Information Delivery -- A Project Management Approach

Kim Foster, CoreTech Consulting Group, Inc., King of Prussia, PA

ABSTRACT

The purpose of this paper is to provide a strategic approach to managing a data warehousing project. The content is intended for the project manager, team leader, or business manager charged with the challenging role of managing the implementation process for data warehouse or data mart design and development. Special attention will be given to the balancing act required between planning and producing results. Practical suggestions are provided that give insight on the effective information delivery.

INTRODUCTION

Pick up any trade magazine, and you will read that anywhere from 25 - 40% of technology based projects are canceled, lose steam, or are put on hold within the first 18 - 24 months. Although the reasons range from 'lack of funding' to 'changes in organizational structure,' you will find that in most cases these situations could have been avoided.

Have you ever been faced with one or many of the following scenarios?

- ◊ Chasing people down the hall to get updates on tasks or processes
- ◊ Playing referee (or participant) in a heated discussion, debate, or argument about unclear data requirements, data quality, budget constraints, or lack of resources
- ◊ Find deadlines slipping on your project plan
- ◊ Say to yourself, "what's a project plan?"
- ◊ Fear taking a vacation (or new job) because the project might die
- ◊ Wonder why you took/were given this job to begin with
- ◊ Find it nearly impossible to meet expected goals or objectives you set for the project

Welcome to the world of project management!

Don't worry, these situations do not mean the project is bound for failure. But, do take them as warning signs that larger problems exist.

For those of you in business areas who have done project management before, you know the grueling road ahead. In fact, you may actually have a process down pat and not had to experience these types of problems in your most recent efforts. The difference is that now you have to apply these techniques to a support group made up of programmers, data management gurus and a very non-technical user group (I'd use a bridge metaphor here if I didn't think you'd gag from overuse of the reference). No sweat, right? Right.

If you are from the technology side of the house, you may not be very familiar with the process of managing a project, much less providing users with services they haven't asked for yet. Here's the wake up call: **most users don't care about what a data warehouse is or does**. Most don't even want you messing around with their current processes unless you plan on improving those processes by tomorrow morning. So now you have to manage a project, where the users are not too comfortable with what you're doing and you're not too comfortable with their discomfort. Go figure.

Roll up your shirt sleeves because you know you like a good challenge and here comes a big one.

THE BASICS - DATA WAREHOUSING

Anyone who has read the work of Inmon or any of his disciples knows the fundamental principle behind data warehousing. But, just in case you have not, here is an overview as identified in W.H. Inmon's book *Using the Data Warehouse*:

- **subject oriented,**
- **integrated,**
- **time-variant,**
- **non-volatile**

collection of data supporting management's decision process.

In most cases, the goal is to move operational or transaction based data into more functional and accessible data stores that allow the users to get to data (historical, current, detailed and/or summarized) more efficiently and preferably more effectively. Why? To provide automated, reusable, and manageable procedures accessing data across different platforms, using different query/decision support tools and front-end applications, and freeing up that valuable programmer time currently being spent on generating ad-hoc reports for the marketing department.

This venture can be a critical one because it will allow different levels of users to access consistent views of dollars, percentages, and totals over time by product, subject area, geographic region, even service division. But even more critical is that the business needs are identified, not assumed.

Users may not care about warehousing but they do care that they are overly dependent on 'techies' to get access to their data. Or, that it is difficult (sometimes, even impossible) to get a true picture of a customer or product line because the information is on one too many systems with one too many compatibility issues. When the business lines cannot put a finger on the pulse of their business it is usually due to data access related issues. And this will ultimately effect the bottom line. Are these your organization's business needs? Are you sure you understand them? Let us say they are and you do. Keep in mind that you want to hit as much business need one subject area at a time. You may find it necessary to sacrifice breadth for depth when it comes to making the initial data selection choices. The 80/20 rule works well here...20% of the data helps determines 80% of the business decisions.

Make sure you have the appropriate executive and management sponsors. A sponsor is someone who provides both monetary and management support and is willing to identify the project as a high priority within the organization. This includes your management (1-2 levels above you) as well as an executive who has the corporate and financial clout to back it. Ideally, you will obtain the crux of this support from the business side because that is where the expected benefit will be. Don't be surprised, however, if your business support goes into sticker shock once the cost estimates are provided. These groups may be unwilling or unable to foot up front costs given the lack of immediate rewards (remember, all business is about making money and this requires that they spend it...a lot of it). You may find it necessary to obtain financial backing from the technology (or IT) organization. Then, you can plan for billing back to specific business areas based on usage and demand over time.

Plan to establish several phases that include continuous user communication and short-term, tangible deliverables. It may not always be feasible to provide a complete source of data for the warehouse every three to four months. You will still need to plan to provide users with something they can see and touch. Get evaluation copies of front-end query or decision support tools and let the users test them out and give feedback. Have more proficient users assist with data quality assurance and testing. Keep communication flowing so that they don't think the project has slowed down.

Although it is very important to find out what data stores or marts already exist, try not to get caught up in departmental data mart strategies. Keep to the outline for the overall architecture. You may find a great deal of compromise and inconsistency in data integrity from department to department. Use them to get a sense of what data is needed and utilized most often. Design standardized data models for the data by subject area as opposed to department or division.

Be wary of staying on the bleeding edge. To some extent you may be forced to utilize whatever technology your organization has in place. Don't fight too big of a battle too early on. Utilize this to your project's advantage. If you have a DB2 mainframe environment and beta copies of some slick new data extraction software that runs extremely well on UNIX, think before jumping. It may be best to stick with DB2 tables for the first six to eight months and convert once you have had the opportunity to do some road tests. You will have more internal support and fewer situations where the source of problems may be unclear. Weigh the cost with the benefit. Is it more costly to build in a new environment now and retrain/hire new support teams or wait and convert later? You may find that there are numerous tools out there that can access data wherever it resides and do convertions somewhat painlessly. Depending on what data access and transformation tools you use, it could just mean purchasing additional components.

Try to overestimate for size, performance, capacity and costs. It may be necessary to then double or even triple them over the term of the project. Many data warehousing project management teams have found it necessary to purchase additional space, processors, and hardware to make up for underestimating original requirements. A transaction based data store five gigabytes in size could end up being 40-50 gigabytes in the warehouse due to multiple summarizations, indices, and historical data (that may have previously been archived on tape). Don't forget external data feeds, images, sound, or even video data sources.

Maintain consistency with data definitions. Most of the battle on this front will be between the different user areas. Have a tough-minded, no nonsense individual (better known as the DW data administrator) get users and, if necessary, the source system administrators in a room and lock the door. Don't let them leave until they arrive at a consensus. Do this over and over again until you get results. This is an area where you simple cannot compromise. Without consistency there will be a lack of confidence in the data. If the business line isn't confident about the accuracy of the data, they won't use it. If they don't use it, you stand the chance of losing executive and management support. Then, start checking out the want ads because the project will have become obsolete.

Once you have obtained a high (if not complete) level of agreement for the data definitions, publish them. Have each business and technical area represented

sign-off on the new meta data structure and wallpaper it everywhere. You will have to do this for each subject area as data is added to the warehouse, so be prepared.

Don't get caught up in trying to replace hard copy reports with the exact same information on-line. This gives cause for inaccurate or incomplete extracts. Your team will then end up having to go back and revisit the same subject or data sources again. If you don't ask, don't make the assumption that users will tell you everything they want. Most user generated reports were created due to a lack of information. By providing weekly or monthly 'snapshots' or critical areas for review (both good and potentially bad) by market area or geography, you may be able to negate five hard copy reports and give users twice the information at their fingertips.

If you see a potential gap, ask the users. Let them see complete lists of data fields (and definitions) so that they have some idea of what they are missing. Have them check off what they need and add any summarizations, data hierarchies or classifications that they may not see. Encourage questions if they are unsure as to what a specific data element contains. Use reports as a guideline to start off the discussion in a non-threatening, non-technical way. This will give you (or those in the DW team) and opportunity to better understand how to satisfy the business need.

There will be a need to review some major steps and outlined a potential strategy for moving forward. You will want to do some reading about data warehousing development and design as well. For now, let's go over a generic drill for building the warehouse:

- ◊ Identify what data is there/ what data is needed
- ◊ Identify systems, platforms, constraints
- ◊ Design data models
- ◊ Extract data elements/Test data models
- ◊ QA/Test/Cleanse
- ◊ Build meta data tables
- ◊ Convert/Transform/Massage/Sort data elements
- ◊ QA/Test/Cleanse
- ◊ Update meta data tables
- ◊ Summarize/Build hierarchies/Sort data elements
- ◊ QA/Test/Cleanse
- ◊ Update meta data tables
- ◊ Build table structures
- ◊ Load data into table structures (test environment)
- ◊ QA/Test/Cleanse
- ◊ Update meta data tables
- ◊ Move tables into production environment
- ◊ QA/Test/Cleanse
- ◊ Finalize & automate all programs
- ◊ QA/Test/Cleanse
- ◊ Update meta data tables
- ◊ Make data available to users

Hint: The difference between QA and Test is simple:

QA - Is the data valid?
(ex. 300,000 customers and only 500 social security numbers show up...better check it out)
Test - Do the results make sense?
(ex. Delinquency numbers that are summed over time instead of averaged...doesn't make sense)

Let us make the assumption that you have read the books, magazine articles, attended seminars and classes and know the drill on getting to the data. You've selected the tools (or they've been selected for you), have resources and funding identified, and even built a strong prototype with some customer account information. Looks like you're ready to start attacking data, right? Well, not just yet. Let's take a look at the other side of this see-saw.

THE BASICS - PROJECT MANAGEMENT

Project management uses methods and techniques to keep track of processes and tasks based on objectives defined within a specific scope. The methods and techniques with which you manage a set of processes can be utilized across projects, departments, even industries. Keep in mind that the more specific processes are not as generic. You will need to tailor priorities, milestones, resource allocation, etc. to the company, people, environment, and project at hand.

The important thing to remember is that clearly defined objectives, time frames, budgetary constraints, tasks and resource allocations need to be established **up front**. In addition, limitations on tools, technologies, or systems infrastructure need to be identified as factors that can affect a project at all times.

Given these fundamental issues, you certainly have your hands full. But wait, there's more. Remember the users? Well, don't forget that collecting and keeping track of their requirements is an ongoing process. It is all too easy to get so caught up in technical fires or successes and wander off the proverbial track. You're role is to keep the team focused on the objectives. This includes both short-term and long-term.

There will be a need to manage priorities, your people skills, different levels of user training, vendor relationships, and a team consisting of many individuals, some of whom will not report directly to you. Try not to let this overwhelm you. Take the time to do it right the first time and keep track of where you and the project have been. This will help you stay focused on where it needs to go and help avoid repeating mistakes. And, there will be mistakes. Just learn from them and move on.

There are many different methodologies to follow. The author strongly suggests that if you haven't had some extensive project management experience or training that you enroll in a minimum two-day seminar. Remove yourself from your every day environment and focus on getting organized and disciplined. Ever hear the phrase, 'it takes 21 days to make a habit and twice as long to break one?' It can be very easy to fall into old processes or patterns.
As a project leader you need to recognize when this occurs and correct it quickly.

It is usually assumed that funding or resources determine success or failure, especially for a data warehouse or data mart project. Yet, the major downfall is usually communication or lack thereof. Every individual who touches or is touched by this project is important. No one less or more than any other. One of your goals is to ensure that all are treated fairly and with respect at all times.

Many people can feel overwhelmed by the process once it begins rolling. And, more often, people can feel very threatened by change. When someone feels that he or she is not being heard or that his or her ideas are not being considered, the project can begin to lose footing. Through bi-weekly meetings and monthly email bulletins or formal updates, ensure that everyone involved has an opportunity to be empowered and accountable for different aspects of the process.

Utilize your management and executive sponsors to support these communications so that constant recognition is provided to members of the team. No one wants to work on a project where it appears that management doesn't care or understand what its purpose is. Any executive or senior manager can sign an invoice. *It takes a strong leader to follow through and provide ongoing, visible support, coaching, and guidance.* Make sure you choose at least one of these types of leaders to support your team.

Stay in control of the project, utilize all available resources, and document processes. After every major milestone of the project, go back and review the successes and problems. This will help the team learn from past pitfalls and move forward more effectively.

THE LETHAL COMBINATION

The design and implementation of automated, reusable, and manageable processes for a data warehouse or data mart project must incorporate a methodical strategy that includes iterative planning. Say that three times fast. Better yet, just memorize the words **METHODICAL** and **ITERATIVE**.

The goal here is to track your path so that you can keep an eye on where the project has been and where it is going. You should also look for potential problems down the road (like the need to juggle funding or resources). Remember, there should be a constant flow of communication to and from all who touch this project. Everything from successful milestones and task delays to changes in priorities or time frames need to be tracked and communicated.

Sound like a lot of work? Imagine being the person who has to steer this ship right after the resident expert walks out the door...with everything in his/her head and nothing documented. Those who have been there know it is a long and difficult clean up process. Some have found it easier to start over to avoid wasting additional time and money.

This type of project can be a bear to maintain if frequent changes in products, software, processes, or even resources occur. Plan to delegate many of the processes identified wherever possible. Have different members of your team be responsible (empowered *and* accountable) for different aspects of the project. Review these responsibilities on a regular basis to ensure there is a comfort level with all participants.

Keep all processes and documentation up-to-date and be prepared for sudden changes that you cannot control. An example may be changes in management sponsorship, which in some industries can happen every 12-18 months like clock work. If you maintain a thorough tracking system including a charter or executive summary, there is less of a need to rebuild key support or financial backing each time such changes occur.

At the end of this paper is an outline of some key areas important to managing your data warehousing project. Apply your own personal style to an approach but stick to it. You'll find your life to be quite a bit less stressful in the long run.

CONCLUSION

The role of project manager, at its high point, provides a great deal of recognition and opportunity to expand one's leadership abilities. It can open doors to exciting, new professional experiences. At its low point, this role can mirror that of a high paid, neurotic baby-sitter. As a data warehousing project manager, you have the additional challenge of keeping the reins on a very volatile environment. A sometimes dark place where players change often, progress can take an extensive amount of time, data is never clean, and funds are stretched thin more often than not. The true measure of success is not how much data is in your repository but how much it becomes relied on by the business lines. By making consistent, accurate progress and being a disciplined and effective leader, you are at making it easier for someone to pick up where you left off at any point in the process.

Document Processes	Start to finish, establish an overview of what the project objectives are, when, where, how and why the project was established. Include overviews, technology reviews, tool selection processes, cost analyses & justification, as well as resources (identify who will be involved with the project). Be sure to include who was involved in the recommendation and approval process.
Document System Architecture	Draw the big picture. Identify the current environment and where the data patterns currently fit. Then, diagram the envisioned infrastructure. Be careful not to try and replace or rebuild everything. Work within the environment in place as much as possible, otherwise you will be biting off more than you can chew...and losing the support of valuable systems resources in the process.
Document Data Flows	Use CASE tools or visual diagrams to track where data is coming from and where it is going. CASE tools can also be used when building the meta data environment.
Maintain Contracts	Review all contracts and confidentiality agreements in detail. Send copies to your internal legal contact for review as well. Don't hesitate to make adjustments where necessary. Remember, most contracts are negotiable and you are the customer. Keep copies on file locally and within your legal or administrative department.
Document Tool Inventories	This includes hardware, software, and applications. Keep an inventory of hardware and software versions, installation procedures and disks or CDROMs. Also, keep track of all upgrades or replacements.
Track Costs & Expenditures	Keep copies of invoices for consultants, software, hardware, space, resources, major supplies, etc. Develop a tracking system and keep it updated. When it comes time to do return on investment analysis and annual budgeting, you will have everything you need.
Establish Contact Lists	Have a list of names, numbers (home if necessary), and roles for all individuals involved with the project both internally and externally. This should include staff, vendors and consultants. An outline of who does what will be useful especially for your back-up when you take that much needed vacation.
Maintain Meta Data	All data definitions related to user defined data, application based data, and system administration needs to be complete at all times. Make sure that the appropriate individuals get copies of updates or changes as they occur.
Update & Use Project Plans	A bi-weekly tracking system for all aspects of the project should be maintained at all times. This should include major milestones, start and finish dates (estimated and actual), priorities, resources, costs, and additional comments where necessary. Gantt charts are useful for tracking many tasks over time. Make sure that those individuals accountable for tasks are also empowered to help establish priorities and time frames where applicable. Distribute this plan to all sponsors, users, and team members so that there are no surprises regarding expectations.
Maintain All Code	This documentation will need to include copies of source code, logistical flows, programs, custom applications, and maintenance procedures. If you are utilizing consultants, make sure they document their work. They may not be on-site in six months when you run into a product modification or business acquisition.
Have Disaster Recovery & Back Up Procedures	These should be complete as well as concise. Keeping back up copies of meta data, applications, extraction programs, and data tables is great. Just make sure you include easy to follow procedures for how to put it all together from scratch. Keep copies of all documentation, files, and applications off-site or at another location within the organization. There are numerous companies that specialize in back up storage facilities if your organization does not have one.

REFERENCES

Inmon, W.H. and Hackathorn, R.D. (1994), *Using the Data Warehouse*, New York: John Wiley & Sons, Inc. 1-3.

USEFUL RESOURCES

Adriaans, Pieter and Zantinge, Dolf (1996*), Data Mining*, New York: Addison-Wesley Publishing Company, Inc.

Gill, Harjinder S. and Rao, Prakash C. (1996), *The Official Client/Server Computing Guide to Data Warehousing*, Indianapolis, IN: Que Corporation.

Inmon, W. (1992), *Building the Data Warehouse*, New York: QED Technical Publishing Group.

Mattison, Rob (1996*), Data Warehousing Strategies, Technologies, and Techniques*, New York: Magraw-Hill, Inc.

SAS Institute Inc. (1995), *Data Warehousing A Blueprint For Success*, Cary, NC: SAS Institute Inc.

Whitten, Neal (1995), *Managing Software Projects Formula for Success*, New York: John Wiley & Sons Inc.

AUTHOR CONTACT

Kim Foster
Application Technologies
CoreTech Consulting Group, Inc.
1040 First Avenue, Suite 400
King of Prussia, PA 19406-1336
Phone: 610-992-0800 ext. 3542
Fax: 610-337-2333
email: kfoster@coretech-group.com