REAL WORLD CASE 1

Argosy Gaming Co.: Challenges in Building a Data Warehouse

When you've got half a dozen riverboat gambling operations, it's important that everyone plays by the same rules. Argosy Gaming Co. (www.argosycasinos.com), with headquarters in Alton, Illinois, and a fleet of six Mississippi riverboat casinos, had decided that bringing all customer data together would enhance management's view of operations and potentially help strengthen customer relationships. To accomplish those goals, though, the company needed to access a variety of databases and develop an extract, transform, and load (ETL) system to help construct and maintain a central data warehouse.

Jason Fortenberry, a data-warehousing analyst, came aboard at Argosy just as the company's data warehouse project started in 2001. His job was made easier, he says, by the adoption of Hummingbird Ltd.'s Genio ETL software tool, which helped bridge systems and automate processes. But like others going through such projects, he learned the hard way that preparing for the ETL process is just as important as having the right software.

The riverboats each had unique and incompatible ways of defining a host of operational activities and customer characteristics—in essence, the floating casinos were each playing the same game but with different rules. But those problems remained hidden until reports from the company's data warehouse began to turn up inconsistent or troubling data. That's when Fortenberry and his staff discovered conflicting definitions for a wide range of data types—problems he wishes he had identified much earlier. Fortenberry's troubles—and his successes—are typical of ETL, the complex and often expensive prelude to data warehouse success.

ETL is often problematic because of its inherent complexity and underlying business challenges, such as making sure you plan adequately and have quality data to process. Analysts, users, and even vendors say all bets are off if you don't have a clear understanding of your data resources and what you want to achieve with them. Then there are choices, like whether to go for a centralized architecture—the simplest and most common configuration—or a distributed system, with ETL processing spread across various software tools, system utilities, and target databases, which is sometimes a necessity in larger, more complicated data warehouses. Even if you navigate those waters successfully, you still need to ensure that the ETL foundation you build for your data warehouse can meet growing data streams and future information demands.

As the term implies, ETL involves extracting data from various sources, transforming it (usually the trickiest part), and loading it into the data warehouse. A transformation could be as simple as reordering the fields of a record from a source system. But as Philip Russom, a Giga Information Group analyst, explains, a data warehouse often contains data values and data structures that never existed in a source system. Since many analytical questions a business user would ask of a data warehouse can be answered only with calculated values (like averages, rankings or metrics), the ETL tool must calculate these from various data sources and load them into the warehouse. Similarly, notes Russom, a data warehouse typically contains "time-series" data. The average operational application keeps track of the current state of a value such as a bank account balance. It's the job of the ETL tool to regularly add new states of a value to the series.

For his year-long ETL project, Argosy's Fortenberry says Hummingbird's Genio Suite, a data integration and ETL tool, quickly became the project's "central nervous system," coordinating the process for extracting source data and loading the warehouse.

But for Argosy, getting all that data into the warehouse didn't produce immediate usable and dependable results. "The lesson was that people thought that they were talking about the same thing, but they actually were not," says Fortenberry. For example, he explains, riverboats calculated visits differently. One riverboat casino would credit a customer with a visit only if he actually played at a slot machine or table. Another had an expanded definition and credited customers with visits when they redeemed coupons, even if they didn't play. So identical customer activity might have one riverboat reporting 4 player visits and another reporting 10. "This type of discovery was repeated for everything from defining what a 'player' is to calculating a player's profitability," says Fortenberry.

IT played a lead role in identifying problems and helping to hammer out a consensus among the business units about how to define and use many categories of data, he says. Now, the data warehouse is running smoothly and producing dependable results for business analysis and management reporting, so the number of problem-resolution meetings has dropped dramatically. Still, Fortenberry reckons that three-quarters of the meetings he attends nowadays have a business focus. "For our part, we now know better what questions to ask business users as we continue with the data warehouse development process," he says.

Case Study Questions

1. What is the business value of a data warehouse? Use Argosy Gaming as an example.

2. Why did Argosy use an ETL software tool? What benefits and problems arose? How were they solved?

3. What are some of the major responsibilities that business professionals and managers have in data warehouse development? Use Argosy Gaming as an example.