AN INTRODUCTION TO PROJECT DATA WAREHOUSES

Cliff Brandon
January 2003

AN INTRODUCTION TO PROJECT DATA WAREHOUSES
By: Cliff Brandon 8/31/2002

TABLE OF CONTENTS

COPYRIGHTS and CAVEATS

INTRODUCTION

Data warehousing and reporting, once the domain of large organizations, can be rescaled to meet the needs of small to medium size organizations. Project-centric organizations in particular have an opportunity to construct truly integrated project environments using this technology.

The topics covered in this paper address 'Why' questions. Why create a data warehouse for projects and what are the specific benefits of using this approach to control projects? Other papers in this series will cover the specifics of creating project data warehouses.

WHY PROJECT CONTROL?

All organizations execute projects. The relatively recent emergence of PMOs (Project Management Office), Project Management as a distinct discipline, and project portfolio programs indicates a growing interest in the selection and tracking of projects. Given the recent track record of project failure (particularly for IT initiatives) and current economics this interest is probably more than just a passing fad. This paper looks at a technical mechanism for integrating and distributing information in the project environment with the goal of improving project performance.

AUDIENCE

This paper focuses on the information needs of small to medium-size, project driven organizations. Examples of business practices that tend to be project driven include: construction, engineering and design, consulting, software/hardware development, commercial art, environmental remediation, job shop manufacturing, event management, exploration, law, medicine and many others. A project driven environment exists where work is defined in discrete units by schedule, scope, and budget.

This paper is written for managers, organizational strategists, technical architects and those who manage computer systems technology (IT).

This is not to say that large organizations or organizations that have more operations-centric practices would not benefit from the ideas presented here. In fact many of the ideas presented could benefit organizations considering activity based accounting (ABC). However, this paper is written primarily from the perspective of implementation

and delivery of value to smaller, project driven organizations. Appendix A is provided to help determine if this technology is right for your organization.

ECONOMIES OF SCALE

What strikes most people when they pick up almost any book or article on data warehousing is the scale of the undertaking: Terabytes of data in complex enterprise scaled systems, millions of dollars in implementation and maintenance cost, and hordes of consultants. Can this technology provide real value to small and medium size project driven companies?

Businesses are using data warehousing to consolidate data, improve access to information, and provide for decision support at all levels of the organization. This is not just a consideration for large organizations. Small and medium size organizations have similar needs. The difference is scale. As data warehousing and its associated tools become more commoditized, new opportunities will open up for smaller organizations to take advantage of these technologies. Scaling the technology down is far easier than scaling it up.

Below is a list of some factors that contribute to the cost of large data warehouses and some of the advantages that being smaller bring:

- Large data sets spread out nationally or internationally in different time zones that have to be coordinated regularly… What if your data had a more limited distribution or was all local?
- Terabytes of data requiring specialized, expensive storage… What if the five year projected size of your data warehouse was less than 100Gb and it could comfortably reside on an off-the-shelf server?
- Multiple operating systems and database platforms with both active and legacy databases… What if you had a much more limited set of operating systems and database platforms?
- Multiple data marts or cubes serving distinct functional areas… What if most organizational information could be served from a single data warehouse source?
- Complex ERP accounting and management systems requiring specialists to compose and maintain the data warehouse… What if your accounting and management systems were simpler and architecturally open? What if your IT staff could build and maintain it themselves?

I would suggest that lower complexity results in lower the cost and that lower complexity may also promote greater coordination and better functionality.

Data warehouses and their related reporting offer small to medium-size organizations the opportunity to significantly improve/integrate the project control process and can create new, unique selling propositions for their businesses in a crowded market. Specific technical benefits are the subject of the next section.

PROJECT DATA WAREHOUSE BENEFITS

While there are a number of benefits that can be gained from using this technology, there are two that stand out from the rest in a project driven environment. The first are gains an organization can realize from organization-wide consolidated, timely reporting. The second is the leverage that can be gained in the marketplace by making this information readily available to clients. Both of these points are explored in the sections that follow along with a list of other benefits.

Consolidated Reporting

Consolidated, timely reporting from a data warehouse permits all areas of the organization to see and report from a single, integrated data source. Below are some common business reporting challenges:

- It is often difficult for users to get at business information that is relevant to their role in the organization.
- Considerable time is often spent compiling information from various sources for routine and ad hoc reports.
- Meetings where different departments show up with various 'aged' sets of information are not uncommon.
- Problems are handled reactively versus proactively based on available information.
- A month or more may pass once a project has been completed before it can be determined if it came in under budget or made a profit.
- Project managers and subordinates cannot timely identify and analyze problem areas for cost overruns and schedule delays to prepare and implement corrective action.

It would be wrong to say that a data warehouse and associated reporting is the answer to these challenges. Business process, culture and other factors also contribute. The data warehouse is an

5

information resource. If the data warehouse information is widely available, can be easily rearranged, meets the functional needs, and is current then many of the listed challenges can be met.

Information as a Unique Selling Proposition (USP)

Regardless of whether your clients are internal or external to your organization, client reporting is critical to the success of any project. Clients and project sponsors want to know if the project is on schedule, scope, and budget. Below are some common client reporting challenges:

- Project managers and account executives spend excessive amounts of time on preparing client reports. Can a project manager or account executive push a button and generate a complete project status report?
- Project information is not current when presented to the client. How often are clients briefed on project status after they receive an invoice for the work?
- Project information does not match the contract and is not complete for schedule, scope, and cost.
- Reporting formats vary from project-to-project and client-to-client making it difficult for project management and clients to follow more than one project.
- Clients must wait for hard copy project reports. Would your clients find it valuable to be able to retrieve a project status report at a time of their own choosing? Would this type of reporting change your credibility with your clients? Are your competitors offering this type of project visibility?

Again, data warehouses and associated reporting do not answer all these challenges. However, the same reasoning used in the reporting section applies. The same data warehouse used for your core business can deliver select project information to Project Managers and clients alike.

6

Other Project Data Warehouse Benefits

Other reasons exist for constructing a data warehouse and a related reporting environment. Seven additional reasons summarized below:

Accounting Data Security: Accounting systems can be good sources of project data. Direct access to accounting systems by individuals without direct accounting responsibilities or indirect access to accounting data using tools like Excel can be problematic because accounting systems control and track real costs and payments. A data warehouse holds a copy of select accounting data and is accessed for reporting. Changes or damage to the data warehouse do not flow backward to the accounting system. The system is more secure.

Accounting Application Cost Reduction: Most accounting systems are licensed and billed by the 'seat'. The cost (per seat or per server) to access a client/server database where a data warehouse would reside is typically quite a bit less than the cost per seat to access an accounting system. Having more than a few individuals without direct accounting responsibilities in the accounting system can be cost prohibitive.

Ad Hoc Reporting: The data warehouse provides structure, access, and content for constructing standard and add hoc reports. There are a fixed number of standard reports available in an accounting system. There is always a need for more diverse, extensive, and flexible reporting including the ability to create ad hoc reports. Specialty reports can often provide critical analyses vital to managers who are trying to keep their projects "on-track".

Static versus Dynamic Reporting: The data warehouse is not continuously updated and so provides a snapshot view of the organization's information. A new snapshot is produced each time the warehouse is updated. Because there is constant activity in the accounting system, static coordinated views of the information are difficult to obtain directly from an accounting system. These types of views are often ideal for decision support in situations where all users need to see the same information for a specific period of time.

Data Compilation: Data warehouses provide a common repository for an organization's data. Data (other than accounting data) from other sources often needs to be integrated and reported with the accounting information. Scheduling, documents, and other shadow systems can be coordinated in the data warehouse using this technology.

Shadow System Coordination: Data warehouses can provide a common information source for other applications. Content for shadow system controls and lookups can originate in the data warehouse. Systems that use a consistent set of data can be created across the enterprise.

OLTP vs. OLAP Transactions: It is possible to create queries and associated specialty reports directly from an accounting database. This is not necessarily a good idea however. Accounting systems perform the majority of their activities using OTLP methods (on line transaction processing) that optimize individual transactions (transactions from purchase orders, contracts, etc.). Most external reporting systems have an OLAP (on line application processing) engine that selects sets of data from the system. Processes for individual transactions and queries can conflict with one another (can create locking problems in the accounting system). Depending on the accounting system, a conflict like this could wreck havoc on the company's books or perceptibly slow the system. These types of processes should be separated. One way to accomplish this is to load accounting system data into a data warehouse.

Any of the benefits listed, either singly or in combination, may be sufficient to consider this technology. Appendix A summarizes the benefits listed and provides a mechanism to allow you to assess the value of the technology for your organization. At this point it is time to look at how the data warehouse is structured and where it gets its content.

THE PROJECT DATA WAREHOUSE MODEL

The introduction indicates the architecture for a project data warehouse will be discussed at some length in other white papers. Having said that, it is not really possible to understand the value of such a system without presenting at least the basic architecture of a project control system and some of the assumptions that go along it.

Figure 1 depicts a possible structure and relation between an organization's systems and the data warehouse. Of course there are many possible variations on this approach but the one portrayed is very likely the one that would be used in small to medium size organizations.
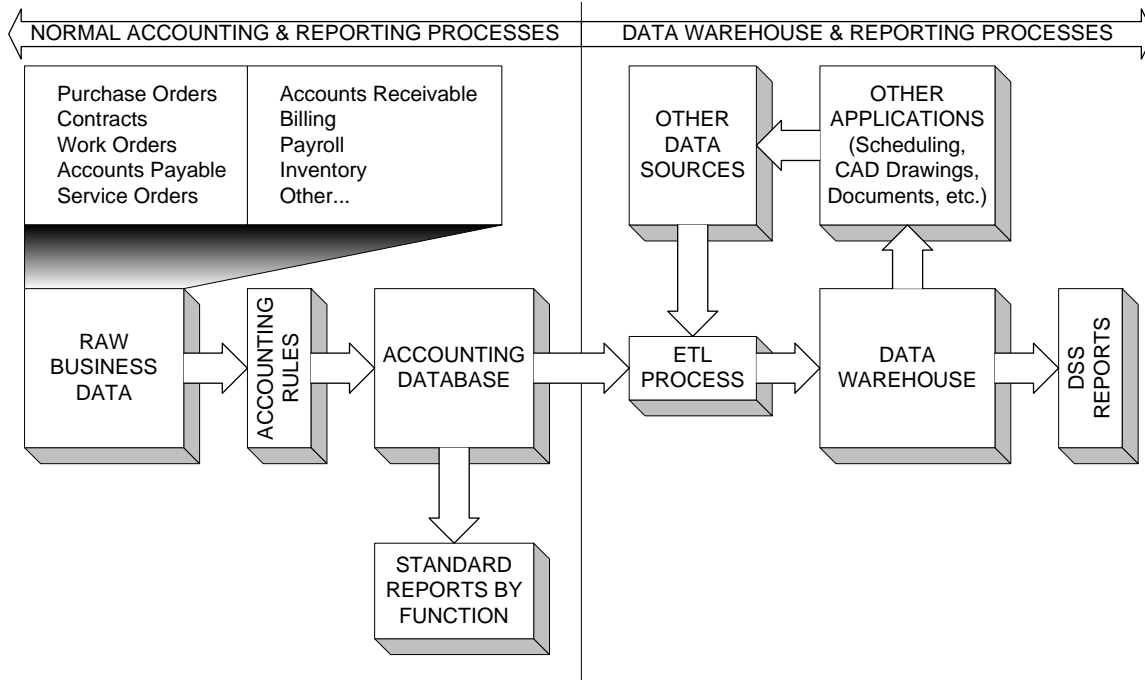
Figure 1. General model for a project control data warehouse system

In Figure 1 the architecture is divided into two parts by a solid vertical line. The normal organization's accounting and reporting processes is to the left of the line and data warehouse and reporting processes are to the right.

The left side of the diagram depicts a simplified common accounting structure. Various types of transactions specific to the company (raw business data) are entered into an accounting system based on the system rules. These transactions wind up in an accounting repository (accounting data base). Standard reports can then be generated from this repository and are usually specific to function (purchasing, accounts payable, billing, etc.).

The right side of the diagram shows the data warehouse and its supporting components. The Extract, Transform, and Load (ETL) process usually operates on a fixed schedule (often at night) to extract select information from the accounting database and other sources. The ETL process transforms data into structures that better suit reporting needs. Finally, the ETL process loads the transformed data into the data warehouse. Various reporting and business intelligence tools can then be used to access the data warehouse and provide specialized and cross functional reporting for decision support at all levels of the organization.

Accounting systems can be a rich source of information. Consider that all purchases, labor, contracts, billing, and payment have to pass through and conform to the rules of this system. Three reasons for using accounting system data as one of the primary content sources for data warehouse content are:

1. The accounting system is usually the single source of financial information for an organization,
2. The controls, business processes, and application rules should provide a relatively clean, integrated data set for the data warehouse, and
3. Project control codes, necessary for reporting in a project based environment, should be part of the accounting information set and should be passed to the data warehouse.

Alignment of project reporting and accounting is critical for decision making and credibility. Having said that, this technology should be used with a few caveats. First, the accuracy of the information will depend on the completeness of the information and how the content is treated. To report cost and hours, all sources of cost and hours have to be identified and captured. Deciding how to treat committed cost is a content decision that will impact the reporting. Second, it should be understood that reports from data warehouses are not the accounting system but are designed to be a good reflection of the accounting information and other relevant systems at a moment in time. Based on need, the system should be tuned to the level of accuracy required and users expectations should be adjusted to understand the meaning and use of the information.

A word or two needs to be said about the 'Other Data Sources' shown in Figure 1. The project environment often has a number of other information resources that may or may not have direct ties to accounting or cost issues. Take scheduling as an example. Scheduling may be able to deliver resource and event information to the warehouse and may, in turn, consume resource cost information (cost hierarchy for example). Likewise, the warehouse could also be a good repository for links between documents or drawings and projects depending on an organization's needs. It is also possible to grow additional functionality as need and opportunities present themselves. Additional functionality and extending the data warehouse architecture will be covered in more detail in another white paper.

The data warehouse is only a repository, a holding pen if you will, for data. What is the actual value of the project data warehouse and how is it delivered?

DELIVERING PROJECT DATA WAREHOUSE VALUE

The general benefits of data warehousing were previously presented in the Data Warehousing Benefits section. But how are these benefits actually delivered?

The data warehouse has one single purpose: to collect and retain business data and provide it as common resource to the enterprise. Reporting systems attached to the data warehouse are the delivery mechanism for the data. An idealized system with data warehouse and reporting components is depicted in Figure 2.

NORMAL ACCOUNTING & REPORTING PROCESSES   DATA WAREHOUSE & REPORTING PROCESSES

Purchase Orders   Accounts Receivable
Contracts   Billing
Work Orders   Payroll
Accounts Payable   Inventory
Service Orders   Other...

OTHER DATA SOURCES

OTHER APPLICATIONS (Scheduling, CAD Drawings, Documents, etc.)

RAW BUSINESS DATA → ACCOUNTING RULES → ACCOUNTING DATABASE → ETL PROCESS → DATA WAREHOUSE → DSS REPORTS

P/O Register
Invoice Register
Aged Receivables
Inventory Listing
Check Register
Other...

STANDARD REPORTS BY FUNCTION

Key Performance Indicator Reports
Project Managers Reports
Client Reporting
Vendor History Reporting
Sales and Marketing Reports
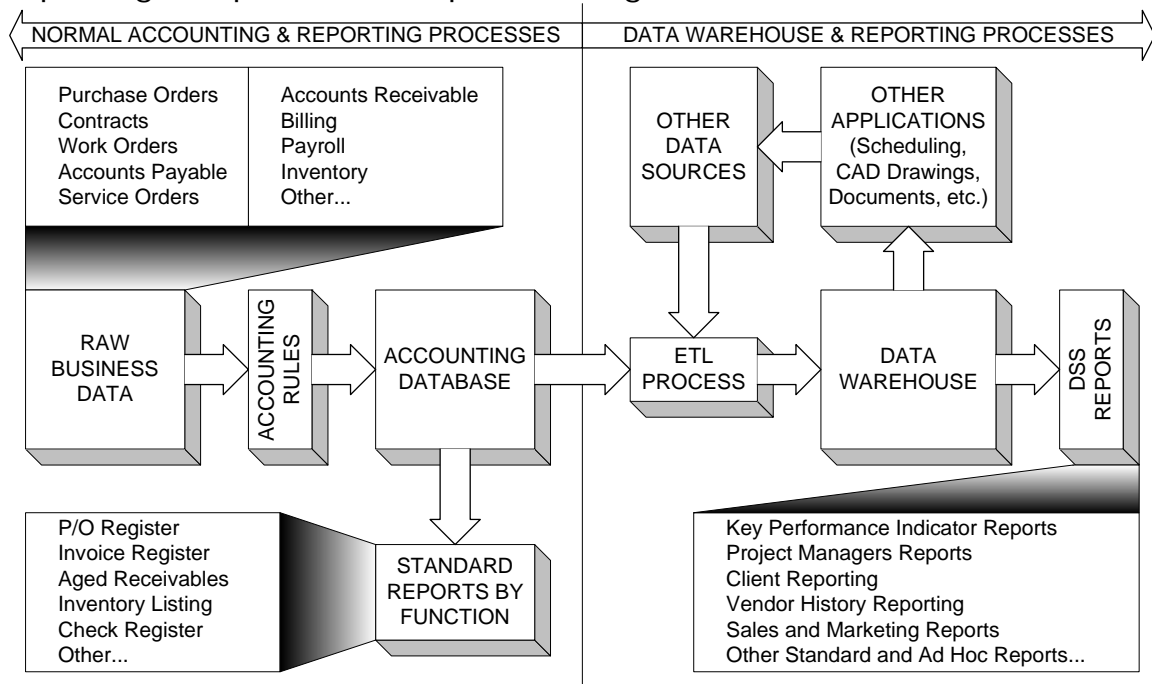Other Standard and Ad Hoc Reports...

Figure 2. General model for a project control data warehouse system

At the bottom of figure 2 there are two sets of reports separated by the solid vertical boundary. These reports and some of the supporting components on each side have been broken out of the diagram and are reconfigured in Figure 3 for readability.

11

```
  NORMAL ACCOUNTING & REPORTING PROCESSES    DATA WAREHOUSE & REPORTING PROCESSES

        ACCOUNTING                              DATA
         DATABASE                             WAREHOUSE

                                                 ↓
          ↓

         STANDARD                              DECISION
        REPORTS BY                              SUPPORT
         FUNCTION                              REPORTING


        P/O Register                    Key Performance Indicator Reports
       Invoice Register                     Project Managers Reports
      Aged Receivables                          Client Reporting
       Inventory Listing                   Vendor History Reporting
       Check Register                     Sales and Marketing Reports
     Other Standard Reports...         Other Standard and Ad Hoc Reports...
```
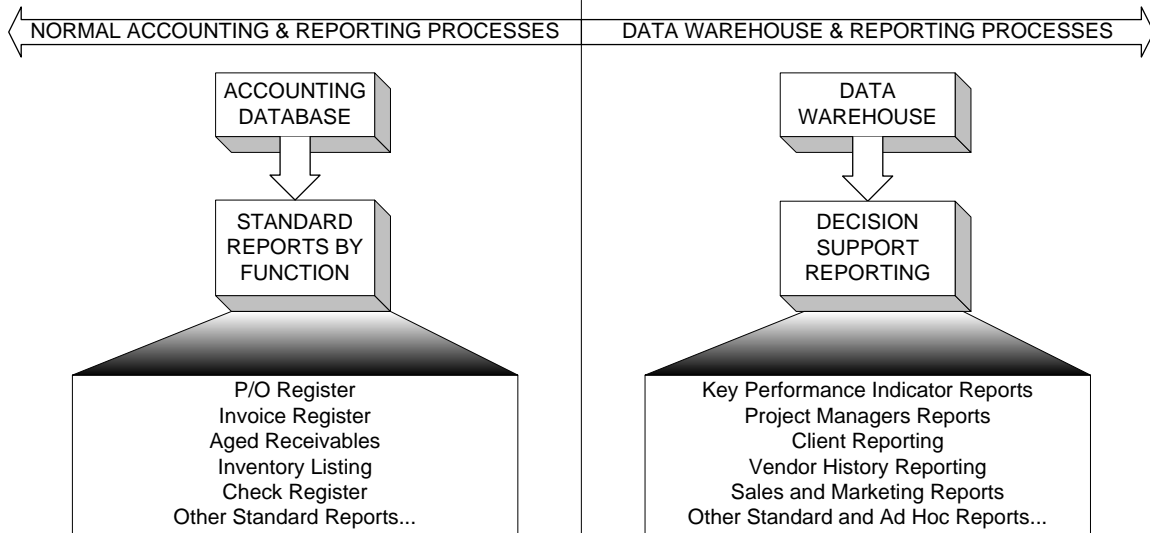
Figure 3. Report content differences between an accounting system and a data warehouse system

Accounting reports (left side of Figure 3) are typically generated by accounting function. The reports from the data warehouse (right side of Figure 3) can either support traditional functional reports or more elaborate cross-functional reports.

Accounting Reports

Functional accounting reports have their place and are quite useful in the accounting arena but are somewhat limited in other applications. To illustrate the limitations, consider the following common situation. A Project Manager (PM) goes to accounting and makes a request for a report on all the transactions for the project. This is a common and deceptively simple request. Because accounting is commonly and necessarily divided by function, it is unlikely that there will be a single individual in accounting who will be able to fulfill this request in its entirety. The PM may have to make the rounds to A/P, A/R, Purchasing, Contracts, and Payroll to retrieve the needed information. The PM may not even be able to access a complete set of information because some information is considered sensitive (payroll for example) and is restricted as a matter of policy. Acquisition of complete project information can be difficult and time consuming. How often would you predict a Project Manager would go through the exercise described above to get an understanding of the financial health of the project?

Given that accurate, timely project cost information is critical during the execution of a project. What is the ROI for access to current and complete project information by a PMs and management? Run through

this compilation exercise with some of the other reports listed on the right side of Figure 3 and you will probably begin to develop a sense of the limitations of standard accounting reports.

<u>Decision Support System (DSS) Reports</u>

The reports on the right side of Figure 3 differ from the standard accounting reports in that they are often cross functional and contain data aggregates (summaries). They tend to be used to present information rather than raw data. Much of the information in these reports ultimately has the same source (accounting database). The data has simply been restructured in the data warehouse to allow for more comprehensive reporting. The class of reports with content originating in the data warehouse is broadly identified here as Decision Support System reports (DSS). DSS reporting is commonly regarded as an executive tool. The term's use here is broadened to include users outside the executive domain. This makes sense in smaller, flatter organizations, where decision making and responsibility are placed closer to where work is executed.

Since we have looked at the basic architecture of a data warehouse it is now time to look at the report generation and delivery process.

REPORTING FROM THE DATA WAREHOUSE

A data warehouse project delivered without an effective report creation and delivery mechanism cannot succeed. The reason is simple. Management that pays for a data warehouse project cares about the content not about the content container. Deliver the content and succeed. Focus on the container and fail. This simple truth is often missed in data warehouse development projects largely because the majority of the effort is spent on data warehouse container.

Reporting and delivery is critical to the success of this type of project. It is so critical, the mechanism for report creation and delivery should be identified and approved at the beginning of the data warehouse project and should be developed and implemented in parallel with the data warehouse development. Figure 4, extracted and expanded from Figure 1, identifies some of the different mechanisms for reporting and delivering the data warehouse content.
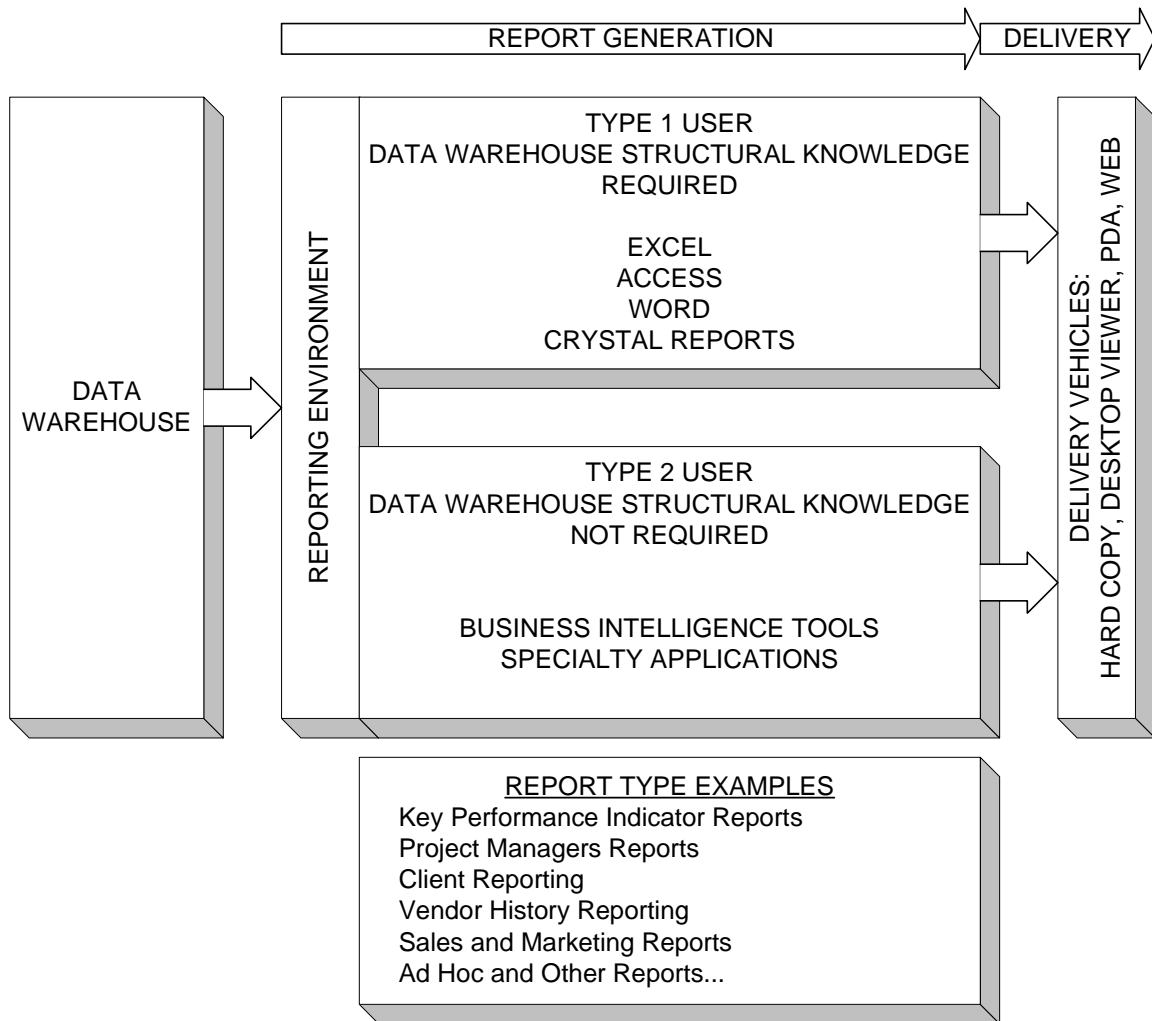
13

Figure 4. Data warehouse reporting and report delivery options.

The content delivery process is divided into two parts: reporting, which controls format/content, and delivery which describes how users receive the reports. Both elements are more fully described below.

Reporting Environment

Reporting is further divided into two separate approaches based on the type of user accessing the data warehouse content to create the reports.

Type 1: The Type 1 user is usually either a Data Base Administrator, Programmer, or Power User that understands both the data warehouse architecture and the table/field names. Type 1 users are often assigned to IT and create standard reports with specific data base knowledge using tools like the ones listed in the diagram.

Type 2: The Type 2 user can have a range of titles but usually are Business Analysts and Cost Engineers in a project driven organization. These individuals do not typically have intimate technical knowledge of databases, database mechanics, or queries. Type 2 users are very familiar with the content however and have a driving need to bend, twist, and cut project data in an ad hoc manner to evaluate various aspects of the business. Business Intelligence tools and internally developed specialty applications can serve as intermediaries for the Type 2 user. This approach can provide Type 2 users with an opportunity to create standard, specialty, and ad hoc reports.

Both reporting mechanisms, those used by Type 1 and Type 2 users, can exist together in the same system.

Report Delivery

The report generation from a data warehouse creates an electronic document. This document has to be delivered, in some format, to the end user or users of the document. Delivery to the users can be either in hard copy or electronic form. If the delivery mechanism is to be by hard copy, the report is printed and copies are sent to the various recipients. If the delivery mechanism is electronic, a report is made available to users as a file or 'live-link' via an electronic device such as a desktop computer, PDA, cell phone or other device. This delivery mechanism allows a user to view or print reports from whatever device is available to them. There are a number of options to consider when sending electronic documents:

Format: The type of document sent (example: Word, Excel, PDF, etc.).
Bursting: The same document can be sent electronically to many individuals. Document content can be identical or customized to meet individual needs.
Push vs. Pull: Documents can be pushed to individuals on a one-time or subscription basis (example: through email) or it can be pulled from the system by a requestor.
Digital Dashboards: This personalized environment is being used by some companies to allow users to compile role and project related information on their desktops. Static or linked reports could be added to this environment
Document Repositories: Common location for static or linked reports.

Business Intelligence

Business Intelligence tools need special mention in the reporting context and should not be overlooked in the report decision making process. Business Intelligence tools offer an impressive strategic opportunity to move the "Report Request and Fulfillment" process from IT into the hands of the requestors. Considering the resource cost for IT, development time, and the iterative cycle time for report creation, Business Intelligence tools have the potential for creating a significant ROI for an organization.

The basic architecture of a project data warehouses, their content, and reporting have now been covered. In the next section specific challenges of the project environment and the role data warehouse reporting plays in project success will be examined.

THE NEED FOR SPEED…THE PROJECT ENVIRONMENT

How sensitive is the need for information in a project driven environment? The answer to this question will doubtless vary by organization and may ultimately drive decisions about whether data warehouse technology would benefit the organization. An example from the construction industry may provide some insights into this question. Figure 5 charts the progress of several cost components over the life of an idealized construction project. Of course other tracking measures can be used but the following four were chosen to keep the example reasonable.
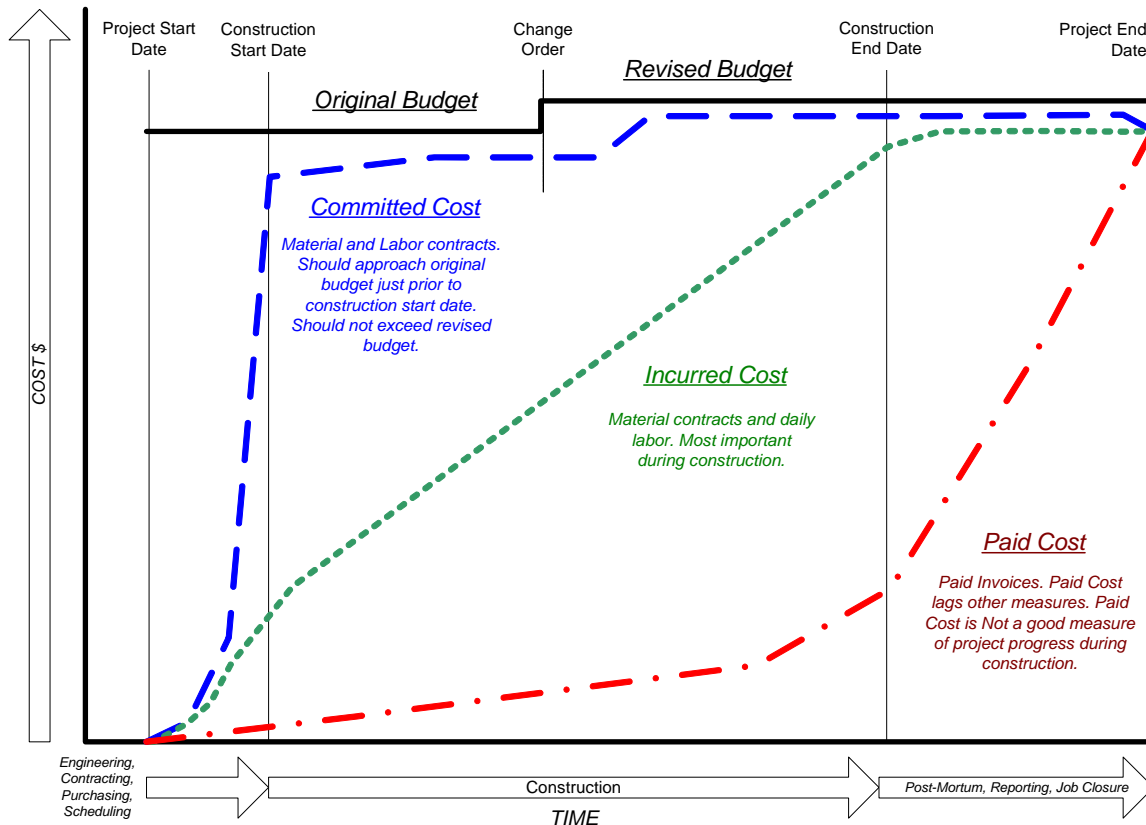

Figure 5. For an idealized construction project; the cumulative cost for various cost elements over time.

The chart in Figure 5 shows cumulative cost values over time for budget, committed, incurred, and paid cost. Relative to the budget and each other, each of the cost components has different meaning depending on which phase of the project is being executed at any given time. A list of terms is provided below with some basic functional definitions and uses for the measures.

Original/Revised Budget: Spending limit for the project. Project costs should not exceed the budget. Note that budgets may be adjusted during the course of the project.
Committed Cost: Important at the onset of the project and is used to determine if procurement has contracted for all the material and labor necessary to complete the project.
Incurred Cost: Important during the actual construction to determine the 'burn rate' for the project. The greatest exposure to risk of exceeding the budget is during the construction phase.
Paid Cost: Becomes important in the project wrap-up. Paid cost is not a good measure of project progress prior to project completion.

It is clear that accurate, current information would be valuable at any point in the project. However, if the goal is to control the project by identifying and correcting variances, the construction period is the most critical. This is because costs are being incurred at a rapid rate in this phase of the project. In order to find and correct problems or potential problems it is necessary to update and report project information regularly and as close in time to the construction events as possible.

Given proper regular input to the warehouse and reasonable warehouse refreshes, tracking various types of specific costs issues can be accomplished with the data warehouse and related reporting tools.

The specific cost profiles for any given project will vary depending on project type and industry. This construction project example was provided to illustrate the need for timely and accurate information.

CONCLUSION

I want to return to the goal of project performance improvement stated earlier in the paper. Can the technology outlined here improve project performance? Technology alone does not change performance. What this technology can do is make a common set of information available to project managers, management, and all functional areas of the organization irrespective of accounting system, operating system, or database. By making project information more available to those who need it, project issues can be addressed more quickly thus allowing for improvements in project performance.

ABOUT THE AUTHOR

Cliff Brandon

Cliff has worked as independent consultant in Information Technology since 1985 designing and building data management systems for clients. Over the past seven years he has concentrated his efforts on designing, building, and supporting a project control system from the ground up for a large mining firm's engineering and construction group. This paper reflects many of the lessons learned and applied from that experience.

Questions may be directed to Cliff Brandon at:
cliff.brandon@asaservices.biz