

Typical IT R&R for the Data Warehouse

Sung-Soo Kim
Data Management Research Section
Electronics and Telecommunications Research Institute
128 Gajeong-ro, Yuseong-gu
Daejeon, South Korea
sungsoo@etri.re.kr

ABSTRACT

Data warehousing is not new. Most large organizations have been investing in data warehousing for years. Currently, cost-effective technology is creating more possibilities for small and medium-size companies to build and deploy data warehouse solutions too.

There are many different roles on a data warehouse project that are filled by systems professionals. Many roles are consistent regardless of what kind of system is being developed. In this technical report, we describe the IT roles and responsibilities (R&R) for a data warehouse project.

1. INTRODUCTION

There are many stories about wild successes, and just as many about failed projects. With so much buzz about data warehousing, it is often assumed that everyone already knows the basics. However, many people are being exposed to these concepts for the first time. To ensure a common understanding, it is worth taking the time to boil things down to the essence of data warehousing.

A *data warehouse* (DW) is the collection of processes and data whose overarching purpose is to support the business with its analysis and decision-making. In other words, it is not one thing per se, but a collection of many different parts. Before looking more closely at the specific parts of a data warehouse environment, it is helpful to compare the characteristics and purpose of a data warehouse with an operational application system.

Differences Between Operational and DW Systems Applications that run the business are called *online transaction processing systems* (OLTPs). OLTP systems are geared toward functions such as processing incoming orders, getting products shipped out, and transferring funds as requested. These applications must ensure that transactions are handled accurately and efficiently. No one wants to wait minutes to get cash from an automated teller machine, or to enter sales orders into a company's system. In contrast, the purpose and characteristics of a data warehousing environment are to provide data in a format easily understood by the business community in order to support decision-making processes. The data warehouse supports looking at the business data over time to

identify significant trends in buying behavior, customer retention, or changes in employee productivity.

Figure 1 shows the correct order to successfully design and implement a data warehousing environment. Both the technical and business team members play a role throughout.

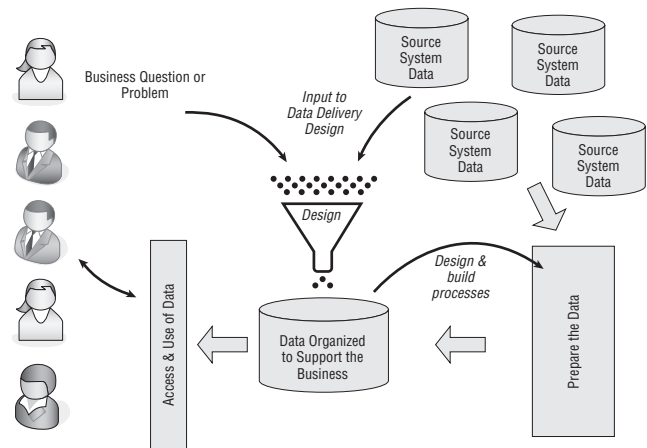


Figure 1: Optimal data warehouse design and development sequence

The input of the business community is invaluable to a project team. Likewise, a data warehouse cannot be built without the talent, technical expertise, and rigor that the systems staff offers. The data warehousing industry has developed technologies and methods to design and implement robust environments that support businesses today and into the future. You must trust the technical team members to construct components to fit together to provide you with an environment that will support your business requirements. Let the systems staff do what they do best: make technical design decisions and build the capabilities. You really should not care which hash algorithm is used to load large tables into a database as long as you can get the data you need in a timely manner.

2. CIO/IT EXECUTIVE SPONSOR

The data warehouse should have one executive sponsor from the IT organization that will partner closely with the executive business sponsor. This is typically the CIO, but in large organizations this could be a senior manager in the IT organization. This person should be ultimately responsible for the technology side of the data warehouse and should have management responsibility over the IT staff on the data warehouse. This person usually shares the funding responsibility with the executive business sponsor and will

own control for the technical aspects of the data warehouse environment. It is very important that the executive sponsors from the two organizations are seen as equals and that they are on the same page in terms of goals and objectives for the data warehouse [2].

3. DATA WAREHOUSE MANAGER

The data warehouse managers role represents the driving force behind the data warehouse environment from the IT organization. This individual is often given direct responsibility for the funds for the data warehouse project. This is the primary visionary for data warehousing in the organization, and the lead systems professional who is responsible for building and sustaining the data warehouse. The data warehouse manager continues to have responsibility for the data warehouse after a project is complete and oversees support, maintenance, and growth.

Data warehousing focuses on the management of five primary data flows, namely the *inflow*, *upflow*, *downflow*, *outflow*, and *metaflow* [1]. The data flows within a data warehouse are shown in Figure 2. The processes associated with each data flow include:

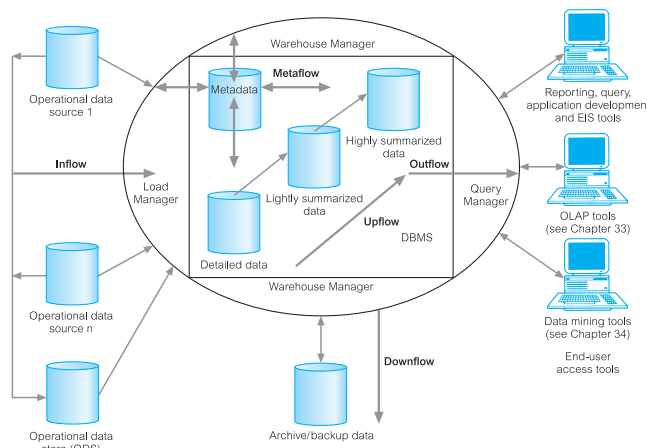


Figure 2: Information flows of a data warehouse

- **Inflow** Extraction, cleansing, and loading of the source data.
- **Upflow** Adding value to the data in the warehouse through summarizing, packaging, and distribution of the data.
- **Downflow** Archiving and backing-up the data in the warehouse.
- **Outflow** Making the data available to end-users.
- **Metaflow** Managing the metadata.

The data warehouse manager must set the overall direction of the data warehouse. This is done in conjunction with key business management. This includes making sure that a sound technical and data architecture is developed. The data warehouse manager is the focal point for blending industry best practices with internal best practices to develop an overall methodology and techniques that will ensure the long-term success of data warehousing within the organization. This includes adapting *system development life cycle* (SDLC) practices and the definition of standard deliverables to meet the unique data warehousing needs.

If there are many data warehousing projects underway, the data warehouse manager would then be involved in coordinating and leveraging resources across the projects. These resources could include *people* as well as *technology*.

DESIRED CHARACTERISTICS OF THE DATA WAREHOUSE MANAGER

- Sound analytical and reasoning skills
- Strong presentation and communication skills
- Able to educate others
 - IT: About the business and uses of the data
 - Business: About datawarehousing, how the data looks, and what applications/reports are needed
- Understands the business
- Possesses project management experience
- Has exposure to the technologies involved
- Understands the organizations culture and how things get done
- Exhibits leadership
- Understands and can manage risk
- Able to manage and coordinate multiple projects and priorities

4. BUSINESS SYSTEMS ANALYST

The business systems analyst role is similar to the business analyst, but this person needs to have sound knowledge of data warehousing design principles. The business systems analyst will participate in, or even drive, the business requirements gathering process and will be a key player in building the partnership with the business community. The business systems analyst will also be expected to do the following: Participate in the development of the dimensional model Help research data-related issues Be a liaison between systems and the business communities for daily project tasks Develop and oversee testing plans Develop project documentation and deliverables The business systems analyst role is often blended with the business analyst role. This is effective only if a single person can understand the business clearly enough and delve deeply enough into the details of data warehouse design and development.

DESIRED CHARACTERISTICS OF THE BUSINESS SYSTEMS ANALYST

- Excellent problem-solving skills
- Understanding of business processes and functions
- Strong technical background
- Able to facilitate and communicate technical issues to a variety of audiences
- Understands the system development life cycle
- Good writing skills
- Quick learner
- Persistence to work through complex issues and design challenges
- Enthusiastic, motivated, and curious about learning the business

NOTE The most important characteristic to look for in candidates for the position of business systems analyst for a data warehouse project is curiosity. If a person is sincerely interested in what the business does and how things work, then this will feed their ability to gather requirements and then translate them into a solution.

5. SOURCE SYSTEM ANALYST

The source system analyst is not a full-time member of the data warehouse project team but plays a critical role. The data warehouse team must have one source system analyst designated for each source system that will feed the data warehouse. This person needs to have experience and knowledge of the underlying systems that will be used to populate the data warehouse. The source system analyst will also be expected to do the following:

- Support the development of the dimensional model
- Provide file layouts, data models, and data element definitions that exist for the source systems
- Research data-related issues
- Serve as the subject matter expert for the source system

When there is limited documentation for source systems, the source system analyst is often the only person who knows what each data element actually means. The names and how the source systems use each field may be a mystery without the active participation of this highly valuable resource.

DESIRED CHARACTERISTICS OF THE SOURCE SYSTEM ANALYST

- Strong problem-solving skills
- Persistence to work through complex issues and design challenges
- Patience to support detailed, field by field analysis
- Willing to put forth extra effort to help the data warehouse initiative
- Understands the critical nature of his or her knowledge
- Able to juggle multiple priorities, including communicating to management concerns about the scheduling of their own time
- Shows interest in learning about data warehousing

6. DATA MODELER/DATA ARCHITECT

The data architects responsibilities are to develop a *big picture strategy* for how data will be handled, stored, and processed to support the requirements of a specific project and to accommodate the requirements across the enterprise. This is the heart of the data warehouse. The data architect sets this vision. This includes driving or participating in the following:

- Defining the data structures and philosophy for the staging area
- Defining the data structures and philosophy for the data warehouse
- Participating in the requirements gathering process
- Understanding specific project requirements

- Gathering or researching enterprise data requirements and standards
- Ensuring that the data warehouse architecture fits in the overall enterprise data vision

Data modeling takes things to a lower level of detail. This entails understanding individual data elements and how they relate to each other; and then applying what was learned during the requirements gathering process in order to develop data models to support the business. There may be a need for detailed data modeling to support the staging and preparation processes. There is also a need for a dimensional model to reflect the specific data needed to support the business objectives of the project. The model must also be flexible enough to be extended as new data and requirements are discovered over time. The basics of a well-designed dimensional model should be able to sustain a data warehouse for years. The data modeler will be responsible for the following:

- Designing data models and databases to meet current and future needs Understanding how the data model will support the business reporting and analyses Researching data content and availability to ensure that the model is grounded in reality Researching and recommending approaches to handle data integration Performing detailed data analysis to understand the data as it currently exists Developing logical database designs

NOTE Dimensional modeling for a traditional data modeler can prove to be very challenging. If a data modeler or architect is fighting standard data warehousing practices, take action. Education may not be enough to change their ways. If this continues, consider reassigning that individual to a traditional project and get someone who will effectively support the data warehouse. That way, you can still leverage their skills and experience without causing delays and poor design choices for the data warehouse.

DESIRED CHARACTERISTICS OF THE DATA MODELER/DATA ARCHITECT

- Understands the business requirements and the business point of view
- Understands the impact of data design on the rest of the project
- Understands data warehouse design best practices and approaches
- Understands how BI tools work and present information to users
- Understands the benefits and purpose of dimensional modeling and is an expert modeler
- Curious and motivated
- Able to communicate effectively with both business and technical groups

7. ETL DEVELOPER(S)

One of the most complex, challenging, and time-consuming parts of a data warehouse project is preparing the data for access and analysis. The main steps are to extract, transform, and load (ETL) the data. There are usually multiple systems professionals who perform the design and development of the ETL processes to populate

the data warehouse. The lead ETL developer may sit in on requirements gathering sessions. It is also beneficial for the lead ETL developer to participate in the creation of the dimensional data model. This helps to ensure a full understanding of the complete model.

ETL developers work with the rest of the project team to define rules about how to clean, validate, and integrate the data. This often involves working directly with the business community. In some organizations, this must be done by working through the business system analyst. The ETL developers spend a lot of time researching and resolving issues discovered in the data as it is prepared for the data warehousing environment.

DESIRED CHARACTERISTICS OF THE ETL DEVELOPER

- Ability to adapt when changes occur
- Understand the full data warehouse life cycle
- Appreciation and desire to work through data quality issues
- Experienced system developer
- Ability and interest to learn new techniques and approaches
- Creative problem solver

8. BUSINESS INTELLIGENCE APPLICATION DEVELOPER

The business intelligence application developer is responsible for creating and maintaining queries, reports, and applications. This usually involves using business intelligence (BI) tools. These tools need to be configured or set up properly to access the data mart. Each BI tool has a semantic layer, or metadata, that it uses to traverse the data and manage defined metrics. Once the BI tool is configured, a wide range of activities need to be performed, including the following:

- Designing and developing a suite of report templates
- Designing and developing a deployment strategy for reports
- Designing and developing performance dashboards or scorecards
- Designing and developing BI report and application documentation and training materials
- Supporting business users to help them increase their knowledge and use of the BI tool

NOTE There may be two different types of BI developers. The first has deep technical skills and is responsible for installing and configuring the BI technology. They also design and deploy complex reports and analyses. The second type of BI developer is less technical and may be filled from the business community. This second role is responsible for creating reports and delivering analyses to support the business community. This second type often directly supports requests from senior management.

DESIRED CHARACTERISTICS OF THE BI APPLICATION DEVELOPER

- Strong understanding of the business requirements
- Interested in helping the business leverage the data warehouse

- Flexible and able to adapt to ongoing changing requests for information
- Works well with the business community
- Good communication skills
- Able to translate business requests into BI report specifications
- Strong analytic abilities

9. OTHER SUPPORTING ROLES

A number of other technical roles are required for successful completion of a data warehousing project. These include a variety of technical functions, including database administration, security, network administration, systems architecture, training, testing, and operations support. These people are involved in making sure that the foundation technologies are set up and working properly and that the project is deployed successfully. The roles for the business and IT communities have been described here individually, but the real strength of a data warehouse emerges when these people all work together. Ideas for helping to build and maintain this relationship are covered next.

10. REFERENCES

- [1] T. M. CONNOLLY and C. E. BEGG. *Database Systems, A Practical Approach to Design, Implementation, and Management, Fourth Edition*. Pearson Education Limited, Edinburgh Gate, 2005.
- [2] L. L. Reeves. *A Managers Guide to Data Warehousing*. Wiley Publishing, Inc., Crosspoint Boulevard, 2009.