H&H Technologies
1636 Century Road
RR1 Kars, ON, Canada
K0A 2E0

Internet: www.h-htech.com
Telephone: (613) 799-7020
E-mail: info@h-htech.com

1. Data Warehouse Lifecycle

The roles and responsibilities required to build and maintain a Data Warehouse are consistent throughout its lifecycle: the responsibilities coincide with the implementation steps. Once the tasks have been identified, the roles associated to these responsibilities can be mapped.

The number of staff required to fulfill the roles greatly depends on the status of the Data Warehouse. The Data Warehouse can be in one of two phases: growth or maintenance:

> *Growth Mode* is defined as any phase where the warehouse is being actively built or expanded. This includes the initial building initiatives, and integration of the business units, but also includes any major release-based enhancements.
>
> *Maintenance Mode* deals with the day-to-day operations of the Data Warehouse and includes activities such as monitoring, error recovery, data validation and correction, operational support, and help desk.

Once the first iteration of the Data Warehouse has been completed there is an overlap present as maintenance and operational support is required for the installed business areas, even as the warehouse is expanded for the next business unit.

1.1. Implementation Steps and Responsibilities

In order to build a successful Data Warehouse, a hybrid Top-Down and Bottom-Up approach must be used. This approach keeps the projects simple, focused, and short. Return on Investment is realized in three or four months instead of the two or three years of a Top-Down (enterprise wide) implementation, yet provides a consistent organization-wide view of the data.

The Top-Down Approach dictates that the central data warehouse is build first. All common elements, attributes, and dimensions are identified and collected. All levels of aggregation and facts are placed in the corporate Data Warehouse and then *exact* subsets of the data are created in the distributed, subject area datamarts. The problems with this approach are the massive effort and lengthy time required to define the corporate data needs. This often causes data warehousing efforts to stall and ultimately fail.

The Bottom-Up Approach establishes distributed, subject-orientated datamarts first, using a multi-phase approach, and then combines the data from all datamarts into a corporate data store. The benefits to this methodology are four month turn-around on the datamarts, quick return on investment, and a chance to leverage lessons learned as the project team moves from installation to installation. The downside to using the Bottom-Up Approach in its purist sense, is preventing the creation of stovepipe datamarts. It is very easy when using a Bottom-Up approach to make design or quality concessions within the individual datamarts that will make future integration into the corporate Data Warehouse very difficult. It is not uncommon for further data quality, integrity, and business rules being required to populate the Data Warehouse from these stovepipe datamarts. In this way, the datamarts become a second staging area for the corporate Data Warehouse instead of an end user reporting facility.

The Hybrid Approach ensures that the individual business units' needs are focused upon, answered, and delivered within four month windows as per the Bottom-Up Approach The twist comes with defining and then using a framework to facilitate corporate integration. The three major steps to the iterative Hybrid approach are planning, implementation, and support (see Figure 1-1).

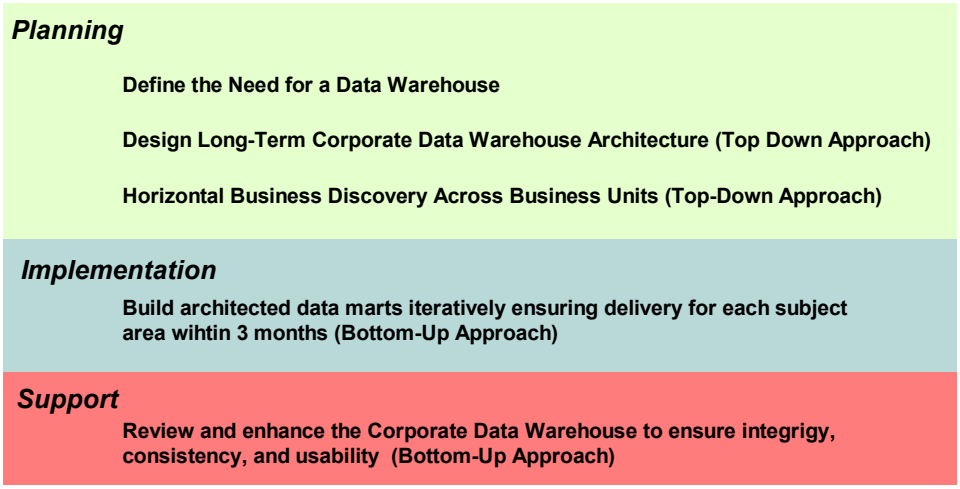


Figure 1-1: Summary Hybrid Approach to DW Architecture

Using this hybrid approach solves the integration problem as all data marts will share common business rules, semantics, and definitions are enforced by the Enterprise Meta Data Repository (EMDR). As each project is implemented they use the existing corporate standards and develop new ones as necessary. Having the EMDR is critical in this regard, as it will be used to maintain the object-oriented principles of reuse and modularity. Because the EMDR becomes the central repository of reusable code, the enforcer of standards and procedures, and dictates the corporate Data Warehouse standards, it is extremely important to ensure that all Back Room and Front Room applications can interact with the meta data repository.

Building from the bottom up will lay the foundations for future projects and will reduce the overall cost, spread the risk, and allows for fuzzy factors (e.g. database size and performance) to be addressed when they surface. The iterative project cycle encourages problems to be identified early enough in the process that they cause less re-work and result in fewer delays than the "waterfall" approach to project management. Waterfall project management, as used in most Top-Down implementations tend to leave the complicated issues to the end and consequently fall prey to changing requirements, technology problems, and unexpected business rules that cause enormous amounts of redesign, rework, and ultimately lengthy delays.

The medium and long term goal of the organization should be to have a "hub-and-spoke" like Data Warehouse Architecture. This means having a central, common data warehouse of relational design storing a combination of detailed atomic data and summarized facts feeding a series of subject-orientated datamarts. The corporate Data Warehouse will support organization-wide analysis and reporting, while the numerous datamarts will supply context sensitive information to localized or regional user communities. The central coordination and management of the Data Warehouse will be maintained by the DW/BI group and enforced by the maintenance of an Enterprise Meta Data Repository.

Detailed Steps to Hybrid Approach to DW Development, as shown in Figure 1-2, follow an iterative cycle. Figure 1-3 graphically represents the Data Warehouse System Development Lifecycle (SDLC) as it relates to the Hybrid Approach.

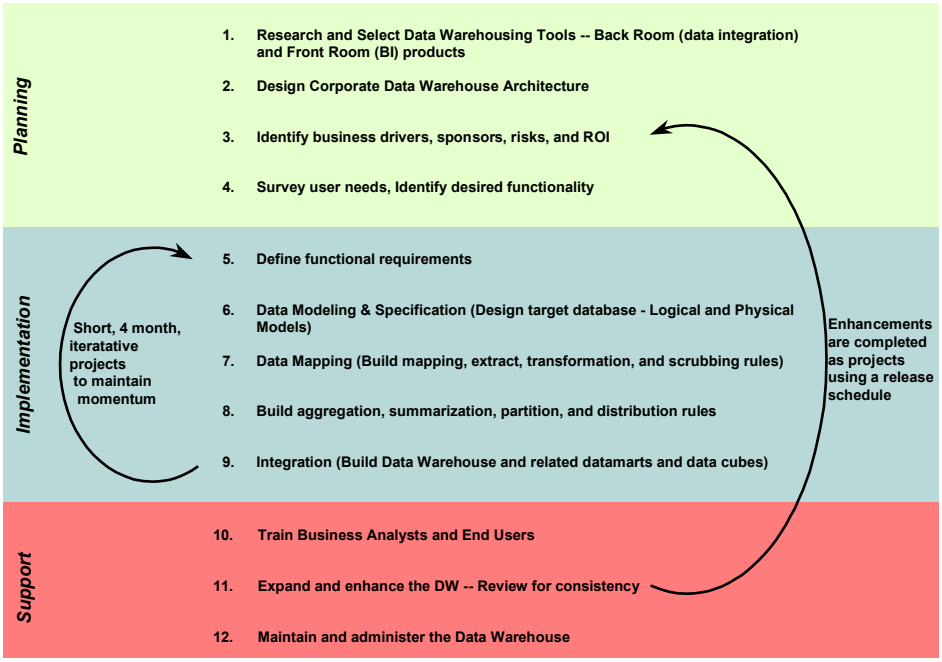
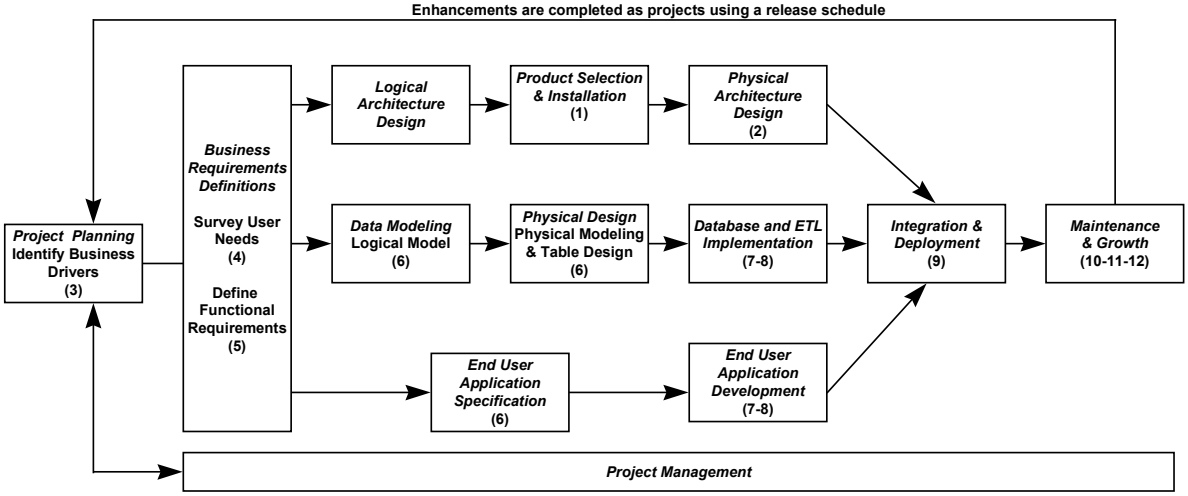


Figure 1-2: Hybrid Approach to Data Warehouse Implementation



NOTE

The lifecycle diagram does *not* reflect absolute timelines. The boxes are equally sized; the tasks and resource requirements are *not* equal

The numbers in parethenthsis () relate the Data Warehouse SDLC back to the Hybrid Approach to Data Warehouse Implementation

Figure 1-3: Graphical Representation of DW System Development Lifecycle

1.1.1. Research and Select Data Warehousing Tools

Selecting the Back and Front Room tools is critical to the detailed physical design of the Data Warehouse Architecture. The tools that are chosen and the way they interact will have a dramatic impact on the policies and procedures defined in step 2 (Designing the Corporate Data Warehouse Architecture). In order to be able to choose the tools best suited for the organization, careful requirement gathering is needed. Tool selection is discussed in greater detail in "Selecting Business Intelligence and Data Warehousing Tools".

1.1.2. Design Corporate Data Warehouse Architecture

The high-level design should be completed in a 2 or 3 day Joint Application Development (JAD) workshop including the key Information Technology/Information Management personnel, all DW/BI team members, and one or two of the more technical Business Analysts. The Data Architect can then take the high-level design and fully specify the physical architecture and relationships.

The goal of this step is to formally identify the relationships and interactions between the ETL, Meta Data Repository, database administration, and Business Intelligence toolsets. The logical design can be found in Section 7 (Data Warehouse Architecture), however the physical design needs to identify:

- server configurations
- application interactions and interfaces
- support processes
- knowledge transfer (from DW/BI group to appropriate support IM/IT staff)
- user documentation (when to use the what products – especially with the Business Intelligence suite)
- procedures and processes used to create code, test, integrate, etc.

1.1.3. Identify the Business Drivers, Sponsorships, Risk, and ROI

This step is to identify the business area requiring a Data Warehouse or Business Intelligence solution. Just as a data warehouse is built to solve a corporate business problem, so too must be the individual datamarts or subject area implementations. These problems must be specific, and painful, in that the department must be able to significantly alter the way users are doing business in order to meet the organization's goals of achieving greater efficiency and enabling the department to provide better, more effective reporting and data analysis capabilities.

The problems must be specific in order to identify the cause, effect, and solution. For example, the problem within ABC company may be the inability to efficiently collect and analyze safety related data to allow for the development of new policies or procedures. The cause is the multitude of stove-pipe legacy systems, each collection a portion of the data required for the 'big picture'. The effect is long wait times for analytical requests. Business Analysts are spending much of their time integrating data sources, cleaning up inconsistent or missing data, and sifting through duplicate data. The solution (as identified in the strategic direction document) is to create a Data Warehouse that integrates the various source systems and provides a consistent set of reporting tools for data analysis and retrieval. This same principal must be applied for each local business or subject area. If a reporting problem does not exist or the cause can not be identified, incorporating the data into the Data Warehouse will not produce a return on investment.

The problems must be painful because of the high costs associated with Data Warehousing. The initial implementation of a Data Warehouse is only the beginning of the cost curve. Ongoing maintenance, hardware, software, training, and operational support and enhancements all add up to make Data Warehousing an expensive operation. The business must be able to show a return on investment for the Data Warehouse project.

A steering committee should be created to guide the Data Warehousing effort. This committee will be responsible to establishing the order of business area implementations and for the assignment of resources to the projects. The steering committee will assign the Project Mangers (Business and Technical) and work with them throughout the project to ensure the milestones are being met and the project is on-time. The committee should be composed of a cross-section of business and IM/IT personnel and meet on a monthly basis.

During this phase the essential business questions must be asked:

What is the need?
Who is going to sponsor the project?
What is the risk of doing nothing versus attempting a data warehousing effort?
What is the expected return on investment? Worst-case, best-cast, most realistic?
Who will be the Business Project Manager?
Who will be the Technical Project Manager?

1.1.4. Survey User Needs, Identify Desired Functionality

This step clearly identifies the mission and responsibility of the sponsoring business units and begins the requirement gathering. The challenges identified in step three (Identifying the Business Driver) need to addressed in further detail. Examination of the current business processes, decision support tools, and the deficiencies in the current environment will provide valuable information on how to design a more appropriate and effective information delivery solution. Identification of the user types (viewers, casual users, super users, etc.) should be completed as business rules, definitions, facts, and dimensions are listed. Personnel need to be assigned to the key roles (DBA, BI Specialist(s), Business Analysts, etc.) during this phase.

The easiest method for the completion of this phase is through a series of half-day Joint Application Development (JAD) sessions involving the data architect, the Business Intelligence specialist, the business analysts, and appropriate user-expert representatives (a.k.a Business Analysts). Questions to consider are:

- What level of detail is required?
- What amount or length of history (how long does the data need to be retained)?
- Where is the data is being sourced from?
- What are the reporting requirements? Show samples!
- What are the current deficiencies that need to be answered or solved?
- Who is going to be assigned to the various roles?
- What additional resources are likely to be required?
- What are the security rules and implications?
- Any data integrity issues? Is the data clean and consistent?

Asking the users "how do you want to slice or analyze the data?" will provide the dimensions while collecting the measurable attributes will provide the facts.

It is important to note that it is possible a Data Warehouse solution is *not* required after this step has been completed. If the deficiencies in the business processes or technical implementations can be corrected within the source systems and Business Intelligence products can be either fixed or introduced, then the Data Warehousing effort should stop and move onto the next business area. The time and effort required to perform the JAD sessions should still be considered valuable as the end product is a better serviced user community that has access to new information.

1.1.5. Define Functional Requirements

This step builds on the survey needs (step 4) and takes each requirement and drills down to the "atomic" level. Every element, attribute, and dimension is fully qualified and documented. Information that should be obtained for each element include: source, data type, expected values, data quality (identify if there is a problem), constraints, referential integrity rules, and update frequency.

Further requirement planning JAD workshops should be used to define these exact functional requirements. Tasks, phases, milestones, project plan, schedule, budget, and resources are also identified and discussed. The JAD approach is used as it gains a consensus on all the issues surrounding the data warehousing project and helps minimize the politics.

Following the resolution of all issues and functional specifications, effort begins immediately on the implementation. "Strike while the iron is hot" in order to keep the momentum.

The first project will need an elongated requirements gathering phase, as the scope is enlarged to touch on enterprise-wide issues. The idea is to highlight as many of the common requirements, attributes, facts, and dimensions as possible. Specify the attributes and facts required by the current project and provide detailed implementation plans. Requirements identified, but not needed for the current project, should be fleshed-out (high level design and documentation only) and associated to the first future project with that particular requirement.

There is no need to define every corporate dimension, attribute, and fact – highlighting the top eighty or ninety percent will suffice as there will be some retrofitting necessary as a byproduct of the ongoing iterative project cycle. Dealing with the major dimensions such as make and model, customers, human resource requirements, and financial facts will decrease the amount of rework in each successive project. This type of approach also gives the DW/BI team an opportunity to improve the overall design during each new cycle – a feature that the Top-Down Approach can not provide as all steps are done once, in successive fashion.

A sample of current reports will provide the means to define required entities. Commonly used summarizations should be identified (e.g. sales by region by customer or budgets by department by time). Reports should be fully specified into common templates that can be later used by the BI specialists and Business Analysts for programming.

During this phase, the implementation team is further refined and augmented as necessary. Depending on the number of attributes listed, the ETL processes, data quality, and the number of report requests, the implementation team may need additional resources.

1.1.6. Data Modeling and Specifications

The design of the database specifically deals with Entity-Relationship, Logical, and Physical modeling. Tools such as Oracle Designer and Computer Associates ERWin help create representations of data facts, dimensions, attributes, and sources.

Due to it's potential data volumes, the Corporate Data Warehouse will be of modified third normal form. This means the majority of relationships will be normalized into one to many (1:M) keyed relationships. The Data Warehouse structure ends up being a series of overlapping subject areas (Figure 1-4). The common attributes, facts, and dimensions provide these overlaps, while the fact that each business area will be implemented separately (hybrid approach) will create a more subject orientated data warehouse than a true Top-Down approach.

The datamarts and cubes should be designed as star schemas and denormalized in order to simplify end user reporting and navigation. The datamarts will be further refined subject areas and only have subsets of the data located within the Data Warehouse.

The involved parties at this stage include the Data Architect, the Database Administrator, and to a lesser degree the Business Intelligence Specialist. These resources need to work as a team to ensure the database is designed to facilitate the storage of large volumes of data, the reuse of as many dimensions as possible, and the ease of data extraction for datamarts and cubes.

The Business Intelligence specialists and Business Analysts are continuing to specify the reports during this phase.

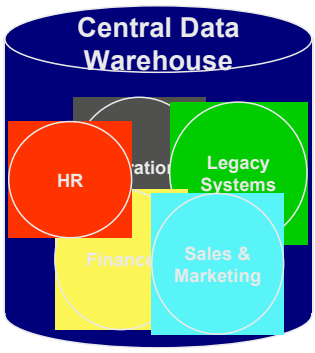


Figure 1-4: Overlapping Subject Areas Create a DW

1.1.7. Build Data Mapping, Extract, Transformation Rules

The data sources for all required elements need to be identified and mapped to the target dimensions and attributes within the physical model established in step six (Data Modeling and Specifications). Data quality should be closely examined and cleansing rules established to deal with missing or inconsistent data.

It is critical not to populate the data warehouse with inconsistent, "dirty", or incomplete data.

Extraction, transformation, and conversion rules are applied. These rules include name changes, key changes, value mapping, filters, defaults, masks, and multiple-choice (many to one) selections. It is strongly recommended that the transformation and mapping rules use referential integrity checks against clean dimension tables in order to populate the data warehouse. Any records that produce errors should be manually inspected and error routines invoked that allow new business rules to be established or data values to be correctly mapped under user intervention. Unit testing of each rule or process is mandatory and should be completed before moving onto the aggregations, partitions, and distribution phase.

This is the most time consuming step in the Data Warehousing effort and requires approximately 50% of the total time. It is imperative to identify data that is inconsistent, missing, incomplete, duplicative, or populated with incorrect values. Creating rules for standardizing values across data sources (e.g. customer numbers/names) is only slightly less time consuming than dealing with data integrity rules to populate inconsistent or missing data.

This step will ultimately define the successfulness and ROI of the data warehouse.

Reports and other BI requirements (data dictionaries, models, etc.) begin construction during this stage of the project. The Business Analysts should be involved as much as possible in the writing of reports, as this will provide hands-on training and improve their understanding of the tools they will be using to support their local user communities.

1.1.8. Build Aggregations, Summarization, Partitions, and Distribution Rules

Once the data has been loaded into the database and the ETL and data cleansing processes have been unit tested, the commonly-used aggregations, pre-computed data cubes and partitions are created. These summary levels of data are necessary to increase reporting performance. Additional tables, database views, indexes, and SQL hints are also options for making end-user reporting simpler and more efficient. Many of the ETL and OLAP tools on the market provide functionality to create these pre-calculations using their point-and-click GUIs. Unit testing of each aggregation or summarization is mandatory.

1.1.9. Build Data Warehouse and Related Datamarts and Data Cubes

Simply stated this is the construction and integration testing stage. Once the individual process and code components have been constructed and unit tested, they need to be integrated into a working model. This entails placing the ETL and aggregation processes into job streams, scheduling the jobs to populate the database, and have the Business Intelligence elements produce the desired reports and/or data cubes. Once integration testing has been completed, the final product is rolled out and the system comes online in a production capacity. At this point, the system architecture document should be completed and handed off to the operations group who will be responsible for the monitoring of the new system.

1.1.10. Train Business Analysts and End Users

The investment in people, technology, and data will all be wasted if the information is not delivered to the people who need it to drive, manage, and measure the organization. Training is a key role that will ensure the highest return on the data warehousing investment is realized by delivering data to the people who asked for the information in the first place. As there are many levels of users, providing different levels of training is a requirement to company-wide acceptance of the new warehouse environment. There are three objectives to training:

1. Make users comfortable browsing and accessing the new "canned" standard reports and data sources
2. Demonstrate to the users that the new reports provide the same information integrity as previously available (status quo)
3. Prove that there is a dramatic improvement in quantity and quality of data and that the new toolset provides better delivery, exploration, and presentation options.

1.1.11. Expand and Enhance the Data Warehouse – Review for Consistency

At the end of the major project implementations and before the first round of enhancements, the DW/BI team should review the resulting architecture and evaluate the consistency, integrity, and performance of the overlapping subject areas. Even though the team will have had several chances during each successive project to review, modify, and enhance the existing architecture there will always be inconsistencies. Taking the time before any major enhancements to do this review will significantly improve overall performance, reduce on-going maintenance costs, and allow for shorter, cheaper enhancement projects.

Enhancements should be completed using a version/release discipline where a number of small enhancements or additions are grouped together and integrated as a project. The Business (not IM/IT) should drive the release schedule. Each enhancement project should be scoped for approximately four months of work, allowing for an additional full month of integration and acceptance testing. By the time the Data Warehouse is mature enough to accept enhancements, the business will have come to rely heavily on its information. Improper regression testing will undermine user confidence and jeopardize user support of the Data Warehouse if data quality and timeliness are affected.

1.1.12. Maintain and Administer Data Warehouse

On-going monitoring and administration is required on any Data Warehouse. It is the responsibility of the DW/BI team to review the type of activities, report requests, analysis, and usage patterns. Monitoring the business issues facing end users will allow for predictive planning and allow IM/IT to offer proactive services instead of only reacting to user requests.

Maintenance also includes the day-to-day activities of setting up user accounts, security authorization, scheduling, operational support, help desk, and troubleshooting completed by the entire IM/IT organization.

1.2. Roles

This section couples roles identified for various H&H Technologies clients with diagrams and additional roles and responsibilities from industry sources such as Gartner Group, DM Review, and Data Warehouse Institute.

The organization and make up of the Data Warehouse/Business Intelligence team can be configured in numerous ways depending the organizational processes, management structure, and distribution of resources. Whether the groups number eight, nine, ten, or more also depends upon the size of the organization and the warehouse initiative. Though the staffing levels will change depending on the status of the Data Warehouse, the responsibilities and roles required remain constant.

Business and Information Management/ Information Technology (Figure 1-5) is the first major breakdown in the roles and responsibilities. This is not intended to pit the groups against each other, as their relationships must be symbiotic in order to have a successful data warehouse, but the groups serve different purposes.

The business team is responsible for ensuring the accuracy of the requirements, testing, data validation, business process reengineering, and communicating with the IM/IT staff to guarantee their requirements are met. The DW/BI group, a new department within IM/IT, is charged with taking the business requirements and providing cost effective, efficient, and scaleable solutions, that will be easy to maintain and can be supported to the level

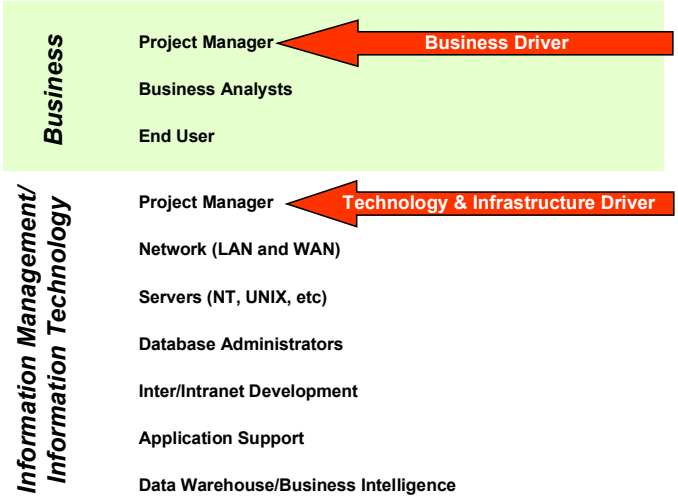


Figure 1-5: Major Roles in Data Warehousing

demanded from the business.

Data Warehousing is a collaborative effort and requires dedication and commitment from a key group of specialists, often for the entire length of the project – from the initial requirements gathering all the way through implementation and support. The allocation of these resources is directly related to the success of the data warehouse effort.

One of the critical lessons most organizations learn during their pilot projects is the need for dedicated resources. The pilot projects often make do with 'time as convenient' from many of their key resources including the IM/IT project manager, the Database Administrators, and the Business Intelligence experts. The basic design, data mapping, and loading mechanisms were completed but the pilot project experienced difficulties and delays during the extraction and loading processes due to data quality. End user reporting, specification, and delivery was pushed back and the project, as a whole, missed the end-target date. Having dedicated resources will minimize the risk of delays.

1.2.1. Business Project Management

A critical role, and the area that initiates a warehouse project, is the Business, who are the "customers" of the data warehousing project: Information Management/Information Technology may be considered the "vendors". The business' Project Manager has the role of 'Business Driver'. This individual defines the purpose and target user group, assigns business analysts to gather requirements, monitors funds, assesses project status, and works with the IM/IT Project Manager to smoothly develop the Data Warehouse. Many organizations use a steering committee to designate the order in which business units are implemented. Using the steering committee to provide guidance will fairly allocate and schedule the implementations across the organization. The steering committee should select the most productive projects early in the Data Warehouse lifecycle.

It is imperative to have a single project manager responsible for the subject area implementation as it is necessary to ensure the specific, documented requirements and data needs are met. The project manager can report back to a steering committee but the organization should not fall into the trap of having the steering committee become a substitute for a project manager, as it will lead to confusion, debates, and delays. The steering committee, as a cross section of the organization, will not be a position to effectively decide on the day-to-day issues of the project. A single project manager, familiar with the business area being implemented, can quickly make the necessary decisions that will remove roadblocks, assign resources, and manage the scope of requirements in a manner that is more likely deliver successful results.

1.2.2. Business Analysts

This group is responsible for identifying and defining the requirements as gathered from the end users. Good requirements will ensure that the warehouse will meet the strategic objectives of the business. Data warehousing experts agree on one thing: "if you don't know what and who you're building for, the project is doomed."

The Business Analyst(s) will also be responsible for programming some of the canned reports. The Business Intelligence Specialists will often do the majority of project-based reporting requirements but having the Business Analyst assume some of the responsibility is an excellent learning situation as they will have an opportunity to learn some of the tips and tricks of report writing that are not available in training classes. This approach will also put the Business Analysts in a better position to train the end users as they will have intimate knowledge of the reporting packages.

Training and interacting with their end users is essential for the Business Analysts. Their small, focused user community will look to them for simple help desk functions, reporting questions, or for on the job training needs. After each major implementation the Business Analyst will work with one or more of the Business Intelligence Specialists to provide complete end user training. This training will ensure the end users are capable of utilizing the tools and understanding how the data is organized.

1.2.3. End Users

The end users are the target audience of the Data Warehouse and consequently they need to be involved early in the cycle – especially for requirements gathering and initiation of the project. The Business Analysts will be responsible for polling appropriate end-users for report layouts, information requirements, and business processes that will be required for complete and accurate requirement definitions.

The end users are also essential during the implementation phase to complete acceptance testing (validate data and review output formats) and ensure their needs are being met. Training is also required as part of this role. It is imperative that the end users are trained on the reporting tools and troubleshooting procedures. Additionally, the users will need to know where to find their canned reports and how to access them. Involving the end users from the beginning will reduce on-going support, maintenance, and help desk activities significantly and will give the Data Warehouse a sense of community spirit as everyone will have been involved at some level.

The investment in people, technology, and data can all be wasted if information is not delivered to the people who need it to drive, manage, and measure the organization. Training will insure that the highest return on the data warehousing investment is delivered to the people who asked for the information in the first place – the end users.

1.2.4. IM/IT Project Management

Managing the technical aspects of a data warehouse implementation is an awesome responsibility due to the breadth and depth of interactions that must be monitored and acted upon. Sponsorship at the executive level, and a willingness to make data warehousing a priority within the organization, are necessary components for successful projects. The IM/IT Project Manager, in conjunction with the Business Project Manager (i.e. business driver) must be able to demonstrate returns on investment in short three or four month time frames in order to keep ongoing sponsorship. Project management of the warehouse team is also needed to ensure that the warehouse is aligned with the business needs and to develop project plans which technically support the warehouse as it matures and expands.

1.2.5. Information Systems Services (Networks, Servers, Inter/Intranet Development, Application Support)

Resources from network services, operations, and enterprise application systems [e.g. ERPs such as Oracle Financials, SAP, CODA] must go beyond defining the initial technical architecture and play an active role in the expansion, maintenance, and support of the Data Warehouse. The issues in this realm are large and complex as multiple vendors are brought together in every facet of information technology to build and support the Data Warehouse. Individuals from each of the areas are not required as full-time resources during the implementations; however, the Data Warehousing effort should be a priority for these groups. Fast turn-around on questions and designs to facilitate the collection, integration, and delivery of the information will be necessary to ensure projects are completed on time and within the organization's documented policies.

1.2.6. Database Administrators

As the size of databases has quickly moved from hundreds of giga-bytes to the multiple-terabyte range, the stress of the technology is often transferred to the database administration (DBA) staff. The DBA's skills and techniques are being tested along with the tools. Techniques in distribution, security, segmentation, array processing and other techniques that improve performance at all levels are required to meet these increasing demands. The DBAs are the primary IM/IT team to assist the DW/BI staff in the Data Warehouse implementations. One or more DBAs will be required on a full-time basis for eighty percent of the project life-cycle. The DBAs are critical team members during the definition of the technical architecture, data modeling and design, and for database implementation and loading. During the report building phases of the project the DBAs will be concerned with security, performance, and integration activities.

1.2.7. Data Warehouse/Business Intelligence (DW/BI) Team

The Data Warehouse/Business Intelligence Staff is composed of:

- **Data Architect(s)**
- **Business Intelligence Specialist(s)**
- **ETL & Data Validation Specialist(s)**
- Technical Writer
- Tester

The core DW/BI team positions (in bold) are absolutely necessary for any organization with a Data Warehouse. The Technical Writer and Tester are classified as auxiliary or optional resources that can be brought in during projects to supplement the team.

1.2.7.1. Data Architects

Data collection, transformation, distribution and loading must be logically defined by the data architect(s). This individual or team must also possess the skills and experience to define the data models that form the basis of the Data Warehouse. The initial data model, which supports the business requirements, continues to evolve as legacy application data is mapped to the physical model during the pilot phase of the warehouse, and continually changes with modifications and mapping that occur during the warehouse life-cycle.

In addition to the business requirements, detailed data requirements should be fleshed out before, or as, the technical details are defined. Data collection rates, volumes, and timing from every point to point in the warehouse will dictate the network capacity, transfer rates, storage volumes and growth rate of the warehouse. The data architecture and model will dictate each of these factors, and is the responsibility of the Data Architect.

1.2.7.2. Business Intelligence Specialists

There are often several Business Intelligence specialists on a DW/BI team. This role interacts heavily with the Business Analysts as end user requirements and specifications are transformed into reports and datasets (cubes or marts). The BI Specialists may have specific focuses (query, OLAP, data mining and forecasting, etc.), but should be at least familiar with all other BI tools being implemented within the organization. They are responsible for the coding of the BI data dictionaries, OLAP cubes, designing the deployment structures, implementing the BI web solution, and training.

Working with the Data Architect(s) and the ETL Specialist(s) to model and map the data is a required element for this role. It is important to involve the BI team in the up-front work, as the design and often the efficiency (or ease of creation) of the data cubes and dictionaries depend heavily upon the database schema.

The BI Specialists will also mentor the Business Analysts during canned report creation. Support for report optimization, best coding practices, tips and tricks, and advanced help desk are functions that need to be performed. The more effort put forth in the training of the Business Analysts, the better the long-term relationship, and the more self-reliant the Business Analysts will become.

The last major role for the Business Intelligence staff is to perform business, context-sensitive, tool-specific training to the end user communities. This training is done in conjunction with the roll-out of the Data Warehouse and also involves the Business Analysts. Training the end users immediately following the roll-out will increase the success of the Data Warehouse, as it will assist end users in becoming comfortable with the new system. Periodic training courses or seminars should be held as refresher courses or for new employees to help maintain the positive momentum.

1.2.7.3. ETL and Data Validation Specialists

The individuals within this role are adept at using the ETL and Data Validation tools. They will also work with the Data Architect and Database Administrators to model the target database, map data, and are ultimately responsible for building, scheduling, and monitoring the code required to extract, transform, aggregate, cleanse, and then load the data. This role will also provide the code for the extraction of DW data into datamarts (cubes are the realm of the BI specialists).

Maintenance and administration of the Enterprise Meta Data Repository (EMDR) normally falls within this group. They, in conjunction with the Database Administrators, protect and enforce the integrity of the 'data about data', and ensure that all ETL, BI, database, and modeling components are complete and fully documented with the EMDR.

1.2.7.4. Technical Writer

Although Technical Writers are not required during maintenance or operational activities, during project implementations or enhancement releases (i.e. during growth mode), a Technical Writer is an invaluable asset to the team. Data Warehouse projects are inherently short bursts of furious activity by highly skilled and technical individuals. Each of these team members have a different writing style and varying degrees of word processing proficiency. The use of a technical writer ensures consistent, well-constructed documents that would otherwise be left to the end of the project (if they were in fact ever written). Technical Writers take roughly formatted developers notes and create concise, user-friendly text and free up the more expensive DW/BI developers from having to deal with the tedious, and often times frustrating, "wordsmithing" and formatting issues. The developers are still responsible for checking the content and accuracy of the polished documents, just as the Technical Writer is responsible for ensuring all sections of the documents are completed by the appropriate team member.

The Technical Writer would be responsible for the collection of all system architecture documentation and for the production of an end user guide (written with assistance from the Business Analysts and the Business Intelligence Staff).

1.2.7.5. Tester

The Tester role is required during the last stages of Data Mapping and Implementation phases of the project or enhancement release. Their role is to assist the Business Intelligence and ETL specialists perform unit and integration testing. Short, aggressive schedules tend to sacrifice good testing principles: having a dedicated tester is an excellent way to combat this tendency. Data quality is critical because the users must trust what they receive from the Data Warehouse, otherwise they will not use the data and the entire Data Warehouse investment will be in jeopardy. Testing the code and finding glitches or inconsistencies while still in development will reduce the overall cost of the implementations and will ultimately mean less work is required during each subsequent release.

1.3. Roles and Responsibilities Summary

The initial project in the iterative project cycle will need an extended window as it will take longer to create the initial database design and ETL/data cleansing rules as many of the dimension tables, relationships, and attributes will be established for the first time. However, it is still important that this first project be kept short; no more than four or five months from the first requirements gathering to implementation and training. Each successive project should be scoped to be completed within four months. As more subject areas are added to the Data Warehouse it is expected that the amount of overlap will increase and fewer dimensions will need to be defined.

Each completed project should be used as a learning experience; the procedures, tools, and methodologies adapted for the next iteration.

In short, there a few "Do" and "Don't" recommendations that should be observed when implementing a Data Warehouse:

DO

- Identify Business Drivers and Technology Drivers as independent resources
- Use Data Warehouse to solve only strategic business problems that can't be fixed at the source system level
- Ensure the Data Warehouse is business driven, not technology driven
- Build the Data Warehouse using a Hybrid approach (identify 80% of the common attributes and dimensions first, but implement by subject/business area).

- Ensure the Data Warehouse is scaleable and meets performance expectations

DON'T

- Load "dirty" data into the warehouse as it will severely and adversely affect return on investment
- Build stove-pipe datamarts in isolation.
- Develop code where Commercial Off-The-Shelf (COTS) products exists: ensure all products integrate with a central meta data repository

Following these steps will show an immediate reduction in the time it takes for the Business Analysts to gather requirements and produce reports. Having a clean, consolidated Data Warehouse or datamart in an accessible format for the Business Analyst will see a significant reduction in the work turn around time, information delivery time, and the accuracy of decision making.

1.4. Growth Mode Staffing

Using the iterative (hybrid) approach means the Data Warehouse implementations are short but extremely intense periods of effort completed in quick succession. For each subject area implementation, approximately 11 to 15 full time resources will be required. Figure 1-6, following, shows the levels of commitment for each role during the implementation cycle. There are minimum of seven full time resource requirements for each project cycle:

1. Business Project Manager (Business Driver)
2. Business Analyst
3. Database Administrator

Core DW/BI Team Roles

4. Technical Project Manager (Infrastructure and Architecture Driver)
5. Data Architect
6. Business Intelligence Specialist
7. ETL/Data Validation Specialist

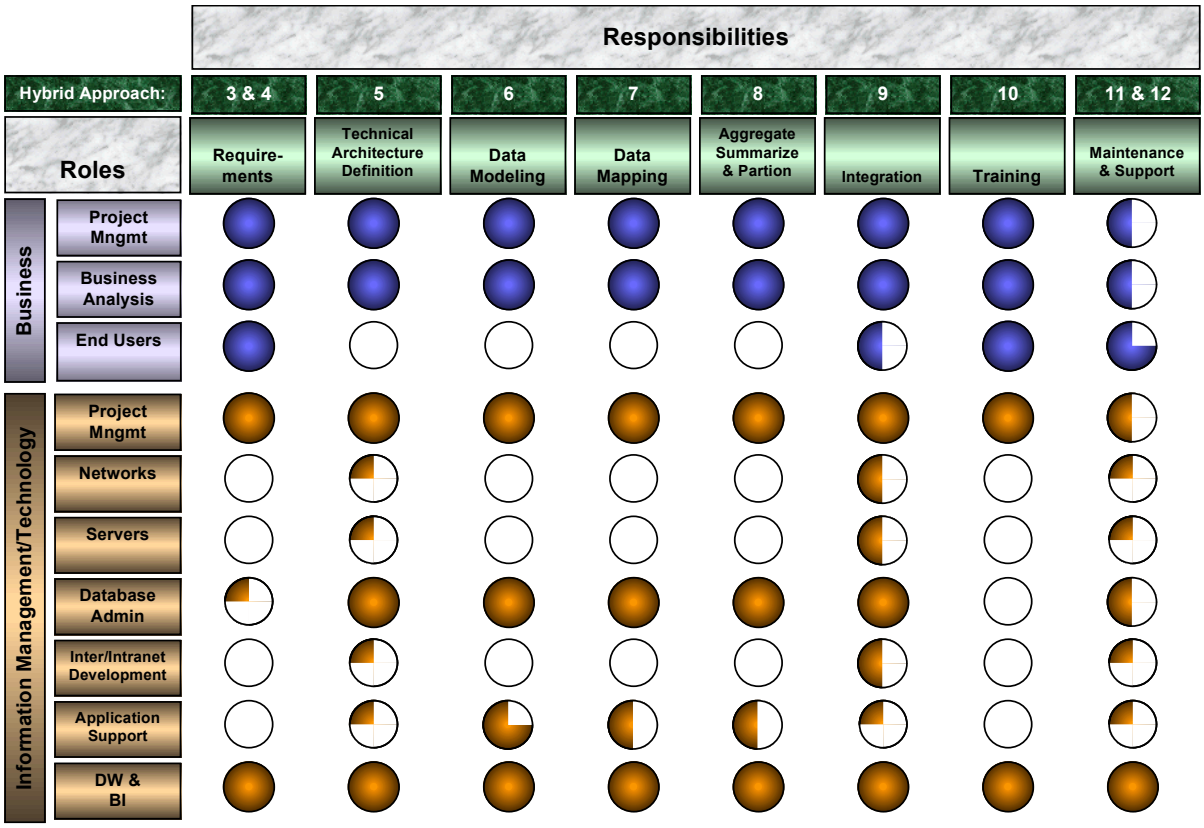


Figure 1-6 Data Warehousing Roles and Responsibilities

It is imperative that a core DW/BI team be established with a set of consistent, dedicated members. At least one resource in each of these key roles should be present during all Data Warehousing projects. By being involved in multiple projects, this "swat team" will bring a sense of continuity and purpose to the corporate Data Warehouse. The optimum "swat team" in many organizations are composed of employees or long-term contract personnel as the DW will require a consistent implementation approach that can only be achieved through longevity and organization-specific implementation experience. Depending on the business area being implemented, the core team will need to be augmented by additional resources. These resources are typically identified during the requirements phase (Step 5: Define Functional Requirements) and can be contract or permanent resources.

The Business Project Manager, Business Analysts, and Database Administrator roles are required full time during any project, but are often project specific resources. The organization will want to assign these resources based on their detailed knowledge of the source systems being consolidated.

Many projects will need 2 to 4 BI specialists depending on the reporting and data cube requirements, and up to 3 ETL specialists to create the data extraction and transformation rules. The essential rule to remember is: *delivery within four months*. The projects should be small enough to ensure this tight schedule can be met, but it is also important to anticipate problems early and staff accordingly. During the larger, more aggressive projects, a technical writer maybe added to the staff, in order to reduce the amount of word processing and documentation for which the programmers are responsible. It is absolutely necessary to have concise and legible system documentation in place by the end of the project in order to provide effective maintenance and operational support.

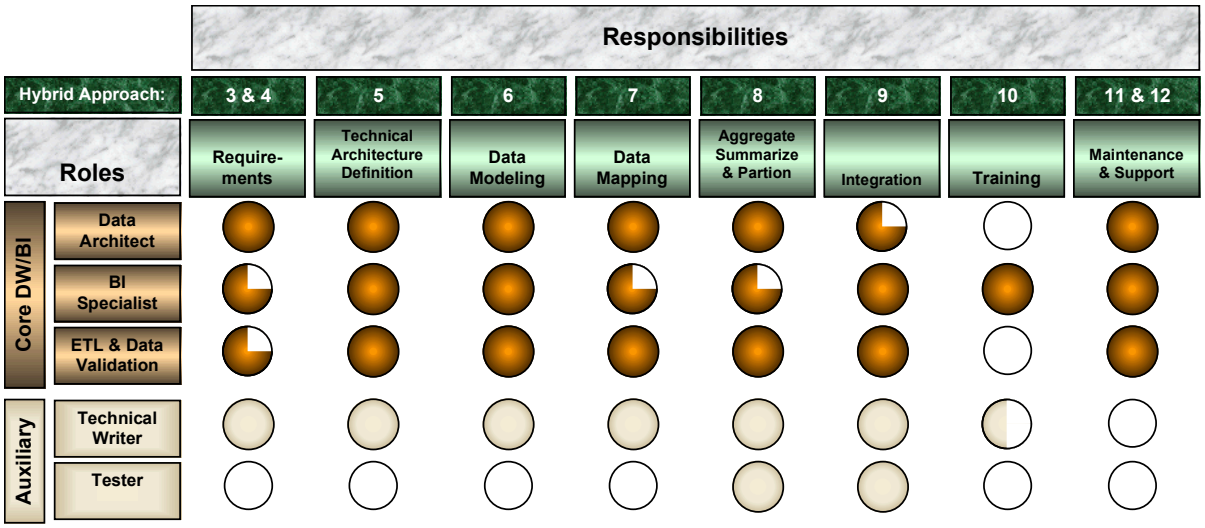


Figure 1-7: DW/BI Team's Roles and Responsibilities

Therefore a typical team implementing a Data Warehouse project would require:

Quantity	Team Member Title
1	Business Project Manager (Business Driver)
2-3	Business Analysts
1	Technical Project Manager (Infrastructure and Architecture Driver)
1	Database Administrator
1	Data Architect
2-4	Business Intelligence Specialist
2-3	ETL/Data Validation Specialist
1	Technical Writer
11-15	**Full Time Resources**

There are also part-time resources required during the completion of a Data Warehouse phase. It is imperative to the success of these short implementation bursts that resources from the other IM/IT departments and key end users (needed for requirement gathering, testing, and acceptance) be made available at appropriate times. Having these resource available to answer questions, find solutions, or provide assistance in a timely manner will ensure that milestones are met and the four month window is not jeopardized. Obtaining the support of the various IM/IT departments is the job of the IM/IT Project Manager.

1.5. Maintenance Mode Staffing

During periods where the Data Warehouse is not expanding, operational support is completed by a number of resources distributed throughout the business areas and Information Management/Information Technology.

The Business Analysts, as per the three tier hierarchy are responsible for assisting their localized group of end users. The analysts will support report execution, level 1 help desk activities, new report creation, and specification development for new data requirements. Business Analysts should not be spending more than one half (50%) of their time supporting these types of activities as they must also be able to perform their basic business analysis function. Activities that contribute to the workload are the number of users the analysts are supporting, the number and complexity of report requests fielded, and the quality of data within the data warehouse. Analysts spending too much time programming should be assisted and a second Business Analyst role be created to produce an even more localized set of end users.

Operationally, IM/IT will have to support the Data Warehouse. This involves the everyday activities of monitoring and support of the NT and Unix Servers, the network, databases, web front end, etc. Each of the IM/IT departments can expect to have one-quarter to one-half of a resource dedicated to supporting the corporate Data Warehousing initiative.

The Database group will be required to supply the equivalent of one or two resources full time on the performance, security, maintenance, and support of the various databases involved in the initiative (corporate DW plus all of the distributed datamarts). Most organizations will divide the responsibility between several DBAs. The caveat to this system is ensuring the DW/BI staff have documented escalation procedures that clearly delineates which DBA is responsible for each system.

The Data Warehouse/Business Intelligence core staff are composed of one data architect, one ETL specialist, and at least one BI specialist. These resources will be dedicated full time to the maintenance and support of the Data Warehouse. During non-growth periods, optimizations, reviews, and enhancements will be the work orders. This team will also supply level 2 support for both the business and the IM/IT departments. When the Business Analyst has problems with a report (e.g. long running query), experiences a technical glitch, or needs support, the DW/BI staff with work to solve the issue. The DW/BI staff will also be on call for the support of ETL and BI job streams problems. Should a job fail to execute and Operations can't recover as per the backup and recovery plans, DW/BI members will assist in correcting the issue. The DW/BI staff should also monitor the query usage and patterns, security requirements, and data needs in order to become a *proactive* service provider instead of only *reacting* to customer demands and requests. Continual performance enhancements, addition of new, summarized, or partitioned data, and additional canned reports will only improve the quality of information delivery and increase the effectiveness of the organization's Data Warehouse.

1.6. Conclusion

Creating a Data Warehouse is a organizational effort. Interactions between the Business units and IM/IT must be established as openly as possible. A new IM/IT team, the Data Warehouse/Business Intelligence group, must be established to act as trusted advisors to the business and integrators of information technology. This small, permanent group of 3 or 4 individuals does not displace any of the current IM/IT or business functions – they simply facilitate the building and operation of a comprehensive and accurate information store and delivery mechanism.

The role of the central DW/BI team is to augment, facilitate, integrate, and maintain the momentum of the data warehousing project – not to displace IM/IT or Business Analysts.

During major projects, management at must dedicate a minimum team of 11-15 people to each iterative Data Warehouse and mandate four month delivery schedules. Being able to deliver reports every four months increases return on investment and will encourage end users to utilize the system as it proves its capabilities of delivering on the information promise.