

## What is Data Mart?

A **Data Mart** is focused on a single functional area of an organization and contains a subset of data stored in a Data Warehouse. A Data Mart is a condensed version of Data Warehouse and is designed for use by a specific department, unit or set of users in an organization. e.g. Marketing, Sales, HR or finance. It is often controlled by a single department in an organization.

Data Mart usually draws data from only a few sources compared to a Data warehouse. Data marts are small in size and are more flexible compared to a Datawarehouse.

## Why do we need Data Mart?

- Data Mart helps to enhance user's response time due to reduction in volume of data
- It provides easy access to frequently requested data.
- Data mart are simpler to implement when compared to corporate Datawarehouse. At the same time, the cost of implementing Data Mart is certainly lower compared with implementing a full data warehouse.
- Compared to Data Warehouse, a data mart is agile. In case of change in model, data mart can be built quicker due to a smaller size.
- A Datamart is defined by a single Subject Matter Expert. On the contrary data warehouse is defined by interdisciplinary SME from a variety of domains. Hence, Data mart is more open to change compared to Datawarehouse.
- Data is partitioned and allows very granular access control privileges.
- Data can be segmented and stored on different hardware/software platforms.

## Types of Data Mart

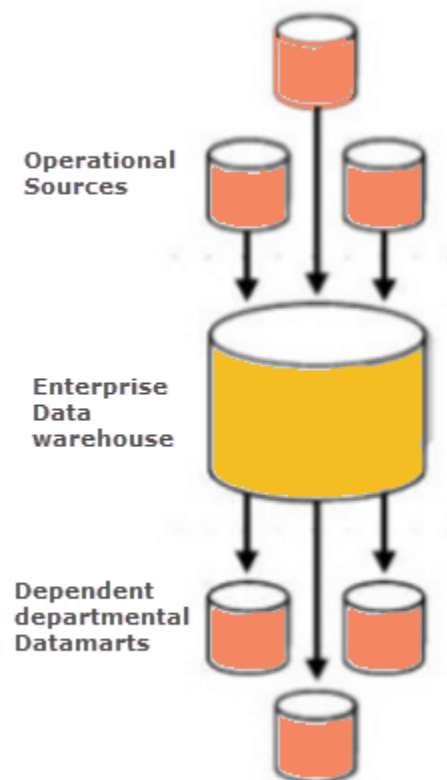
There are three main types of data mart:

1. **Dependent:** Dependent data marts are created by drawing data directly from operational, external or both sources.
2. **Independent:** Independent data mart is created without the use of a central data warehouse.
3. **Hybrid:** This type of data marts can take data from data warehouses or operational systems.

## Dependent Data Mart

A dependent data mart allows sourcing organization's data from a single Data Warehouse. It is one of the data mart example which offers the benefit of centralization. If you need to develop one or more physical data marts, then you need to configure them as dependent data marts.

Dependent Data Mart in data warehouse can be built in two different ways. Either where a user can access both the data mart and data warehouse, depending on need, or where access is limited only to the data mart. The second approach is not optimal as it produces sometimes referred to as a data junkyard. In the data junkyard, all data begins with a common source, but they are scrapped, and mostly junked.



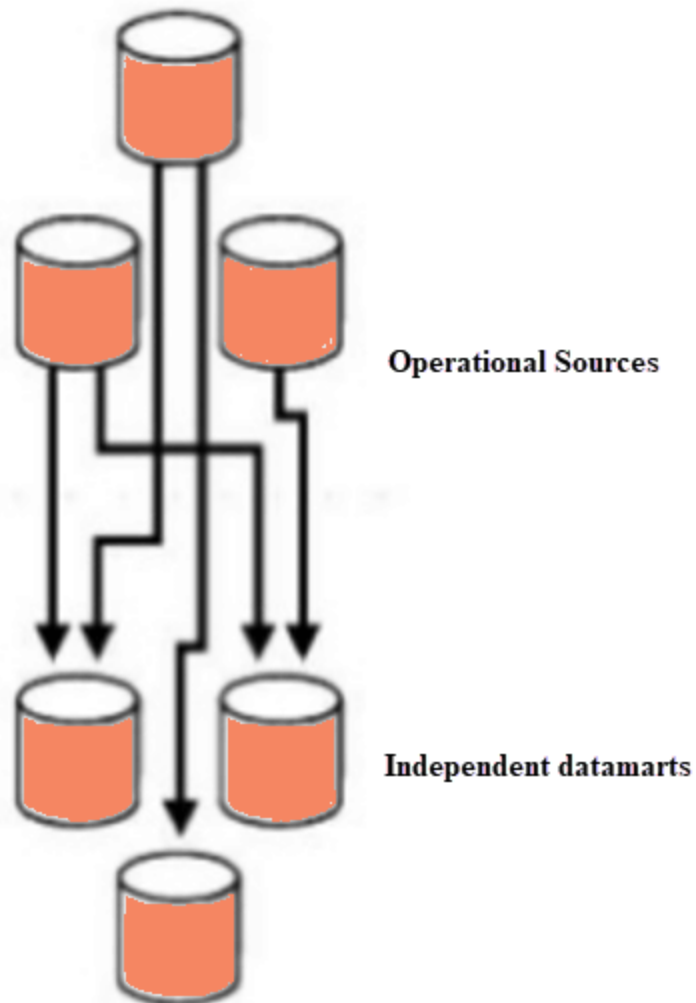
**Dependent Data Mart**

## Independent Data Mart

An independent data mart is created without the use of central Data warehouse. This kind of Data Mart is an ideal option for smaller groups within an organization.

An independent data mart has neither a relationship with the enterprise data warehouse nor with any other data mart. In Independent data mart, the data is input separately, and its analyses are also performed autonomously.

Implementation of independent data marts is antithetical to the motivation for building a data warehouse. First of all, you need a consistent, centralized store of enterprise data which can be analyzed by multiple users with different interests who want widely varying information.

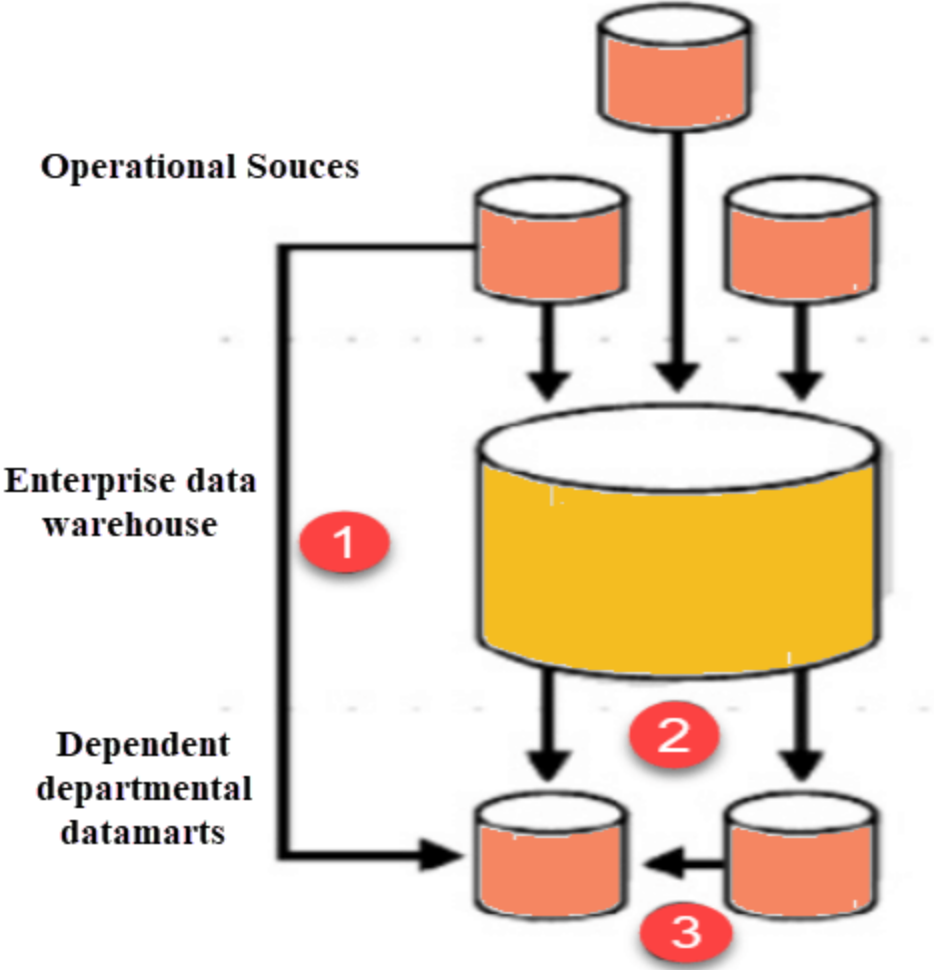


**Independent Data Mart**

### **Hybrid Data Mart:**

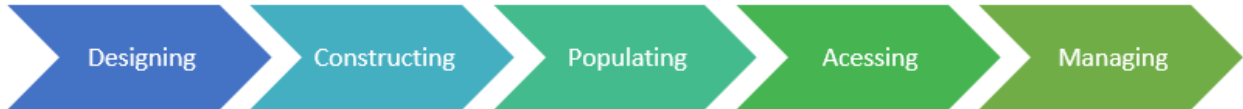
A hybrid data mart combines input from sources apart from Data warehouse. This could be helpful when you want ad-hoc integration, like after a new group or product is added to the organization.

It is the best data mart example suited for multiple database environments and fast implementation turnaround for any organization. It also requires least data cleansing effort. Hybrid Data mart also supports large storage structures, and it is best suited for flexible for smaller data-centric applications.



Hybrid Data Mart

### Steps in Implementing a Datamart



## Designing

Designing is the first phase of Data Mart implementation. It covers all the tasks between initiating the request for a data mart to gathering information about the requirements. Finally, we create the logical and physical Data Mart design.

The design step involves the following tasks:

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

Data could be partitioned based on following criteria:

- Date
- Business or Functional Unit
- Geography
- Any combination of above

Data could be partitioned at the application or DBMS level. Though it is recommended to partition at the Application level as it allows different data models each year with the change in business environment.

### What Products and Technologies Do You Need?

A simple pen and paper would suffice. Though tools that help you create UML or ER diagram would also append metadata into your logical and physical designs.

## Constructing

This is the second phase of implementation. It involves creating the physical database and the logical structures.

This step involves the following tasks:

- Implementing the physical database designed in the earlier phase. For instance, database schema objects like table, indexes, views, etc. are created.

### What Products and Technologies Do You Need?

You need a relational database management system to construct a data mart. RDBMS have several features that are required for the success of a Data Mart.

- **Storage management:** An RDBMS stores and manages the data to create, add, and delete data.
- **Fast data access:** With a SQL query you can easily access data based on certain conditions/filters.
- **Data protection:** The RDBMS system also offers a way to recover from system failures such as power failures. It also allows restoring data from these backups in case of the disk fails.
- **Multiuser support:** The data management system offers concurrent access, the ability for multiple users to access and modify data without interfering or overwriting changes made by another user.
- **Security:** The RDMS system also provides a way to regulate access by users to objects and certain types of operations.

## Populating:

In the third phase, data is populated in the data mart.

The populating step involves the following tasks:

- Source data to target data Mapping
- Extraction of source data
- Cleaning and transformation operations on the data
- Loading data into the data mart
- Creating and storing metadata

## What Products and Technologies Do You Need?

You accomplish these population tasks using an ETL (Extract Transform Load) Tool. This tool allows you to look at the data sources, perform source-to-target mapping, extract the data, transform, cleanse it, and load it back into the data mart.

In the process, the tool also creates some metadata relating to things like where the data came from, how recent it is, what type of changes were made to the data, and what level of summarization was done.

## Accessing

Accessing is a fourth step which involves putting the data to use: querying the data, creating reports, charts, and publishing them. End-user submit queries to the database and display the results of the queries

The accessing step needs to perform the following tasks:

- Set up a meta-layer that translates database structures and objects names into business terms. This helps non-technical users to access the Data mart easily.
- Set up and maintain database structures.
- Set up API and interfaces if required

### What Products and Technologies Do You Need?

You can access the data mart using the command line or GUI. GUI is preferred as it can easily generate graphs and is user-friendly compared to the command line.

## Managing

This is the last step of Data Mart Implementation process. This step covers management tasks such as-

- Ongoing user access management.
- System optimizations and fine-tuning to achieve the enhanced performance.
- Adding and managing fresh data into the data mart.
- Planning recovery scenarios and ensure system availability in the case when the system fails.

### What Products and Technologies Do You Need?

You could use the GUI or command line for data mart management.

## Best practices for Implementing Data Marts

Following are the best practices that you need to follow while in the Data Mart Implementation process:

- The source of a Data Mart should be departmentally structured

- The implementation cycle of a Data Mart should be measured in short periods of time, i.e., in weeks instead of months or years.
- It is important to involve all stakeholders in planning and designing phase as the data mart implementation could be complex.
- Data Mart Hardware/Software, Networking and Implementation costs should be accurately budgeted in your plan
- Even though if the Data mart is created on the same hardware they may need some different software to handle user queries. Additional processing power and disk storage requirements should be evaluated for fast user response
- A data mart may be on a different location from the data warehouse. That's why it is important to ensure that they have enough networking capacity to handle the Data volumes needed to transfer data to the data mart.
- Implementation cost should budget the time taken for Datamart loading process. Load time increases with increase in complexity of the transformations.

## Advantages and Disadvantages of a Data Mart

### Advantages

- Data marts contain a subset of organization-wide data. This Data is valuable to a specific group of people in an organization.
- It is cost-effective alternatives to a data warehouse, which can take high costs to build.
- Data Mart allows faster access of Data.
- Data Mart is easy to use as it is specifically designed for the needs of its users. Thus a data mart can accelerate business processes.
- Data Marts needs less implementation time compare to Data Warehouse systems. It is faster to implement Data Mart as you only need to concentrate the only subset of the data.
- It contains historical data which enables the analyst to determine data trends.

### Disadvantages

- Many a times enterprises create too many disparate and unrelated data marts without much benefit. It can become a big hurdle to maintain.
- Data Mart cannot provide company-wide data analysis as their data set is limited.



## Summary:

- Define Data Mart: A Data Mart is defined as a subset of Data Warehouse that is focused on a single functional area of an organization.
- Data Mart helps to enhance user's response time due to a reduction in the volume of data.
- Three types of data mart are 1) Dependent 2) Independent 3) Hybrid
- Important implementation steps of Data Mart are: 1) Designing 2) Constructing 3) Populating 4) Accessing and 5) Managing
- The implementation cycle of a Data Mart should be measured in short periods of time, i.e., in weeks instead of months or years.
- Data mart is cost-effective alternatives to a data warehouse, which can take high costs to build.
- Data Mart cannot provide company-wide data analysis as data set is limited.