

## A Data Mart Concepts

This chapter reviews some basic concepts relating to data marts and establishes some working definitions for use in the rest of this book. Although there is a lot of agreement among users and vendors on the definitions and terminology, they have not yet reached complete consensus. If you talk to a dozen people, you are likely to hear about half a dozen similar but slightly differing answers for even something as basic as "What is a data mart?" This chapter takes a quick look at some definitions and explains what a data mart is (and is not).

This chapter contains the following topics:

- [Section A.1, "What Is a Data Mart?"](#)
- [Section A.2, "How Is It Different from a Data Warehouse?"](#)
- [Section A.3, "Dependent and Independent Data Marts"](#)
- [Section A.4, "What Are the Steps in Implementing a Data Mart?"](#)

### A.1 What Is a Data Mart?

A data mart is a simple form of a data warehouse that is focused on a single subject (or functional area), such as Sales, Finance, or Marketing. Data marts are often built and controlled by a single department within an organization. Given their single-subject focus, data marts usually draw data from only a few sources. The sources could be internal operational systems, a central data warehouse, or external data.

### A.2 How Is It Different from a Data Warehouse?

A data warehouse, unlike a data mart, deals with multiple subject areas and is typically implemented and controlled by a central organizational unit such as the corporate Information Technology (IT) group. Often, it is called a central or enterprise data warehouse. Typically, a data warehouse assembles data from multiple source systems.

Nothing in these basic definitions limits the size of a data mart or the complexity of the decision-support data that it contains. Nevertheless, data marts are typically smaller and less complex than data warehouses; hence, they are typically easier to build and maintain. [Table A-1](#) summarizes the basic differences between a data warehouse and a data mart.

**Table A-1 Differences Between a Data Warehouse and a Data Mart**

Category	Data Warehouse	Data Mart
Scope	Corporate	Line of Business (LOB)
Subject	Multiple	Single subject
Data Sources	Many	Few
Size (typical)	100 GB-TB+	< 100 GB
Implementation Time	Months to years	Months

### A.3 Dependent and Independent Data Marts

There are two basic types of data marts: dependent and independent. The categorization is based primarily on the data source that feeds the data mart. Dependent data marts draw data from a central data warehouse that has already been created. Independent data marts, in contrast, are standalone systems built by drawing data directly from operational or external sources of data, or both.

The main difference between independent and dependent data marts is how you populate the data mart; that is, how you get data out of the sources and into the data mart. This step, called the Extraction-Transformation-and Loading (ETL) process, involves moving data from operational systems, filtering it, and loading it into the data mart.

With dependent data marts, this process is somewhat simplified because formatted and summarized (clean) data has already been loaded into the central data warehouse. The ETL process for dependent data marts is mostly a process of identifying the right subset of data relevant to the chosen data mart subject and moving a copy of it, perhaps in a summarized form.

With independent data marts, however, you must deal with all aspects of the ETL process, much as you do with a central data warehouse. The number of sources is likely to be fewer and the amount of data associated with the data mart is less than the warehouse, given your focus on a single subject.

The motivations behind the creation of these two types of data marts are also typically different. Dependent data marts are usually built to achieve improved performance and availability, better control, and lower telecommunication costs resulting from local access of data relevant to a specific department. The creation of independent data marts is often driven by the need to have a solution within a shorter time.

### A.4 What Are the Steps in Implementing a Data Mart?

Simply stated, the major steps in implementing a data mart are to design the schema, construct the physical storage, populate the data mart with data from source systems, access it to make informed decisions, and manage it over time.

This section contains the following topics:

- [Section A.4.1, "Designing"](#)
- [Section A.4.2, "Constructing"](#)

- [Section A.4.3, "Populating"](#)
- [Section A.4.4, "Accessing"](#)
- [Section A.4.5, "Managing"](#)

## A.4.1 Designing

The design step is first in the data mart process. This step covers all of the tasks from initiating the request for a data mart through gathering information about the requirements, and developing the logical and physical design of the data mart. The design step involves the following tasks:

- Gathering the business and technical requirements
- Identifying data sources
- Selecting the appropriate subset of data
- Designing the logical and physical structure of the data mart

## A.4.2 Constructing

This step includes creating the physical database and the logical structures associated with the data mart to provide fast and efficient access to the data. This step involves the following tasks:

- Creating the physical database and storage structures, such as tablespaces, associated with the data mart
- Creating the schema objects, such as tables and indexes defined in the design step
- Determining how best to set up the tables and the access structures

## A.4.3 Populating

The populating step covers all of the tasks related to getting the data from the source, cleaning it up, modifying it to the right format and level of detail, and moving it into the data mart. More formally stated, the populating step involves the following tasks:

- Mapping data sources to target data structures
- Extracting data
- Cleansing and transforming the data
- Loading data into the data mart
- Creating and storing metadata

## A.4.4 Accessing

The accessing step involves putting the data to use: querying the data, analyzing it, creating reports, charts, and graphs, and publishing these. Typically, the end user uses a graphical front-end tool to submit queries to the database and display the results of the queries. The accessing step requires that you perform the following tasks:

- Set up an intermediate layer for the front-end tool to use. This layer, the metalayer, translates database structures and object names into business terms, so that the end user can interact with the data mart using terms that relate to the business function.
- Maintain and manage these business interfaces.
- Set up and manage database structures, like summarized tables, that help queries submitted through the front-end tool execute quickly and efficiently.

## A.4.5 Managing

This step involves managing the data mart over its lifetime. In this step, you perform management tasks such as the following:

- Providing secure access to the data
- Managing the growth of the data
- Optimizing the system for better performance
- Ensuring the availability of data even with system failures