

Who in the World Needs a Data Warehouse?

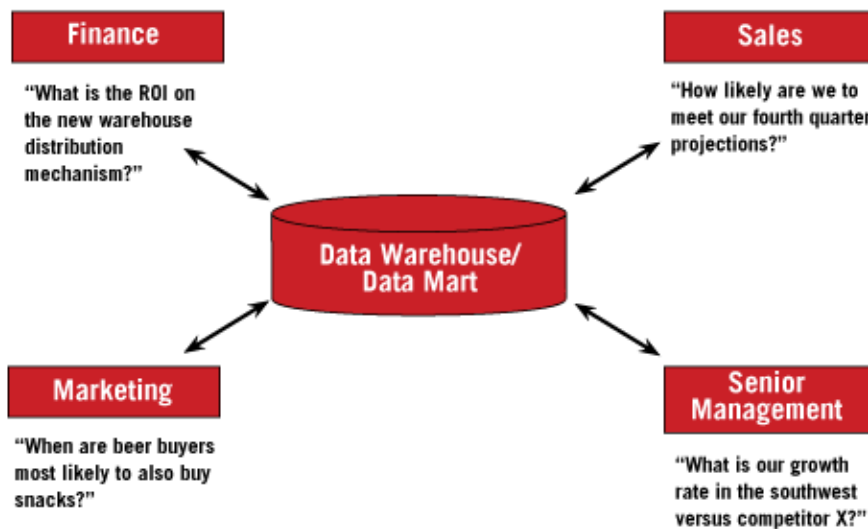
By [Shaku Atre](#)

JAN 27, 2011 3:00pm ET

A data warehouse is a cross-organizational decision support environment. It is not a product, and as such, it cannot be bought off the shelf as a unitized whole. Instead, it must be custom-designed to solve specific business problems and give the organization a competitive edge.

This type of decision support environment serves every business unit in the organization; it is not as easy as building a standalone system for one individual user or one department. It also cannot be built in one big bang - it must evolve and grow over time. Some of the typical data warehousing questions are depicted in Figures 1 and 2.

Figure 1: Typical Data Warehouse Questions



Source: Atre Group, Inc.

Figure 2: Fundamental Business Questions

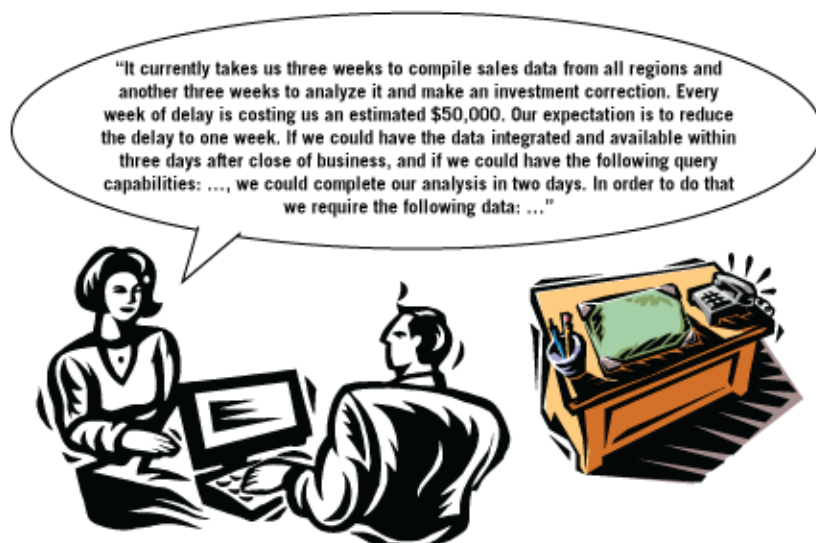


Source: Addison-Wesley, "Business Intelligence Roadmap: The Complete Project Lifecycle for Decision-Support Applications," by Larissa T. Moss and Shaku Atre

Data warehouse projects are time-consuming for various reasons:

Changes to scope. This can include staff, budget, technology, business representatives and sponsors, all of which can severely impact the success of the project. "Scope creep" can be one of the largest roadblocks in completing a long-term project. At the outset, certain requirements need to be established, defining the information that needs to be provided. After a week or so, more requests tend to pop up: "Would you please do this too? I really need this information!" The requirements keep expanding, and the scope of the project keeps stretching, resulting in scope creep (see Figure 3). Data warehouse projects are very prone to scope creep because these projects try to provide data to many users from numerous departments within the organization.

Figure 3: Requirements Definition Statement



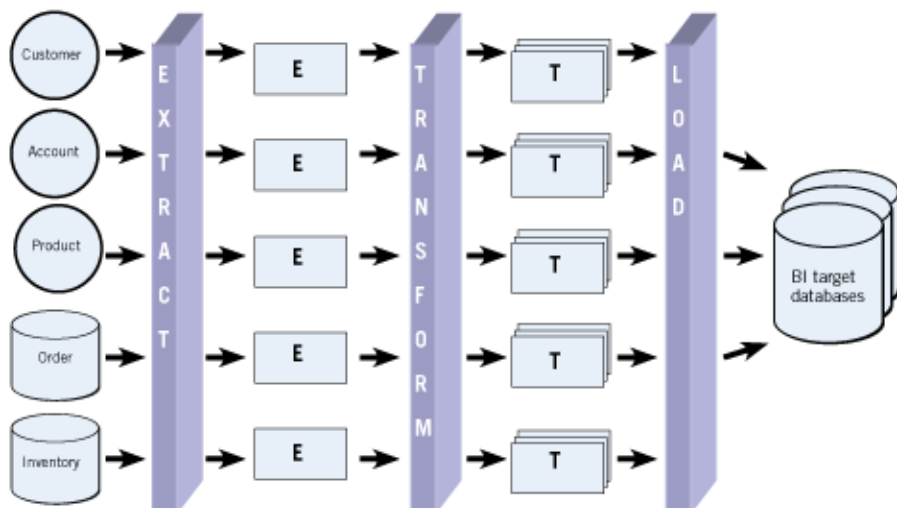
Source: Addison-Wesley, "Business Intelligence Roadmap: The Complete Project Lifecycle for Decision-Support Applications," by Larissa T. Moss and Shaku Atre

Data quality. Source data quality can be one of the biggest challenges for all data warehouse projects. Before the data warehouse initiative, data was limited to the view of one line of business, and the data was never reconciled with other views in the organization. As a result, data integration and data cleansing steps require a significant percentage of time out of the project schedule.

Extract, transform and load design. ETL is the most complicated process of a data warehouse project. The extract stage requires extracting data from the source; the transform stage requires transforming data from its original format to fit into the format to be stored in the data warehouse; and the load stage is when you actually load the transformed data into the data warehouse.

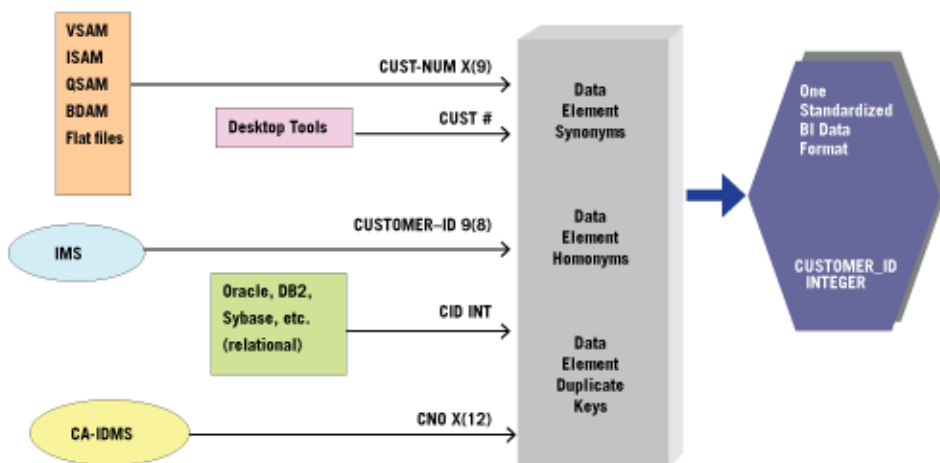
ETL processing time frames, or data staging windows, are typically small. Because the source data is often of poor quality, a lot of time is required to run the transformation and cleansing programs. Finishing the ETL process within a designated time frame can pose a challenge to most organizations. In general, the time requirement necessary to build a data warehouse is the biggest problem for any organization. And one of the big reasons is the long and winding ETL process (see Figures 4, 5 and 6).

Figure 4: ETL Processes

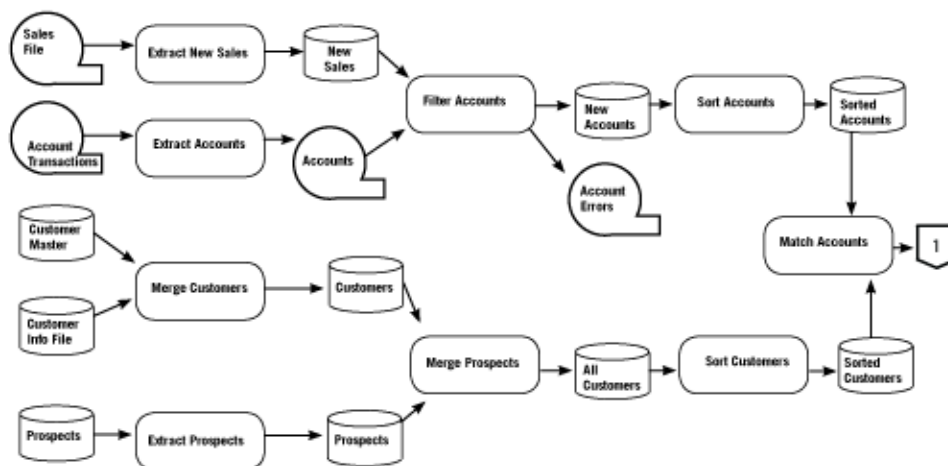


Source: Addison-Wesley, "Business Intelligence Roadmap: The Complete Project Lifecycle for Decision-Support Applications," by Larissa T. Moss and Shaku Atre

Figure 5: Resolution of Duplicate Primary Keys



Source: Addison-Wesley, "Business Intelligence Roadmap: The Complete Project Lifecycle for Decision-Support Applications," by Larissa T. Moss and Shaku Atre

Figure 6: ETL Process Flow Diagram

Source: Addison-Wesley, "Business Intelligence Roadmap: The Complete Project Lifecycle for Decision-Support Applications," by Larissa T. Moss and Shaku Atré

The Difficulty of Meeting User Demands

Two other factors that bog down and complicate data warehouse projects are data integration and data presentation. The users in a business are a diverse group with diverse needs. As a result, the data they want comes from many disparate sources, such as order entry, sales, marketing, human resources and finance, as well as from various electronic documents, printed documents or multimedia. (I'll address unstructured data sources in a future column.)

Business users want up-to-date information from their online transaction processing systems. Figure 7 illustrates the various systems from which organizations retrieve data. All of these disparate data sources have to be integrated.

Figure 7: Users Want to Have Data from a Number of Operational Sources

Types of Data Sources

Operational data – internal operational data from OLTP and batch systems
Examples are:

- Financial systems
- Logistics systems
- Sales systems
- Order entry systems
- Personnel systems
- Billing systems
- Research and engineering systems

And many more – but the data has to be up to date.

Source: Atré Group, Inc.

Determining user needs can be ambiguous. Often, users don't know what they want in advance - only after they see it do they recognize the need. What's more, user requirements for information delivery tend to be dictated by whatever technology they are using at home or on the job, and as technology changes, the tools change. These days, users engage with user-friendly websites when ordering merchandise online, banking online or even ordering food online. As a result, they expect the technology at their jobs to be equally accessible. This means the solutions have to be intuitive and should require no training. After all, no user training is required for consumers to order books from Amazon, clothes from Land's End or food from FreshDirect. Why should there be a need for training at work? Indeed, if too many bells and whistles are added to the tools, they become too complex and inaccessible.

Online technology also increases user demand for the most immediate data possible. They want current, actionable information. This notion works counter to the data warehouse, which generally provides older data.

Data Warehouse Realities

Although data warehouse vendors don't like to admit that building a data warehouse is a long and winding road, there are many challenges to building and maintaining one.

As mentioned, data that is housed in the data warehouse is often either incorrect or inconsistent. Second, determining which data should be stored in a data warehouse is a major issue, and cleaning up that data can be even more so. Frequently, this activity is not practiced or is relegated to "benched" personnel.

Third, easy user access to corporate data is still a mirage after building (or failing to build) a data warehouse. And lastly, most budget decisions are based on the corporation's quarterly performance. A data warehouse usually becomes a multiyear project; to sustain its construction and reach completion, the project must show progress and results at the end of each quarter for many quarters. As a result, any nondesirable quarterly performance bodes bad news for the budget supporting construction of a data warehouse.

Alternatives for Providing Actionable Information

Thankfully, there are other ways to provide users with the information they need to make good business decisions.

Start working on in-memory analytics: With increased size of available memory (also called random access memory, or RAM), one can store data in the memory. Accessing memory is much faster than accessing the hard drive. As a result, the searches of data stored in-memory are unbelievably faster than accessing the same data from the hard drive. (A future column will go into more detail on this subject.) Accessing the large volumes of data typically required for analytics is also improved in-memory.

Start small: Develop a list of small successes by providing information that was promised to users. Seek out one particularly influential user to whom you can offer useful, actionable information.

Create a roadmap of small successes: Realize business benefit quickly, which helps encourage further investment.

Recognize your data strategy: Identify the most important data – it is probably only 10 percent of all your collected data. Determine how accurate and consistent it is, and what percentage of it needs to be updated. If an unusually large amount of data needs to be cleaned up, it is possible that the newer data entering the system is dirty. If this is not the case, you can use the newer data to provide information to users. After all, they want the most current data.

Build metadata as you observe data usage: Metadata describes an organization in terms of its business activities and the business objects on which the business activities are performed. Consider, for example, the sale of a product to a customer by an employee. The sale is a business activity, and the product, customer and employee are the business objects on which the sale activity is performed. These activities and objects, and the relationships and rules that govern them, provide the context in which the businesspeople use the business data every day. Metadata helps metamorphose business data into information, and it ensures the correct interpretation (based on activities, objects, relationships and rules) of what the business data actually means. Since metadata provides the business context in which business data is used, it can be viewed as a semantic (interpretive) layer of the BI decision support environment.

Consider data objects: In order to achieve much better performance with in-memory analytics, one can use the technology of data objects. A data object is a data store where information is stored from multiple and disparate sources. For ad hoc queries and reports, the data can be modeled as one or more flat or multidimensional structures. These data objects don't need a data warehouse or the metadata layer; however, they can co-exist with data warehouses. In-memory architecture should enable users to drill down to other objects stored on the same or different machines. This could be the implementation of cloud computing.

(Read more about the cloud in my column, "[Who in the World Doesn't Want to Reach for the Clouds?](#)") Data objects should be available for on-demand or parameterized reports, ad hoc queries and reporting tools.

Clearly, a data warehouse is not the ultimate solution for companies that want to move into making better, more data-driven decisions. These systems consume time and resources to build, with unresolved issues around their development. Increasingly, companies will turn to alternative ways, such as in-memory analytics, of presenting users with the actionable information they need to make the best business decisions.

This is the second in a series of articles by Shaku Atre. Click on the titles to read the other recent articles: "[Who in the World Uses Only Words and Numbers in Reports?](#)"; "[Who in the World Wants to Stay Locked Up?](#)"; "[Who in the World Doesn't Want to Reach for the Clouds?](#)"; "[Who in the World Wouldn't Want a Collaborative BI Architecture?](#)"; "[Who in the World Wants More Data?](#)"; "[Who in the World Wouldn't Want to Evaluate BI Products?](#)"; "[Who in the World Needs a Hard Drive?](#)"; "[Who in the World Wants to Just Be Structured?](#)"; "[Who in the World Needs a Hard Drive?](#)"; and "[Who in the World Wants to Just Be Structured?](#)"

Shaku Atre is president of Atre Group Inc. and of a leading consulting, training and publishing company specializing in business intelligence. An acknowledged expert in the data warehousing and database field and an exceptional speaker, she captures the attention of audiences and maintains their interest while guiding her listeners through sophisticated material. Before heading her present company, Ms. Atre was a partner with PricewaterhouseCoopers. She also has 14 years of experience in various fields with IBM. As a prolific writer, she has published more than 500 columns in various publications and is the author of a number of books, including "Database: Structured Techniques for Design, Performance and Management," "Information Center: Strategies and Case Studies," "Database Management Systems," "Distributed Databases, Cooperative Processing & Networking," "Atre's Roadmap for Data warehouse/Data mart Implementations," and "Business Intelligence Roadmap: The Complete Project Lifecycle for Decision-Support Applications." She can be contacted at shaku@atre.com. Download some of her previously published columns at www.atre.com/forms/survey1.html.



© 2013 [SourceMedia](#). All rights reserved.