

A McKnight Associates, Inc. White Paper:

## **Effective Data Warehouse Organizational Roles and Responsibilities**



Numerous roles and responsibilities will need to be acceded to in order to make data warehouse efforts successful and generate return on investment. For the technical personnel (application programmer, system administrator, database administrator, data administrator), it is recommended that the following roles be performed full-time by dedicated personnel as much as possible and that each responsible person receive specific Data Warehouse training. The data warehouse team needs to lead the organization into assuming their roles and thereby bringing about a partnership with the business. Management also needs to make actionable plans out of these directives and make sure the staff executes on them.

The following are team and extended team member composition and roles and responsibilities suggestions.

### **Manager/Director**

The responsibilities of the Manager or Director of the Data Warehouse team should be to:

- Ensure support for the data warehouse program at the highest levels of the organization
- Understand high level requirements of the business
- Present the business with the possibilities available to them through data warehousing
- Staff the team
- Establish and ensure adherence to a set of guiding principles for data warehousing
- Communication of key milestone status to IT management
- Ensuring the remainder of the team accede to their responsibilities as enumerated below
- Liaise with strategic vendors
- Establishing partnerships with key IT partners in support of data warehousing initiatives

### **Project Manager**

The perceived strength of data warehousing within an organization will be the sum of the strength of the Project Managers. Project Managers must deliver commitments and must deliver on time. They will do this by culling resources from within the data warehouse team and from consultancy as necessary and establishing partnerships with other internal support organizations required to support a data warehouse iteration. A Project Manager delivers by:

- Maintaining a highly detailed plan and obsessively caring about the progress on it.
- Applying personal skill and judgment to everything on the project. This is a real value-add of the Project Manager. It is the Project Manager's job to exercise relevant discretion.
- Matching team member's skills and aspirations as closely as possible to tasks on the plan.
- Tracking all relevant metrics for each iteration:
  - Project Plan milestones
  - Issues list
  - Adherence to change control practices
  - Adherence to source code control practices
  - Documentation fit for users and support personnel
  - Architectural components adherence to fit for purpose and standards
  - Regression testing performed and tests updated based on changes
  - Team members fit for tasks and career-enhanced

## **Chief Architect**

The Manager/Director will need to rely on a Chief Architect (or similar title) position, as one of his/her direct reports, to work on complex issues of architecture, modeling, and tools.

The person should be able to quickly qualify as an authority in data warehousing within the organization and have mastery of the data warehousing paradigm, both current and emerging technologies. His/her knowledge of the business needs to be just as great. A business technologist, s/he will meet business objectives with existing or emerging technologies and work on issues with broad technical or strategic implications. This person would have significant interface with the internal clients and increase their confidence in the data warehouse organization.

## **Data Steward** (in user community - by subject area)

Data Stewardship appointment should be made at the subject area to management-level personnel in the business areas most impacted by the subject area. For the core subject areas, this will be the person on the working team by appointment of the executive sponsor.

Data stewards, unlike other users, have read and write access to their area of stewardship. Other users have read-only access to the Data Warehouse areas that they have been approved to read by the data stewards.

The responsibilities of a Data Steward include:

1. Arbitrating the transformation rules
2. Entering data or determining data population method
3. Verifying the data after load
4. Contribution of the business metadata
5. Approving new users
6. Supporting the user community on the data
7. Participation on a decision support steering committee
8. Data Quality

A brief description of each of these responsibilities follows.

The Data Steward is responsible for the transformation rules used in the process of moving data from source to target for sourced data. The specific tasks are:

- Assuring transformation rules keep the data meaningful and consistent
- Arbitrate differences of opinion on different interpretations of value as to how the data will be represented in the Data Warehouse
- Make the call on the initial rules and all subsequent changes

The Data Steward is responsible for the timely and accurate population of data in the area of stewardship. This is perhaps the most important of the Data Steward's tasks. This could mean manual entry through the developed user interface or the selection of the system that will provide the data (ie. The "source"). The specific tasks are:

- Determine the systems that will feed the data to the data warehouse
- Ensure feeds are developed to feed to the data into the data warehouse as soon as it is practical to do so on a regular basis

The Data Steward, as the responsible party for the quality of the data, will verify the data after it is loaded from the operational sources. The specific tasks are:

- Confirm the data was loaded and that the transformational rules were properly applied
- Formally give the approval to the greater user community for query and analysis of the data

The Data Steward will contribute the business metadata to the metadata repository. The business metadata consists of the business definition written in terms the business can understand.

The Data Steward will broker requests for new usage of the Data Warehouse. The specific tasks are:

- Knowledge of the data sources, transformation rules, and uses of the data for the area of stewardship
- Knowledge of the workload limitations of the Data Warehouse system for the area of stewardship
- Approve new users and their authority levels (usually read only)

It should be a requirement for new users of the Data Warehouse to undergo training internally on both the data model and the data access tool. The Data Steward will be responsible for the training and ongoing support of the user community on the data model and the data itself. The specific tasks for the data steward for this is:

- Participating in the regular training sessions given to new users
- Help-desk style support of the user community on a regular basis

Meetings of the decision support steering committee should be scheduled regularly to make strategic discernment and prioritization over the major additions of usage, subject areas, and data sources to the Data Warehouse. The specific tasks for the Data Steward for the committee is:

- Meeting attendance and leadership
- Judgment on the issues presented

The consistency, accuracy, and timeliness of the data are the responsibility of the data steward. The data steward's area of stewardship must receive data according to recognized, systemic methods. Data quality issues must be addressed immediately. Unattended perceptions of uncleanness quickly tear down the value of the Data Warehouse.

## **End User**

The Data Warehouse is being built for End User query and reporting. Nonetheless, there are responsibilities that an end user must meet to get maximum benefit from the Data Warehouse.

The responsibilities of an End User include:

1. Attending training before receiving an ID and password on the Data Warehouse system
2. Browsing the technical and business metadata for information on data sourcing and data definitions
3. Providing feedback to the Data Warehouse team

The End User must attend training to familiarize himself or herself with the Data Warehouse environment especially the metadata, the query tool, and the data model. The metadata will consist of tables containing data about the user data. The End User should find most of his or her data-related questions answered in the metadata.

The Data Warehouse will potentially disrupt but should ultimately increase the organization's ability to generate sales and lower costs. To get to that point, however, requires that the users actively use the Data Warehouse as instructed and, since it should be a user-centric Data Warehouse, actively provide feedback to the Data Warehouse team on issues of:

- Performance
- Functionality
- Data quality and completeness
- Data sources
- Metadata quality and completeness

### **Executive Sponsor** (by data warehouse iteration)

The Data Warehouse must have high-level and sustainable sponsorship. The Executive Sponsor must be politically viable and be able to garner and retain adequate resources for the construction and maintenance of the Data Warehouse.

The responsibilities of the Executive Sponsor include:

1. Knowledge of Data Warehouse systems
2. Leadership in the Decision support steering committee meetings
3. Knowledge of the organization's Data Warehouse environment
4. Keeping the Data Warehouse out of internal cross-fire

The Executive Sponsor, as the responsible executive for the success or failure of the iteration of the Data Warehouse, should be able to articulate and quantify the value of Data Warehouse project and its many elements. The Decision support steering committee is designed as a forum for data stewards, IS, and the Executive Sponsor to make strategic discernment and prioritization over the major additions of usage, subject areas, and data sources to the Data Warehouse. It is imperative to the success of the Data Warehouse that the Executive Sponsor takes leadership and provides vision to the committee.

The Executive Sponsor, as the responsible executive for the success or failure of the Data Warehouse, should be able to articulate and quantify the value of the Data Warehouse program. Specific areas of focus:

- Data Sources
- Subject Areas
- Data Stewardship
- High-level Architecture
- Project Plan and Budget

The Data Warehouse, to be successful, must not fall victim to infighting. To maintain this countenance, the executive sponsor must keep the focus at the positive return-on-investment potential and realization of the Data Warehouse. Frequent status should be delivered to upper level executives.

### **Database Administrator**

A key design point for a data warehouse group is the placement of the database administration function and the division of roles and responsibilities between the support group and the user community.

The responsibilities of a Database Administrator include:

1. Physical database design
2. Database maintenance
3. Backup and recovery
4. Participation on the Decision support steering committee
5. Data Replication
6. Repository Management
7. Security administration
8. Database loading
9. Performance Monitoring and Summary table creation
10. Ad-hoc data manipulation

A brief description of each of these responsibilities follows.

The Data Warehouse Database Administrator will translate the logical database design into a physically implementable design. The specific tasks are:

- Denormalizing the models based on potential queries, reporting, and system feeds that will be generated from the tables
- Implement, as necessary, some of the entities as domains rather than tables
- The addition of timestamp and user ID columns to most tables to identify the person and the time of the row addition or change
- The use of standard abbreviations in the table and column names
- Assignment of data types, nullability, and defaults for each column
- Creating the databases and database objects such as the tables, tablespaces, synonyms, views, triggers, stored procedures, defaults, domains, and indexes

The Data Warehouse Database Administrator will maintain the tables as necessary for optimization. The specific tasks may include:

- Periodic table reorganization
- Structure update with change control

The Data Warehouse Database Administrator will ensure recoverability of the Data Warehouse with minimal loss of information. Specific tasks include:

- Extensive testing of the backup and recovery processes
- Ability to recover databases within the service level agreement
- Ability to articulate any data loss encountered
- Ensure logging and recoverability of database updates since the last backup
- Take frequent backups – full and incremental as needed
- Frequent offsite storage of the latest full backup in the event of a natural disaster

Representation from the Data Warehouse Database Administration staff will participate on the Decision support steering committee.

If the Data Warehouse becomes as a distributed system, a synchronization effort may be needed. The Data Warehouse Database Administrator will be responsible for the synchronization, or replication, process. Specific tasks include:

- Ensuring accurate and complete replication of applicable data during the available windows
- Setup and administration of Oracle Replication Services or other relevant replication tool
- Collision detection to avoid loss of data
- Input to the Worldwide Data Architecture process based on replication functionality

The Data Warehouse Database Administrator will be responsible for the metadata management infrastructure. Tasks will include:

- Database maintenance for the database(s)
- Installation of the repository software, if applicable
- Front-end development
- User training

Only users approved by the data stewards will be granted security privileges on the Data Warehouse. The Data Warehouse Database Administrator will execute the grants or provide the ability for a central security group to do this. Specific tasks include:

- Creating security profiles
- Assuring new users are approved by the data steward(s)
- Assuring new users have taken training
- Granting privileges to the data stewards

For data that is sourced for the Data Warehouse, the Data Warehouse Database Administrator will, if necessary, code and install the programs to load the data into the Data Warehouse.

The Data Warehouse Database Administrator is the point person for issues of performance within the Data Warehouse environment. Specific tasks follow:

- Being first on call for performance issues from the user community
- Setting up the governors on the database systems to stop runaway queries
- Assist the user community in writing well-performing SQL
- Route system and network performance issues to the responsible party
- Spotting trends in queries and creating and maintaining summary tables for low-performing queries where possible
- Making the end user query tool aware of the summary tables

Occasionally, the most expedient way to create or restore consistency to the data in the Data Warehouse is to write insert, update, or delete statements against the database. As the in-house SQL expert, the Data Warehouse Database Administrator will perform these queries. Specifically, following the data steward review of the data load, there may be reconciliation queries to be run to bring the data to a level of consistency.

### **Data Administrator**

A key design point for a data warehouse group is the placement of the data administration function and the division of roles and responsibilities between the DBAs and the data administrators. The Data Warehouse Data Administrator will translate the user requirements into a logical database design.

The responsibilities of a Data Administrator include:

1. Logical data modeling for the data warehouse (sans the data marts)
2. Gathering business data requirements
3. Participation on the Decision support steering committee

## **Application Programmer – ETL Specialist**

The responsibilities of an Application Programmer – ETL Specialist include:

1. Sourcing the data from the operational systems
2. Applying the business transformation rules
3. Preparing a database-loadable file for the Data Warehouse
4. Management of the deployment of the data acquisition tool(s)
5. Contributing the technical metadata to the metadata repository

A brief description of each of these responsibilities follows.

The Data Warehouse Application Programmer will extract data destined for the Data Warehouse from the operational systems that store the data. The specific tasks are:

- Work with the source system analysts to understand the windows available for data extraction
- Program, test, implement and maintain any data extraction programs necessary to extract the data from the operational systems needed to be moved to the Data Warehouse

The Data Warehouse Application Programmer is responsible for applying transformation rules as necessary to keep the data clean and consistent and therefore usable by the user community. Specific tasks include:

- Participation in design sessions chaired by data stewards and/or IT personnel where decisions are made involving the transformation from source to target
- Programming the data acquisition tool with the rules to be applied to the data
- Ensuring the correct application of the business rules through data query after the data is loaded into the Data Warehouse

Following the data extraction and rules application, the file(s) will need to be made ready for loading into the Data Warehouse. This is the responsibility of the Data Warehouse Application Programmer. Specific duties include:

- Obtaining complete knowledge of the physical database schema
- Preparing the files needed to load each table that has been designated to receive files from operational systems rather than direct input from data stewards
- Programming the data acquisition tool(s) with the tables to load the files into
- Working with the Data Warehouse Database Administrator to ensure the file loads properly into the Data Warehouse

The metadata repository will contain technical metadata such as data sources and transformation rules applied. The Application Programmer will be responsible for the entry and maintenance of this information.

## **OLAP Specialist**

The responsibilities of an OLAP Specialist include:

1. Management of the deployment, maintenance, and user support of the data access tool(s)
2. Identification and enablement of the user community to access the data in a manner consistent with their business goals



A brief description of each of these responsibilities follows.

The OLAP Specialist is responsible for the management of the deployment, maintenance, and user support of the data access tools. Specific duties include:

- Working with the System Administration staff to most effectively place server components in the architecture.
- Keeping a version of the OLAP Tool in production that is under support by the tool vendor
- Installation of any client-side software needed on approved users desktops (this could also be the PC support group)
- Maintaining a high-level view of the OLAP environment and ensuring fit-for-purpose tool usage at the organization
- Ensuring access to business and technical metadata from within the OLAP tool

The OLAP Specialist is responsible for the identification and enablement of the user community to access the data in a manner consistent with their business goals. Specific duties include:

1. Profiling the user community early in the data warehouse iteration
2. Obtaining complete knowledge of the physical database schema
3. Training and supporting users on tool usage
4. Identification and scheduling of frequently needed reports

### **System Administrator** (usually by SLA with the System Administration group)

The responsibilities of a System Administrator include:

1. Installing and maintaining the Database Management System (DBMS)
2. Data warehouse architecture (in partnership with the Chief Architect)
3. Monitoring network performance
4. Monitoring DASD utilization

A brief description of each of these responsibilities follows.

The Data Warehouse System Administrator will be responsible for data destined for the Data Warehouse from the operational systems that store the data. The specific tasks are:

- Tuning the operating system for DBMS fit
- Installing and configuring the DBMS
- Maintaining the DBMS levels supported by the DBMS vendor

The Data Warehouse System Administrator is responsible for architecting, with the Chief Architect, the placement of the Data Warehouse and data marts and the adjoining data acquisition and replication strategies. Specific tasks include:

- Participation on a decision support steering committee in decisions on architecting a world-wide data infrastructure that will accommodate the data entry and retrieval needs of the domestic and international user constituencies
- Knowledge of the network capabilities
- Data architecture comprising decision points around number of data stores and the amount and timing of replication and synchronization needed to achieve the objectives of having a single version of the truth for each Data Warehouse element

The Data Warehouse System Administrator is responsible for the performance of data transfers, either in response to a query or as part of a data replication or synchronization effort across the WANs and LANs. Specific tasks include:

- Proactive network monitoring for performance and accuracy
- Recommending and expanding capacity as needed to meet performance service levels

The Data Warehouse System Administrator is responsible for ensuring enough DASD is available and efficiently managed to accommodate existing and upcoming data needs. Specific tasks include:

- Proactive DASD monitoring
- Recommending and expanding capacity as needed to assure availability

**Visit us on the World Wide Web at**  
<http://www.mcknight-associates.com/>



998 E. El Camino Real, Suite 205  
Sunnyvale, CA 94087  
Tel: 408.246.6348 Fax: 800.886.7033