these essential metrics for measuring the value of the data warehouse need to be agreed upon in advance of implementation in order to avoid confusion over value attribution.

The data warehouse used primarily as a report generator has little inherent benefit. However, a data warehouse can enable analysis which provides information and knowledge to generate more and better leads, identify valuable trends, improve asset management, or negotiate better prices. That knowledge in conjunction with new programs and effective action is what generates the value, and it must be recognized that the data warehouse is a necessary enabler, thus a major contributor to cost savings or increased revenue.

2.1 Estimating Tangible Benefits

The tangible benefits are those that have hard dollars associated with them. The following are examples of such tangible benefits.

2.1.1 Revenue enhancement

Improved marketing can result in more revenue per customer resulting from increased spending, and a greater share of the customer's wallet. As a side benefit, the increase in wallet share will be at the expense of your competitors. Selling higher margin products, focusing on the more profitable customers, and turning unprofitable customers into profitable ones will all enhance revenue.

An example of revenue enhancement is the additional state tax receipts produced by a data warehouse that identified businesses that had not paid or had underreported their taxes. The Texas Comptroller of Public Accounts was initially able to collect an additional \$94,000,000 in revenue as a direct result of the data warehouse. 1

2.1.2 Cash flow acceleration

Accelerating cash flow always results in more effective use of working capital which, in turn, results in greater profit to the organization. The data warehouse can accelerate cash flow by more

efficiently managing the supply chain including inventory management, increasing inventory turns, managing suppliers, managing product shipments, backlog control, more accurate demand forecasting, and managing cancellations. Cash flow is accelerated with improved control of accounts receivables with an emphasis on time-to-collect. *Time to value directly impacts cash flow acceleration since a shortened time to deploy a valuable data warehouse application means the benefits accrue earlier.*

3M used their data warehouse to improve demand planning and were able to reduce inventory from four months to a little over three months. This resulted in an accelerated cash flow of \$437,000,000. 3M set a challenging goal to reduce inventory and met the goal. 2

2.1.3 Analyst productivity

In the past, business users, analysts, and knowledge workers had to spend 80% of their time gathering data with only 20% left over to perform the analysis. With some data warehouses, we are seeing those numbers reversed. Depending on the degree of consolidation and integration of the data, and the availability of useful meta data, analysts now may spend only 20% of their time gathering the data. This does not mean firing analysts. Instead, they can now address the backlog of questions that were left unanswered and, in many cases, not even asked. It also means the questions can be answered in a more timely fashion. To be able to make comparisons, it's important to be able to measure the productivity both before and after. A very rough metric is the number of queries, reports, or analyses performed per day, however, this measurement does not account for the value or complexity of the analyses which should be considered elsewhere.

A major factor associated with productivity is query response time. If a query takes too long, the effectiveness and productivity of the analyst may be severely impacted. This is particularly true for iterative queries where the result of one query generates ideas for the next query. For the data warehouse to truly deliver its promised productivity, response time must satisfy the response time service level agreement (SLA). While some queries need a response time within seconds, (for example, call-center inquiries), for others, minutes or even hours are acceptable (for example, monthly sales reports).

If the fully-burdened rate for an analyst is \$100,000/year, and the productivity improvement is estimated at 25%, the annual estimated value of the data warehouse for analyst productivity is \$25,000 times the number of analysts using the data warehouse. Highmark Blue Cross Blue Shield, experienced a 3:1 productivity improvement for their analysts from their new data warehouse, giving them measurable before and after comparisons.

2.1.4 Cost containment

While revenue must be balanced with the cost required to produce that revenue, cost savings flow directly to the bottom line. The data warehouse can help to control costs in a number of areas. In each area, the organization must know its cost of doing business, estimate the savings per situation and extrapolate the cost savings. The following are industry and general business examples of areas where costs can be significantly contained:

Retail

- Having a better understanding of customer purchasing patterns, timing of purchases and the specific products that will be purchased, an organization can minimize its inventory.

Distribution

- Correcting name and address data will reduce the fines from shippers who charge for incorrect zip codes. *For example, using correct addresses and reducing the \$2.00 fine on 100,000 shipments, results in \$200,000/year savings.*

Internal Corporate Costs

- Knowing more about employees, their productivity and contribution to profit, organizations are in a better position to negotiate contracts with their unions and are also in a better position to make Human Resource decisions regarding head count and use of temporary services.
- Having better information about employees, their financial position and desires, organizations are able to structure more cost-effective pension and retirement plans and are better able to negotiate contracts with health maintenance providers.
- Improving the quality of products and services reduces warranty and service costs.
- Having more accurate and complete knowledge of their operations, organizations are able to reduce governmental fines and reduce contractual penalties from suppliers and customers.

Customer Relationship Management

- Knowing more about customers allows organizations to minimize the number of promotional mailers and still achieve target sales. *For example, reducing the number of \$3 mailers by 100,000 results in a cost savings of \$300,000 for each promotion.*
- Using more cost-effective channels to deliver services (e.g. web, telephony). *For example, moving 10,000 customer calls per month from the live person channel, estimated at \$8.00 per call, to web services, estimated at 10 cents per call, results in a \$79,000 savings per month.*

2.1.5 Demand chain management

The data warehouse should improve demand forecasting, and with this capability, companies are able to reduce their inventory costs. Knowing their inventory carrying costs, companies are able to calculate the cost reductions attributable to the data warehouse.

Burlington Northern and Santa Fe Railway was able to significantly reduce crew expenses, specifically overtime expenses, by identifying problem areas and better understanding the work histories of the crews. 3

Anthem Blue Cross/Blue Shield was able to negotiate more favorable rates from their providers (doctors, hospitals) by better understanding claims and costs. 3

2.1.6 Fraud reduction

The data warehouse and specifically data mining have been used to detect fraudulent insurance claims and fraudulent credit card usage. Analysis of claims has identified fraudulent health and workers' compensation claims coming from specific doctors and lawyers. The types and patterns of the claims alert the investigators, who then conduct a more thorough audit to uncover fraud and abuse. Mining current and historical data together has identified profiles of usage that can indicate that a credit card is stolen sooner than it would be detected through simple transaction monitoring. Anthem Blue Cross/Blue Shield has been able to uncover fraudulent claims from providers. 3 Pele-Phone Communications, Ltd. is able to identify changes in calling patterns that signaled the potential for fraud. This gave the company time to take appropriate actions and reduce overall fraud expenses. 3

2.1.7 Customer conversion rates

When a marketing solicitation results in a sale, the prospect has been converted to a customer. The effectiveness of the marketing or sales effort is measured as a conversion percentage. Better understanding the customer and targeting the prospect with the right products, channels, and incentives can dramatically improve the conversion rate.

By using the data warehouse with event-based marketing, Union Bank of Norway was able to convert potential customers to real customers at a 60% rate. The new customers raised deposits by 15% and lending by 11.5%. Quantifying the increased revenue is a simple calculation.

Travelocity.com developed a CRM application based on their data warehouse of detail data. By targeting customers who were most likely to take advantage of specific offers, they were able to raise their conversion rate of their 30 million personalized emails to 7.3% from 2%. They were able to lift booking conversions of event driven campaigns (post booking and new member welcome) by 35%. They increased their booking conversions for newsletter subscribers by 84%. They estimated their CRM data warehouse contributes a minimum of \$2,000,000/month to their revenues. 4

If each conversion is worth \$10 and the data warehouse is able to increase the customer conversion rate by 3% on 1,000,000 promotions or solicitations, the benefit is:

$\$10 \times .03 \times 1,000,000 = \$300,000$ for each promotion

2.1.8 Customer attrition/retention rates

By knowing which customers are likely to leave, and knowing their relative profitability, you can take appropriate action to control attrition. The value of retention programs is a function of the percentage increase in retention.

Pele-Phone Communications, Ltd was able to effectively monitor phone usage, identify customers likely to leave and from this information develop programs, promotions and services tailored to those customers and was able to generate a positive ROI by reducing the number of defecting customers. 3

National Australia Bank was able to hold onto its existing customers at an unprecedented 98.4% versus the typical retention rate of approximately 90% in Australia. The average revenue for a retail bank customer in the United States is just short of \$500. Against this revenue are interest expense, loan loss reserve, and the cost of servicing the customer. If an average banking customer is worth \$100/year in bottom line profit and if the bank was able to retain 20,000 more customers by better understanding their needs and tailoring services to those needs, the increased revenue from retention enabled by the data warehouse is \$2,000,000 per year.

2.1.9 Marketing campaign selection and response rates

Marketing campaigns associated with advertising and give-aways are expensive. Some campaigns are successful, as measured by the response to the campaign, the revenue generated from the campaign, and the number of new customers. Profits can be improved by avoiding ineffective marketing programs. By testing out different marketing campaigns with different customer segments, marketing is able to find the right approach for each customer segment. This reduces marketing costs and generally results in successful marketing and sales campaigns.

Medco Health Solutions produced eight direct mail programs designed to convert their members to internet prescription fulfillment. By using the data warehouse and a scoring model to fine tune communications, they were able to increase their response rate by 15%. They were able to send mailers to only 30% of their members - those most likely to respond – saving 70% of the marketing expense and achieved almost the same number of responses as if the mailers were sent to all their members.

RBC Group, formerly Royal Bank of Canada, was able to tailor its promotions to the appropriate customer segments and achieved some response rates as high as 40%. Industry averages for banking promotions are normally less than 4%. 1

2.1.10 Better relationships with suppliers and customers

The data warehouse provides an organization with an additional means of communicating needs and product status with suppliers and customers throughout the supply chain. Manufacturing and retail organizations are giving their suppliers selected access to a data warehouse where the suppliers can track inventory levels, sales, and product quality. This gives the suppliers the information they need to minimize stock-outs, trim inventory, reduce overproduction, and improve their own asset turns and improve product quality. This also gives the data warehouse owner the ability to negotiate favorable terms, conditions, and discount levels which should flow directly to the bottom line.

Suppliers are giving their wholesale customers selected access to a data warehouse where the customers can track inventory, orders, and shipping information. This information capability helps to lock in customers who find the data warehouse information available only from this supplier. It also minimizes stock-outs and better on time delivery, resulting in more product sold to these customers. In addition, this capability lowers the customer's overall costs, and gives the customer better information to help them manage their end of the supply chain.

2.1.11 Data mart consolidation

As seen in the section on costs, maintaining data marts is expensive. The cost includes the cost of the multiple hardware platforms and RDBMS licenses, including maintenance and upgrades, internal personnel to support the multiple systems and redundant ETL processes, increased network costs, additional software, outside consultants and contractors, costs associated with operations and systems administration, and maintaining redundant data. Redundant data typically accounts for 50% of the data in data marts and this percentage goes up as the number of data marts increases. The cost specific to each data mart is a function of the size of the data mart, the number and complexity of the source files, the cleansing required, the number of users, and the DBAs assigned to the data mart. Studies have shown the annual costs to be between one and two million dollars to maintain each data mart.

Besides the obvious costs of maintaining multiple disparate data marts, carrying redundant data on customers and suppliers results in a costly and disruptive operation: synchronizing data and the reconciliation of different outcomes. No one has studied or measured the cost and frustration associated with the question from management, “Why are these numbers different? Would someone please tell me what are the true results?” While not all data marts can be consolidated (there will always be political, regulatory, or security reasons to keep some separate), an organization can significantly reduce their on-going data mart costs by consolidating.

An organization was able to consolidate six data marts resulting in a three year \$11,500,000 savings. The cost for the consolidation was \$1,900,000. If we assume the savings were evenly distributed over the three years, the savings per year is \$3,833,333.

Dividing 1,900,000 by 3,833,333, we get .49, which means the payback period is ½ year or six months. It should be noted that even though the savings were only counted for three years, the savings should continue to accrue for as long as those data marts were planned to be supported; in most cases, they tend to proliferate, further increasing the costs. In another data case, a large bank estimated savings of \$65,000,000 from consolidating multiple data marts.

2.2 Estimating Intangible Benefits

Intangible benefits are, by definition, difficult to quantify. The following list contains examples of such intangible benefits:

2.2.1 Public relations, reputation, and impact on shareholders

When an organization is in crisis, it’s always embarrassing if they can’t answer questions about the critical situation. It’s also embarrassing if they don’t have enough information to tell the press and to inform their employees how they plan to deal with the crisis. A data warehouse can provide those answers and give upper management the information they need to respond to the

crisis. A manufacturing quality problem may have gone unnoticed. As the news is disclosed, it becomes apparent that senior management was unaware of the situation. The question is always why upper management didn't know what's going on. The data warehouse can provide an early warning system to alert management to these soon-to-be-uncovered disasters.

Organizations are evaluated in a number of ways. The quality of management is given great weight as Wall Street analysts determine a company's value. The technological capability and the ability of an organization to make effective use of information has become a major topic in computer and industry publications, and this has had a strong impact on the organization's reputation. The price of a company's stock is strongly affected by analysts who read these articles. A number of CEOs have indicated the way they are measured is by the price of the stock. The data warehouse is a vehicle for providing information effectively and has been identified as both a competitive edge for certain companies and as an indicator of the strength of the company's technological capabilities and of management's competence.

Harrah's was able to attract the attention of Wall Street with its ability to understand and effectively market to its clientele. These capabilities were far beyond those of its competitors. This resulted in a stock analyst raising the buy recommendation by two levels.

2.2.2 Competitive effectiveness

Competitors who are using their data warehouse effectively put you in peril. They are in a position to steal your best customers and produce higher quality products at a lower cost. Your competitors with a financial data mart will be able to tap the capital markets more effectively and deliver a higher price/earnings ratio on their stock. Your competitors will be able to develop successful products and get them to market more quickly than you can. Companies that provide meaningful and accurate data analysis to employees across the organization enable their employees to make optimal decisions, improve their productivity, their effectiveness, and their morale. It is very difficult to quantify the edge you would have to relinquish to the competition but it is definitely a major factor in many decisions to initiate a data warehouse.

A handful of companies in the health care industry have integrated claims data warehouses and provided data at the detail level. Due to their centralized efficiency, they were able to easily achieve early HIPAA compliance and they credited their data warehouse with providing competitive differentiation by streamlining costs while maintaining and even improving healthcare quality.

2.2.3 Better and faster decisions

If decision makers had better access to more accurate and more timely information, would they be able to make better decisions and to make these decisions more quickly? Would these decisions result in more sales and more revenue and profit? Would responding to a customer more quickly close a sale that might otherwise be lost? While we may not be able to quantify these results, most of the people who are now getting this more complete, higher quality, and more up-to-date data believe it is highly valuable.

Organizations sometimes operate without sufficient timely information about their operations. Lack of information about product profitability, quality control, customer satisfaction, competitive pricing, and other critical information robs the organization of what they need to make the tactical and strategic decisions that spell the difference between profitability and loss. Questions needing to be answered in a timely manner means in time for them to take effective action. Very often management doesn't even bother asking their IT people or their analysts for answers because they know that, by the time the question is answered, it will be too late to act. There is usually a visible backlog of unanswered questions and an even bigger backlog of questions that have not been posed.

A baseline measure of user satisfaction should be taken before the data warehouse is rolled out to the user community. Three months after roll-out, the questionnaire should again be distributed and should continue to be sent out twice a year as long as the system is in place. We have seen significant increases in user satisfaction but your own measurements will point you to where you are doing well and where you can improve. The questionnaire would include questions about comfort with the access and analysis tool, adequacy of training, and perceptions of data quality, performance, availability, and user support. 5

Capital One Financial Corporation felt that cutting the time executives have to wait to get their answers will deliver an ROI of two to three times what the system is expected to cost. 6

2.2.4 Better customer service

Most people are sensitive to the service provided by his or her bank, broker, auto mechanic, retailer, airline, doctor, and financial adviser. We choose our providers based on cost but very much on how we are treated. We want to believe that the provider knows us and is able to accommodate our specific needs and is able to satisfy our desires. Better customer service translates into improved customer satisfaction, higher retention, and an increase in cross-sell and up-sell ratios.

Anthem Blue Cross/Blue Shield was able to give their members information about providers' quality of care, as measured by mortality, for coronary arterial bypass surgery. This resulted in lowering their member mortality for this procedure from 4% to 1%.

Travelocity.com was able to retain an exceptionally high percentage of their customers even as competitors were offering promises of lower prices. They credit the high retention to excellent personalized customer service that was supported by their robust CRM system.

2.2.5 Employee empowerment

By giving employees access to better information, the employees become more productive and the requirement for supervision becomes that much less. Employees always feel better about their jobs and the organizations they work for when they are given the right tools and the right information. This can lead to suggestions from employees for additional ways to leverage the data warehouse which may result in additional tangible benefits.

2.2.6 Impact on internal champion

There are a number of installations where the data warehouse conferred stardom on the data warehouse champion (the primary sponsor). In a case referenced in this paper, the champion was promoted to CEO as the benefits of the data warehouse became apparent and critical to the success of the organization.

2.3 Post Implementation Benefits Measurement

As part of cost justification, the organization estimated its benefits but post-implementation, it's important to measure just what those benefits actually were. This measurement is crucial to determine the ROI from the project and also to help with future goal-setting, benefits estimates, and proposed expansions or enhancements. The measurement can highlight areas where benefits were overestimated and also uncover benefits that were serendipitous and not anticipated. The intangible benefits are difficult to actually measure but all the stories and testimonials associated with those stories should be captured, documented, and disseminated.

It would be unusual if the resulting benefits were as predicted. It is important to review the benefits to determine the accuracy of the predictions. Armed with the results of the review, more accurate predictions can be made in each category and the prediction process can be improved with more accurate benefits projected for future projects. The data warehouse lends itself to iterative development and this then allows us to measure the costs and benefits of each new iteration. More accurate benefits analysis can aid in prioritizing data warehouse projects. Even if you didn't estimate costs and benefits, post implementation measurement will still be beneficial. You will still be able to determine if your project has a positive ROI and you will still be able to determine if any existing data warehouses should be abandoned.

Ace Hardware developed a data warehouse that includes a focus on retail pricing using existing data and price elasticity analysis. Ace had projected an ROI for the project but their post-implementation measurement demonstrated an ROI that was much higher in some application areas due to acceleration of their implementation. Earlier payback has a measurable and positive impact.

Section 3: Full Cycle Assessment of Value

Management wants to know that the data warehouse is delivering a positive, acceptable ROI and would like to see it with real numbers. Astute management will ask penetrating questions about how the ROI was derived and those presenting the numbers need to be sure of how the ROI is calculated. Management will also want to know how the ROI measured up to the ROI that was estimated when the project was proposed. This will help determine if the project was a success.

The decision of whether to present the results in a report or in presentation format is totally dependent on the way management normally expects to view their information. Use the type of ROI calculation that the CFO and the sponsors like to see. If payback period is most often used, that's what to present but it is also advisable to present a discounted cash flow technique such as net present value or internal rate of return. ROI should be calculated and presented for cost justification to ensure management commitment for each data warehouse project. ROI should be measured and presented to management to gain continued support and budget for existing systems as well as future systems.

Appendix A – ROI Calculation Process 7

Traditionally, when IT professionals and managers calculated the ROI of an IT investment, they focused only on financial benefits. Today however, business leaders also consider the non-financial benefits of IT investments such as improved customer satisfaction and shorter cycle time.⁸

Most organizations use one or more “financial metrics” which they refer to individually or collectively as ROI. These metrics include:

Net Present Value (NPV): The sum of the net yearly benefits discounted by the cost of capital less the initial investment. A project with an estimated NPV greater than zero may be a candidate for acceptance. One with an estimated NPV less than zero would probably be rejected.

Internal Rate of Return (IRR): If you have an investment that requires and produces a number of cash flows over time, the internal rate of return is defined to be the discount rate that makes the net present value of those cash flows equal to zero. Some companies set a minimum acceptable IRR (or hurdle rate) based on their own cost of capital and the minimum percentage return they'd like to see from their investments.

Payback Period: The point at which the yearly benefits of a project cover the costs.

ROI Example

We will use an example to demonstrate each of these calculations. They are not mutually exclusive – all three will be used for evaluation. While each is valid, the usefulness of any one of them alone is limited. For example, payback does not include a measure of value after the investment has broken even. There are assumptions embedded in each model. Net present value assumes that you'll reinvest cash flows at the cost of capital. Internal rate of return assumes that you'll reinvest cash flows at the internal rate of return. It is advisable to use a combination of discounted cash flow techniques to ensure that you get as complete a view as possible.

In our example, there is an initial outlay for the data warehousing system of \$1,000,000. The yearly benefit is \$800,000 while the cost to maintain the system will be \$300,000 for every year the system is in production. This leaves a net benefit of \$500,000 per year. The conservative CFO demands that the system be estimated to only be used for three years following delivery. Assume a cost of capital of 10%.

Cost of Capital

An understanding of the cost of capital is necessary for determining rate of return and net present value. Any project must earn a return that's greater than its cost of capital, or the project would diminish the organization's value. Organizations with good credit ratings are blessed with a lower cost of capital than those on the verge of bankruptcy. Expressed as a rate of interest, it is a number the CFO in your organization fully understands and will want you to use as you perform ROI calculations; be sure to ask. For the purposes of illustration, I'll use 10% as the cost of capital or the discount rate.

Risk

Risk is only important in the cost justification process. Once the system is in, running, and is being measured, most bullets have already been dodged. There are substantial risks to any data warehouse project and any cost justification must consider the risk of the system not going in at all, the risk of it being over budget, the schedule slipping, important functions not being delivered, the system not performing, the risk of the users not making effective use of the system, and the risk of the expected benefits not materializing. Any estimate of benefits should include the expected range of improvement. A conservative approach that helps manage expectations dictates using estimates in the lower range of improvement. If the expected yearly cost savings from improved inventory management ranges from \$500,000 to \$1,000,000 with a most likely savings estimate of \$800,000, it would be wise to use a lower number such as \$700,000 for the ROI calculation. It's always better to deliver more than you promised.

Net Present Value

The present value is equal to the sum of the net yearly benefits discounted by the cost of capital less the initial investment. In our example, the net yearly benefits are \$500,000 for each of the three years, and the cost of capital is 10%.

$$\text{Net Present Value} = \Sigma (\text{Year 1 Benefits} - \text{Costs}) / (1 + \text{discount rate}) + \\ (\text{Year 2 Benefits} - \text{Costs}) / (1 + \text{discount rate})^2 +$$

$$(\text{Year 3 Benefits} - \text{Costs}) / (1 + \text{discount rate})^3 - \text{initial investment}$$

$$= \$500,000 / 1.10 + \$500,000 / 1.21 + \$500,000 / 1.331 - \$1,000,000$$

$$= \$454,545 + \$413,223 + \$375,657 - \$1,000,000 = \$243,425$$

The ROI as represented by the net present value of this project is \$243,425.

Internal Rate of Return

The internal rate of return (IRR) is the discount rate for which the present value of the cash flow benefits over time are equal to the amount of the investment. In other words, the IRR is the discount rate that produces a NPV equal to zero. In the example, the IRR is calculated as follows:

$$\text{Net Present Value} = 0 = \Sigma (\text{Year 1 Benefits} - \text{Costs}) / (1+X) +$$

$$(\text{Year 2 Benefits} - \text{Costs}) / (1+X)^2 +$$

$$(\text{Year 3 Benefits} - \text{Costs}) / (1+X)^3 - \text{initial investment}$$

where $X = \text{IRR}$.

The easiest way to determine the rate of return is to use a spreadsheet with financial functions such as those found in Microsoft Excel. There are other alternatives including a calculator with

financial functions including IRR (Internal Rate of Return) and NPV (Net Present Value) or a book that contains tables with “Present value of \$1 received per period.”

In the example, use Excel’s IRR function with the argument =IRR(A1:A4) and with the cells filled in as follows:

A1	\$(1,000,000)
A2	\$500,000
A3	\$500,000
A4	\$500,000

The IRR for this example is 23.38%.

Payback Period

The payback period (breakeven analysis) is the easiest to calculate and to explain. It resonates well with management and is the most frequently used measure of ROI for data warehouse applications. It’s the time required for the cash stream resulting from the data warehouse investment to equal the original cost of the warehouse. Management will usually reject any application where the payback period exceeds some predetermined number of years. As with the other methods, management will use payback period to rank data warehouse alternatives. The payback period for our example is two years ($\$1,000,000/\$500,000$).

Appendix B - Cost Calculation Template

Cost Category	Calculation	Dollars
Hardware		
Maintenance		
Internal personnel costs		
Network		
Maintenance		
Internal personnel costs		
Desktops/Laptops		
Maintenance		
Internal personnel costs		
Software		
RDBMS		
Maintenance		
Internal personnel costs		
Modeling tool		
Maintenance		
Internal personnel costs		
Query/Report		
Maintenance		
Internal personnel costs		

ETL		
Maintenance		
Internal support		
Other tool 1		
Maintenance		
Internal support		
Other tool 2		
Maintenance		
Internal support		
Contracting		
Consulting		
Help Desk Support		
IT training		
User training		
Operations & Systems Administration		
Total cost		

Appendix C - Benefits Calculation Templates

Tangible Benefit	Calculation	Dollars
Revenue enhancement		
Cash flow acceleration		
Analyst productivity		
Customer conversion rates		
Customer attrition/retention rates		
Marketing campaign selection and response rates		
Cost containment		
Fraud reduction		
Data mart consolidation		
Demand chain management		
Better relationships with suppliers and customers		
Total tangible benefits		

Intangible Benefit	Explanation or Narrative	Dollars or other benefit
Public relations, reputation and shareholder value		
Competitive effectiveness		
Better and faster decisions		
Better customer service		
Employee empowerment		
Impact on internal champion		
Total intangible benefits		

Appendix D – Glossary

Attrition: When a customer leaves you and gives his or her business to your competitor.

Business drivers: The financial metrics against which the business is measured.

Business value: The benefit to the business from such major drivers as revenue enhancement, cost containment, cash flow acceleration, and risk mitigation.

Cost/benefit analysis: The process by which the value of a project is estimated based on the expected costs compared to the tangible benefits usually expressed as an increased revenue or reduced cost.

Cross-selling: Selling an additional category of products as a result of the customer's original purchase.

Customer segmentation: Separating customers by factors such as age, gender, educational background, and yearly income.

Fully burdened: The total cost of the employee including health benefits, pension funding, workers' compensation, taxes, social security, vacation time, cost of office space, telephone, supplies, administrative support, and all other costs associated with maintaining an employee.

Hard-dollar benefits: tangible benefits that can be measured. Hard-dollar benefits can result from an increase in revenue or a reduction in cost.

Internal champion: The person within the organization who sponsors the data warehouse project. The internal champion is usually from the business side.

Justification: The process by which each project is evaluated to determine if there is financial viability in its implementation. The justification process also allows management to prioritize projects.

Market Penetration: The percentage of the market owned by a company as represented by share of revenue.

Metrics: Any type of measurement. Metrics could include business results, quantification of system usage, average response time, benefits achieved, etc. The measures that an organization believes is vital for its success.

OLTP: Online transaction processing. The operational system.

Pain: An unfulfilled business need that jeopardizes the success of the organization.

Power users: Knowledge workers who are capable of writing complex queries and reports with little need for help. Power users are usually more active than the other users.

Retention: The ability of keep customers from defecting to the competition expressed as percent retained.

ROI – return on investment: Usually represented as a percentage of tangible monetary value in relation to the cost of the system.

Scalable: Ability to increase the number of users, the size of the databases and the complexity of the queries and reports without having to replace the existing platform or architecture.

Silos: Data marts that do not talk to each other and can only be joined and integrated with considerable effort.

Sponsor: The person in the organization, usually from the business side – who supports the project. This person should be someone with power, money and commitment to the project.

Strategic business initiatives: The action plans that support the long-term goals of the organization.

Supply chain: The management of the components, manufacturing and distribution of a manufactured commodity. The supply chain management includes warehousing and tracking inventory.

Timely: Data is valuable and useful to analysts only if it represents organizational activities that are reasonably current. Timeliness is a function of the users' requirements for currency and is consistent with user expectations. Timeliness is usually measured by how soon the data is available after some distinctive end-of-period such as "two days after the close of the month." The act of getting the data to the users at the most opportune time.

Total Cost of Ownership: The cost to the organization for the initial implementation and the maintenance of the system.

Up-Selling: Selling a customer products or services that are more expensive and more profitable.

User: A knowledge worker, a business analyst, a statistician, or a business executive who will access the data in the data warehouse to perform some type of business analysis.

Value added: The notion of additional benefit being provided by some activity or service.

References

Adelman, Sid, and Larissa T. Moss. *Data Warehouse Project Management*. Boston, Addison Wesley, 2000.

Adelman, Sid, Joyce Bischoff, Jill Dyche, Douglas Hackney, Sean Ivoghli, Chuck Kelley, David Marco, Larissa Moss, and Clay Rehm. *Impossible Data Warehouse Situations: Solutions from the Experts*. Boston, Addison Wesley, 2002.

Bierman, Harold Jr., and Seymour Smidt. *The Capital Budgeting Decision*. New York, The Macmillan Company, 1992.

Business Week, "3M: Glued to the Web," November 20, 2000

English, Larry, P. *Improving Data Warehouse and Business Information Quality*. New York, Wiley, 1999.

Farrell, Vickie, and Carrie Ballinger, Positive Payoff, Teradata Magazine, 1Q, 2002.

Hoffman, Thomas. Despite Downturn, CIOs Push Projects, ComputerWorld, November 25, 2002.

Smith, Caroline. *Teradata Customers are Growing*. Teradata Partners Conference, 2002

Swift, Ronald. *Accelerating Customer Relationships*. Upper Saddle River, NJ, Prentice Hall, 2001.

Sid Adelman
818 783 9634
sidadelman@aol.com
www.sidadelman.com
