

Publication: Data Management Review

Column Title: Data Warehouse Delivery

Author: Douglas Hackney

Headline: What is a Data Mart?

Issue: February 1997

Submitted via: e-mail & fax

[begin copy]

What is a Data Mart?

As the data mart has emerged from the shadow of the enterprise data warehouse a plethora of definitions and criteria have emerged as litmus tests to qualify data sets as a legitimate "data mart". Almost everyone has a cut at what defines a data mart. Each has a unique perspective, with unique criteria for establishing legitimacy. All assume that the data mart is designed using the basic principles of data warehousing. Beyond that, there remains considerable debate as to what, exactly, defines a data mart.

Criteria

The various definitions of data marts usually revolve around the following criteria:

- target application(s)
- size
- number of users
- audience
- location/topology
- level of aggregation

Target application(s)

The term "data mart" was primarily driven into legitimacy through the efforts of application vendors who needed a legitimate platform to host their three tier applications. The initial drivers were the MOLAP (Multidimensional On-line Analytical Processing) vendors who needed a legitimate discrete location for their MDDB (Multi Dimensional Data Base) and its associated analytical processing engines.

Because of this heritage, some analysts are still application focused in their criteria for data mart legitimacy. In their view, if a data mart does not have a dedicated application for its data set, it does not qualify.

In the quickly evolving world of data marts, this criteria is losing its relevancy, as data marts become the general purpose DSS (Decision Support System) information repository it becomes increasingly difficult to assign specific dedicated applications to specific dedicated data sets.

Size

It has also been popular historically to segment data marts by size. "Less than 20 gigabytes" was a common breakpoint for some time. Today, however, you don't have to look far to find data marts of all sizes, from the classic 2-20 gigabyte up to 250 gigabytes and well beyond. Clearly, it is not possible to identify and segment data marts by size alone.

Number of users

It has also been popular to attempt to attach a segmentation based upon the number of users of a data set to qualify it as a legitimate data mart. When data marts were primarily used to host

specific MOLAP applications, with tightly segmented knowledge worker constituencies, this was somewhat appropriate.

Today's data mart, however, is a vital information resource for everyone from work groups of five to functional divisions of thousands. As the data mart has evolved away from the application centric definition, the number of users criteria has lost its validity and relevancy.

Level of aggregation

In Bill Inmon's "Building the Data Warehouse", he defined a data mart as being composed of highly summarized information derived from the detail in the parent enterprise data warehouse. This provided a natural starting point for the application vendors who soon followed with their aggregate dependent solutions.

As we have seen, however, a criteria that excludes detail does not last long in the realm of real world requirements, where a myriad of business needs for detail exist among data mart users. The quickly evolving world of distributed and virtual data warehouses composed of multiple data marts demands detail resident in the data mart. As a result aggregation level based criteria are no longer applicable to determining the definition of a data mart.

Location/Topology

Another popular criteria, especially among RDBMS (Relational Data Base Management System) and server vendors is physical location or topology of the data set. In this view, any data set that is physically separate from the enterprise data warehouse or central location is a data mart.

While this makes labeling of topology diagrams much easier, and is an easily understandable criteria for non-technical audiences, it ignores the prerequisite of a valid design and architecture to the data set. This criteria allows any subset of data, regardless of architecture and level of integration to inherit the mantle of "data mart". This is a dangerous precedent, as anyone who has been tasked with integrating non-architected data marts can attest.

Audience

The last, and perhaps only valid criteria, is audience. Data marts must be driven by a clear business need in order to succeed. Because of this, the only valid segmentation and qualification of "data marts" is business user based.

Data marts are defined by users that share one or more of the following:

- common business problems
- common semantics
- common metrics
- common roles/levels
- common mission
- common domain
- common geography

You undoubtedly have many data sets that have users that pass one or more of these criteria. However the data set is either not architected, not integrated, or does not reflect the industry standard data warehouse design criteria. In this case you are the proud owner of a classic data island.

Any data set that reflects a valid design, is integrated, and/or is developed under a valid enterprise architecture, and has users that meet these criteria can legitimately be called a "data mart".

Excerpted from "Understanding and Implementing Successful Data Marts", a forthcoming book from Addison Wesley Longman Publishing. Further information is available at www.entergroup.com.

[end copy]

Copyright © 1997, Douglas Hackney