

Lakshmi Mohan - University at Albany

Building a data warehouse is a complex process that is beset by problems. It takes persistence to see the project through to the end. And as with any data warehouse project, the HIMS project was a data odyssey.

The objective of a data warehouse project is to create an integrated database of relevant data from multiple sources to support managerial decision-making. The challenge is to define relevant data, which must be done in the context of how the data will be used by management for problem-finding and problem-solving. The relevant data are determined by focusing on what you must do and know, rather than what is nice to do and know. Since you can't use every possible piece of data, you must identify the data that is necessary for your business needs. The key is to get actionable information.

Another powerful concept that must be applied in a data warehouse project is the "satisficing" concept. Better and more data will entail higher costs and more development time. However, the value of the project depends on the impact of the data on managerial actions. The satisficing approach to building data warehouses advocates that management only needs information that is "good enough" (satisficing) for decision making, because "perfect" information would cost too much and take too much time.

The process of actually building the data warehouse should employ an evolutionary approach. In this approach, the data warehouse evolves over time with users getting value in each phase because they actually use the data. To ensure that the data have value, a prototype system should be built in each phase of the project.

In the first phase, you must start with the core data elements that are critical for managerial decision-making, and develop the prototype using real data. The prototype is the only way to ensure that the system built on this data meets user needs. It gives users a real feel for the information produced by the system, allows them to give constructive feedback, and identifies potential data problems that could lead to implementation failure. The prototype must demonstrate the value of the data to users. If they do not see the value, then you should discard it before more time and money is wasted on building a data warehouse that will not be used. But if users believe that the system will help them do their jobs better, then you should move forward by converting the prototype into version 1.0 and install it in the users' workstations.

In the second phase, version 1.0 is already in full operation. Users can now provide input on new data to be added to the system that will enhance its utility. These enhancements to version 1.0 should also be prototyped to determine their added value. Implement version 2.0 and continue, if users desire it, with version 3.0. Continue this same process with version 3.0 and so on. The data warehouse should expand in time with its actual use. The key is to start small and keep going until you have a comprehensive data warehouse that meets business needs.

The data are the foundation of the HIMS prototype. The integration of data from several sources posed several challenges in this project.

- A comparative analysis of data elements from the files of two shelter providers revealed gaps in the records. BSS had to send staffers out to the providers to collect the missing data from paper files on clients. The team focused on gathering data about demographics and the following services: housing preparation, employment, child care, mental health, substance abuse, and living skills.
- Some of the information was inconsistent between the shelters. For instance, there were more than a dozen terms used to describe clients' ethnicity, and neither shelter used all the same vocabulary. Common data definitions had to be created to solve these kinds of issues.
- Relevant data from the state's Welfare Management System also had to be incorporated into the prototype. This involved gaining a thorough understanding of the system's functions and features to determine which payment and benefits information were necessary to meet the business needs of the users.
- Key metrics were defined to help sort out the data. For example, the team devised definitions for "first time" and "repeater" clients. And the group created a formula to compute the length of stay for current residents.

Defining the business rules for handling a variety of data problems encountered in the real data was the key to integrating the data from the different sources. The business rules were defined on the basis of the system design that was created to convert the data into an interactive on-screen reporting system. The system design was critical for converting raw data in the data warehouse into actionable information that management could use. Without the system, the data warehouse would be a worthless luxury.

Once the database was defined and the system design was specified, the next step was to actually build the prototype.