

13. DATA MARTING

Why Do We Need a Data Mart?

Listed below are the reasons to create a data mart:

- ☐ To partition data in order to impose **access control strategies**.
- ☐ To speed up the queries by reducing the volume of data to be scanned.
- ☐ To segment data into different hardware platforms.
- ☐ To structure data in a form suitable for a user access tool.

Note: Do not data mart for any other reason since the operation cost of data marting could be very high. Before data marting, make sure that data marting strategy is appropriate for your particular solution.

Cost-effective Data Marting

Follow the steps given below to make data marting cost-effective:

- ☐ Identify the Functional Splits
- ☐ Identify User Access Tool Requirements
- ☐ Identify Access Control Issues

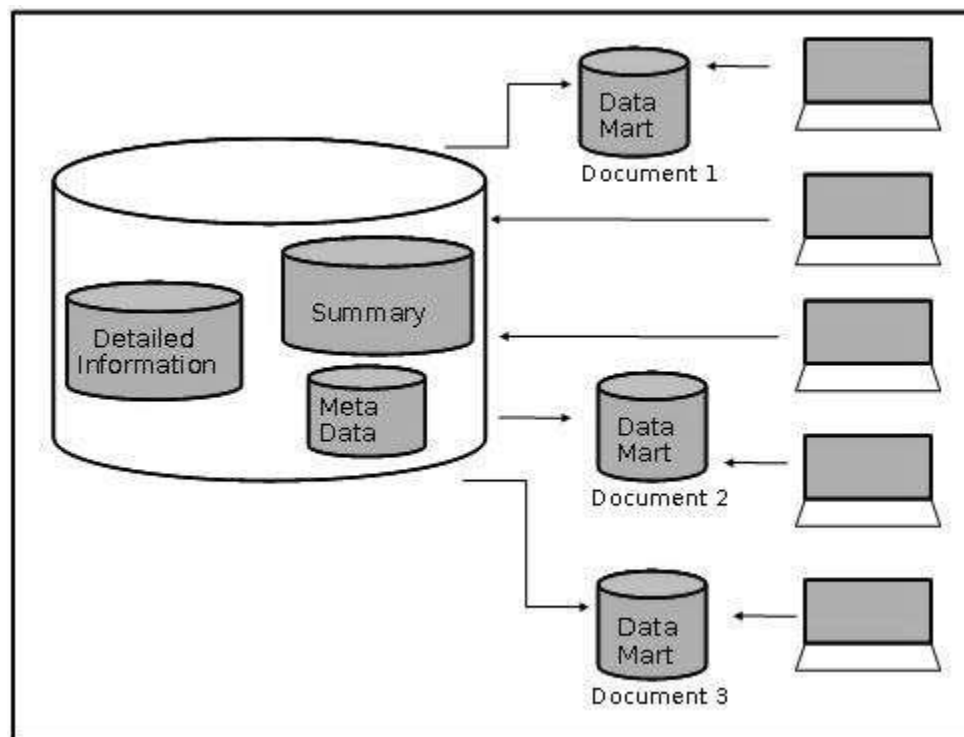
Identify the Functional Splits

In this step, we determine if the organization has natural functional splits. We look for departmental splits, and we determine whether the way in which departments use information tend to be in isolation from the rest of the organization. Let's have an example.

Consider a retail organization, where each merchant is accountable for maximizing the sales of a group of products. For this, the following are the valuable information:

- ☐ sales transaction on a daily basis
- ☐ sales forecast on a weekly basis
- ☐ stock position on a daily basis
- ☐ stock movements on a daily basis

As the merchant is not interested in the products they are not dealing with, the data marting is a subset of the data dealing which the product group of interest. The following diagram shows data marting for different users.



Given below are the issues to be taken into account while determining the functional split:

- ☐ The structure of the department may change.
- ☐ The products might switch from one department to other.
- ☐ The merchant could query the sales trend of other products to analyze what is happening to the sales.

Note: We need to determine the business benefits and technical feasibility of using a data mart.

Identify User Access Tool Requirements

We need data marts to support **user access tools** that require internal data structures. The data in such structures are outside the control of data warehouse but need to be populated and updated on a regular basis.

There are some tools that populate directly from the source system but some cannot. Therefore additional requirements outside the scope of the tool are needed to be identified for future.

Note: In order to ensure consistency of data across all access tools, the data should not be directly populated from the data warehouse, rather each tool must have its own data mart.

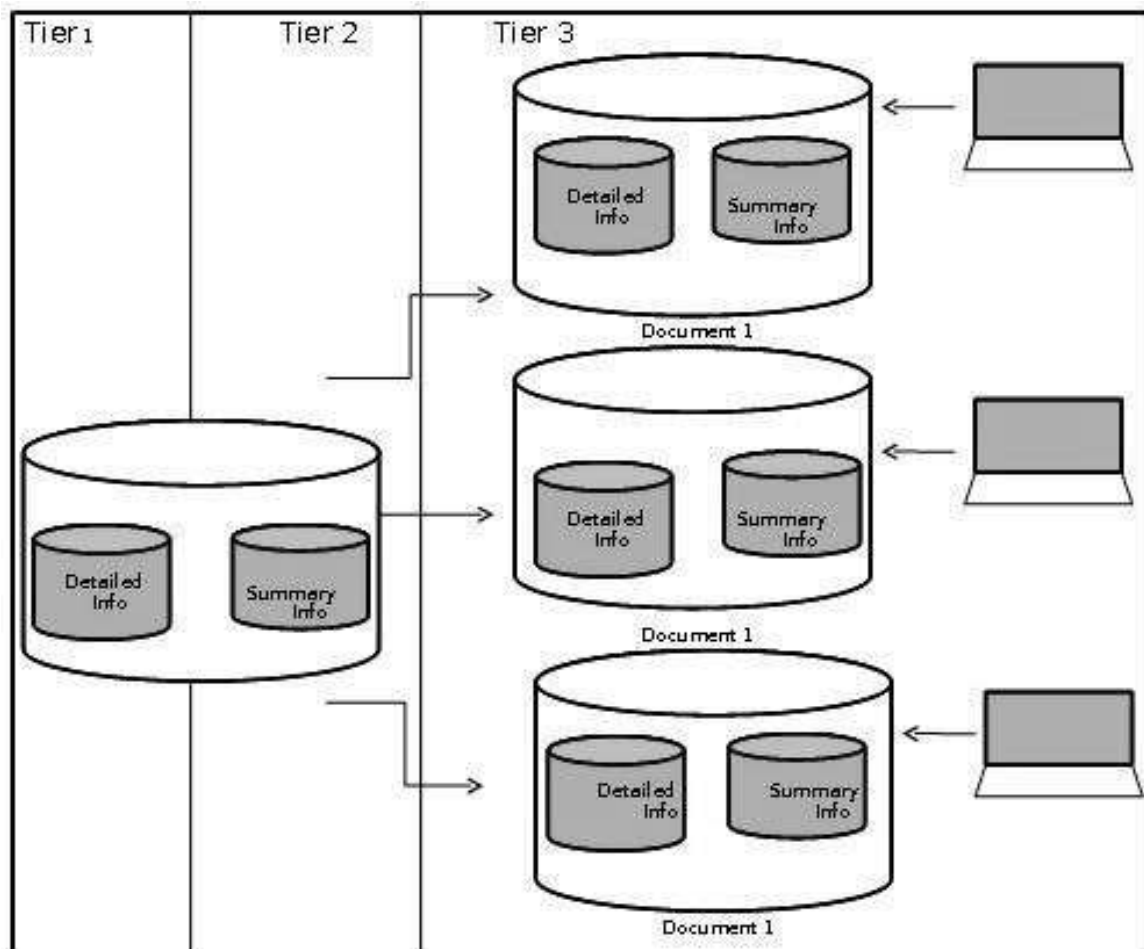
Identify Access Control Issues

There should be privacy rules to ensure the data is accessed by authorized users only. For example, a data warehouse for retail banking institution ensures that all the accounts belong to the same legal entity. Privacy laws can force you to totally prevent access to information that is not owned by the specific bank.

Data marts allow us to build a complete wall by physically separating data segments within the data warehouse. To avoid possible privacy problems, the detailed data can be removed from the data warehouse. We can create data mart for each legal entity and load it via data warehouse, with detailed account data.

Designing Data Marts

Data marts should be designed as a smaller version of starflake schema within the data warehouse and should match with the database design of the data warehouse. It helps in maintaining control over database instances.



The summaries are data marts in the same way as they would have been designed within the data warehouse. Summary tables help to utilize all dimension data in the starflake schema.

Cost of Data Marting

The cost measures for data marting are as follows:

- ☐ Hardware and Software Cost
- ☐ Network Access
- ☐ Time Window Constraints

Hardware and Software Cost

Although data marts are created on the same hardware, they require some additional hardware and software. To handle user queries, it requires additional processing power and disk storage. If detailed data and the data mart exist within the data warehouse, then we would face additional cost to store and manage replicated data.

Note: Data marting is more expensive than aggregations, therefore it should be used as an additional strategy and not as an alternative strategy.

Network Access

A data mart could be on a different location from the data warehouse, so we should ensure that the LAN or WAN has the capacity to handle the data volumes being transferred within the **data mart load process**.

TimeWindow Constraints

The extent to which a data mart loading process will eat into the available time window depends on the complexity of the transformations and the data volumes being shipped. The determination of how many data marts are possible depends on:

- ☐ Network capacity
- ☐ Time window available
- ☐ Volume of data being transferred
- ☐ Mechanisms being used to insert data into a data mart