

1. Understanding Data Warehousing and Business Intelligence

1.1. Background – End to End Information Delivery

Data Warehousing, as a process, has become mainstream primarily because it allows decision-makers access to data and advanced navigation of “informational systems”. The word *informational* is key because it implies that data is stored to optimize the ability to access and analyze data for decision-making purposes. Traditionally, data has been stored in formats which optimize *processing* and *storing* efficiency; not for reporting or analytic purposes.

“Experts have estimated that only a small fraction of the data that is captured, processed, and stored in the enterprise is actually available to decision makers. While technologies for the manipulation and presentation of data has literally exploded, it is only recently that those involved in developing IT strategies for large enterprises have concluded that large segments of the enterprise are ‘data poor’.

Recently, a set of significant new concepts and tools have evolved into a new technology that make it possible to attack the problem of providing all the key people within the enterprise with access to whatever level of information is needed for the enterprise to survive and prosper in an increasingly competitive world.

‘Informational systems’ have to do with analyzing data and making decisions, often major decisions about how the enterprise will operate, now and in the future. And not only do informational systems have a different focus from operational ones, they often have a different scope. Where operational data needs are normally focused upon a single area, informational data needs often span a number of different areas and need large amounts of related operational data, often from many disparate and stand-alone legacy systems.”¹

Ralph Kimball, a data warehousing forefather, used a modified Zachman framework and divided data warehousing technical tasks into two categories – the *Back Room*, and the *Front Room* (Figure 1-1). The Back Room is composed of the processes and technologies required to structure the data into a common data store. The Front Room deals with the analytical reporting and presentation component of the data warehouse process and is often referred to as *Business Intelligence* (BI).

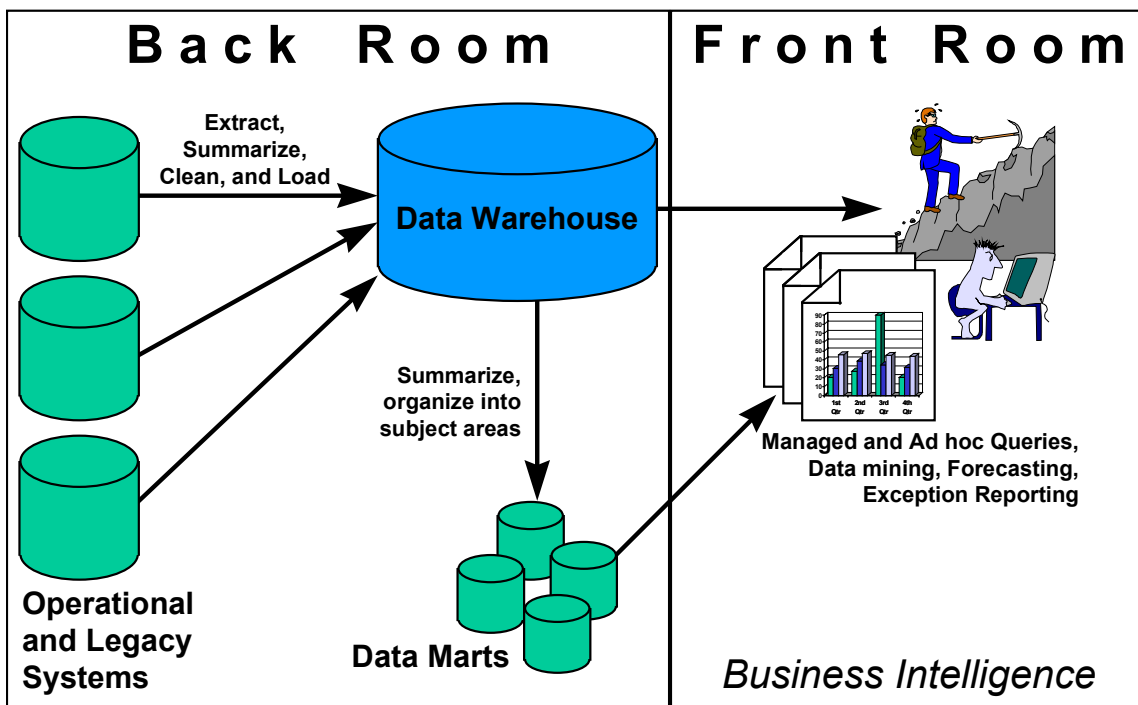


Figure 1-1: Kimball-Zachman Technical Tasks

¹ Orr, Ken, Data Warehousing Technology, Ken Orr Institute, 1997

Tools and methodologies associated with the Back Room include physical and logical data modeling, extract, transform, and load (ETL), databases, meta data repositories, and data cleansing. Front Room or Business Intelligence (BI) applications include ad hoc query, multi-dimensional reporting, data mining, forecasting, statistical analysis, and exception reporting. Portals and dashboard type products are also associated with the delivery of warehoused information and therefore are also part of the Front Room.

1.2. The Data Warehouse

A data warehouse has been defined in many ways by various vendors in the software industry. In 1992, Bill Inmon, a recognized expert in the data warehousing field, popularized the term and provided the following definition:

“A data warehouse is a subject orientated, integrated, time-variant, non-volatile collection of data in support of management decisions.”

Bill Inmon, *Building the Data Warehouse* (John Wiley & Sons)

Expanding upon this definition, each characteristic can be further described as follows:

- ◆ **subject-orientated:** Data is organized according to subject instead of application e.g. an insurance company using a data warehouse would organize their data by customer, premium, and claim, instead of by different products (auto, life, etc.). The data organized by subject contains only the information necessary for decision support processing.
- ◆ **integrated:** When data resides in many separate applications in the operational environment, encoding of data is often inconsistent. For instance, in one application, gender might be coded as “m” and “f”, in another by 0 and 1. When data is moved from the operational environment to the data warehouse the assumption is a consistent coding convention e.g. all gender data is transformed into “M” and “F”.
- ◆ **time-variant:** The data warehouse contains a place for storing data that is five to 10 years old, or older, to be used for comparisons, trends, and forecasting. This data is not updated.
- ◆ **non-volatile:** Data is not updated or changed in any way once they enter the data warehouse.

Less technical definitions also exist as noted in Simon’s *Data Warehousing for Dummies*:

“Data warehousing is the process of creating an architected information-management solution to enable analytical and informational processing despite platform, application, organizational, and other barriers.”

The key words underlined are important because they emphasize key activities and expected results of a data warehouse:

- ◆ **process:** A data warehouse is not a “technology” or an “entity”, it is a process. It is the process of integrating data from multiple sources in order to facilitate better data access and data navigation.
- ◆ **enable:** The data warehouse is an enabler of analytical and informational processing. It creates new possibilities within the organization as data previously inaccessible to users can be navigated and used for decision support activities. This enabling result is what must occur in order to achieve ROI (return on investment) in a data warehousing effort; simply attaining faster access to the same information previously available is prohibitively expensive as the main focus for a data warehouse. A good Business Intelligence implementation will provide the focus necessary produce new decision support data.
- ◆ **despite:** Users do not, and should not need to, care where the data comes from – the application, hardware, network, etc. Accessing, navigating, and using the data should be their primary focus – not understanding the environments from which the data is sourced and the processes in which it is cleansed, aggregated, and updated.

1.3. **Business Intelligence**

The term Business Intelligence (BI) was first coined in September 1996 by a Gartner Group report which indicated a “Information Democracy will emerge in forward-thinking enterprises, with Business Intelligence information and applications available broadly to employees, consultants, customers, suppliers, and the public.]..[The key to thriving in a competitive marketplace is staying ahead of the competition. Making sound business decision based on accurate and current information takes more than intuition. Data analysis, reporting, and query tools can help business users wade through a sea of data to synthesize valuable information from it – today these tools collectively fall into a category called ‘Business Intelligence’.”

Business Intelligence can thus be defined as a broad category of integrated applications and technologies for gathering, storing, analyzing, and providing rapid and effective access to business orientated data to help enterprise users make better business decisions.

The key words underlined are important because they emphasize key activities and expected results of business intelligence:

- ◆ **integrated applications:** A single application can not provide the breadth of analytical capabilities required for Business Intelligence. Query and reporting, online analytical processing (OLAP), statistical analysis, forecasting, and data mining are all tools that must work cohesively as a unit to satisfy decision support activities.
- ◆ **rapid and effective access:** Information must be quickly available at key times. From a user’s perspective, the information is only valuable if it can be accessed quickly when they need it most. Delays caused by poor programming, slow networks or the unavailability of the data during critical business times such as month and year end will cause Business Intelligence installations to falter.
- ◆ **business orientated data:** Business Intelligence software summarizes, translates, and organizes many operational datasets into subject areas that are designed for reporting. For example, an insurance company would have a consolidated customer subject area that included relevant policy information from each of the individual transactional systems that collect vehicle, house, and life insurance information.

1.4. **The Three Tier Data Warehouse Methodology**

1.4.1. Information Delivery: A Paradigm Shift

When Management Information Systems (MIS) were popularized in the late 1980s and early 1990s, the first BI tools were created as self-service applications. It was expected that providing high functionality, reporting toolsets to the masses would create an analytical revolution – an Information Democracy. Vendors guaranteed that this self-service model would allow users to easily browse data sources, build reports, and perform complex analytical queries and consequently reduce the need for a central reporting team. It was also predicted that users wanted self-service reporting and would be clamoring for the new technology as it provided access to data never before available.

By the late 1990s, it was becoming increasingly apparent that the analytical revolution was not proceeding as predicted. Data Warehousing and Business Intelligence was certainly creating an awareness and desire for data availability, but most users did not want self-service reporting. Instead, the complexity of the Data Warehouse, caused by the depth and breadth of the systems they incorporated, had caused an increased need for a central reporting team that understood how the data was retrieved, could gather the requirements from the users, and program the reports using the complex Business Intelligence products.

Until early 1999, Business Intelligence products were predominately a “thick” client application that required a client-server relationship. Each PC in the organization had one or more BI tools (such as query, multi-dimensional, and/or forecasting software) resident on their hard-disks that accessed a central data repository (cube, database, snapshot, etc.). These PC resident programs were fully functioning, high-capability tools that allowed users to browse data sources, build their own reports, and perform complex analytical operations. Although it was *assumed* that providing these tools would create an information democracy, a vast majority of users (over 90% in a typical organization) never used the ad hoc or analytical capabilities provided by these tools. The users instead preferred to execute pre-programmed or batched reports. Even with prompting and filter capabilities being used to isolate problem areas, only 20% of the products’ functionality were used by the majority of the information viewers. The exceptions to this rule were the handful of business analysts who requested training on the tools, knew the business, and routinely explored and reported on the data.

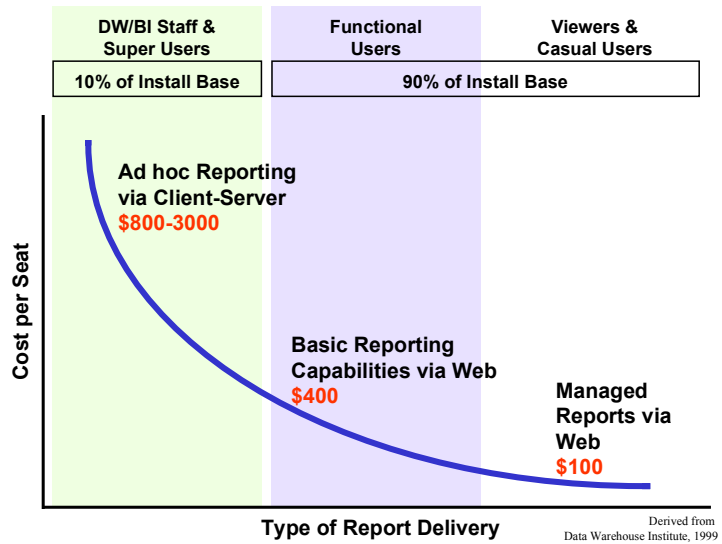


Figure 1-2 Typical Report Delivery Costs Per Seat

Recognizing this behaviour, the associated maintenance costs, and web/e-commerce initiatives, the major BI vendors have produced browser-based reporting clients. These zero-footprint (no software except the resident internet browser is required) tools offer a reduced, yet functional, set of report-viewing capabilities to the end user. These web-based products typically have prompting, slice and dice, and drill-down capabilities; however, they have only limited support for report creation. Additionally, the front end components of these browser-based applications are significantly cheaper (Figure 1-2) and require less training for the majority of a customer’s install base.

The current view of BI data needs and privileges is represented by Figure 1-3, which also indicates the role of a central BI support staff. Notice that 90% of the install base use only the basic report viewing functionality.

Business Intelligence Needs and Privileges

	User Category	Type of User	Information Needs	
Browser Enabled Software	Viewers	Executives Managers	<ul style="list-style-type: none"> Summarized information Predefined reports and parameters Ability to view online or print to a local printer 	More than 90 % of the users
	Casual Users	Managers Supervisors	<ul style="list-style-type: none"> Require next level of detail from the viewers Ability to execute and refresh pre-programmed reports Ability to enter parameters 	
	Functional Users	Managers Supervisors Analysts	<ul style="list-style-type: none"> Does detailed research & analysis which needs transactional data. Has the ability to create basic reports (ad hoc and OLAP) 	
Client Resident Software	Super Users	Analysts	<ul style="list-style-type: none"> Strong understanding of <i>both the business and the technology</i> has access to transactional data full privileges to explore and analyze the data creates and maintains reports 	5-8% of users
	Business Intelligence Support Staff	IM/IT Staff	<ul style="list-style-type: none"> Very strong understanding of the technology Moderate understanding of the business Relies on specifications and Analyst support to build and maintain reports for all levels of users 	2-5 % of users

Based on DM Review and Gartner Reports

Figure 1-3: BI Needs and Privileges

Over time, users may migrate between the classifications. Functional users, as they learn the data and become familiar with the tools, may become super users and require the installation of a client-resident program, just as a change in job may move an analyst away from the day-to-day exploration and analysis and into a functional or even casual role, allowing a client license to be recovered.

The key to success is to have a Data Warehouse/Business Intelligence team that is capable of fostering excellent communication with key Business Analysts and the ability to become a conduit of information between IM/IT departments.

1.4.2. DW/BI Group as Trusted Advisors and Integrators

Although there is a distinction between the Back Room (loading of data) and the Front Room (presentation of information) in a Data Warehousing environment, it is imperative for the success of the Data Warehouse that these skills be maintained within a central team. The professionals found in this team have complimentary skill sets and data warehousing by its very nature, requires a tremendous amount of interaction and group participation in order to successfully retrieve and cleanse source data, load consistent data into the central database and then create reports and/or data marts for user consumption.

The recommended configuration for a Data Warehousing effort is therefore the three tier methodology (Figure 1-4) with a close-knit DW/BI group at the core, interacting with a select set of business analysts (a.k.a. super users) and system specialists within the various IM/IT departments. The Business Analysts are responsible for interacting with their small, localized sets of end users. This interaction involves two way communication as the end users pass requirements, data requests, and issues to the analysts, who act upon the demands by programming reports and procuring new data sources.

The IM/IT departments are charged with the day to day maintenance and operations of the Data Warehouse and Business Intelligence system components such as the network, databases, desktops, web servers, etc.

The DW/BI team's responsibility is keeping the entire process working smoothly and ensuring all the necessary components are present. The core team is also responsible for identifying roadblocks, enforcing the procedures and policies that produce successful data warehouses, and for the coding and maintenance of data staging processes. Data Warehouse and datamart design, ETL construction, and the coding of complex reporting requirements also fall within the purview of the DW/BI team. Working closely with the Database Administration group to ensure the integrity and efficiency of the reporting databases is a role that must be embraced by the DW/BI team as effective reporting relies heavily on effective, efficient, and well maintained databases.

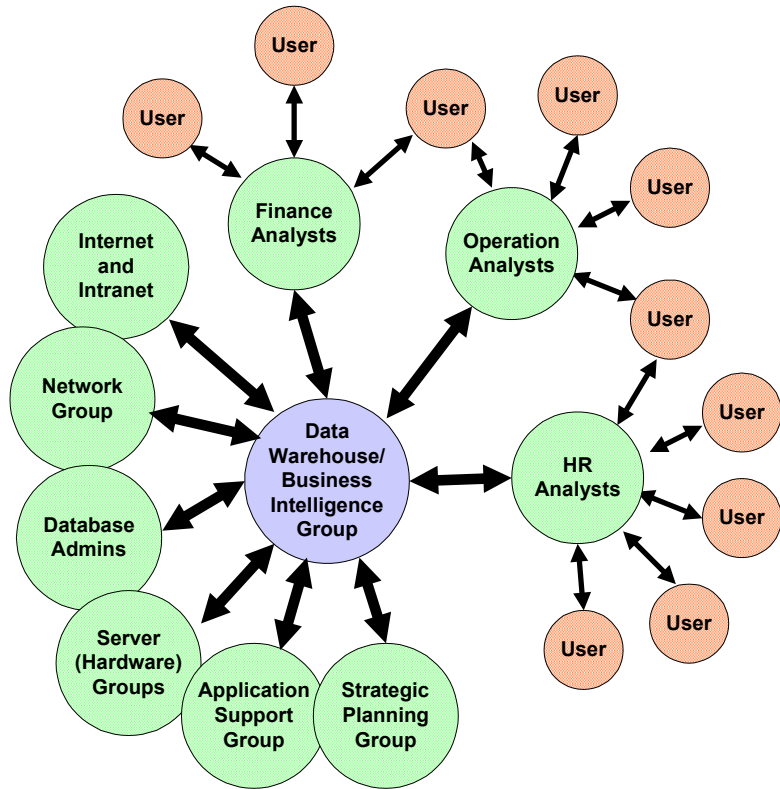


Figure 1-4: Three Tier Data Warehouse Methodology

As the information delivery paradigm shifts from self-service reporting towards full-service delivery of managed reports, the centralized Data Warehouse/Business Intelligence group plays an increasingly important role in the corporate IM/IT department.

Creating the three tiered approach to information delivery will allow the IM/IT department to provide more proactive, customer focused services. Thinking of these services solely as report programming and coding will regress the IM/IT department back into the pre-warehouse days where the users blame the IM/IT department for delays and misinformation. Changing the focus of DW/BI services from report programming to *information delivery* allows a select set of Business Analysts (super users) the ability to create context sensitive reports for localized user communities (Figure 1-4). Information delivery encompasses roles of requirement gathering, modeling, design, coding, and construction of not only the reports, but the data structures (i.e. the Data Warehouse and datamarts) that will provide the users with clean, consistent data.

Assuming a certain level of tool specific training, these Business Analysts can produce reports much quicker than a programmer within the central DW/BI group because they fully understand the business. The system specialists within the IM/IT departments are focused on the day-to-day upkeep of the enterprise's many production systems and infrastructure.

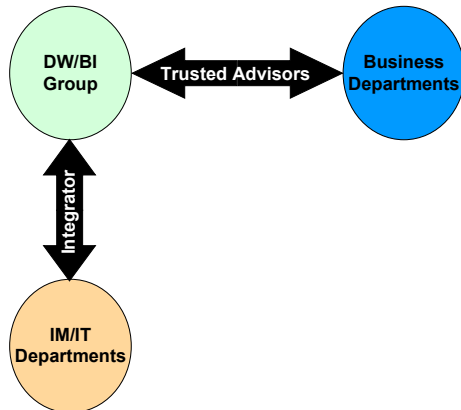


Figure 1-5: DW/BI Group as Trusted Advisors and Technology Integrators

So if the business analysts are coding the majority of reports, and IM/IT departments are maintaining the system components, what are the DW/BI programmer-analysts doing? The central DW/BI group act as *trusted advisors* to the business and *integrators* of information technology (Figure 1-5). Their role is to act as an intermediary between the business and the myriad of technical difficulties that are encountered in data collection and manipulation. They are responsible for taking the users' data requests (via the Business Analysts), locating the source system, extracting, transforming, cleansing and loading the data into the central data warehouse and updating any of the necessary data structures or datamarts.

Placing this central group in a position to solve technical questions (advanced help desk function), resolve performance issues, deal with software and infrastructure inconsistencies, and develop new data sources will enable a mature and robust, end-

to-end information delivery system – the true goal of data warehousing.

In short, the DW/BI group facilitates solutions to technical problems that allow the business analysts to answer business questions.

For this reason The Data Warehouse/Business Intelligence staff cannot be mere programmers who take specifications and code reports for the end users. Today, these professionals must be analysts who possess a very high degree of interpersonal skills: communication, negotiation, interviewing, listening, and synthesizing are critical elements in an effective DW/BI analyst or architect. The ideal candidate is also outgoing and must be capable of interfacing with both the business and technical teams. Proficient knowledge of the business processes, terminology, and business drivers as well as in-depth knowledge of the Data Warehousing and Business Intelligence tools are also essential for the DW/BI team member.

The Data Warehouse/Business Intelligence team is often the only department in IT/IM that interfaces with all others. Information delivery relies on Unix and NT teams for server support, e-commerce and web teams, network administrators, database administrators, business analysts, application support experts, and many of the other teams to support their information distribution activities. When a report or data is erroneous, customers (i.e. the business via the Business Analysts) will call on the DW/BI staff to help solve the problem. Customers do not see “server not available”, “database error”, or “application error”, they simply know the *report is not returning the proper results*. As such, the Data Warehouse/Business Intelligence staff need to be placed in a position to facilitate, through interactions with the rest of the IM/IT organization, solutions to any reporting problems the customer may be experiencing.

The primary responsibility of the central DW/BI group is to work with the various IM/IT departments to supply the business with clean, timely, and consolidated data.

The Data Warehouse/Business Intelligence group fulfills it's responsibility by performing the following roles or functions:

- ensuring a standard environment is established
- maintaining meta data
- optimizing data stores and reports as necessary
- requirement gathering and defining functional specifications
- data modeling
- coding and integrating ETL modules
- mentoring and training Business Analysts and end users

In order to complete these roles, the DW/BI group must have a basic understanding of most, if not all, elements within Information Management/ Information Technology. They should be able to converse comfortably on the high-level theories and methodologies of networking, database design and management, web functionality, server setup and maintenance, and host of other related technical activities. The appropriate IM/IT departments will handle the detailed work, but the DW/BI will have initiated and coordinated the activities pertaining to reporting and information delivery. In addition to general IM/IT knowledge, the DW/BI staff must have in-depth experience and knowledge of the data warehousing design, implementation, and delivery tools (see Figure 1-6) in order to be effective.

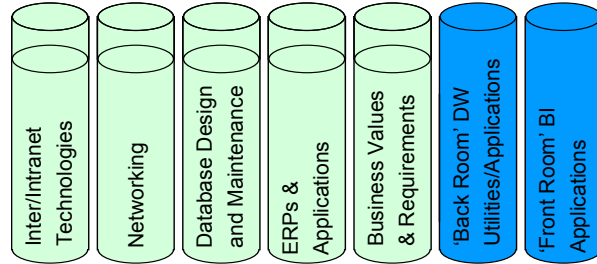


Figure 1-6 DW/BI Knowledge and Experience Requirements

1.4.3. Business Benefits to the Three Tier Infrastructure

Moving away from a pure self-service reporting concept and into three tiered approach to information delivery has several key benefits to the business community. End users who do not wish to explore the data or understand the underlying technology, nor have time to produce their own reports will be serviced by a set of knowledgeable business analysts. Having business people service business people is the most effective way to ensure the data is context sensitive, correct, and on-time. The Gartner Group estimates that the majority of organizations, in non-data warehousing environments, have their business analysts working more almost one hundred percent of their time performing data collection activities that could be automated by a suite of canned reports. Training and empowering the business analysts to code reports and work as decision support programmers, will actually free up their time to do more of the one-time analysis, data mining, and data exploration that is required for new and innovative insights into the business. This time becomes available as canned reports are produced, allowing the end users to have up-to-date information on an as-needed basis, but without the need for end user programming or data manipulation. The time Business Analysts used to require for collection, integration, and presentation of information requests can be used for true analysis such as data mining, statistical analysis, or forecasting.

1.5. The Politics of Data Warehousing

Data Warehousing is one of the most politically oriented projects any company or government agency will undertake. Project managers and senior management often shy away from this topic due to its sensitive nature and the ‘black cloud’ it seems to generate. However, the fact of the matter is, data warehousing is by nature rife with political issues. Choosing to accept this fact, identify the contributing factors, and find solutions will help the organization’s Data Warehousing project become a success. Ignoring the socio-political influences will jeopardize the effort, and greatly increases the chance that the Data Warehouse initiative will become one of the forty-plus percent that fail.

Establishing the central DW/BI group as hub for Business Analysts and the IM/IT departments will increase political sensitivity, but in the medium and long term will ensure that a consistent approach to data warehousing is established, instead of the stove-pipe like implementations associated with legacy systems.

1.5.1. Getting into Politics

Included here is a condensed version of *The Politics of Data Warehousing* by Marc Demarest. <http://www.hevanet.com/demarest/marc/dwpol.html> further details the pitfalls to avoid during data warehousing initiatives.

“Information technology is, for better or for worse, social these days. The good old days of batch and online transaction processing systems design and deployment are gone; we buy those things now, from independent software vendors. The systems we have to build – decision support systems, computer-supported collaborative work environments, workflow systems, intranets, extranets, whathaveyou-nets – are all deeply and inextricably social applications of IT: computing applied to groups of people with power, status, and a network of relationships.



That means, for better or for worse, that politics is an integral part of the IT projects from here on in. Or out, depending on your perspective. And that, in turn, means IT Professionals have no choice but to get into organizational politics, understanding the forms, shapes, and paths organizational politics takes, and become astute at navigating in a political environment. Not because politics is cool, or fun, but because politics is a feature of the landscape: the beast standing between us and the gate marked "successful project conclusion"

Data warehousing projects are always potentially political because:

- they cross organizational treaty lines
- they change both the terms of data ownership and data access
- they expose the often-checked history of data management in the IT organization
- they affect the work practices of highly autonomous and powerful user communities within the firm

TYPICAL DIFFICULTIES ENCOUNTERED:

Loss of Power – Data Ownership and Data Access – If there is any rule that does apply across organizations regardless of their market focus or structure, it is this: power accrues to those who:

- gather data
- control access to that data

Visibility to Bad News/Bad Data – The real political problem with data warehousing is not the loss of data ownership that such projects imply, for every organization is asked to contribute to the warehouse, but a loss of control over access to the raw data itself, something frightening for groups with:

- dirty data
- ambiguous data
- unflattering data

Altering the Decision-Making Power – The work practice of decision-making has been done historically outside the IT infrastructure of the firm. Data warehousing projects threaten this long-standing practice. And they create, in knowledge workers, what Thorsten Veblen called in another context, "the conscious withdrawal of efficiency": passive-aggressive behaviour on the part of knowledge worker communities that includes:

- an unwillingness to participate in requirements gathering, schema design activities, and pilots
- failure to use deployed warehouses and marts
- endless, and pointless, micro-analysis of the "quality" or "real meaning" of the data provided by warehouses and marts.

MASTERING THE POLITICAL: 10 COUNTERMEASURES

- 1. Change your mindset.** As a warehouse designer, you are a sociologist, marketer, diplomat and technologist, probably in that order. Spend more time thinking about your constituencies and their needs, how you'll market to those constituencies, and how you'll establish treaties and working relationships with those constituencies than you spend drawing data models and technology architecture diagrams.
- 2. Get comfortable with being frank,** particularly with data set owners. Expect, and say that you expect, to find dirty, inconsistent, incomplete data. Help the data owners get comfortable with this as state-of-nature, rather than some failing on their part. Help them clean it up; in fact, point out to them that the warehousing or maring project is an ideal opportunity for them to get out from under their data burden, to get clean.
- 3. Do the sociological analysis first.** Know all your constituencies, data and access, technical and business, what their risk/reward profiles are, what they have to give up for the project to succeed, and what they gain if the project is successful.
- 4. Develop the internal marketing plan** before you do the first design. Know who you have to sell the project to, what the value proposition for each target audience is, and how you'll know when they're on board.
- 5. Establish clear bilateral treaties or contracts with all the data-owner constituencies** involved in the project before the design phase of the project is complete. Each treaty should specify who gets what from whom when and under what circumstances.



6. **Regularly repeat and reset expectations with each constituency, face to face.** You'll know you're doing your job when the constituencies begin to notice that you're repeating yourselves.
7. **Spend at least 10 minutes of every project meeting discussing the political climate** surrounding the project, changes in the constituencies, status of internal marketing efforts, and late-breaking gossip and hearsay. When team members are uncomfortable discussing the politics of the project in company, help them to get over that discomfort.
8. **Spend at least 15 minutes of every meeting reviewing the work in progress** in terms of (a) the business objectives for the project (which must be measurable in terms of the firm's income statement) and (b) the user constituencies' needs and their work environment.
9. **Have a formal process for logging**, reviewing and either rejecting or accepting (with cost and time changes formally noted) all changes to project scope.
10. **Spend a lot of time in the face of the senior technical and business sponsors for the project.** Make sure they know, every week, what the status of the project is, what the organizational roadblocks are, and what the project team expects the sponsors to do about those roadblocks. Use your sponsors like air cover to break down, via organizational fiat, resistances in the organization that the project team cannot remove through diplomacy.²

² Demarest, Marc, The Politics of Data Warehousing, 1997