# Data Warehouse Implementation Factors

## Preview

Organization-wide information system development is subject to many constraints. Some of the constraints are based on available funding. Others are a function of management's view of the role played by an IS department and of the extent and depth of the information requirements. Add the constraints imposed by corporate culture, and you understand why no single formula can describe perfect data warehouse development. Rather than proposing a single data warehouse design and implementation methodology here, this appendix identifies a few common factors to consider in implementing data warehousing.

## Data Files and Available Formats

| MS Access | Oracle | MS SQL | My SQL | | MS Access | Oracle | MS SQL | My SQL |

There are no data files for this appendix.

*Data Files Available on* **cengagebrain.com**

## O-1 Remember the Data Warehouse Is an Active Decision Support Framework

Perhaps the first thing to remember is that a data warehouse is not a static database. Instead, it is a dynamic framework for decision support, that is almost by definition a work in progress. Because the data warehouse is the foundation of a modern BI environment, designing and implementing a data warehouse means that you are involved in designing and implementing a complete database system development infrastructure for company-wide decision support. Although it is easy to focus on the data warehouse database as the central BI data repository, you must remember that the decision support infrastructure includes hardware, software, people, and procedures, as well as data. The argument that the data warehouse is the only *critical* component of BI success is as misleading as the argument that a human being needs only a heart or a brain to function. The data warehouse is a critical component of a modern BI environment, but it is certainly not the only critical component. Therefore, its design and implementation must be examined in light of the entire infrastructure.

## O-2 Be Aware of Organizational Components and Solicit User Involvement

When you design a data warehouse, you are given an opportunity to help develop an integrated model of the data that are considered to be essential to the organization, from both end-user and business perspectives. Data warehouse data cross departmental lines and geographical boundaries. Because the data warehouse represents an attempt to model all of the organization's data, you are likely to discover that organizational components (divisions, departments, support groups, and so on) often have conflicting goals, and it certainly will be easy to find data inconsistencies and damaging redundancies. Also, because information is power, the control of data sources and uses is likely to trigger turf battles, end-user resistance, and power struggles at all levels. Building the perfect data warehouse is not just a matter of knowing how to create a star schema; it requires managerial skills to deal with conflict resolution, mediation, and arbitration. In short, the designer must:

- Involve end users in the process.
- Secure end users' commitment from the beginning.
- Solicit continuous end-user feedback.
- Manage end-user expectations.
- Establish procedures for conflict resolution.

## O-3 Satisfy the Trilogy: Data, Analysis, and Users

Great managerial skills are not, of course, solely sufficient. The technical aspects of the data warehouse must be addressed as well. The old adage of input-process-output repeats itself here. The data warehouse designer must satisfy:
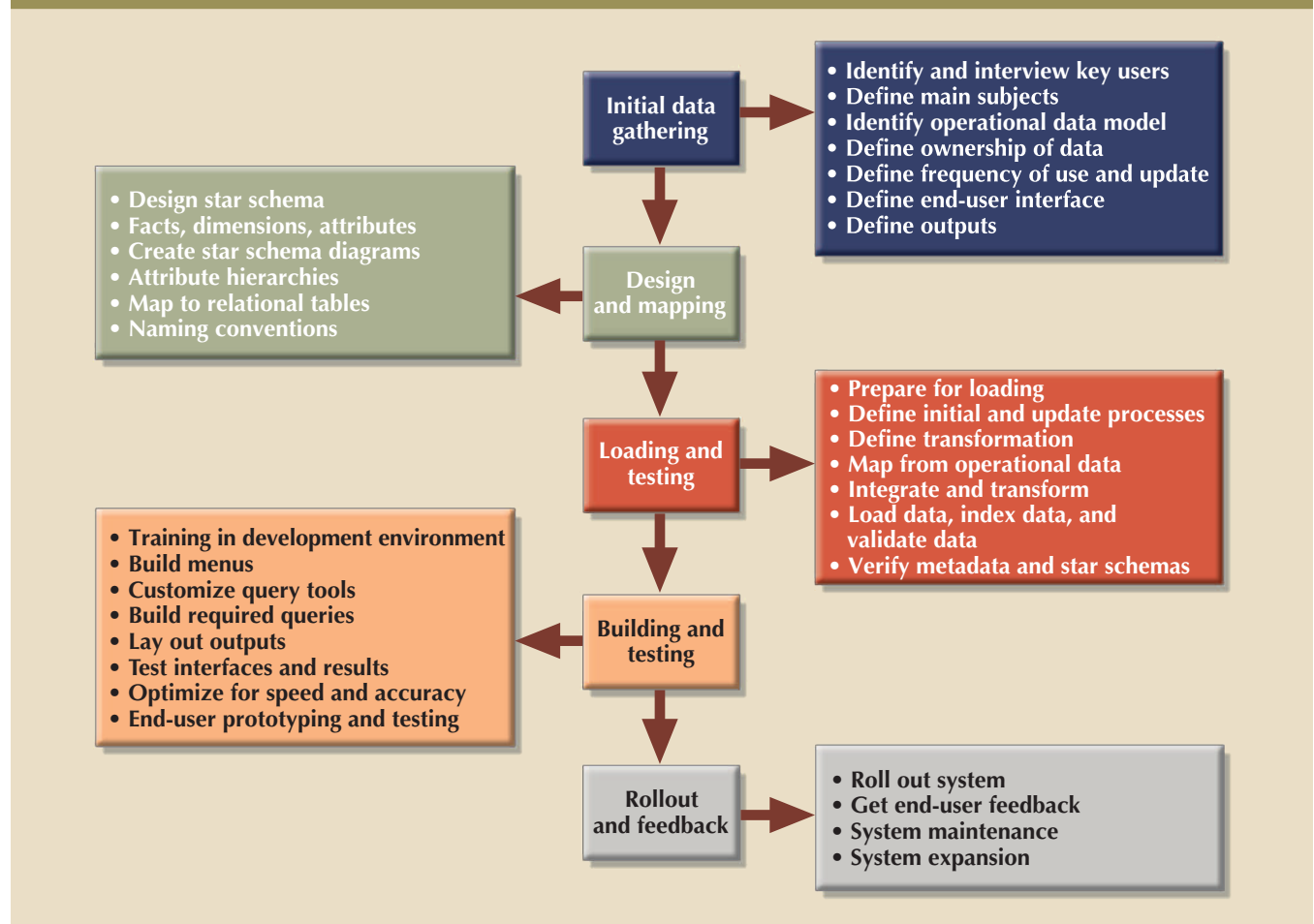
- Data integration and loading criteria.
- Data analysis capabilities with acceptable query performance.
- End-user data analysis needs.

The foremost technical concern in implementing a data warehouse is to provide end-user decision support with advanced data analysis capabilities—at the right moment, in the right format, with the right data, and at the right cost.

## O-4  Apply Database Design Procedures

You learned about the database life cycle and the database design process in Chapter 9, Database Design, so perhaps it is wise to review the traditional database design procedures. These design procedures must then be adapted to fit data warehouse requirements. If you remember that the data warehouse derives its data from operational databases, you will understand why a solid foundation in operational database design is important. (It's difficult to produce good data warehouse data when the operational database data are corrupted.) Figure O.1 depicts a simplified process for implementing the data warehouse.

**FIGURE O.1   DATA WAREHOUSE DESIGN AND IMPLEMENTATION ROAD MAP**



**Initial data gathering**
- Identify and interview key users
- Define main subjects
- Identify operational data model
- Define ownership of data
- Define frequency of use and update
- Define end-user interface
- Define outputs

**Design and mapping**
- Design star schema
- Facts, dimensions, attributes
- Create star schema diagrams
- Attribute hierarchies
- Map to relational tables
- Naming conventions

**Loading and testing**
- Prepare for loading
- Define initial and update processes
- Define transformation
- Map from operational data
- Integrate and transform
- Load data, index data, and validate data
- Verify metadata and star schemas

**Building and testing**
- Training in development environment
- Build menus
- Customize query tools
- Build required queries
- Lay out outputs
- Test interfaces and results
- Optimize for speed and accuracy
- End-user prototyping and testing

**Rollout and feedback**
- Roll out system
- Get end-user feedback
- System maintenance
- System expansion

As noted, developing a data warehouse is a company-wide effort that requires many resources: human, financial, and technical. Providing company-wide decision support requires a sound architecture based on a mix of people skills, technology, and managerial procedures that is often difficult to find and implement. For example:

- The sheer and often mind-boggling quantity of decision support data is likely to require the latest hardware and software—that is, advanced computers with multiple processors, advanced database systems, and large-capacity storage units. In the not-too-distant past, those requirements usually prompted the use of a mainframe-based system. Today's client/server technology offers many other choices to implement a data warehouse.

- Very detailed procedures are necessary to orchestrate the flow of data from the operational databases to the data warehouse. Data flow control includes data extraction, validation, and integration.

- To implement and support the data warehouse architecture, you also need people with advanced database design, software integration, and management skills.