RELIEF FOR THE DATA MAPPING HEADACHE

Exeros helps companies discover the relationship between data in different databases

by Fern Halper,
Partner

A typical large company may have hundreds of databases that were built for various purposes such as to support ERP, CRM, or e-commerce initiatives - but does the company understand the relationship of the data in one database to another database? The answer is, not likely.   There are a number of reasons why this understanding is critical including speeding up data integration initiatives, deploying new applications, or building master data. And, if a company doesn't know what data it has and how it relates, how can it build an enterprise-wide metadata repository? This lack of understanding can result in huge dollars spent on manually determining relationships between data, which can hold up activities that impact decision-making and have a direct impact on both top and bottom line revenue. All in all, a major headache.   California-based Exeros has developed software called DataMapper that is targeted at alleviating this all too common problem.

Companies Don't Know What They Have

There are many reasons why companies may have lost track of the data that they have in their databases. Design specifications get lost or outdated, subject matter experts leave companies, databases and business rules get changed without updating documentation, mergers and acquisitions wreak havoc on databases, all leading to a company not knowing exactly what they have. Metadata associated with the source database might also not be enough.   The end result is inconsistent data.

Consider the following simple example: A company runs an online billing application that allows its customers to view and pay their bills online. The person(s) that developed the application is no longer with the company. The data for this application comes from three different sources, but the application is supported by a local instance of a database. Customers start to complain that the application isn't working properly. The IT group knows that the application didn't change, and determines that one of the databases that feeds the application must have changed. However, they cannot map the data in the source database(s) to that in the target (local) database. The result is that the IT group needs to manually research and understand the tables and the business processes between the systems – a time consuming task that will incur high costs and decreased customer satisfaction.

Various options exist in the market for profiling and mapping data. Data profiling tools help end-users understand their source data, for example by profiling values in different database rows. These tools often work off of one dataset and are used in conjunction with data quality tools. Data mapping tools can take the source data profile and use it to match it to an existing or new target database. This is important because in any integration, migration, or new application deployment effort, companies need to understand where the data fields are coming from and where they are going to.

The Exeros Solution

Exeros has developed a patent-pending solution that automatically discovers data relationships such as pair-wise maps. The software, called DataMapper, does this by looking at data values between source and target systems and discovering patterns between the two using a combination of heuristics and statistics. Key features include:

- An intuitive user interface that lets analysts map tables and determine the relationship between them. For example, an analyst might look at four tables from one database and three tables from another to understand the relationship between the two. DataMapper provides the transformation in standard sequel that informs the analyst how the columns are related. Exeros has determined that anything above a 20% match rate is not an accident. The company's philosophy is to provide users with types of interfaces they are used to working with so they don't have to learn another set of tools.
- DataMapper also identifies discrepancies in the mappings by providing a hit and miss ratio between transformations and allows analysts to drill down into the data to understand why.
- Output can be directly integrated with a metadata repository or ETL tools. Exeros currently generates Informatica PowerCenter mappings and is in the process of integrating with IBM and Business Objects ETL software.

Exeros has helped users solve problems dealing with identifying and understanding sensitive data, developing master data, successfully deploying new applications, and dealing with ETL migrations. Typically, it can discover between 80-90% of the relationships between two datasets.

The Value of the Approach

Understanding the relationship between source and target data systems can be very valuable for a company. What is interesting about the Exeros approach is

that it can automatically discover relationships between data across databases, utilizing data values.

"Automating the mapping process can also speed up various integration efforts resulting in a faster time to market."

Benefits of the approach include:

- **Cost Reduction** – If Exeros can determine 80% of the relationships between data in different databases, this is a big improvement over manually having to discover all of the relationships. This results in savings in time and cost.
- **Faster time to market** – Automating the mapping process can also speed up various integration efforts resulting in a faster time to market.
- **Improved data governance** – it is far easier to deal with the guidelines and policies that a company has set up to deal with its data issues (e.g. ownership, compliance, security) when it understands the data that it has.
- **Better decision making** – If a company better understands its data, it can develop better applications, faster which can result in better business decisions.

Fern Halper is a Partner at Hurwitz & Associates and can be reached at fern.halper@hurwitz.com and marcia.kaufman@hurwitz.com.

Hurwitz & Associates is a consulting, research, and analyst firm focused on the customer benefits derived when advanced and emerging software technologies are implemented to solve pragmatic business problems. The firm's research concentrates on understanding the business value of software technologies, such as SOA and Web services, and how they are successfully implemented within highly distributed computer environments.