



## I.T. News & Features

Shortcuts:

Go!

home

event info

sign up

exhibit now

i.t. news

press room

find it

help



### A Taxonomy of Corporate Data Warehouses

By Aaron Zornes

Executive VP, Application Delivery Strategies, META Group

The proliferation of data marts will drive IT toward a corporate data warehouse (DW) architecture. Differing styles of corporate DWs will drive interoperability with their respective data marts.

**META Trend:** During 1995/96, DW architectures will enable component-level integration of OLAP access with corporate OLTP applications and data. Through 1996/97, key challenges for large-scale DWs include lagging support for metadata synchronization, information catalogs, and DW-smart database design tools and methodologies.

In ongoing client briefings on DW architecture, we see large companies continuing to struggle with the relationship between data marts (DMs) and centralized corporate DWs. As discussed in an earlier Delta (ADS Delta 422, 30 Nov 95 -- for convenience, we are repeating its DW and DM definitions in Figure 1, below), IT will face significant issues in constructing corporate DWs. Moreover, a series of misconceptions about overall corporate DW strategy and the relationship between DMs and corporate DWs are emerging:

**Myth No. 1:** Corporate DWs are mandatory as part of an overall decision support strategy. We have continued to emphasize the importance of data marts being constructed for sound business reasons. The "If we build it, they will come" philosophy rarely works, and end users must both pay for the data mart and maintain active participation in the iterative construction process. Similarly, a corporate DW requires the same type of business justification, and will be driven either by the simplification of data distribution or the aggregation of data spanning business units; hence it supports cross-divisional analysis.

**Myth No. 2:** Corporate DWs are larger than data marts. In many cases, this will be true. However, it is entirely possible that business unit analysis requires greater historical perspective than cross-sectional

analysis, and in the latter case, two years of data may be required, while for corporate purposes, six months might suffice.

**Myth No. 3:** Data in data marts must be represented in the corporate DW. Breadth of data in both data marts and corporate DWs is driven by the needs of their respective business owners. Consequently, unless these data requirements are complementary, data marts may quickly become the primary sources of data, requiring IT to manage individual backup strategies.

During 1996, systems integrators (SIs) will improve their overall DW methodologies to define business requirements and design for both data marts and interoperable corporate DWs. This expertise will continue to come from traditional SIs as well as "biased" hardware and software vendors, which can offer an end-to-end solution, including their individual product offerings. By the first half of 1997, middleware and replication software vendors (e.g., Sybase/MDI, Information Builders, Praxis) will mature to provide faster data distribution to data marts, heterogeneous joins across marts, and advanced catalogs to facilitate key data location. Business information directory technology (i.e., the ability to identify core data elements, their definitions, and how they are used in a variety of queries/reports) will reach maturity in 1998 and will be combined with directory services provided by middleware vendors.

We can identify a variety of corporate DW "styles," each with different interoperability approaches for their respective data marts:

**Cross-functional Data Warehouse:** This category represents "traditional" corporate DWs, which are built for various business reasons. In many cases (e.g., banking), these DWs provide a centralized view of the customer, while the customer is served by various business departments. It is important to realize, however, that centralized customer DWs are valuable only if the organization has the opportunity to cross-sell these customers across disparate business units. Cross-functional DWs are often a logical aggregation of data stored in individual data marts. They serve two important functions: 1) With heterogeneous joins across databases still immature, these DWs provide a vehicle for manageable cross-functional analysis; and 2) These DWs are often politically correct -- in fiercely decentralized organizations, IT can maintain a central backup strategy, providing nightly refreshes to business data marts.

**Distribution Data Warehouse:** As data marts proliferate (most companies will have three or more data marts by the first half of 1997), distribution DWs serve a purpose identical to distribution centers supplying retailers from a central warehouse. For example, imagine an organization with four data marts. Conceivably, these data marts could all require feeds from a variety of centralized operational systems (e.g., order management, accounting, customer billing, and sales analysis). If 10 data sources are required to feed four data marts, 40 individual replications are required. Conversely, the 10 data sources could be replicated into a distribution DW (10 replications), and the distribution DW could then perform four replications into the respective data marts (for a total of 14 replications). In this case, by creating a distribution DW, IT can save 26 replications nightly -- in addition to the aggregate value of a central data store. For distribution DWs, however, it is highly likely that individual data marts may contain more data than the central DW, since central DW data may outlive its utility shortly after replication.

**Operational Data Stores:** Operational data stores (ODSs) provide a centralized view of near real-time data from operational systems. Although our research shows most DWs are refreshed daily (the warehouse data is of daily periodicity), there are situations (e.g., inventory movement, freight balancing) where quick analysis is required, and, if the data exists in separate files, a central ODS may facilitate this analysis. In addition, the ODS can also serve as a replacement for change logs (to refresh other DSS files in the enterprise).

**Figure 1 -- Data Mart and Data Warehouse Defined**

**A data mart is a subject- or department-oriented data warehouse. It can include data duplicated from a corporate data warehouse and/or local data. A corporate data warehouse is a process by which related data from many operational systems is merged to provide a single, integrated business information view that spans all business divisions.**

**Bottom Line:** IT and end users need to justify both data marts and a variety of corporate DWs. The degree of data redundancy between data marts and corporate DWs will depend entirely on ongoing analytical needs and overall DW backup strategies.

**Aaron Zornes** is featured at [DCI's Data Warehouse World](#).

[\[home\]](#) [\[event info\]](#) [\[sign up\]](#) [\[exhibit now\]](#) [\[i.t. news\]](#) [\[press room\]](#) [\[find it\]](#) [\[help\]](#)

© Copyright DCI Consulting, Inc. (978) 470-3880  
All event names are trademarks of DCI or its clients.  
Comments? [webmaster@dciexpo.com](mailto:webmaster@dciexpo.com)