



## TenStep Supplemental Paper

---

1 January 2004

### Data Warehouses

Data warehousing was proclaimed the solution to the management information dilemma. However, the term "data warehouse" has become one of the most used and abused terms in the IT vocabulary. Ask a variety of vendors and professionals for their vision of what a data warehouse is and how it should be built. The ambiguity of the term will quickly become apparent.

The concept of "data warehousing" dates back at least to the mid-1980s, and possibly earlier. In essence, it was intended to provide an architectural model for the flow of data from operational systems to decision support environments. It attempted to address the various problems and high costs associated with this flow. In the absence of such architecture, there usually existed an enormous amount of redundancy in the delivery of management information.

Many people define a data warehouse to be any collection of summarized data from various sources, structured and optimized for query access using OLAP (on-line analytical processing) tools. The vendors of OLAP tools originally propagated this view. To others, a "logical" data warehouse is virtually any database containing data from more than one source, collected for the purpose of providing management information. This definition is not helpful since such databases have been a feature of decision support solutions long before the coining of the term "data warehouse." Some people also define a data warehouse narrowly to only include solutions that utilize data warehouse hardware and software.

In larger corporations, it was typical for multiple decision support projects to operate independently, each serving different users but often requiring much of the same data. The process of gathering, cleaning, and integrating data from various sources, often legacy systems, was typically replicated for each project. Moreover, legacy systems were frequently being revisited as new requirements emerged. Each of these required a subtly different view of the legacy data. Data warehouses were intended to help resolve this mess.

Based on analogies with real-life warehouses, data warehouses were intended as large-scale collection/storage/staging areas for legacy data. From here, data could be distributed to "retail stores" or "data marts" that were tailored for access by decision support users. The data warehouse was designed to manage the bulk data from many sources. Data marts could be used to handle the organization and storage of narrower collections of data that fulfilled specific user requirements.

This definition is still valid today, although some vendors put various other spins on the definitions to highlight strengths with their products. However, if you have trouble visualizing what a data warehouse is, try to imagine instead a physical warehouse receiving, storing and disseminating goods. In a very general sense, a data warehouse fulfills a similar role with data.